KAKENHI Symposium

MEXT KAKENHI (A) No. 23244011. Waseda University, Masanobu Taniguchi.
(Asymmetric, Nonlinear Statistical Theory and Its Applications to Economics and Biosciences)

Recent Advances in Statistical Theory and Applications for High Dimensional Data Analysis and Related Topics

Organizer, Otaru University of Commerce, Qingfeng Liu.
Sponsor, Research Group for Regional Studies, Otaru University of Commerce.
Sponsor, Saturday Research Meeting, Department of Economics, Otaru University of Commerce.

PROGRAM

Venue: Conference Room 1, Administration Bldg 2F, Otaru University of Commerce.

	5th September 2013	**Chair Katsumi Shimotsu**
14:00-14:50	Naoya Sueishi Kyoto University	A New Interpretation of Empirical Likelihood for Time Series Models and Its Application to Model Selection Testing
15:00-15:50	Chihiro Hirotsu Meisei University	A Unifying Approach to the Shape and Change-point Hypotheses
16:00-16:50	Wenjie Wang Kyoto University	GMM Model Averaging for Conditional Moment Restrictions, by Qingfeng Liu and Wenjie Wang.
	6th September 2013	**Chair Ryo Okui**
10:00-10:50	Ryo Okui Kyoto University	Asymptotic Inference for Dynamic Panel Estimators of Infinite Order Autoregressive Processes, by Yoon-Jin Lee, Ryo Okui and Mototsugu Shintani
11:00-11:50	Katsumi Shimotsu University of Tokyo	Testing the Number of Components in Finite Mixture Models
11:50-14:00		*Lunch Time*
14:00-14:50	Yang Feng Columbia University	Consistent Cross-Validation for Tuning Parameter Selection in High-Dimensional Variable Selection
15:00-15:50	Andrey R. Vasnev University of Sydney	Interpretation and Use of Sensitivity in Econometrics with an Illustration to Forecast Combinations, by Jan R. Magnus and Andrey L. Vasnev.
16:00-16:50	Jau-er Chen National Taiwan University	Intertemporal Relation between Risk and Return: Panel Quantile Regression Approach
18:00–		*Dining*
	7th September 2013	**Chair Qingfeng Liu**
09:00-09:50	Fumiya Akashi Waseda University	An Empirical Likelihood Approach toward Discriminant Analysis for Non-Gaussian Vector Stationary Processes
10:00-10:50	Hiroaki Ogata Waseda University	Consideration on a Serial Correlation
11:00-11:50	Shinya Tanaka Takasaki City University of Economics	Identification of Approximate Factor Models through Heteroskedasticity

A new interpretation of empirical likelihood for time series models and its application to model selection testing

Naoya Sueishi
Kyoto University

Abstract

The empirical likelihood estimator loses its interpretation as a nonparametric maximum likelihood estimator when it is applied to time series. This paper gives a new information -theoretic interpretation of empirical likelihood for time series models specified via moment restrictions. We show that the empirical likelihood estimator of Kitamura (1997, Annals of Statistics) minimizes the Kullback-Leibler information criterion of the joint distribution from the model to the true data generating process. As an application, we also propose a Vuong-type test for comparing possibly misspecified dynamic models.

1 Model

Let $\{Y_t\}_{t=1}^{T}$ denote observations of a finite-dimensional stationary and strong mixing process $\{Y_t\}_{t=1}^{\infty}$ that is on a probability space $(\Omega, \mathcal{F}, \mu)$, where $\Omega = \mathbb{R}^{d_y \infty} = \prod_{t=1}^{\infty} \mathbb{R}^{d_y}$ and $\mathcal{F} = \mathcal{B}(\mathbb{R}^{d_y \infty})$. We consider the model

$$E[m(Y_t;\theta_0)] = \int_{\Omega} m(y^t;\theta_0) d\mu = 0, \quad (t = 1, 2, \cdots)$$

where $y^t \in \mathbb{R}^{d_y}$ is the t-th coordinate of $\mathbb{R}^{d_y \infty}$.

Kitamura (1997) proposes an efficient estimator of θ_0 on the basis of the empirical likelihood. However, an interpretation of his estimator is rather unclear. This study gives a new information-theoretic interpretation to the Kitamura's (1997) estimator.

2 KLIC minimization problem

The Kullbuck-Leibler information criterion (KLIC) defines a pseudo-distance between the model and the true DGP. Let $\mathbf{M}_B$ be the joint probabilities on $(\mathbb{R}^{dyB}, \mathcal{B}^{d_y B})$. We define $\mathcal{P}_\theta^B = \{P_B \in \mathbf{M}_B : \int m(y^t;\theta) dP_B = 0, t = 1, 2, \ldots, B\}$ and $\mathcal{P}_B = \cup_{\theta \in \Theta} \mathcal{P}_\theta^B$. $\mathcal{P}_B$ is a set of joint probabilities that are compatible with the moment restriction.

Let μ_B be the true joint probability that is induced from μ. The KLIC from $\mathcal{P}_B$ to μ_B is

$$D^B(\mu_B \| \mathcal{P}_B) = \min_{P_B \in \mathcal{P}_B} D^B(\mu_B \| P_B),$$

where

$$D^B(\mu_B \| P_B) = \begin{cases} -\int \log\left(\frac{dP_B}{d\mu_B}\right) d\mu_B & \text{if } P_B \ll \mu_B \\ \infty & \text{otherwise.} \end{cases}$$

After some manipulation, we obtain

$$D^B(\mu_B \| \mathcal{P}_B) = \min_{\theta\in\Theta} \max_{\lambda\in\Lambda} \int \log\left(1 + \lambda' \sum_{i=1}^{B} m(y^i;\theta)\right) d\mu_B.$$

The sample analogue of the saddle point problem is

$$\min_{\theta\in\Theta} \max_{\lambda\in\Lambda} \frac{1}{T-B+1} \sum_{t=1}^{T-B+1} \log\left(1 + \lambda' \sum_{i=0}^{B-1} m(Y_{t+i};\theta)\right).$$

The resulting estimator is the same as the blockwise EL estimator of Kitamura (1997). Thus, we can obtain the blockwise EL estimator as the solution to the KLIC minimization problem.

3 Application to model selection test

Suppose that there are two competing dynamic models that are specified by moment restrictions: $E[m^{(1)}(Y_t;\theta_0^{(1)})] = 0$ and $E[m^{(2)}(Y_t;\theta_0^{2)})] = 0$. Both models are misspecified. We propose an empirical likelihood-based model selection test. Our test is similar to that of Kitamura (2001). Let $\mathcal{P}_B^{(j)} = \cup_{\theta\in\Theta}\{P_B \in \mathbf{M}_B : \int m^{(j)}(y^t;\theta)dP_B = 0, t = 1,2,\ldots B\}$ for $j = 1,2$. We test the null hypothesis:

$$D^B(\mu_B \| \mathcal{P}_B^{(1)}) = D^B(\mu_B \| \mathcal{P}_B^{(2)})$$

for some B.

Our test is based on the sample analog of the difference of two KLICs:

$$\hat{D}^B(\mu_B \| \mathcal{P}_B^{(1)}) - \hat{D}^B(\mu_B \| \mathcal{P}_B^{(2)})$$

where

$$\hat{D}^B(\mu_B \| \mathcal{P}_B^{(j)}) = \frac{1}{T-B+1} \sum_{t=1}^{T-B+1} \log\left(1 + \hat{\lambda}_B^{(j)\prime} \sum_{i=0}^{B-1} m^{(j)}(Y_{t+i};\hat{\theta}_B^{(j)})\right).$$

Under the null hypothesis, the test statistic converges in distribution to a normal distribution.

