平成２３年度科学研究費基盤研究　(A)
研究課題番号　: 23244011
研究課題名　　: 非対称・非線形統計理論と経済・生体科学への応用
研究代表者　　: 谷口　正信（早稲田大学）

経験尤度法と判別・分類解析の理論と応用

2011 年 12 月 2 日～4 日

和歌山ビッグ愛　８０２会議室

Theory and Applications for Empirical Likelihood and Discriminant and Cluster Analysis

2-4 Dec 2011

Wakayama Symposium

＜開催責任者＞
天野　友之（和歌山大学）
松田　忠之（和歌山大学）
Cheng, M. Y. (国立台湾大学)
Chen, C. W. S. (台中大学)

＜Local Organizer＞
Tomoyuki Amano (Wakayama Univ.)
Tadayuki Matsuda(Wakayama Univ.)
Cheng, M. Y.(National Taiwan Univ.)
Chen, C. W. S.(Feng Chia Univ.)

Program

<2 Dec (Fri)>

*10:30～13:00 (Pre Discussion)

*13:00～13:55 (Lunch)

*13:55～14:00 (The opening of a meeting)

-------------------------------**[Chair]JunichiHirukawa(NiigataUniv.)**---------------------

*14:00～14:35 Shuichi Shinmura (Seikei University)

Fisher 以降の判別分析の新世界

*14:35～15:10 Mitsuru Tamatani (Shimane University)

Kanta Naito (Shimane University)

高次元小標本における正準相関に基づく多群判別

*15:10～15:45 Mei Saito (Tokyo University of Science)

Takuma Sumikawa (Tokyo University of Science)

Kazuyuki Koizumi (Tokyo University of Science)

Takashi Seo (Tokyo University of Science)

On some tests for assessing multivariate normality based on sample skewness and kurtosis

*15:45～15:55 (Coffee break)

-------------------------------**[Chair]HiroakiOgata(WasedaUniv.)**--------------------------

*15:55～16:30 Tomoyuki Amano (Wakayama University)

Empirical likelihood approach to discriminant analysis for stationary processes

*16:30～17:05 Alexandre Petkovic (Waseda University)

Double unit root process and non stationary variance

-------------------------------**[Chair]TomoyukiAmano(WakayamaUniv.)**-----------------

*17:05～17:40 Kenta Hamada (Waseda University)

Masanobu Taniguchi (Waseda University)

Multistep ahead portfolio estimation for dependent return processes

*17:40～18:15 Hiroaki Ogata (Waseda University)

Optimal portfolio with generalized empirical likelihood

*19:00～ (Dinner)

<3 Dec (Sat)>

----------------------[Chair]MarcHallin(Universite'　libre de Bruxelles)-----------

*10:00～10:35　Yoshihide Kakizawa (Hokkaido University)

Bartlett correctability of empirical likelihood ratio test
for a parameter subvector in the over-identified case

*10:35～11:10　Junichi Hirukawa　(Niigata University)
Kentaro Kobayashi (Niigata University)

Asymptotic properties of time series non-life insurance model

*11:10～12:00　Siegfried Hormann (Université libre de Bruxelles)
R. Gabrys (University of Southern California)
P. Kokoszka (University of Utah)
L. Horvath (University of Utah)
R. Reeder (University of Utah)

Using FDA techniques for econometric time series

*12:00～13:30　(Lunch)

-------------------------[Chair]TomoyukiAmano(WakayamaUniv.)-------------------

*13:30～14:20　David Veredas　(Université libre de Bruxelles)
Roxana Halbleib (University of Konstanz)
Matteo Barigozzi (London School of Economics)

Which models to match?

*14:20～15:10　Cathy Chen (Feng Chia University)

Classification in segmented regression problems

*15:10～16:00　Ming-Yen Cheng (National Taiwan University)
Hau-Tieng Wu　(Princeton University)

Local Linear Regression on Manifolds and its Geometric Interpretation

*16:00～16:20　(Coffee Break)

-----------------------------[Chair]HiroshiShiraishi(JikeiUniv.)--------------------

*16:20～17:10　Ngai Hang Chan (Chinese University of Hong Kong)

Statistical Arbitrage and Fractional Cointegration

*17:10～18:00　Marc Hallin (Université libre de Bruxelles)

Skew-symmetric Families and Degenerate Fisher Information

<4 Dec (Sun)>

--------------------------[Chair]YoshihideKakizawa(HokkaidoUniv.)------------------

*10:00～10:35　Toshihiro Hirano　(Tokyo University)

Yoshihiro Yajima　(Tokyo University)

Covariance Tapering for Prediction of Large Spatial Data Sets in Transformed Random Fields

*10:35～11:10　Hideatsu Tsukahara　(Seijo University)

Asymptotics of L-statistics with dependent data and their applications to risk measure estimation

*11:10～11:20　(Coffee Break)

------------------------------[Chair]YoshihiroYajima(TokyoUniv.)--------------------

*11:20～11:55　Ryozo Miura　(Hitotsubashi University)

Asymptotic Normality of Estimators derived from Rank Statistics for Generalized Lehmann's Alternative Models when the Observations are a sequence of weakly dependent random variables: a General Model and special cases including Skew Symmetric models

*11:55～12:30　Takeaki Kariya　(Meiji University)

Empirically Effective Bond Pricing Model and Analysis on Term Structures of Implied Interest Rates in Financial Crisis

*12:30～12:35　(The closing of a meeting)

Fisher 以降の判別分析の新世界

成蹊大学　新村秀一

1936 年に Fisher は，説明変数 X_1，…，X_p の線形結合（$Z=a_1*X_1+\cdots+a_p*X_p$）を Z 軸上へ射影を行い，群間分散と群内分散の分散比の最大化で，判別係数 a_1，…，a_p を決定する線形判別関数を提案した．この定式化によれば，様々な分野の問題に 2 群の多変量正規性や等分散共分散性の仮定（「フィッシャーの仮説」）を満たさなくとも線形判別関数（Linear Discriminant Function,LDF）の利用が正当化される．その後，「フィッシャーの仮説」から簡単に同じ LDF が得られ，近年この定式化で考えることが多い．その後 70 年以上にわたり，2 次判別関数(Quadratic Discriminant Function,QDF)，数量化Ⅱ類，名義ロジスティック回帰分析，決定木分析といった豊饒な統計的判別分析の理論が構築された．

一方，郵便番号読み取り装置に見るパターン認識や，数理計画法の分野でも判別分析の研究が行われている．そして，パターン認識，医学診断，経済と経営分野の各種の格付け，最近ではゲノム診断の分野で広く応用されている．

しかし，判別分析には次のような 4 つの問題があり，これが誤分類数最小化(Minimum Number of Misclassifications, MNM) 基準による最適線形判別関数(Optimal Linear Discriminant Function, OLDF)ですべて解決できた(Shinmura(2011),新村(2010))．

(1)　線形判別関数 $f(\mathbf{x})$ が $f(\mathbf{x})>0$ であれば群 1，$f(\mathbf{x})<0$ であれば群 2 と判別するが，判別得点が $f(\mathbf{x})=0$ の判別超平面上のケースの帰属が判定不能なまま放置されてきた．この問題が解決できないと正しい誤分類数が分からない．この問題を正しく処理できるのは，最適線形判別関数の手法の一つである改定 IP-OLDF だけである．

(2)　現実のデータは「Fisher の仮説」を満たすものは少ない．この点は判別分析の研究者も当初から意識しており，判別境界を動かすことで誤分類数が少なくなる事実を「正規性からの乖離」といってきた．それを解決するため，2 次判別関数やマハラノビスの汎距離による多群判別，さらに田口の MT 法が提案された．仮にデータがこの仮説を満たせば，LDF の誤分類数は最小の MNM になる．すなわち，最初から MNM 基準で定式化すればよかったが，MNM 基準による最適線形判別関数は，現時点では整数計画法(Integer Programming,IP)による改定 IP-OLDF(Revised Optimal Linear Discriminant Function by IP)による最適線形判別関数でしか実現できない．最近では，医療分野や金融分野で名義ロジスティック回帰が用いられている．これは経験的に LDF や 2 次判別関数に比べて，名義ロジスティック回帰の誤分類数（誤分類確率）が少ない点である．これは，名義ロジスティック回帰が正規分布を仮定せず，分析データに依存していることが理由と考えられる．

(3)　LDF や 2 次判別関数は，正規分布を仮定しているにもかかわらず，誤分類確率や判別係数の信頼区間がわかっていなかった．LDF や 2 次判別関数は，推測統計学とは無縁の手法であり，2 群が正規分布であることを仮定する必要がなかった．

(4)　誤分類数と判別係数の関係が不明であったが，IP-OLDF でこの関係が分かった(新村(1998),新村他(1999))．

最適線形判別関数の研究で，以下のことが分かった．

スイス銀行紙幣データ(Flury et al.(1988))は，6 個の計測値で真札と偽札を 2 群判別する問題である．新村は初めてこのデータが 2 個の計測値 *{X4, X6}* で *MNM=0* であり，MNM の単調減少性（*p* 変数の最適線形判別関数とそれに任意の 1 個の説明変数を加えた最適線形判別関数で，$MNM_p \geqq MNM_{(p+1)}$ の関係がある）から，この 2 変数を含むすべての判別関数で *MNM=0* であることを発見した．

しかし，LDF や 2 次判別関数は，線形分離可能(線形超平面で 2 群が判別可能，すなわち *MNM=0*)なデータを一般的に認識できない．また，名義ロジスティック回帰は，線形分離可能なデータで必ず回帰係数の推定が不安定になり，95%信頼区間は 0 を含む大きな区間幅をもつ(Firth(1993)）．さらに，逐次変数選択法，AIC，Cp 統計量は，線形分離可能な判別係数の空間より，より高次の空間を選ぶという重大な瑕疵があることが分かった(新村(2007a))．

試験の合否判定は，自明な *MNM=0* の判別データを提供してくれる．筆者は，2010 年度と 2011 年度の成蹊大学経済学部の 1 年次生を対象とした必修科目の統計入門を担当し，10 択 100 問の中間試験と期末試験を行った．この判別データを用いて，LDF，2 次判別関数，名義ロジスティック回帰が最小の説明変数で合否判定できないことを示す．また，線形分離可能な最小な説明変数（設問）の情報が，試験の出題の質を説明できるか否かを検証した(新村(2011))．

その成果を受けて 2011 年に大学入試センターから 4 年間 15 教科 105 個の試験データの提供を受けて実証研究を行うことにした．本研究では，大学入試センターの 105 個の試験データの実証研究を行うに際し，2010 年度と 2011 年度の統計入門のデータで，合否判定における判別分析の問題点と試験問題の質保証に関する方法論の開発を行った．この結果に基づいて，大学入試センター試験の分析を行う予定である．

文献

Firth,D.(1993). Bias reduction of maximum likelihood estimates, Biometrika, 80, 27-38.

Flury, B. & Rieduyl, H. (1988). Multivariate statistics : A Practical Approach, Cambridge University Press.

Shinmura,S. (2011).Beyond Fisher’s Linear Discriminant Analysis -New World of Discriminant Analysis -, ISI2011 Proceedings, 1-6.

新村秀一(1998).数理計画法を用いた最適線形判別関数,計算機統計学,11/2,89-101.

新村秀一,垂水共之(1999). 2 変量正規乱数データによる IP-OLDF の評価,計算機統計学 12/2,107-123.

新村秀一（2007a）. 数理計画法による判別分析の 10 年,計算機統計学, 20/1&2, 59-94.

新村秀一(2007b).改定 IP-OLDF による IP-OLDF の問題点の解消,計算機統計学,19/1,1-16.

新村秀一（2009）. 線形計画法による改定 IP-OLDF の計算時間の改善,計算機統計学, 1/21,37-57.

新村秀一(2010）.最適線形判別関数,日科技連出版社.

新村秀一(2011).合否判定データによる判別分析の問題点,応用統計学, 40/3, 157-172.

高次元小標本における正準相関に基づく多群判別

玉谷　充[1]，　内藤　貫太[2]
[1] 島根大学大学院総合理工学研究科　　[2] 島根大学総合理工学部

1　はじめに

判別分析の研究の起源は Fisher の群内と群間の変動の比を最大にする手法から始まり，近年の研究では，高次元小標本における判別手法の構築およびその理論評価が行われている．Bickel and Levina (2004) は，高次元における Fisher の判別式の性能が不十分であることを示し，Fan and Fan (2008) は，高次元小標本における特徴選択手法を提案した．

本講演では，高次元小標本における多群判別関数を正準相関の観点から構築し，多群へと拡張したときに得られる漸近評価及び数値的結果を報告した．

2　設定

次元を d とし，データ数を n とし，k クラスの判別問題について考える．このときに得られるクラス $\ell \in \{1,\ldots,k\}$ の n_ℓ 個の観測値ベクトルを $\boldsymbol{X}_{\ell i} = (X_{\ell i1},\ldots,X_{\ell id})$，$i \in \{1,\ldots,n_\ell\}$ とする．各観測値ベクトルは $\boldsymbol{X}_{\ell i} = \boldsymbol{\mu}_\ell + \boldsymbol{\varepsilon}_{\ell i}$ を満たすと仮定する．ただし，$\boldsymbol{\mu}_\ell \in \mathbb{R}^d$ は平均ベクトルで $\boldsymbol{\varepsilon}_{\ell i} \sim N_d(\boldsymbol{0},\Sigma)$ は誤差ベクトルである．

ここでは，高次元小標本の設定を考えていることから，$n = o(d)$ とする．

3　パターン認識手法

3.1　Fisher の線形判別関数と naive Bayes

2 群の判別問題について考える．等分散性を仮定している場合は，Fisher の線形判別関数

$$\delta_{\mathrm{F}}(\boldsymbol{x}) = \left(\boldsymbol{x} - \frac{\boldsymbol{\mu}_1 + \boldsymbol{\mu}_0}{2}\right)^T \Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$$

がベイズルールとなり最適であることが知られている．しかしながら，$n < d$ では分散共分散行列 Σ に対する不偏推定量 $\widehat{\Sigma}$ を求めたとしても，逆行列は一般に存在しない．そこで，δ_{F} における Σ の代わりに $D = \mathrm{diag}\Sigma$ を用いた naive Bayes によって判別関数を構築する．この場合，$\widehat{D}$ は対角成分が非零である限り，逆行列は存在する．

3.2　正準相関に基づく判別分析

正準相関に基づいて判別関数を構築することを考える．正準相関分析では，$\boldsymbol{X}$，$\boldsymbol{Y}$ を共に中心化された確率変数ベクトルと定めていたが，特に $\boldsymbol{Y}$ を中心化せずに $C = \Sigma^{-1/2}E[(\boldsymbol{X} - \boldsymbol{\mu}_X)\boldsymbol{Y}^T]E[\boldsymbol{Y}\boldsymbol{Y}^T]^{-1/2}$ を考える．ここで，$\boldsymbol{Y}$ は k 値を取る離散型確率変数ベクトルとし，$P(\boldsymbol{Y} = \boldsymbol{e}_\ell) = \pi_\ell$ を満たす．ただし，$\boldsymbol{e}_\ell$ は第 ℓ 成分のみ 1 で他は 0 の k 次元のベクトルと