形状制約と変化点仮説への総合的接近法

明星大学　広津千尋

１．　序論

主題の総合的とは幾つかの意味を含んでいる．最も重要なのは，従来統計学の二つの独立な流れで研究されてきた形状制約と変化点仮説を同時に扱う事である．それは，単調仮説と段差変化点仮説，凸性仮説とスロープ変化点仮説，そしてＳ字性仮説と変曲点モデル等を含んでいる．この総合化は，例えば，副作用自発報告のモニタリングにおいて，増加傾向と同時にその変化点にも興味が持たれることから，応用上も重要な意味を持つ．次に，基礎とする確率分布を正規分布に限定せず，１変量指数分布族を一般的に扱う．さらに、1元配置で自然に導かれる手法を一貫した方法で2元配置に拡張する。

この分野で最もよく知られているのは，単調仮説に対する制約付き最尤法(isotonic regression)であるが，それはこのようにパラメータースペースに制約がある場合に対し明らかな最適性は持たず，直感的に導入された物に過ぎない．また，それは 1 元配置正規分布モデルを超えて一般化するには困難が伴う．本論はそれを一般化し，かつ，制約仮説に対する検定の完全類を基本としているため，その導入から最適性に立脚している．

２．検定の完全類

今，データ $\boldsymbol{y}$ が確率分布（尤度）$L(\boldsymbol{y},\boldsymbol{\theta})$に従うものとする．ここで，$\boldsymbol{\theta}=(\theta_1,\dots,\theta_K)'$に関する制約仮説

$$\mathrm{H}\colon\ A'\boldsymbol{\theta}\geq\mathbf{0} \tag{1}$$

を考える．ただし，不等式(1)は要素ごとに適用し，少なくとも一つの不等号は厳密とする．すべて等号が成立する場合が帰無仮説: $A'\boldsymbol{\theta}=\mathbf{0}$である．このとき，検定の完全類は次の補助定理で与えられる．

補助定理(Hirotsu, 1982). 制約仮説(1)に対する検定の完全類は$(A'A)^{-1}A'\boldsymbol{\nu}(\widehat{\boldsymbol{\theta}}_{\mathbf{0}})$の各要素に関し単調増大な y の凸領域で与えられる．ただし，$\boldsymbol{\nu}(\widehat{\boldsymbol{\theta}}_{\mathbf{0}})$は帰無仮説の下で評価した$\boldsymbol{\theta}$に関する efficient score vector である．以下本論では，y_kが互いに独立に 1 母数指数分布族

$$L(\boldsymbol{y},\boldsymbol{\theta})=\prod_k a(\theta_k)b(y_k)\mathrm{e}^{\theta_k y_k}$$

に従うことを仮定する．このとき，$\boldsymbol{\nu}(\widehat{\boldsymbol{\theta}}_{\mathbf{0}})=\boldsymbol{y}-\widehat{\mathrm{E}}(\boldsymbol{y}|\,\boldsymbol{\theta}_{\mathbf{0}})$である．ここで$\widehat{\mathrm{E}}(\boldsymbol{y}|\,\boldsymbol{\theta}_{\mathbf{0}})$は帰無仮説下における $\boldsymbol{y}$ の期待値の最尤推定量であり，完備十分統計量の関数となる．そこで，この項は相似検定を構成する際の十分統計量を与えた条件付き推測では定数として扱うことが出来，$\boldsymbol{\nu}(\widehat{\boldsymbol{\theta}}_{\mathbf{0}})$は本質的に観測値ベクトル $\boldsymbol{y}$ と同等になる．すなわち，$\boldsymbol{t}=(A'A)^{-1}A'\boldsymbol{y}$が完全類を構成する key vector となる．一方，係数行列$(A'A)^{-1}A'$の各行は制約式(1)で定義される凸錐の corner vector を表す．すなわち，

$$A'\boldsymbol{\theta} \geq \mathbf{0} \leftrightarrow \boldsymbol{\theta} = A(A'A)^{-1}\boldsymbol{c},\ \boldsymbol{c} \geq \mathbf{0}$$

が成立する．言い換えると，制約仮設 H を満たす任意の$\boldsymbol{\theta}$は$A(A'A)^{-1}$の列の一意な正係数線形結合で表される．従って，検定の完全類は $\boldsymbol{y}$ を凸錐の corner vector に射影したすべての成分に関して単調増大，かつ凸な領域で与えられるという意味を持つ．問題によってこの corner vector は簡明，かつ解釈の容易な形を取ることがある．

例１． 単調仮説

$\boldsymbol{\theta}$に関する単調仮説：$\theta_1 \leq \cdots \leq \theta_K$は，(1)において，$A'$を差分行列

$$A' = D'_K = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & \ldots\ldots. & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & \ldots\ldots. & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & \ldots\ldots. & 0 & 0 & 0 \\ & & & & \ldots\ldots\ldots & & & & \\ 0 & 0 & 0 & 0 & 0 & \ldots\ldots. & 0 & -1 & 1 \end{bmatrix}_{(K-1)xK} \tag{2}$$

とすることによって定義される．帰無仮説は$\boldsymbol{\theta}$の一様性$\mathrm{H}_0{:}\theta_1 = \cdots = \theta_K$である．このとき，corner vector は陽に

$$D_K(D'_K D_K)^{-1} = \frac{1}{\mathrm{K}}\begin{bmatrix} -(K-1) & -(K-2) & & -1 \\ 1 & -(K-2) & & -1 \\ 1 & 2 & & -1 \\ & & \ldots\ldots\ldots. & \\ 1 & 2 & & -1 \\ 1 & 2 & & (K-1) \end{bmatrix}_{K\mathrm{x}(K-1)} \tag{3}$$

と表され，key vector

$$(D'_K D_K)^{-1} D'_K y = \begin{bmatrix} (Y_K/K) - Y_1 \\ 2(Y_K/K) - Y_2 \\ \ldots\ldots. \\ (K-1)(Y_K/K) - Y_{K-1} \end{bmatrix}$$

の各要素が，本質的に累積和統計量

$$Y_k = y_1 + \cdots + y_k,\ k = 1, \ldots, K-1$$

で代表される．ただし，Y_K/Kは総平均であり，帰無仮説H_0の下での十分統計量である．すなわち，完全類は累積和Y_k，$k = 1, \ldots, K-1$，に関して単調減少な凸領域で与えられる．単調仮説H_Mを表すのに，一見，差分を定義する(2)式が基本的に思われがちだが，実は(3)式の各列こそが，すべての単調増大ベクトルを一意の正係数線形結合で表すという意味で本質的な規定単調ベクトルである．完全類がこれらの規定ベクトルが作る対比に関して単調増大な関数から成るというのは直観的にも大変合理的である．一方，(3)式の各列は段差変化点仮説の要素を表す．このことから，単調仮説と段差変化点仮説の間に密接な関係のあることが分かる．このクラスに属する検定統計量はいくらでも考えられるが，取り扱いが

簡単で検出力も高いためよく用いられるのが，最大対比型(max acc. t)と二乗和型(累積χ^{*2})である．とくに前者は，段差変化点仮説

$$\mathrm{H_D}: \theta_1 = \cdots = \theta_k \leq \theta_{k+1} = \cdots = \theta_K,\ k = 1, \ldots, K-1$$

に対する efficient score 検定となる．

凸性仮説は２階差分行列，S 字性仮説は 3 階差分行列で定義され，例 1 と同様な議論により，それぞれスロープ変化点仮説，および変曲点モデルとの関係が導かれる(Hirotsu & Marumo, 2002)．また，基礎統計量として 2 重累積和，および 3 重累積和が導かれる．

３．最大対比統計量

最大対比統計量は単調仮説と変化点仮説の両方に対応する統計量であり，片側，両側どちらにも用いられる．とくに累積和に基づく方法は，変化点，および変化量の効率の良い同時信頼区間が得られる(Hirotsu et al., 1992，Hirotsu and Srivastava, 2000，Hirotsu et al., 2011)こと，および成分統計量のマルコフ性から正確，かつ高速な確率計算アルゴリズムが得られる(広津他，1997)等の利点から，近年よく用いられる．一方，2 重累積和に基づく方法に関して最近発展が見られた．正規分布モデルについては，Hirotsu & Marumo (2002) である程度の結果が得られているが，最近一般の指数分布族への拡張がなされた．