し，π_ℓ はクラス ℓ に属する事前確率であり，$\sum_{\ell=1}^k \pi_\ell = 1$ を満たす．このとき，固有値問題 $CC^T\boldsymbol{p} = \nu^2\boldsymbol{p}$ について考えると，Fisher の評価基準に帰着し，中心化されたデータに対して，$\boldsymbol{b}_j = \Sigma^{-1/2}\boldsymbol{p}_j$，$j \in \{1,\dots,k-1\}$ を用いた判別関数を構築すれば良い．しかし，ここでも $n < d$ だと $\widehat{\Sigma}$ の逆行列が得られないため，X を $d \times n$ のデータ行列，Y を $k \times n$ のクラス行列とし，行列 C の推定量を以下によって与える：

$$\widehat{C} = \widehat{D}^{-1/2}\left(\frac{1}{n}XY^T\right)\left(\frac{1}{n}YY^T\right)^{-1/2}.$$

4　漸近的性質及びシミュレーション

以下では $k > 2$ とし，判別関数に関する漸近的性質の拡張を考える．多群における判別関数 $\widehat{g} : \mathbb{R}^d \to \{1,\dots,k\}$ は，$\widehat{B} = [\widehat{\boldsymbol{b}}_1,\dots,\widehat{\boldsymbol{b}}_{k-1}]^T$ を用いて

$$\widehat{g}\,(\boldsymbol{X}) = \underset{\ell\in\{1,\dots,k\}}{\arg\min}\,(\boldsymbol{X} - \widehat{\boldsymbol{\mu}}_\ell)^T\,\widehat{B}^T(\widehat{B}\widehat{D}\widehat{B}^T)^{-1}\widehat{B}\,(\boldsymbol{X} - \widehat{\boldsymbol{\mu}}_\ell)$$

によって与えられる．そこで K を一つ固定し，パラメータ空間

$$\Theta_K = \left\{(\boldsymbol{\mu}_1,\dots,\boldsymbol{\mu}_k,\Sigma)\,\middle|\,\min_{\ell\neq K}(\boldsymbol{\mu}_K - \boldsymbol{\mu}_\ell)^T D^{-1}(\boldsymbol{\mu}_K - \boldsymbol{\mu}_\ell) \geq C_d, \lambda_{\max}(R) \leq b_0, \min_{1\leq j\leq d}\sigma_{jj} > 0\right\}$$

を定義する．ただし，$\Sigma = (\sigma_{ij})$，$R = D^{-1/2}\Sigma D^{-1/2}$ であり，$\lambda_{\max}$ は最大固有値を表す．ここで，任意の $\theta \in \Theta_K$ に対して，観測ベクトル $\boldsymbol{X}$ がクラス K にも関わらずその他のクラスに判別してしまう確率

$$W_K(g,\theta) = P\left(g\,(\boldsymbol{X}) \neq K\,|\boldsymbol{X}_{\ell i},\ i = 1,\dots,n_\ell, \ell = 1,\dots,k\right)$$

を評価する．本講演では予めいくつかの条件を与え，そのもとで領域の包含関係を用いて誤判別確率の上界や，$\widehat{\boldsymbol{p}}_j$，$\widehat{\boldsymbol{b}}_j$ に関する一致性の有無の議論を行った．得られた結果は Fan and Fan (2008)，Tamatani et al. (2011) で与えられた $k = 2$ のときに帰着することが示される．シミュレーションでは，3 群における判別方向ベクトルの一致性，すなわち，なす角 $\angle(\widehat{\boldsymbol{b}}_j, \boldsymbol{b}_j)$ の数値的挙動を調べた．このとき，予め与えた条件かつデータ数が次元を超えないという条件のもとで，角度は 0 へと向かう結果が得られた．

参考文献

Bickel, P. J. and Levina, E. (2004). Some theory for fisher's linear discriminant function, "naive bayes" and some alternatives when there are many more variables than observations. *Bernoulli*, **10** 989–1010.

Fan, J. and Fan, Y. (2008). High-dimensional classification using features annealed independence rules. *The Annals of Statistics*, **36** 2605–2637.

Tamatani, M., Koch, I. and Naito, K. (2011). Pattern recognition based on canonical correlations in a high dimension low sample size context, submitted.

On some tests for assessing multivariate normality based on sample moments

東京理科大学・理・院　齋藤　めい
東京理科大学・理・院　澄川　琢磨
東京理科大学・理　　　小泉　和之
東京理科大学・理　　　瀬尾　　隆

本報告では, 平均ベクトル　, 分散共分散行列 Σ をもつ確率分布から得られた連続値データが正規分布に従っているかを調べる問題を考える．Koizumi et al. (2009) は，Srivastava (1984) による多変量歪度，尖度を用いた総括的な多変量正規性検定統計量を提案している．しかし，この統計量の極限分布である χ^2 分布の自由度が次元数 p に依存しているため，p が大きいときに $b_{S,1}^2$ の影響が強くなることが懸念される．よって，本報告では極限分布の自由度が次元数 p に依存しない統計量を提案する．

標本共分散行列 S の固有値を $\omega_1, \omega_2, \ldots, \omega_p$ $(\omega_1 > \omega_2 > \cdots > \omega_p > 0)$ とすると, ある直交行列 $H = (\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_p)$ が存在して $H'SH = D_\omega = \mathrm{diag}(\omega_1, \omega_2, \ldots, \omega_p)$ となる．Srivastava (1984) は，多変量標本歪度を $b_{S,1}^2 = \frac{1}{N^2 p} \sum_{i=1}^{p} \omega_i^{\quad 3} \left\{ \sum_{j=1}^{N} (y_{ij} \quad \bar{y}_i)^3 \right\}^2$, 多変量標本尖度を $b_{S,2} = \frac{1}{Np} \sum_{i=1}^{p} \omega_i^{\quad 2} \sum_{j=1}^{N} (y_{ij} \quad \bar{y}_i)^4$ として与えている．ただし，$y_{ij} = \boldsymbol{h}_i' \boldsymbol{x}_j$, $\bar{y}_i = (1/N) \sum_{i=1}^{N} y_{ij}$ とする．これより, Koizumi et al. (2009) は, 次の総括的な多変量正規性検定統計量

$$MJB = Np \quad \frac{b_{S,1}^2}{6} + \frac{(b_{S,2} \quad 3)^2}{24}$$

を提案している．この統計量は十分大きな N に対して，自由度 $p+1$ の χ^2 分布に従う．

また χ_ν^2 は自由度 ν の χ^2 分布を表す確率変数とする．このとき,

$$\sqrt{2\chi_\nu^2} \quad \sqrt{2\nu \quad 1}$$

は ν が大きいとき，漸近的に $N(0,1)$ に従う．そこで本報告では，次の MJB_{ssk} を考えた．

$$MJB_{ssk} = (\sqrt{2z_1} \quad \sqrt{2p \quad 1})^2 + \frac{Np}{24}(b_{S,2} \quad 3)^2$$

ここに，$z_1 = Np(b_{S,1}^2/6)$ である．このとき，MJB_{ssk} は, $\forall x \in \mathbb{R}$ に対して,

$$\lim_{p\to\infty} \lim_{N\to\infty} \Pr(MJB_{ssk} \quad x) = G_2(x)$$

が成り立つ．ここに，$G_2(x)$ は自由度 2 の χ^2 分布の累積分布関数である．

さらに，Seo and Ariga (2011) では,

$$z_{NT} = \frac{\sqrt{Np}\{ \quad e^{\quad b_{S,2}} + e^{\quad 3} + 6e^{\quad 3}(1+\frac{2}{p})/N\}}{\sqrt{24}e^{\quad 3}}$$

は十分大きな N に対して, 漸近的に $N(0,1)$ に従うことが導かれている．よって, 本報告では，MJB_{ssk} をさらに改良した

$$MJB_{ssk}^* = (\sqrt{2z_1} \quad \sqrt{2p \quad 1})^2 + \frac{Np}{24} \quad e^{\quad b_{S,2}+3} + 1 + \frac{6}{N} \quad 1 + \frac{2}{p} \quad ^2$$

を考えた．このとき, $\forall x \in \mathbb{R}$ に対して

$$\lim_{p\to\infty}\lim_{N\to\infty} \Pr(MJB^*_{ssk} \quad x) = G_2(x)$$

が成り立つ.

　次に，Miyagawa et al. (2011) による多変量尖度を考えた．Miyagawa et al. (2011) は，多変量標本尖度を

$$b_{MS} = \frac{1}{Np^2}\sum_{j=1}^{N}\left(\sum_{i=1}^{p}\frac{y_{ij} \quad \overline{y}_i}{\sqrt{\omega_i}}\right)^4$$

と定義している．Miyagawa et al. (2011) では正規性のもとで，Σ が既知の場合における b_{MS} の期待値, 分散を導出している．本報告では，正規性のもとで，Σ が未知の場合を考えた．

　$\boldsymbol{y}_\alpha = (y_{1\alpha}, y_{2\alpha}, \dots y_{p\alpha})'$, $\overline{\boldsymbol{y}} = (\overline{y}_1, \overline{y}_2, \dots \overline{y}_p)'$, D_ω の従属性を避けるために

$$\begin{aligned}
D_\omega &= \frac{1}{N}\sum_{j=1}^{N}\left(\boldsymbol{y}_j \quad \overline{\boldsymbol{y}}\right)\left(\boldsymbol{y}_j \quad \overline{\boldsymbol{y}}\right)' \\
&= \frac{N \quad 1}{N} \quad D_\omega^{(\alpha)} + \frac{1}{N}\left(\boldsymbol{y}_\alpha \quad \overline{\boldsymbol{y}}^{(\alpha)}\right)\left(\boldsymbol{y}_\alpha \quad \overline{\boldsymbol{y}}^{(\alpha)}\right)'
\end{aligned}$$

とする．ただし

$$\begin{aligned}
\overline{\boldsymbol{y}}^{(\alpha)} &= (\overline{y}_1^{(\alpha)}, \overline{y}_2^{(\alpha)}, \dots \overline{y}_p^{(\alpha)})', \\
D_\omega^{(\alpha)} &= \frac{1}{N \quad 1}\sum_{j=1, j\neq\alpha}^{N}\left(\boldsymbol{y}_j \quad \overline{\boldsymbol{y}}^{(\alpha)}\right)\left(\boldsymbol{y}_j \quad \overline{\boldsymbol{y}}^{(\alpha)}\right)'
\end{aligned}$$

は，それぞれ α 番目の観測値 $\boldsymbol{y}^{(\alpha)}$ を除いて計算する推定量である．さらに，一般性を失うことなく，　$= \mathbf{0}$，$\Sigma = I_p$ としてよいから

$$\begin{aligned}
\boldsymbol{z} &= \sqrt{N \quad 1}\,\overline{\boldsymbol{y}}^{(\alpha)}, \\
Z &= \sqrt{N \quad 1}(D_\omega^{(\alpha)} \quad I_p)
\end{aligned}$$

とおいて，b_{MS} を摂動展開することにより，b_{MS} の期待値を計算すると

$$E[b_{MS}] = 3 \quad \frac{6}{N} \quad \frac{15}{Np} + o(N \;\, ^1)$$

が得られた.

　また, 数値実験により, 提案した統計量の近似精度を調べた.

Empirical likelihood approach to discriminant analysis for stationary processes

Tomoyuki AMANO (Wakayama University)

Abstract

Empirical likelihood is a non-parametric method and it does not need the knowledge of the distribution which the data comes from and it is widely used. However this method is not usually applied to discriminant analysis. Hence we shall apply this method to discriminant analysis and propose an empirical discriminant function. Then we prove its consistency, which means misclassification probability converges to 0. Furthermore under contiguous hypotheses the limit of its misclassification probability is derived and its lowerbound is obtained. We also evaluate its properties by some statistical methods and its interesting features are obtained by simulation.

References

[1] Anderson, T. W. (1984) *An Introduction to Multivariate Statistical Analysis, 2nd ed.* New York: Wiley.

[2] Brillinger, D. R. (2001) *Time Series Data Analysis and Theory.* Philadelphia: Society for Industrial and Applied Mathematics.

[3] Brockwell, P. J. & Davis, R. A. (1991) *Time Series: Theory and Methods, 2nd ed.* New York: Springer-Verlag.

[4] Monti, A. C. (1997) Empirical likelihood confidence regions in time series models. Biometrika 84, 395-405.

[5] Ogata, H. & 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, 451-468.

[6] Owen, A. B. (1988) Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75, 237-249.

[7] Owen, A. B. (1990) Empirical likelihood ratio confidence regions. Ann. Statist. 18, 90-120.

[8] Taniguchi, M. & Kakizawa, Y. (2000) *Asymptotic Theory of Statistical Inference for Time Series*. New York: Springer.

Double Random Walk and Non Stationary Variance

Alexandre Petkovic*

December 26, 2011

1 Introduction

Modeling the volatility of a time series as a nonlinear function of an integrated time series remains a relatively unexplored approach in modern econometrics. A first study was made by Hansen (1995) in the context of linear regression model. In an empirical paper Park (2002) modeled the variance of assets returns as a function of an integrated time series. Following Hansen (1995) Chung and Park (2007) studied linear regression model with integrated and stationary regressors when the error term volatility is a nonlinear function of an integrated time series. Petkovic (2010) studied a linear regression model with deterministic regressors when the volatility of the error term depends on an integrated time series.

Following the above work we propose in this note a new family of time series model. The model is a local to unit root autoregressive process but where the variance of the error term is a function of an integrated time series. Such a process can be refereed to as double unit root process. We then derive the distribution of the ordinary least squares estimator of the autoregressive coefficient. Local to unit root autoregressive processes were introduced by Phillips (1987) who derived the asymptotic distribution of the ordinary least squares estimator. The results of this note can be seen as a generalization of those derived by Phillips (1987) .

Thought this note $\to_d$ will stand for convergence in distribution while $\to_p$ for convergence in probability and $\mathcal{D}[0,1]$ will stand for the space of right continuous functions with left limit over $[0,1]$ endowed with the Skorokhod topology.

2 The Model and the Assumptions

Consider the following time series model

$$y_{n,t} = \alpha_n + \theta_n y_{n,t-1} + \epsilon_{n,t}, \qquad t = 1, \ldots, n \tag{1}$$

with $y_0 = 0$ where $\theta_n = e^{c_2/n}$, $\alpha_n = c_1/c(n)$ where $c(n)$ is function of the sample size whose properties will be given bellow. $\epsilon_{n,t}$ is the error term which is modeled as

$$\epsilon_{n,t} = \sigma(z_t)u_{n,t},$$

*Waseda University, International Center for Science and Engineering Programs, Faculty of Science and Engineering, Bldg 51 4F, 3-4-1 Ohkubo, Shinjuku (169-8555), Tokyo, Japan, email: apetkovi@aoni.waseda.jp.

where $u_{n,t}$ is a martingale difference sequence with unit variance with respect to a filtration $\mathcal{F}_{n,t}$, z_t is of the from

$$z_t = z_{t-1} + w_t, \tag{2}$$

where w_t is some stochastic process with $w_{-1} = 0$. σ a function whose properties will be specified bellow. We assume that z_t is measurable with respect to $\mathcal{F}_{n,t-1}$ implying that $(\epsilon_{n,t}, \mathcal{F}_{n,t})$ is a martingale difference sequence satisfying

$$E(\epsilon_{n,t}^2 | \mathcal{F}_{n,t-1}) = \sigma^2(z_t).$$

Definition 1. *We say that the function $\sigma \in \mathcal{H}$ if*

$$\sigma(\lambda x) = \nu(\lambda)\tau(x) + \eta(x, \lambda)$$

where τ is locally Rieman integrable and η can be decomposed as

$$\eta(x, \lambda) = a(\lambda)A(x),$$

where $a(\lambda) = o(\nu(\lambda))$ and $A(x)$ is locally and exponentially bounded (i.e. $A(x) = O(e^{c|x|})$ as $|x| \to \infty$).