一般に，このような最大対比統計量の有意確率計算は，多重積分，あるいは多重和分を要し，水準数が多いと計算が困難になる．ところが，2 重累積和の最大対比については，相続く成分統計量の２階マルコフ性から効率の良い正確計算法が導かれる．今，2 重累積和を

$$S_k = ky_1 + (k-1)y_2 + \cdots + y_k,\ k = 1, \ldots, K-2$$

と表す．この時，同時分布は２階マルコフ性により，$G(\boldsymbol{y}) = \prod f_k\left(S_k \middle| S_{k+1},\ S_{k+2}\right)$と表され，$f_k$は，例えば Poisson 分布の場合に，線形性の帰無仮説の下で具体的に

$$f_k\left(S_k \middle| S_{k+1},\ S_{k+2}\right) = C_{k+1}^{-1}\left(S_{k+1},\ S_{k+2}\right) C_k\left(S_k,\ S_{k+1}\right)\{(S_{k+2} - 2S_{k+1} + S_k)!\}^{-1}$$

という形で表される．C_{k+1}は規準化定数であり，確率変数S_kが右辺最後の項の他，1 ステップ前の規準化定数C_kにも含まれていることに注意する．規準化定数C_{k+1}は漸化式

$$C_{k+1} = \sum_{S_k} C_k\left(S_k,\ S_{k+1}\right)\{(S_{k+2} - 2S_{k+1} + S_k)!\}^{-1},\ k = 1, \ldots, K-2,$$

によって計算される．ただし，$\mathrm{C}_1 = \{\mathrm{S}_1!\,(\mathrm{S}_2 - \mathrm{S}_1)!\}^{-1}$と定義しておく．以上により，確率変数$S_k$，$k = 1, \ldots, K-2$の同時分布とその分解が得られた．系列のサイズ Kが大きいとサンプルスペースが爆発し，同時分布に基づく計算は困難となるが，この分解定理のおかげで，モーメントや p-value 計算の効率の良い漸化式が得られる．

４．応用

累積和に基づく最大対比統計量を max acc. $1t$、2 重累積和に基づく最大対比統計量を max acc. $2t$ と呼ぶことにする．今，副作用自発報告のモニタリングにおいて，増加傾向とその変化点検出には max acc. $1t$ が有用である．一方，増加傾向を検出し，適切な措置を取

った後では,増加傾向が減少に転じたこと(down turn)を検証することに興味が持たれ,max acc. 2tの応用が適切となる.

別の応用として，凸性検定は，用量反応曲線等において対数線形性の指向性適合度検定として有用である.対数線形性が成立しない時は,単調性や凸性のみを仮定した max acc. 1t や max acc. 2tに基づくノンパラメトリックな解析法を適用すればよい.

５．議論

従来は生物統計分野への応用を主としてきた．新たにファイナンス問題への応用を考えたいが，一つの論点として，時間推移に伴う変動を平均のゆっくりした変化と捉えるか，あるいは誤差相関と捉えてモデル化するかという問題がある．例えば、生物統計分野で経験した例で，コレステロール低下剤投与によるコレステロール量の推移を１ヵ月ごとに 6 ヶ月間計測したデータと，血圧を 30 分ごと 24 時間計測したデータがある．前者では時間の推移に伴う変化は，平均のゆっくりした単調変化と捉え，計測誤差は独立というモデルが容認される(Hirotsu, 1991)．一方後者では，計測器を装着したままの連続計測値であるから，誤差の系列相関は否定出来無い(Hirotsu et al., 2003)．このどちらのモデルを採用するかは解析上の重要なポイントとなる.

参考文献

Hirotsu, C. (1982) : Use of cumulative efficient scores for testing ordered alternatives in discrete models. Biometrika 69, 567-577.

Hirotsu, C. (1991) : An approach to comparing treatments based on repeated measures. Biometrika 78, 583-594. .

Hirotsu, C., Kuriki, S. and Hayter, A. J. (1992) : Multiple comparison procedures based on the maximal component of the cumulative chi-squared statistic. Biometrika 79, 381-392.

Hirotsu, C. and Srivastava, M. S. (2000) : Simultaneous confidence intervals based on one-sided max t test. Statistics and Probability Letters 49, 25-37.

Hirotsu, C. and Marumo, K. (2002) : Change point analysis as a method for isotonic inference. Scandinavian Journal of Statistics 29, 125-138.

Hirotsu, C., Ohta, E., Hirose, N. and Shimizu, K. (2003) : Profile analysis of 24-hours measurements of blood pressure. Biometrics 59, 907-915.

Hirotsu, C., Yamamoto, S. and Hothorn, L. (2011) : Estimating the dose-response pattern by the maximal contrast type test approach. Statistics in Biopharmaceutical Research 3, 40-53.

広津千尋，西原　健自，杉原　正顕(1997)：最大ｔ法のための有意確率、検出力、例数設計計算アルゴリズム．応用統計学 26, 1-16.

GMM Model Averaging for Conditional Moment Restrictions

Qingfeng Liu[*]
Department of Economics, Otaru University

Wenjie Wang[†]
Department of Economics, Kyoto University

September 05, 2013

In the situation of many moment restrictions, using larger number of restrictions in GMM estimation reduces variances, on the other hand causes larger bias. In order to reduce the MSE of the GMM estimators, we must control the balance between bias and variance. For this purpose, one method is to select a part of the set of all moment restrictions. Recently as an alternative of moment restriction selection the model averaging selection methods have been attempt to applied to the field of many moment restrictions, for example Kuersteiner and Okui (2010) (hereafter denotes as KO) used model averaging method for the first step of 2 stages least square (2SLS) for linear IV models, and Lee and Zhou (2011) (hereafter denotes as LZ) averaged 2SLS estimators with different instrumental variables directly to construct the model averaging estimator. We propose the GMMMA, a new model averaging method for GMM estimation with conditional moment restrictions of i.i.d. random variables. Since model selection can be regarded as a special case of model averaging with restricted weight vector, the GMMMA is expected to over perform the moment selection methods.

The primal aim of this paper is to choose an optimal weight vector of GMMMA to minimize the MSE. The GMMMA estimator is a weighted average of the GMM estimators based on different sets of moment restrictions. As shown by KO and LZ, since the first order MSE of the model averaging estimator dose not depend on the weight vector of model averaging, we need to derive the higher order MSE of the GMMMA estimate. The Nagar's type higher-order expansion of the MSE of GMMMA is derived. We use the adjust version of the estimate of this expansion as the criterion for choosing the optimal weight vector.

The method for the Nagar's type higher-order expansion is based on Donald, Imbens and Newey (2009) (here after denotes as DIN). LZ and KO have

[*]E-mail: qliu@res.otaru-uc.ac.jp
[†]Corresponding author. E-mail: williamjpzw@yahoo.co.jp.

derived the Nagar's type expansion for linear IV models for model averaging, and Eryürük and Hansen (2012) (hereafter denoted as EH) has done for nonlinear conditional moment restrictions selection. The situation for us differ from that for them, our Nagar's type expansion is for nonlinear conditional moment model, including linear IV models as a special case, and for model averaging not selection.

Following DIN we consider a model of conditional moment restriction. Let $x = \{x_1, \cdots, x_n\}$ and $z = \{z_1, \cdots, z_n\}$ denote i.i.d. random sample with sample size n, where z_i is a $p \times 1$ vector of random variables. We have the conditional moment restriction

$$E[\rho(z_i, \beta_0)|x_i] = 0$$

where $\rho(z_i, \beta)$ is scalar, β is $p \times 1$ parameter vector, and we use the subscript 0 for true values of parameters throughout this paper.

Let $q_i^{K_m}(x_i)$ be a set of $K_m \times 1$ functions of x_i to be used as instruments and let $\hat{\beta}(K_m)$ be the GMM estimator using the instruments $q_i^{K_m}$. The GMMMA estimator takes the form, $\bar{\beta} = \sum_{m=1}^{M} w_m \hat{\beta}(K_m)$, where $w_m \in [0,1]$ for $m = 1, \cdots, M$. We define the weight vector as $W = (w_1, \cdots, w_M)$' here for later use. Similar to the setting in DIN, the instruments do not need to form a nested sequence. K_m fills a double role here, as the index of the instrument set as well as the number of instruments.

Our concern is to reduce the MSE of $\bar{\beta}$

$$MSE(\bar{\beta}) = E\left[n(\bar{\beta} - \beta_0)(\bar{\beta} - \beta_0)'|x\right].$$

Following Donald, Imbens and Newey (2009), we first expand the MSE to $O_p\left(n^{-1}\right)$ and then estimate the expansion so that we can chose the optimal weight vector for GMMMA by minimizing the estimate of the expansion of MSE.