The above class of function is called asymptotically homogenous. The class $\mathcal{H}$ was introduced by Park and Philipps (1999), examples of functions in $\mathcal{H}$ can be found in Chung and Park (2007)

Define

$$U_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[rn]} u_{n,t} \qquad V_n(r) = \frac{1}{\sqrt{n}} z_{[rn]} = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} w_t,$$

We also assume the following $(U_n, V_n) \to_p (U, V)$ in $\mathcal{D}[0, 1]$ as $n \to \infty$, where (U, V) is a vector Brownian motion.

3 The Ordinary Least Squares Estimator

We prove the following

Theorem 1. *Let $(\hat{\alpha},\ \hat{\theta})'$ be the ordinary least squares estimator of model* (1). *Then*

$$\begin{aligned} n^{1/2}\nu^{-1}(\sqrt{n})\hat{\alpha} \quad &\Longrightarrow \quad -\frac{(\int_0^1 C_1(s)ds)(C_1^2(1) - C_2 - 2c_2\int_0^1 C_1^2(s)ds - 2c_1\int_0^1 C_1(s)ds)}{2(\int_0^1 C_1^2(s)ds - (\int_0^1 C_1(s)ds)^2)} \\ n(\hat{\theta} - 1) \quad &\Longrightarrow \quad \frac{(C_1^2(1) - C_2 - 2c_2\int_0^1 C_1^2(s)ds - 2c_1\int_0^1 C_1(s)ds)}{2(\int_0^1 C_1^2(s)ds - (\int_0^1 C_1(s)ds)^2)} \end{aligned}$$

where

$$\begin{aligned} C_1(r) \quad &= \quad c_1 \int_0^r e^{c_2 s} ds + \int_0^r e^{c_2(r-s)} \tau(V(s)) dU(s) \\ C_2 \quad &= \quad \int_0^1 \tau^2(V(s)) ds. \end{aligned}$$

Multi-Step Ahead Portfolio Estimation for Dependent Return Processes

by **Kenta Hamada** (Waseda University)
Masanobu Taniguchi (Waseda University)

Let $\{\mathbf{X}_t = (X_{1t}, \ldots, X_{mt})' : t \in \mathbb{Z}\}$ be a zero mean stationary process with $m \times m$ spectral density matrix $\mathbf{g}(\lambda) = 1/(2\pi)\mathbf{A_g}(e^{\mathbf{i}\lambda})\mathbf{K}\mathbf{A_g}(e^{\mathbf{i}\lambda})^*$, $\mathbf{A_g}(e^{\mathbf{i}\lambda}) = \sum_{j=0}^{\infty} \mathbf{A}(j)e^{\mathbf{i}j\lambda}$, where $\mathbf{K}$ is a non-singular $m \times m$ matrix. Let the spectral representation of $\{\mathbf{X}(t)\}$ be $\mathbf{X}_t = \int_{-\pi}^{\pi} e^{-\mathbf{i}t} d\mathbf{Z}(\lambda)$, where $\{\mathbf{Z}(\lambda) : -\pi \leq \lambda \leq \pi\}$ is an orthogonal increment processes. Hannan (1970) (Theorem 1 of Chapter III) described the best linear predictor $\widehat{\mathbf{X}}_t^{\mathrm{best}}$ based on $\mathbf{X}_t, \mathbf{X}_{t-1}, \ldots$ by

$$\widehat{\mathbf{X}}_t^{\mathrm{best}} = \int_{-\pi}^{\pi} e^{-\mathbf{i}t\lambda} \left\{\mathbf{A_g}(e^{\mathbf{i}\lambda}) - \mathbf{A_g}(0)\right\} \mathbf{A_g}(e^{\mathbf{i}\lambda})^{-1} d\mathbf{Z}(\lambda).$$

We can consider the problem of misspecified prediction based on a conjectured spectral density matrix $\mathbf{f}(\lambda) = 1/(2\pi)\mathbf{A_f}(e^{\mathbf{i}\lambda})\mathbf{A_f}(e^{\mathbf{i}\lambda})^*$, where $\mathbf{A_f}(z) = \sum_{j=0}^{\infty} \mathbf{a}_j^{(\mathbf{f})} z^j, (z \in \mathbb{C})$. Then the pseudo linear predictor computed on the basis of $\mathbf{f}(\lambda)$ is given by

$$\widehat{\mathbf{X}}_t^{\mathrm{p-best}}(\mathbf{f}) = \int_{-\pi}^{\pi} e^{-\mathbf{i}t\lambda} \left\{\mathbf{A_f}(e^{\mathbf{i}\lambda}) - \mathbf{A_f}(0)\right\} \mathbf{A_f}(e^{\mathbf{i}\lambda})^{-1} d\mathbf{Z}(\lambda),$$

where $\mathbf{A_f}(0) = \mathbf{I}_m$ and $\mathbf{I}_m$ is the $m \times m$-identity matrix. The prediction error is given by

$$\mathrm{tr}\, \mathbb{E}\left[\{\mathbf{X}_t - \widehat{\mathbf{X}}_t^{\mathrm{p-best}}(\mathbf{f})\}\{\mathbf{X}_t - \widehat{\mathbf{X}}_t^{\mathrm{p-best}}(\mathbf{f})\}^*\right] \propto \int_{-\pi}^{\pi} \mathrm{tr}\left\{\mathbf{f}^{-1}(\lambda)\mathbf{g}(\lambda)\right\} d\lambda.$$

We can apply this to the h-step ahead prediction. Consider the problem of prediction for $\mathbf{X}_{t+h}$ by linear combination

$$\widehat{\mathbf{X}}_{t+h} = \sum_{j=1}^{L} \mathbf{\Phi}(j)\mathbf{X}_{t-j+1},$$

where $\mathbf{\Phi}(j)$'s are $m \times m$ matrices. This problem can be understood that we fit the following matrix

$$\mathbf{f}_{\boldsymbol{\theta}}(\lambda) = \frac{1}{2\pi}\left\{{\mathbf{\Gamma}_{\boldsymbol{\theta}}}^{-1}(\lambda){\mathbf{\Gamma}_{\boldsymbol{\theta}}}^{*-1}(\lambda)\right\} \tag{1}$$

to $\mathbf{g}(\lambda)$, where ${\mathbf{\Gamma}_{\boldsymbol{\theta}}}^{-1}(\lambda) = \mathbf{I}_m - \sum_{j=1}^{L} \mathbf{\Phi}(j)e^{\mathbf{i}(h+j-1)\lambda}$. Here

$$\boldsymbol{\theta} = \ \mathrm{vec}\left(\mathbf{\Phi}(1)\right)', \ldots, \mathrm{vec}\left(\mathbf{\Phi}(m)\right)'\ ' \in \mathbf{\Theta} \subset \mathbf{R}^r.$$

where $r = \dim \boldsymbol{\theta} = L \times m^2$. The best h-step ahead predictor is given by $\widehat{\mathbf{X}}_t^{\mathrm{p-best}}(\mathbf{f}_{\underline{\boldsymbol{\theta}}})$, where $\underline{\boldsymbol{\theta}}$ is defined by

$$\underline{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\mathbf{\Theta}} \int_{-\pi}^{\pi} \mathrm{tr}\left\{\mathbf{f}_{\boldsymbol{\theta}}(\lambda)^{-1}\mathbf{g}(\lambda)\right\} d\lambda.$$

Because $\underline{\boldsymbol{\theta}}$ is unknown we estimate it by Whittle estimator. Suppose that an observed stretch $\{\mathbf{X}_{t-n+1}, \mathbf{X}_{t-n+2}, \ldots, \mathbf{X}_{t-1}, \mathbf{X}_t\}$ is available. Let

$$\mathbf{I}_n(\lambda) = d_{\mathbf{X}}^n(\lambda)d_{\mathbf{X}}^n(\lambda)^*,$$

where

$$d_{\mathbf{X}}^n(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{j=1}^{n} \mathbf{X}_{t-n+j} e^{\mathbf{i}j\lambda}, \ -\pi \leq \lambda \leq \pi.$$