The resulted higher-order asymptotic MSE from the Nagar's type higher-order expansion can be estimated by the sample counterpart. However, a bias term will occur, whose order is $O(1/n)$, the same order of the asymptotic MSE. DIN and EH used parametric methods to corrected this kind of bias, but did not provide theoretical justification. Alternatively, we introduce a bootstrap method to do that. The simulation results shows that the bootstrap method works well.

Asymptotic Inference for Dynamic Panel Estimators of Infinite Order Autoregressive Processes

Yoon-Jin Lee*, Ryo Okui† and Mototsugu Shintani‡

An abstract prepared for the KAKENHI symposiumu at Otaru University of Commerce in September 2013

In this paper we consider the estimation of a dynamic panel autoregressive (AR) process of possibly infinite order in the presence of individual effects. We utilize the sieve AR approximation with its lag order increasing with the sample size. We establish the consistency and asymptotic normality of the standard dynamic panel data estimators, including the fixed effects estimator, the generalized methods of moments estimator and Hayakawa's instrumental variables estimator, using double asymptotics under which both the cross-sectional sample size (N) and the length of time series (T) tend to infinity. We also propose a bias-corrected fixed effects estimator based on the asymptotic result. Monte Carlo simulations demonstrate that the estimators perform well and the asymptotic approximation is useful. As an illustration, proposed methods are applied to dynamic panel estimation of the law of one price deviations among US cities.

Because economic relationships are often dynamic in nature, dynamic panel models have been considered very useful in the analysis of micro economic data. Many estimation methods for simple dynamic panel models have been proposed, and their theoretical properties are investigated in many studies under a fixed T and large N asymptotic framework. More recently, however, an increasing number of panel data with longer T have become available in practice. Motivated by the availability of longer panel data, Hahn and Kuersteiner (2002), Alvarez and Arellano (2003) and Hayakawa (2009), among others, have investigated asymptotic properties of various estimators for "finite order" panel AR models, using an alternative asymptotic approximation when both T and N tend to infinity.

In this paper, we consider the estimation of a general dynamic panel structure in the presence of unobserved individual effects. To this end, we employ a sieve approach to approximate a panel AR model of infinite order by a panel AR model of order p that increases with sample size T and N. Our specification of infinite order AR models covers a very general class of stationary linear processes, which nests standard autoregressive and moving average (ARMA) models of finite order. Therefore, compared to previous studies in the dynamic panel literature, estimation results from our approach are less subject to problems caused by possible model misspecifications. Such an idea of the AR sieve approximation in estimating a general linear model has long been used in the literature of time series analysis. However, a naïve analogy of time series results cannot be directly used, due to several technical issues specific to dynamic panel data analysis under a large T and large N asymptotic framework.

The AR sieve approximation retains the computational simplicity of the finite order AR models, which can be conveniently estimated by a linear regression estimator. In particular,

*Department of Economics, Indiana University, Bloomington, IN 47405-7104, USA. Email: lee243@indiana.edu

†Institute of Economic Research, Kyoto University, Yoshida-Hommachi, Sakyo, Kyoto, Kyoto, 606-8501, Japan. Email: okui@kier.kyoto-u.ac.jp

‡Department of Economics, Vanderbilt University, Box 1819 Station B, Nashville, TN 37235, USA. Email: mototsugu.shintani@vanderbilt.edu.

we consider sieve variants of (i) the fixed effects estimator (Hahn and Kuersteiner, 2002), (ii) the generalized methods of moments (GMM) estimator (Holtz-Eakin, Newey and Rosen, 1988, Arellano and Bond, 1991, and Alvarez and Arellano, 2003) and (iii) an efficient instrumental variables (IV) estimator (Hayakawa, 2009). We show the consistency and asymptotic normality of each estimator. We further construct consistent standard errors for all the estimators and an asymptotically valid automatic lag selection procedure in AR sieve approximation.

Our main theoretical results can be summarized as follows. First, when T is only moderately large, a fixed effects estimator suffers from asymptotic bias. We show that a simple bias correction method, which is analogous to the standard case, works well for the infinite-order AR model. Second, since the number of lags increases with T, the GMM estimator involves many moment conditions even when T is moderately large. Compared to the standard case, N must be much larger relative to T in order for the GMM estimator to behave well without suffering from many moments bias. Third, Hayakawa's IV estimator is shown to be consistent and asymptotically normal under a condition on the relative magnitude of N and T that is weaker than those required for the other estimators. Overall, our theoretical results suggest that the choice among estimators should be based on the relative magnitude of N and T and their finite sample properties.

Our Monte Carlo simulation to evaluate the finite sample properties provides useful guidance for practitioners in choosing among the estimators. Our proposed bias-corrected estimator works well in reducing the bias of the fixed effects estimator without inflating the dispersion of the estimator. When N becomes larger, however, the GMM estimator has a smaller bias than the bias-corrected fixed effects estimator. The bias of the fixed effects estimator is not negligible even when T is fairly large, which illustrates the importance of bias correction. Among all the estimators, Hayakawa's IV estimator has the smallest bias at the cost of larger dispersion. An automatic lag selection procedure also helps to choose the approximation models that produces precise estimates.

Finally, as an empirical illustration, proposed methods are applied to dynamic panel estimation of the law of one price (LOP) deviations among US cities. The speed of individual good price adjustment is evaluated by the estimated sum of AR coefficients using competing estimators. We find that both the fixed effects and the GMM estimator often provide values less than the bias-corrected fixed effects estimator and Hayakawa's IV estimator.

Key Words: Bias correction; Double asymptotic; Fixed effects estimator; GMM; Infinite order autoregressive process; Instrumental variables estimator.

JEL Classification: C13; C23; C26.

References

Alvarez, J. and Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators, *Econometrica* **71**(4): 1121–1159.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economics Studies* **58**: 277–297.

Hahn, J. and Kuersteiner, G. (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large, *Econometrica* **70**(4): 1639–1657.

Hayakawa, K. (2009). A simple efficient instrumental variable estimator for panel AR(p) models when both N and T are large, *Econometric Theory* **25**: 873–890.

Holtz-Eakin, D., Newey, W. and Rosen, H. S. (1988). Estimating vector autoregressions with panel data, *Econometrica* **56**(6): 1371–1395.

Testing the Number of Components in Finite Mixture Models

Hiroyuki Kasahara
Department of Economics
University of British Columbia
hkasahar@mail.ubc.ca

Katsumi Shimotsu
Faculty of Economics
University of Tokyo
shimotsu@e.u-tokyo.ac.jp

Summary

Finite mixture models provide flexible ways to account for unobserved population heterogeneity. Because of their flexibility, finite mixture models have seen numerous applications in diverse fields such as biological, physical, and social sciences.

This paper considers likelihood-based testing of the null hypothesis of m_0 components against the alternative of $m_0 + 1$ components in a finite mixture model. The number of components is an important parameter in finite mixture models. In economics applications, the number of components often represents the number of unobservable types or abilities. In many other applications, the number of components signifies the number of clusters or latent classes in the data.

Testing the number of components in finite mixture models has been a long-standing challenging problem because of its non-regularity. When testing the null of m_0 components against the alternative of $m_0 + 1$ components, the true m_0-component density can be described with many elements of the parameter space in the $(m_0 + 1)$-component alternative model. These elements are characterized by the union of the two parameter subsets: A, where two components have the same mixing parameter that takes component-specific values; and B, where one of the components has zero mixing proportion. In both null parameter sets, the regularity conditions for a standard asymptotic analysis fail because of such problems as parameter non-identification, singular Fisher information matrix, and true parameter being on the parameter space boundary. When the parameter space is compact, the asymptotic distribution of the likelihood ratio test (LRT) statistic has been derived as a supremum of the square of a Gaussian process indexed by the closure of the convex cone of directional score functions; however, it is difficult to implement these symbolic results.

This paper makes three main contributions. First, we develop a framework that facilitates the analysis of the likelihood function of finite mixture models. In the null parameter space A discussed above, the standard quadratic expansion of the log-likelihood function is not applicable because of the singular Fisher information matrix. The existing works handle this problem by resorting to

non-standard approaches and tedious manipulations. We develop an orthogonal parameterization that extracts the direction in which the Fisher information matrix is singular. Under this reparameterization, the log-likelihood function is locally approximated by a quadratic form of squares and cross-products of the reparameterized parameters, leading to a simple characterization of the asymptotic distribution of the LRT statistic.

Second, we derive the asymptotic distribution of the LRT statistic for testing the null hypothesis of m_0 components for a general $m_0 \geq 1$ in a mixture model with a multidimensional mixing parameter and a structural parameter. Under the null parameter set A, the asymptotic distribution is shown to be the maximum of m_0 random variables, each of which is a projection of a Gaussian random vector on a cone. Both the LRT statistic under the null parameter set B and the (unrestricted) LRT statistic are shown to converge in distribution to the maximum of m_0 random variables, each of which is the supremum of the square of a Gaussian process over the support of the mixing parameter. In contrast to the existing symbolic results, the covariance structure of the Gaussian processes is explicitly presented.

Implementing the LRT has, however, practical difficulties: (i) in some mixture models that are popular in applications (e.g., Weibull duration models), the Fisher information for the null parameter space B is not finite; (ii) the asymptotic distribution depends on the choice of the support of the parameter space, and (iii) simulating the supremum of a Gaussian process is computationally challenging unless the dimension of the parameter space is small.

As our third contribution, building on the EM approach pioneered by Li et al. (2009) and Li and Chen (2010), we develop a likelihood-based testing procedure of the null hypothesis of m_0 components against the alternative of $m_0 + 1$ components that circumvents these difficulties associated with the null parameter space B. The proposed modified EM test statistic has the same asymptotic distribution as the LRT statistic for testing the null parameter space A, and its asymptotic distribution can be simulated without facing the curse of dimensionality. Furthermore, the modified EM test is implementable even if the Fisher information for the null parameter space B is not finite. Simulations show that the modified EM test has good finite sample size and power properties.

References

Li, P. and Chen, J. (2010), "Testing the Order of a Finite Mixture," *Journal of the American Statistical Association*, 105, 1084–1092.

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

Consistent Cross-Validation for Tuning Parameter Selection in High-Dimensional Variable Selection

Yang Feng[†] and Yi Yu[*]

[†]*Department of Statistics, Columbia University, New York NY 10027*

[*]*Statistical Laboratory, University of Cambridge, UK*

In recent years, massive high-throughput and high-dimensional datasets are generated as a result of technological advancement in many fields. Such data are featured by the large number of variables p as compared with the sample size n. Usually, it is assumed that there are only a few important variables that contribute to the response. Variable selection is a natural tool to discover the important variables. We refer to Fan and Lv (2010) for an overview of variable selection in high-dimensional feature space.

The first attempt to variable selection was the ℓ_0 type regularization methods, including AIC (Akaike, 1973), C_p (Mallows, 1973) and BIC (Schwarz, 1978), which work well in low dimensional cases, while computationally prohibitive in high-dimensional settings. As a result, numerous efforts have been made to modify the ℓ_0 type regularization to reduce the computational burden. Among them, Tibshirani (1996) proposed LASSO, which is the ℓ_1 penalty, or equivalently, Chen and Donoho (1994) proposed Basis Pursuit. Also, folded-concave penalties such as SCAD (Fan and Li, 2001) and MCP (Zhang, 2010) have been proposed and widely used over the years. All of these variable selection procedures were shown to have good theoretical results. For LASSO, prediction and selection performances were studied in Greenshtein and Ritov (2004), Meinshausen and Bühlmann (2006), Zhao and Yu (2006), Bunea et al. (2007), Zhang and Huang (2008), Meinshausen and Yu (2009), Bickel et al. (2009), Zhang (2010), among others. For folded-concave penalties, such as SCAD and MCP, their theoretical properties were studied in Fan and Li (2001), Fan and Lv (2011), Zhang (2010) and Feng et al. (2013).

It is very desirable to develop efficient algorithms for calculating the solution path of the coefficient vector as the tuning parameter varies. For LASSO, least angle regression (LARS) (Efron et al., 2004), or homotopy (Osborne et al., 2000) is an efficient method for computing the entire path of LASSO solutions in the linear regression case. For folded-concave penalties including SCAD and MCP, Fan and Li (2001) used the Local Quadratic Approximation (LQA); Zou and Li (2008) proposed the Local Linear Approximation (LLA), which makes a local linear approximation to the penalty, thereby yielding an objective function that can be optimized by using the LARS algorithm; Zhang (2010) proposed the penalized linear unbiased selection algorithm (PLUS), which is designed for linear regression penalized by quadratic spline penalties, including LASSO, SCAD and MCP. Recently, coordinate descent methods have received a majority of attention in high-dimensional setting, including Fu (1998), Shevade and Keerthi (2003), Krishnapuram et al. (2005), Genkin et al. (2007), Friedman et al. (2007), Wu and Lange (2008), among others. Other work on penalized linear regression includes Hastie et al. (2004), Daubechies et al. (2004), Kim et al. (2007), Yu and Feng (2012), among others.

In all the existing algorithms for calculating the solution path for penalized estimators, there is a tuning parameter which controls the penalty level. As it turns out, selecting the optimal tuning parameter is both important and difficult. There has been an abundance of research on using certain kind of information criteria to select the tuning parameter. Tibshirani (1996) used generalized cross-validation (GCV) style statistics, Efron et al. (2004) used C_p style statistics. Zou et al. (2007) derived a consistent estimator of degrees of freedom of the LASSO in the C_p, AIC, and BIC criteria. But from simulation experience, all these traditional methods when applied to the LASSO, tend to over select. Chen and Chen (2008) proposed extended-BIC, by adding an extra term with respect to p into the criterion function. Zhang et al. (2010) proposed the generalized information criterion (GIC), which makes a connection between the classical variable selection criteria and the regularization parameter selection for the nonconcave penalized likelihood approaches.

Besides the many information-type criteria, another popular method for selecting the tuning parameter is cross-validation, which is a data-driven method. Shao (1993) gave rigorous proof of the inconsistency of CV(1) for the classical linear regression model, meanwhile he gave the proper size of construction and validation sets in leave-n_v-out cross-validation (CV(n_v)), under which cross-validation achieves the model selection consistency. Here, by *construction* and *validation* datasets we mean the subsets of the complete dataset used to construct and validate the estimators in cross-validation splits. Zhang (1993) studied multifold cross-validation and r-fold cross-validation in classic linear regression models.

To calculate the solution path for high-dimensional variable selection methods, there are several packages available in R, including lars (Efron et al., 2004), glmnet (Friedman et al., 2007), glmpath (Park and Hastie, 2007), plus(Zhang, 2010), ncvreg (Breheny and Huang, 2011), apple (Yu and Feng, 2012), among others. In all these packages, the default tuning parameter selection method is K-fold cross-validation. Nevertheless, researchers have realized that the K-fold cross-validation in high-dimensional settings tends to be too conservative in the sense that it will select plenty of noise variables. As mentioned in Zhang and Huang (2008), the theoretical justification of cross-validation based tuning parameter is unclear for model-selection purposes.

The contribution of the paper is two-fold. (1) A thorough investigation is conducted for the advantages and drawbacks of the commonly used cross-validation methods for tuning parameter selection in the penalized estimation methods. (2) A new cross-validation method is proposed, which is shown to be model selection consistent for a wide range of penalty functions under the generalized linear model framework.

Interpretation and use of sensitivity in econometrics with an illustration to forecast combinations

Jan R. Magnus

Department of Econometrics & Operations Research,

VU University Amsterdam, The Netherlands

Andrey L. Vasnev

University of Sydney, New South Wales, Australia

6 September 2013

extended abstract for Otaru Symposium

Extended abstract

The majority of applied econometric papers concentrates on the fit of the models and the statistical significance of the coefficients. Sensitivity analysis is often not or only tangentially reported. This is unfortunate, because sensitivity analysis is at least as important as diagnostic testing. While diagnostic testing attempts to answer the question: is it true (for example, that a coefficient is zero), sensitivity analysis addresses the question: does it matter (that we set the coefficient to zero). At first glance, the two questions seem to be closely related. But Magnus & Vasnev (2007) showed that this is not the case. In fact, the two concepts are essentially orthogonal.