The Whittle estimator for $\underline{\boldsymbol{\theta}}$ is defined by

$$\hat{\boldsymbol{\theta}}_w = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \int_{-\pi}^{\pi} \mathrm{tr}\left\{\mathbf{f}_{\boldsymbol{\theta}}(\lambda)^{-1}\mathbf{I}_n(\lambda)\right\} d\lambda.$$

Then, we have

Lemma 1 *(Hosoya and Taniguchi (1982)) Under appropriate regularity conditions,*

$$\hat{\boldsymbol{\theta}}_w \xrightarrow{a.s.} \underline{\boldsymbol{\theta}}.$$

As an estimator of $\widehat{\mathbf{X}}_{t+h}^{\mathrm{p-best}}(\mathbf{f}_{\underline{\boldsymbol{\theta}}})$ we can use $\widehat{\mathbf{X}}_{t+h}^{\mathrm{p-best}}(\mathbf{f}_{\hat{\boldsymbol{\theta}}_w})$. Suppose that we are now interested in the problem of portfolio on $\{\mathbf{X}_t\}$. For a utility function we can define the optimal portfolio $\boldsymbol{\alpha}_{\mathrm{opt}} = (\alpha_1, \ldots, \alpha_m)'$ and its estimator $\hat{\boldsymbol{\alpha}}_{\mathrm{opt}}$ based on $\mathbf{X}_t, \ldots, \mathbf{X}_{t-n+1}$. Under natural assumptions we may assume

$$\hat{\boldsymbol{\alpha}}_{\mathrm{opt}} \xrightarrow{a.s.} \boldsymbol{\alpha}_{\mathrm{opt}}, \tag{2}$$

(e.g., Shiraishi and Taniguchi (2008)). Then we will estimate the h-step ahead portfolio value $\boldsymbol{\alpha}'_{\mathrm{opt}}\mathbf{X}_{t+h}$ by

$$\hat{\boldsymbol{\alpha}}'_{\mathrm{opt}}\widehat{\mathbf{X}}_{t+h}^{\mathrm{p-best}}(\mathbf{f}_{\hat{\boldsymbol{\theta}}_w}).$$

In what follows we evaluate the accuracy:

$$\hat{\boldsymbol{\alpha}}'_{\mathrm{opt}}\widehat{\mathbf{X}}_{t+h}^{\mathrm{p-best}}(\mathbf{f}_{\hat{\boldsymbol{\theta}}_w}) - \boldsymbol{\alpha}'_{\mathrm{opt}}\mathbf{X}_{t+h}.$$

Let $PE \equiv \boldsymbol{\alpha}'_{\mathrm{opt}}\widehat{\mathbf{X}}_{t+h}^{\mathrm{p-best}}(\mathbf{f}_{\underline{\boldsymbol{\theta}}}) - \boldsymbol{\alpha}'_{\mathrm{opt}}\mathbf{X}_{t+h}$. Then we have,

Proposition 1 *Under Assumption 1, it holds that*

$$\text{(i)} \quad \hat{\boldsymbol{\alpha}}'_{\mathrm{opt}}\widehat{\mathbf{X}}_{t+h}^{\mathrm{p-best}}(\mathbf{f}_{\hat{\boldsymbol{\theta}}_w}) - \boldsymbol{\alpha}'_{\mathrm{opt}}\mathbf{X}_{t+h} = PE + o_p(1).$$

$$\text{(ii)} \quad \mathbb{E}\{PE^2\} = \boldsymbol{\alpha}'_{\mathrm{opt}} \int_{-\pi}^{\pi} \Gamma_{\underline{\boldsymbol{\theta}}}(\lambda)\mathbf{g}(\lambda)\Gamma_{\underline{\boldsymbol{\theta}}}{}^*(\lambda) d\lambda \boldsymbol{\alpha}_{\mathrm{opt}}.$$

参考文献

E.J. Hannan. *Multiple time series.* John Wiley & Sons, 1970.

Y. Hosoya and M. Taniguchi. A central limit theorem for stationary processes and the parameter estimation of linear processes. *The Annals of Statistics*, pages 132–153, 1982.

H. Shiraishi and M. Taniguchi. Statistical estimation of optimal portfolios for non-Gaussian dependent returns of assets. *Journal of Forecasting*, 27(3):193–215, 2008.

Optimal portfolio with generalized empirical likelihood

Hiroaki Ogata
Waseda University

1 Introduction

The purpose of this talk is to find the optimal portfolio weights. The log-returns of assets are modeled by multivariate stationary processes rather than i.i.d. sequences. Then, the variance of the portfolio is written by the spectral density matrix, as shown in (2) below, and we seek the portfolio weights minimizing it. In practice, we construct the estimating function for the interested parameters (portfolio weights) in frequency domain, and obtain the estimator with the method of generalized empirical likelihood (GEL).

2 Frequency domain estimating function

Here we are concerned with the m-dimensional stationary process $\{\boldsymbol{X}(t)\}_{t\in\mathbb{Z}}$ with mean vector $\mathbf{0}$, the autocovariance matrix $\Gamma(h)$ and spectral density matrix

$$f(\lambda) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty}\Gamma(j)e^{-ij\lambda}, \qquad -\pi \le \lambda < \pi.$$

Suppose that information of an interested parameter $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ exists through a system of general estimating equations in frequency domain as follows. Let $\phi_j(\lambda;\boldsymbol{\theta})$, $(j = 1,\ldots,q)$ be $m \times m$ matrix-valued continuous functions on $[-\pi,\pi]$ satisfying $\phi_j(\lambda;\boldsymbol{\theta}) = \phi_j(\lambda;\boldsymbol{\theta})^*$ and $\phi_j(-\lambda;\boldsymbol{\theta}) = \phi_j(\lambda;\boldsymbol{\theta})'$. We assume that each $\phi_j(\lambda;\boldsymbol{\theta})$ satisfies the spectral moment condition

$$\int_{-\pi}^{\pi}\mathrm{tr}\{\phi_j(\lambda;\boldsymbol{\theta}_0)f(\lambda)\}\,d\lambda = 0 \qquad (j = 1,\ldots,q) \tag{1}$$

where $\boldsymbol{\theta}_0 = (\theta_{10},\ldots,\theta_{p0})'$ is the true value of the parameter. By taking an appropriate function for $\phi_j(\lambda;\boldsymbol{\theta})$, the equation (1) can express the best portfolio weights as shown in Example 2.1 below.

Example 2.1 (Portfolio selection) Let $x_i(t)$ be the log-return of i-th asset $(i = 1,\ldots,m)$ at time t and suppose that the process $\{\boldsymbol{X}(t) = (x_1(t),\ldots,x_m(t))'\}$ is stationary with zero mean. Consider the portfolio $p(t) = \sum_{i=1}^m \theta_i x_i(t)$ where $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_m)'$ is a vector of weights, satisfying $\sum_{i=1}^m \theta_i = 1$. The process $\{p(t)\}$ is a linear combination of the stationary process, hence $\{p(t)\}$ is still stationary and, from Herglotz's theorem, its variance is

$$\mathrm{Var}\{p(t)\} = \boldsymbol{\theta}'\mathrm{Var}\{\boldsymbol{X}(t)\}\boldsymbol{\theta} = \boldsymbol{\theta}'\left(\int_{-\pi}^{\pi}f(\lambda)d\lambda\right)\boldsymbol{\theta}. \tag{2}$$

Our aim is to find the weights $\boldsymbol{\theta}_0 = (\theta_{10},\ldots,\theta_{m0})'$, that minimize the variance (the risk) of the portfolio $p(t)$ under the constrain of $\sum_{i=1}^m \theta_i = 1$. The Lagrange function is given by

$$L(\boldsymbol{\theta},\lambda) = \boldsymbol{\theta}'\left(\int_{-\pi}^{\pi}f(\lambda)d\lambda\right)\boldsymbol{\theta} + \lambda(\boldsymbol{\theta}'\boldsymbol{e} - 1)$$

where $\boldsymbol{e} = (1, 1, \ldots, 1)'$ and λ is Lagrange multiplier. The first order condition leads to

$$(I - \boldsymbol{e}\boldsymbol{\theta}_0')\left(\int_{-\pi}^{\pi} f(\lambda) + f(\lambda)' d\lambda\right)\boldsymbol{\theta}_0 = \mathbf{0} \tag{3}$$

where I is an identity matrix. Now, for fixed $j = 1, \ldots, m$, consider to take

$$\phi_j(\lambda; \boldsymbol{\theta}) = \begin{cases} 2\theta_j(1-\theta_j) & (j,j)\text{-th component} \\ 1 - 2\theta_j\theta_\ell & (j,\ell)\text{-th and } (\ell,j)\text{-th component with} \\ & \qquad \ell = 1, \ldots, m \text{ and } \ell \neq j \\ -2\theta_k\theta_\ell & (k,\ell)\text{-th component with} \\ & \qquad k,\, \ell = 1, \ldots, m \text{ and } k \neq j,\, \ell \neq j \end{cases}.$$

Then, (1) coincides with the first order condition (3), which implies that the best portfolio weights can be solved with the framework of the spectral moment condition. ∎

Based on the form of (1), we set the estimating function for $\boldsymbol{\theta}$ as

$$\boldsymbol{m}(\lambda_t; \boldsymbol{\theta}) = \Big(\mathrm{tr}\{\phi_1(\lambda_t; \boldsymbol{\theta}) I_n(\lambda_t)\}, \ldots, \mathrm{tr}\{\phi_q(\lambda_t; \boldsymbol{\theta}) I_n(\lambda_t)\}\Big)'$$

where $I_n(\lambda)$ is the periodogram, defined by

$$I_n(\lambda) = (2\pi n)^{-1}\left\{\sum_{t=1}^{n} \boldsymbol{X}(t)\exp(it\lambda)\right\}\left\{\sum_{t=1}^{n} \boldsymbol{X}(t)\exp(it\lambda)\right\}^*$$

and $\lambda_t = (2\pi t)/n$, $(t = -[(n-1)/2], \ldots, [n/2])$. Then, we have

$$\frac{2\pi}{n}\sum_{-[(n-1)/2]}^{[n/2]} E[\boldsymbol{m}(\lambda_t; \boldsymbol{\theta})] \to \left[\int_{-\pi}^{\pi} \mathrm{tr}\{\phi_j(\lambda; \boldsymbol{\theta}_0) f(\lambda)\}\, d\lambda\right]_{j=1,\ldots,q} = \mathbf{0}.$$

3 Generalized empirical likelihood

Once we construct the estimating function, we can make use of the method of generalized empirical likelihood (GEL) as in Smith (2004) and Newey and Smith (2004). GEL is introduced as an alternative to generalized methods of moments (GMM) and it is pointed out that its asymptotic bias does not grow with the number of moment restrictions, while the bias of GMM often does.

To describe GEL let $\rho(v)$ be a function of a scalar v that is concave on its domain, an open interval $\mathcal{V}$ containing zero. Let $\hat{\Lambda}_n(\boldsymbol{\theta}) = \{\boldsymbol{\lambda} : \boldsymbol{\lambda}'\boldsymbol{m}(\boldsymbol{\lambda}_t; \boldsymbol{\theta}) \in \mathcal{V}, t = 1, \ldots, n\}$. The estimator is the solution to a saddle point problem

$$\hat{\boldsymbol{\theta}}_{\mathrm{GEL}} = \arg\min_{\boldsymbol{\theta}\in\Theta} \sup_{\boldsymbol{\lambda}\in\hat{\Lambda}_n(\boldsymbol{\theta})} \sum_{t=1}^{n} \rho\ \boldsymbol{\lambda}'\boldsymbol{m}(\boldsymbol{\lambda}_t; \boldsymbol{\theta})\ .$$

The empirical likelihood (EL) estimator (cf. Qin and Lawless (1994)), the exponential tilting (ET) estimator (cf. Kitamura and Stutzer (1997)) and the continuous updating estimator (CUE) (cf. Hansen et. al. (1996)) are special cases with $\rho(v) = \log(1-v)$, $\rho(v) = -e^v$ and $\rho(v) = -1/2(1+v)^2$, respectively. Let $\Omega = E[\boldsymbol{m}(\boldsymbol{\lambda}_t; \boldsymbol{\theta}_0)\boldsymbol{m}(\boldsymbol{\lambda}_t; \boldsymbol{\theta}_0)']$, $G = E[\partial\boldsymbol{m}(\boldsymbol{\lambda}_t; \boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}]$ and $\Sigma = (G'\Omega^{-1}G)^{-1}$. Under regular assumptions, we obtain the following theorem.

Theorem 3.1 $\sqrt{n}(\hat{\boldsymbol{\theta}}_{\mathrm{GEL}} - \boldsymbol{\theta}_0) \xrightarrow{p} N(\mathbf{0}, \Sigma)$.

Bartlett correctability of empirical likelihood ratio test for a parameter subvector in the over-identified case

柿沢 佳秀 (北大経済)

1. はじめに　バートレット調整として知られている LR 統計量の帰無分布のカイ 2 乗近似を改良する話題は, Bartlett (1937;Proc.R.Soc.Lond.) による分散均一性の具体例が始まりであり, 一般論は, Lawley (1956;Biometrika) が $(\theta'_{(1)}, \theta'_{(2)})' \in \mathbf{R}^p$ に関する帰無仮説 $\theta_{(1)} = \theta_{(1)0}$ の LR 統計量の平均 $E^{(N)}_{\theta_{(1)0},\theta^{\dagger}_{(2)}}[\mathrm{LR}^{(N)}] = p_1 + (\mathcal{E}_p - \mathcal{E}_{p-p_1})/N + \mathrm{o}(N^{-1})$ を計算し, さらに $p_1\mathrm{LR}^{(N)}/E^{(N)}_{\theta_{(1)0},\theta^{\dagger}_{(2)}}[\mathrm{LR}^{(N)}]$ の任意次のキュムラントは $\mathrm{o}(N^{-1})$ を無視すれば $\chi^2_{p_1}$ のキュムラントに一致することを示したことによる. 実際, これらの結果は LR 統計量の漸近展開 (Hayakawa (1977,1987 訂正;AISM))

$$P^{(N)}_{\theta_{(1)0},\theta^{\dagger}_{(2)}}[\mathrm{LR}^{(N)} \le x] = G_{p_1}(x) + \frac{A_1}{24N}\{G_{p_1+2}(x) - G_{p_1}(x)\} + \mathrm{o}(N^{-1})$$

に整合する. $\mathrm{BLR}^{(N)} = \mathrm{LR}^{(N)}/\{1 + A_1/(12p_1N)\}$ または $\{1 - A_1/(12p_1N)\}\mathrm{LR}^{(N)}$ の漸近展開の N^{-1} 項が消失する; $P^{(N)}_{\theta_{(1)0},\theta^{\dagger}_{(2)}}[\mathrm{BLR}^{(N)} \le x] = G_{p_1}(x) + \mathrm{o}(N^{-1})$ という事実は, 尤度法の枠組みを超えて Owen (1988;Biometrika, 1990;AS) による経験尤度 (EL) 法で成り立つことが DiCiccio et al. (1991;AS) により証明された. Chen (1993;AISM) は Owen (1991;AS) による独立な三角列に対する EL 法を採用し線形回帰モデルで ELR 統計量のバートレット調整可能性を示した. また, Zhang (1996;JNS) はスカラー母数の M 推定について ELR 統計量がバートレット調整可能であることを示した. さらに, この分野の最先端として, モーメント制約モデルに対する Qin & Lawless (1994;AS) の EL 法を採用し Chen & Cui (2006;Biometrika, 2007;JE) が ELR 統計量のバートレット調整可能性を丁度識別な場合 ($M = p$) のサブベクトル検定/過剰識別な場合 ($M > p$) のフルベクトル検定で証明しており, 本報告では過剰識別な場合のサブベクトル検定へ拡張した.