Magnus & Vasnev (2007) introduced local sensitivity through a Taylor expansion. If the variable (or parameter) of interest, say y, depends on a nuisance parameter, say θ, then $\hat{y}(\theta)$ denotes the estimator of y for each given value of θ. Special cases are the 'restricted' estimator $\hat{y}(0)$ obtained by setting $\theta = 0$, and the 'unrestricted' estimator $\hat{y}(\hat{\theta})$ obtained by setting θ equal to its estimated value $\hat{\theta}$. The function $\hat{y}(\theta)$ provides not only these two special cases, but the whole *sensitivity curve*, given by the estimates of y for each given value of θ.

The first-order Taylor expansion of the sensitivity curve at the restricted point is given by

$$\hat{y}(\theta) = \hat{y}(0) + S\,\theta + O(\theta^2), \tag{1}$$

where

$$S = \left. \frac{\partial \hat{y}(\theta)}{\partial \theta} \right|_{\theta=0} \tag{2}$$

is the first derivative at the restricted point $\theta = 0$, and is called the *local sensitivity statistic* or simply the sensitivity.

Sensitivity is computed for maximum likelihood estimators in Magnus & Vasnev (2007) and, in general, it can be expressed in terms of the Hessian. For the cases of mean, variance, and distribution misspecification the sensitivity statistics allow tractable representations. This is particularly the case for the B_s and D_s statistics of Banerjee & Magnus (1999) and the sensitivity of GLS estimators in panel data derived by Vasnev (2010).

Magnus & Vasnev (2007) provide an overview of the sensitivity literature, and prove formally the asymptotic independence of the commonly-used diagnostic tests and the sensitivity statistic. Diagnostic tests and sensitivity statistics are therefore complementary, and both require our attention when analyzing a model. It is possible to derive sensitivity statistics, and several papers have suggested local or global sensitivity measures. It is, however, more difficult to answer the question when a sensitivity statistic is large or small. This question is addressed in the current paper. The paper gives practical recommendations with regards to how sensitivity statistics can be used. We shall see that the use of sensitivity is context-dependent, as also emphasized by Severini (1996), so that we need to consider sensitivity in relation to the problem under consideration.

In some situations the value of the sensitivity statistic is important, requiring a threshold in order to decide whether the model is sensitive or not. We call this case 'absolute sensitivity'. In other situations only the relative magnitude is important. We call this case 'relative sensitivity'. Essential for both cases is the realization that sensitivity (unlike a diagnostic test) is context-dependent, and will be closely related to the estimator we analyze or the dependent variable we are modeling. To bring out this dependence, we illustrate all concepts introduced in this paper in a specific context, namely forecasting the Euro yield curve. In this context it is natural to consider the sensitivity to autocorrelation and normality assumptions. Different forecasting models are combined with equal, fit-based and sensitivity-based weights, and compared with the multivariate and random walk benchmarks.

The main purpose of combining forecasts is to improve forecast accuracy (Bates & Granger, 1969). The choice of weights, however, is still an open question. Timmermann (2006) provides a thorough overview of the sizeable forecast combination literature, but in practice the optimal weights have to be estimated and this affects their actual performance. The adaptive weights seem to work well in many situations, but sometimes a simple alternative with equal weights gives better results as shown by Stock & Watson (2004). This fact is explained by Winkler & Clemen (1992) as instability of estimated weights used in generating the combined forecast. We show that when several forecasts are available, the weights based on relative sensitivity perform well and are complementary to the fit-based weights. For long-term maturities the sensitivity-based weights perform better than other weights.

References

Banerjee, A. N., & Magnus, J. R. (1999). The sensitivity of OLS when the variance matrix is (partially) unknown. *Journal of Econometrics, 92*, 295–323.

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly, 20*, 451–468.

Magnus, J. R., & Vasnev, A. L. (2007). Local sensitivity and diagnostic tests. *Econometrics Journal, 10*, 166–192.

Severini, T. A. (1996). Measures of the sensitivity of regression estimates to the choice of estimator. *Journal of the American Statistical Association, 91* (436), 1651–1658.

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting, 23*, 405–430.

Timmermann, A. (2006). Forecast combinations. In: Elliott, G., Granger, C. W. J., & Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, (pp. 135–196). Elsevier.

Vasnev, A. L. (2010). Sensitivity of GLS estimators in random effects models. *Journal of Multivariate Analysis, 101*, 1252–1262.

Winkler, R. L., & Clemen, R. T. (1992). Sensitivity of weights in combining forecasts. *Operations Research, 40*, 609–614.

Intertemporal Relation between Risk and Return: Panel Quantile Regression Approach

[Very Preliminary Draft]

Jau-er Chen and Kuei-Fu Lin

Department of Economics
National Taiwan University

This Version: August, 2013

† ©Jau-er Chen. E-mail: jauer@ntu.edu.tw

Address: 21, Hsu-Chow Road, Taipei, Taiwan

Acknowledgments: We are grateful to Heather Anderson, Jaroslav Borovička, and Riccardo Colacito for valuable discussions and suggestions. We also thank Stacey H. Chen, Keng-Yu Ho, Yu-Chin Hsu, Kamhon Kan, Shyan-Yuan Lee, Hsin-Yi Lin, Shih-Chang Lin, Jen-Che Liao, Hui-Chen Wang, and Tsong-Min Wu for comments. This paper has benefited from presentations at NYU Economics Ph.D. Alumni Conference, Academia Sinica, National Chung Hsing University, and National Tsing Hua University. The usual disclaimer applies.

Intertemporal Relation between Risk and Return: Panel Quantile Regression Approach

[Very Preliminary Draft]

Jau-er Chen and Kuei-Fu Lin

National Taiwan University

Abstract

This paper explores the Taiwan stock market's intertemporal relation between risks and expected returns in the context of the Intertemporal Capital Asset Pricing Model. Our panel data models primarily rely on the time-varying conditional covariances among the return of Taiwan 50 Index (market portfolio), each corresponding component stock return, and the state variables including VIX, term spread, and funding liquidity spread. Specifically, the following two-stage econometric procedure is implemented: we first estimate the time-varying conditional covariances by dynamic conditional correlations models, and then treat the estimates as explanatory variables in the second-stage panel quantile regression (PQR) methods to explore the shape of conditional distribution of excess returns. The risk coefficients estimated via PQR are positive over the upper right tail of the conditional distribution of excess returns; the estimation results signify negative risk coefficients over the lower left tail of conditional distribution of excess returns. No significant intertemporal relation between risk and return is identified over the neighborhood of conditional median of excess returns. Robustness checks indicate that our empirical results are robust to the choice of proxies of risk, explanatory variables, and econometric methodologies.

JEL classification: G12, C32, C33

Keywords: Intertemporal relation between risk and expected return; Intertemporal Capital Asset Pricing Model; Dynamic Conditional Correlations; Panel Quantile Regression.

多次元非正規時系列モデルの判別解析に対する経験尤度アプローチ

明石　郁哉
（早稲田大学 大学院 基幹理工学研究科）

2013 年 9 月 7 日

1　時系列モデルに対する判別解析

本報告では、時系列モデルの判別解析に対して、経験尤度法を用いた新たなアプローチを提案し、従来の判別解析に対する改善点を報告する。

あるベクトル値観測系列 $\boldsymbol{X} = (\boldsymbol{X}(1)', \cdots, \boldsymbol{X}(n)')'$ が得られた際に、この観測系列が複数のカテゴリー $\Pi_1, \cdots, \Pi_J$ のうちのどれか 1 つに属することだけが分かっているとする。判別解析では、$\boldsymbol{X}$ が属する正しいカテゴリーをより高い確率で見出すことを目的とする。一般的に時系列分野における判別解析では、データの属するカテゴリーは時系列モデルのスペクトル密度行列で非母数的に規定される。具体的な設定として、長さ n の s 次元時系列データ $\boldsymbol{X} = (\boldsymbol{X}(1)', \cdots, \boldsymbol{X}(n)')'$ が得られた際に、$\boldsymbol{X}$ が以下のような $s \times s$ スペクトル密度行列で記述される 2 つのカテゴリー

$$\Pi_1 : \boldsymbol{g}_1(\omega) \qquad \Pi_2 : \boldsymbol{g}_2(\omega) \tag{1}$$

のどちらから生成されたかを判別する問題を考える。この形の判別問題に対して、Zhang and Taniguchi [1] は、2 次定常な多次元非正規線形過程から生成されたデータ $\boldsymbol{X} = (\boldsymbol{X}(1)', \cdots, \boldsymbol{X}(n)')'$ に対して、Whittle 尤度比型の判別統計量

$$I(\boldsymbol{g}_1 : \boldsymbol{g}_2) := \frac{1}{4\pi} \int_{-\pi}^{\pi} \left[\log \frac{\det \boldsymbol{g}_2(\omega)}{\det \boldsymbol{g}_1(\omega)} + \mathrm{tr} \left[\boldsymbol{I}_{n,X}(\omega) \left\{ \boldsymbol{g}_2(\omega)^{-1} - \boldsymbol{g}_1(\omega)^{-1} \right\} \right] \right] d\omega$$

を提案した。ここで、$\boldsymbol{I}_{n,X}(\omega)$ は観測系列に対するピリオドグラム行列である。

判別統計量の良さを評価する際には、誤判別確率 $\Pr(2|1)$ を評価するのが自然であると考えられる。ここで、$\Pr(2|1)$ は、観測系列は本来カテゴリー Π_1 に属するにもかかわらず、Π_2 に属していると誤って判別されてしまう確率である。Zhang and Taniguchi [1] は、判別問題 (1) のもとで $\Pr(2|1)$、$\Pr(1|2)$ が漸近的に 0 に収束することを示した。このことは、判別統計量 $I(\boldsymbol{g}_1 : \boldsymbol{g}_2)$ が基本的な「良さ」を持っていることを意味している。また、カテゴリーを記述するスペクトルが母数的に近接している場合において誤判別確率を導出することにより、判別統計量の微妙な良さを評価した。Zhang and Taniguchi [1] は近接したカテゴリーの下では、$I(\boldsymbol{g}_1 : \boldsymbol{g}_2)$ による誤判別確率は漸近的に 0 に収束しないことを示し、その確率について明示的な評価を行った。

2　経験尤度に基づく判別統計量の提案

先に述べたように、従来の判別解析はカテゴリーを規定するスペクトル密度関数を用いて、近似的な尤度比を評価することによって行われてきた。。本研究では新たなアプローチとして、時系列モデルの重要指標によって規定されるカテゴリーに対する判別解析を展開する。$\{\boldsymbol{X}(t);\ t \in \mathbb{Z}\}$ という s 次元非正規線形過程が、ある重要指標 $\boldsymbol{\theta}$（p 次元ベクトル）で規定される 2 つのカテゴリー

$$\Pi_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1 \qquad \Pi_2 : \boldsymbol{\theta} = \boldsymbol{\theta}_2$$

のうち、どちらに属するかを判別する問題を考える。重要指標 $\boldsymbol{\theta}$ は、解析者が定める適当なスコア関数 $\boldsymbol{f}(\omega;\boldsymbol{\theta})$ ($s \times s$ 行列) とともに、

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{-\pi}^{\pi} \mathrm{tr}[\boldsymbol{f}(\omega;\boldsymbol{\theta})^{-1}\boldsymbol{g}(\omega)]d\omega = \boldsymbol{0}, \quad \boldsymbol{g}(\omega): \text{ 関与のモデルのスペクトル密度行列}$$

という等式を満たす意味で、カテゴリーを規定しているものとする。$\boldsymbol{f}(\omega;\boldsymbol{\theta})$ のとり方により、$\boldsymbol{\theta}$ として自己相関係数や最良線形予測子の係数などを取ることが可能であり、本質的に $\boldsymbol{g}(\omega)$ は未知、かつ非母数型でよいという利点がある。この仮説に対し、本研究では経験尤度比に基づく判別統計量

$$ELR(\boldsymbol{\theta}_1 : \boldsymbol{\theta}_2) := \frac{2}{n} \log \frac{R(\boldsymbol{\theta}_1)}{R(\boldsymbol{\theta}_2)}$$

を提案する。ただし、$R(\boldsymbol{\theta}_j)$ は、仮説 Π_j のもとでの経験尤度比と呼ばれるもので、

$$R(\boldsymbol{\theta}) := \max_{w_1,\cdots,w_n} \left\{ \prod_{t=1}^{n} n w_t \ ;\ \sum_{t=1}^{n} w_t \boldsymbol{m}(\lambda_t;\boldsymbol{\theta}) = \boldsymbol{0},\ \sum_{t=1}^{n} w_t = 1,\ 0 \le w_t \le 1 \right\}$$

と定義される。本報告では、この判別統計量による判別は漸近的に誤判別確率を 0 とすることを示す。これは判別統計量 $ELR(\boldsymbol{\theta}_1 : \boldsymbol{\theta}_2)$ が基本的な良さを持っているということを表している。また、さらに微妙な良さの指標として、カテゴリーを記述する重要指標が

$$\Pi_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1 \qquad \Pi_2 : \boldsymbol{\theta} = \boldsymbol{\theta}_1 + \frac{1}{\sqrt{n}}\boldsymbol{h}$$

の形で近接している場合の誤判別確率 $\Pr(2|1)$ 及び $\Pr(1|2)$ を明示的に評価する。

経験尤度統計量の特徴として、スコア関数 $\boldsymbol{f}(\omega;\boldsymbol{\theta})$ 及び重要指標 $\boldsymbol{\theta}_0$ は、適当な正則条件を満たしさえすれば解析者が任意に定められるということが挙げらる。また報告者は、近接条件の下で、経験尤度統計量によるアプローチをとった場合、漸近的な誤判別確率が Whittle 尤度比型の統計量によるものよりも小さくなる場合があることを観測した。さらに、非正規ロバスト性という観点から見ても、経験尤度統計量は優れた性質を持つことが示される。当日は数値例を用いて、これらの結果を紹介する。

参考文献

[1] Zhang, G and Taniguchi, M. (1994). Discriminant analysis for stationary vector time series. *Journal of Time Series Analysis*, **15** 117-126.

[2] Ogata, H. and Taniguchi, M. (2010) An empirical likelihood approach for non-Gaussian vector stationary processes and its application to minimum contrast estimation. *Australian & New Zealand Journal of Statistics*, 52 (4) 451-468.

Consideration on a serial correlation

Hiroaki Ogata

Waseda University
hiroakiogata@aoni.waseda.jp

1 Stationary process with Spearman's rho

Let (X, Y) be a random vector with a joint distribution function $F_{(X,Y)}(\cdot, \cdot)$, a copula function $C_{(X,Y)}(\cdot, \cdot)$ and marginal distribution functions $F_X(\cdot)$ and $F_Y(\cdot)$. *Spearman's rho* is one of the well known measures of association between X and Y, and defined as:

$$\rho(X, Y) = 3\left(P[(X_1 - X_2)(Y_1 - Y_3) > 0] - P[(X_1 - X_2)(Y_1 - Y_3) < 0]\right) \tag{1}$$

where $(X_1, Y_1), (X_2, Y_2)$ and (X_3, Y_3) are independent copies of (X, Y) (Kruskal (1958)). We can reexpress (1) as

$$\rho(X, Y) = 12 \int_{u=0}^{1} \int_{v=0}^{1} C_{(X,Y)}(u, v) du dv - 3$$

or

$$\rho(X, Y) = 12 \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} F_{(X,Y)}(x, y) F_X(dx) F_Y(dy) - 3$$

Definition 1 For given $r, s \in \mathbb{Z}$, the *autorho function* $\rho_Y(\cdot, \cdot)$ of the stochastic process $\{Y_t, t \in \mathbb{Z}\}$ is defined by

$$\rho_Y(r, s) = \rho(Y_r, Y_s)$$

Definition 2 The time series $\{Y_t, t \in \mathbb{Z}\}$ is said to be *rho stationary* if

(i) $\mathrm{Med}(Y_t) = \tilde{\mu}$ for all $t \in \mathbb{Z}$

(ii) $\rho_Y(r, s) = \rho_Y(r + t, s + t)$ for all $r, s, t \in \mathbb{Z}$

Remark 1 If $\{Y_t, t \in \mathbb{Z}\}$ is rho-stationary process, we redefine the autorho function as

$$\rho_Y(k) \equiv \rho_Y(k, 0) = \rho_Y(t + k, t)$$

2 Autorho function of linear processes

We consider a serial correlation of time series data with Spearman's rho. Let $U_1, U_2, \ldots$ be a sequence of i.i.d. random variables with distribution function $F_U(\cdot)$ and density function $f_U(\cdot)$. Consider the linear process

$$Y_t = \sum_{i=0}^{\infty} \theta_j U_{t-i} \tag{2}$$

where $\theta_0 = 1$ and $\sum_{i=0}^{\infty} |\theta_i|^2 < \infty$.

Theorem 1 *The linear process defined by (2) is rho stationary. The autorho function at lag k is given by*

$$\rho_Y(k) = 12 \iint_{\mathbb{R}^2} F_k(y_0, y_k) f_Y(y_0) f_Y(y_k)\, dy_0 dy_k - 3$$

where

$$F_k(y_0, y_k) = \begin{cases} \displaystyle\int_{\mathbb{R}^{\mathbb{N}}} F_U\left(\min\left\{\frac{y_0}{\theta_k} - \sum_{i=0, i\neq k}^{\infty} \frac{\theta_i}{\theta_k} u_i,\ y_k - \sum_{i=1}^{\infty} \theta_i u_{k+i}\right\}\right) \\ \displaystyle\quad \times \prod_{i=0, i\neq k}^{\infty} f_U(u_i)\, du_i \qquad (\theta_k > 0) \\ \displaystyle\int_{\mathbb{R}^{\mathbb{N}}} F_U\left(y_k - \sum_{i=1}^{\infty} \theta_i u_{k+i}\right) F_U\left(y_0 - \sum_{i=1, i\neq k}^{\infty} \theta_i u_i\right) \\ \displaystyle\quad \times \prod_{i=0, i\neq k}^{\infty} f_U(u_i)\, du_i \qquad (\theta_k = 0) \\ \displaystyle\int_{\mathbb{R}^{\mathbb{N}}} \min\left\{F_U\left(y_k - \sum_{i=1}^{\infty} \theta_i u_{k+i}\right) - F_U\left(\frac{y_0}{\theta_k} - \sum_{i=0, i\neq k}^{\infty} \frac{\theta_i}{\theta_k} u_i\right), 0\right\} \\ \displaystyle\quad \times \prod_{i=0, i\neq k}^{\infty} f_U(u_i)\, du_i \qquad (\theta_k < 0) \end{cases}$$

and

$$f_Y(y) = \int_{\mathbb{R}^{\mathbb{N}}} f_U\left(y - \sum_{i=1}^{\infty} \theta_i u_i\right) \prod_{i=1}^{\infty} f_U(u_i)\, du_i.$$

References

[1] Kruskal, W.H. (1958): Ordinal measures of association. *J. Amer. Statist, Assoc.* **53**, 814–861.

Identification of Approximate Factor Models through Heteroskedasticity

Shinya Tanaka[1]

JSPS Research Fellow

September 7, 2013

Summary

This paper investigates the identification problem of factor models and proposes a new identification scheme for large-dimensional factor models through heteroskedasiticity of factors.

It is well known that factor models have a serious problem in estimating factors: the "rotation problem." Consider the factor model

$$x_t = \Lambda^0 F_t^0 + \varepsilon_t, \qquad t = 1, 2, \ldots, T,$$

where x_t is an $N \times 1$ observation vector, Λ^0 is an $N \times r$ factor loading matrix, F_t^0 is a stationary $r \times 1$ latent factor and ε_t is an $N \times 1$ idiosyncratic error vector. Note that 0 indicates the true values. Let H^0 be a non-singular $r \times r$ matrix and insert $(H^0)^{-1}H^0$ $(= I_r)$ between Λ^0 and F_t^0 in this model. We thus obtain

$$\begin{aligned} x_t &= \Lambda^0 (H^0)^{-1} H^0 F_t^0 + \varepsilon_t, \\ &\equiv \mathcal{L}^0 \mathcal{F}_t^0 + \varepsilon_t, \end{aligned}$$

where $\mathcal{L}^0 = \Lambda^0 (H^0)^{-1}$ and $\mathcal{F}_t^0 = H^0 F_t^0$. Then, it is obvious that we cannot distinguish $\left(\Lambda^0, F_t^0\right)$ from $\left(\mathcal{L}^0, \mathcal{F}_t^0\right)$ without imposing additional assumptions on the model. That is, we cannot consistently estimate (true) Λ^0 and F_t^0 without imposing any identifying restrictions on the factors and factor loadings in general. We call H^0, $\mathcal{L}^0$ and $\mathcal{F}_t^0$ rotation matrix, rotated factor loadings (matrix) and rotated factors, respectively.

The main assumptions of our model are that the sample period is divided into two regimes and that the variance of the factors changes depending on the regime while the factor loadings are invariant through

[1]Takasaki City University of Economics, 1300 Kaminamie-machi, Takasaki, Gunma, Japan. E-mail: stanaka@wd5.so-net.ne.jp.

regimes. Let S_1 and S_2 be the first and second halves of all sample periods, respectively. Then the process of factors is assumed as follows:

$$F_t^0 = \Phi_1^0(L)F_{t-1}^0 + e_t, \quad E[e_t e_t'] = I_r \qquad t \in S_1, \qquad\qquad F_t^0 = \Phi_2^0(L)F_{t-1}^0 + e_t, \quad E[e_t e_t'] = \Sigma_2^{e0} \qquad t \in S_2,$$

where $\Sigma_2^{e0} = diag(\sigma_{10}^2, \sigma_{20}^2, \ldots, \sigma_{r0}^2)$. On the other hand, this paper assumes that the factor model remains $x_t = \Lambda^0 F_t^0 + \varepsilon_t$ in both regimes. Although it seems that the invariance of factor loadings is unrealistic given that the process of factors varies depending on regimes, we see in this paper that the Great Moderation is one example of this assumption, so it is not an unrealistic one.

Based on the assumptions and asymptotic properties of the PCA estimator, we construct identifying restrictions that link reduced form parameters with structural parameters. Let $\tilde{F}^s$ and $\tilde{\Lambda}^s = [\tilde{\lambda}_{1,s}, \tilde{\lambda}_{2,s}, \ldots, \tilde{\lambda}_{N,s}]'$ be the PCA estimators of $F^{0,s}$ and Λ, respectively, where $F^{0,s}$ is a set of the true factors in S_s. Furthermore, assume that H_s^0 is a rotation matrix in S_s. Since $\tilde{\lambda}_{i,s} - (H_s^{0\prime})^{-1}\lambda_i^0 \xrightarrow{p} 0$ holds for each i and s in a standard approximate factor model, essentially we have the restrictions[2]

$$\tilde{\Lambda}^1 - \Lambda(H_1^0)^{-1} = 0, \qquad\qquad \tilde{\Lambda}^2 - \Lambda(H_2^0)^{-1} = 0.$$

Note that the first terms are obtained from the data and the second terms are structural parameters. Then, we can estimate the structural parameters by minimum distance estimation and obtain the estimators of rotation matrices H_1^0 and H_2^0. Since $\widetilde{F}_t^s - H_s^0 F_t^0 \xrightarrow{p} 0$ holds for $s = 1, 2$ from Bai (2003), we yield the estimator of true factors by

$$\widehat{F}_t = \begin{cases} (\hat{H}_1)^{-1}\,\widetilde{F}_t & \text{if } t \in S_1 \\ (\hat{H}_2)^{-1}\,\widetilde{F}_t & \text{if } t \in S_2 \end{cases},$$

where $\hat{H}_1$ and $\hat{H}_2$ are the minimum distance estimators of H_1^0 and H_2^0 respectively. It should be noted that since the total number of restrictions depends on N and $N \to \infty$ is assumed in standard approximate factor models, the asymptotic properties of the minimum distance estimator are not trivial. This paper investigates the asymptotic properties of the minimum distance estimators and derives the consistency of $\hat{H}_1$ and $\hat{H}_2$.

Monte Carlo simulations confirm that our identification scheme works well and give encouraging evidences that we can precisely estimate the true factors with our estimator as expected.

[2] See the paper for details of more restrictions.