2. 設定　$\mathbf{X}_1, \ldots, \mathbf{X}_N$ は互いに独立で同一分布 (d_X-次元の未知分布 F) に従うとする. ここで, モデルはモーメント制約 $E_F[\mathbf{g}(\mathbf{X}, \theta_0)] = \mathbf{0}_M$ によって記述されている. ただし, $\theta_0 = (\theta'_{(1)0}, \theta'_{(2)0})' \in \mathbf{\Theta} \subset \mathbf{R}^p$, $\mathbf{g}(\mathbf{x}, \theta) = (g_1(\mathbf{x}, \theta), \ldots, g_M(\mathbf{x}, \theta))'$.

Qin & Lawless (1994) によるサブベクトル検定 $\theta_{(1)} = \theta_{(1)0}$ の ELR 統計量について

$$\mathrm{ELR}^{(N)} = 2\sum_{i=1}^{N} \log\{1 + (\widetilde{\lambda}^{(N)})'\mathbf{g}(\mathbf{X}_i, \widetilde{\theta}^{(N)})\} - 2\sum_{i=1}^{N} \log\{1 + (\widehat{\lambda}^{(N)})'\mathbf{g}(\mathbf{X}_i, \widehat{\theta}^{(N)})\} \xrightarrow{d} \chi^2_{p_1}$$

が知られている. ただし $\begin{pmatrix} \widetilde{\theta}^{(N)} \\ \widetilde{\lambda}^{(N)} \end{pmatrix}$ (なお, $\widetilde{\theta}^{(N)} = \begin{pmatrix} \theta_{(1)0} \\ \widetilde{\theta}^{(N)}_{(2)} \end{pmatrix}$ とおく) は

$$\mathbf{0}'_{p_2} = \sum_{i=1}^{N} \frac{(\widetilde{\lambda}^{(N)})'(\partial/\partial\theta'_{(2)})\mathbf{g}(\mathbf{X}_i, \widetilde{\theta}^{(N)})}{1 + (\widetilde{\lambda}^{(N)})'\mathbf{g}(\mathbf{X}_i, \widetilde{\theta}^{(N)})}, \quad \mathbf{0}_M = \sum_{i=1}^{N} \frac{\mathbf{g}(\mathbf{X}_i, \widetilde{\theta}^{(N)})}{1 + (\widetilde{\lambda}^{(N)})'\mathbf{g}(\mathbf{X}_i, \widetilde{\theta}^{(N)})}$$

を満たし, $\begin{pmatrix} \widehat{\theta}^{(N)} \\ \widehat{\lambda}^{(N)} \end{pmatrix}$ は

$$\mathbf{0}_p' = \sum_{i=1}^N \frac{(\widehat{\lambda}^{(N)})'(\partial/\partial\theta')\mathbf{g}(\mathbf{X}_i, \widehat{\theta}^{(N)})}{1+(\widehat{\lambda}^{(N)})'\mathbf{g}(\mathbf{X}_i, \widehat{\theta}^{(N)})}\,, \quad \mathbf{0}_M = \sum_{i=1}^N \frac{\mathbf{g}(\mathbf{X}_i, \widehat{\theta}^{(N)})}{1+(\widehat{\lambda}^{(N)})'\mathbf{g}(\mathbf{X}_i, \widehat{\theta}^{(N)})}$$

を満たす.

3. 主要な結果　EL 法の (高次) 漸近理論は母数モデルの尤度法の (高次) 漸近理論とほぼパラレルに議論することができる. 実際, Newey & Smith (2004;JE) に従うと,

$$-\frac{1}{N}\sum_{i=1}^N \log\{1+(\widehat{\lambda}^{(N)})'\mathbf{g}(\mathbf{X}_i, \widehat{\theta}^{(N)})\} = \max_{\theta\in\Theta} \min_{\lambda\in\Lambda_{\theta,N}} \Big[-\frac{1}{N}\sum_{i=1}^N \log\{1+\lambda'\mathbf{g}(\mathbf{X}_i, \theta)\}\Big]\,,$$

$\Lambda_{\theta,N} = \{\lambda : 1+\lambda'\mathbf{g}(\mathbf{X}_i,\theta) > 0\,,\ i=1,\ldots,N\}$ であるから $\widehat{\eta}^{(N)} = \begin{pmatrix} \widehat{\theta}^{(N)} \\ \widehat{\lambda}^{(N)} \end{pmatrix}$ は

$$\ell^{(N)}(\eta) = \frac{1}{N}\sum_{i=1}^N \log\{1+\lambda'\mathbf{g}(\mathbf{X}_i,\theta)\}\,, \quad \eta = \begin{pmatrix} \theta \\ \lambda \end{pmatrix}$$

の鞍点となり, $\widetilde{\eta}_{(2)}^{(N)} = \begin{pmatrix} \widetilde{\theta}_{(2)}^{(N)} \\ \widetilde{\lambda}^{(N)} \end{pmatrix}$ は $\ell^{(N)}\begin{pmatrix} \theta_{(1)0} \\ \eta_{(2)} \end{pmatrix}$ の鞍点である. ただし $\eta_{(2)} = \begin{pmatrix} \theta_{(2)} \\ \lambda \end{pmatrix}$.

定理　ELR 統計量はバートレット調整可能であるが, EL-based score/Wald 統計量は一般にバートレット調整不可能である (それらのバートレット型調整も議論した).

EL 法は Newey & Smith (2004) による GEL 法に属するので, 一般的な鞍点問題

$$\max_{\theta\in\Theta} \min_{\lambda\in\Lambda_{\theta,N}^{\rho}} \frac{1}{N}\sum_{i=1}^N \rho\{-\lambda'\mathbf{g}(\mathbf{X}_i,\theta)\}$$

も考察した. ここに, $\rho(\cdot)$ は開区間 $\mathcal{V}(\ni 0)$ 上の凸関数で, $\rho(0)=0$, $\rho'(0)=\rho''(0)=1$ を満たし, $\Lambda_{\theta,N}^{\rho} = \{\lambda : -\lambda'\mathbf{g}(\mathbf{X}_i,\theta)\in\mathcal{V}\,,\ i=1,\ldots,N\}$ [EL 法は $\rho(v) = -\log(1-v)$, $\mathcal{V} = (-\infty, 1)$ の場合に他ならない].

定理　ELR 統計量と同様に

$${}_{\rho}\mathrm{GELR}^{(N)} = 2N\{{}_{\rho}\ell^{(N)}\begin{pmatrix} \theta_{(1)0} \\ {}_{\rho}\widetilde{\eta}_{(2)}^{(N)} \end{pmatrix} - {}_{\rho}\ell^{(N)}({}_{\rho}\widehat{\eta}^{(N)})\} \xrightarrow{d} \chi_{p_1}^2$$

を考えるとき, もし $\rho'''(0)=2, \rho''''(0)=6$ ならばバートレット調整可能である. ここに

$${}_{\rho}\ell^{(N)}(\eta) = -\frac{1}{N}\sum_{i=1}^N \rho\{-\lambda'\mathbf{g}(\mathbf{X}_i,\theta)\}\,, \quad \eta = \begin{pmatrix} \theta \\ \lambda \end{pmatrix}\,.$$

したがって, Cressie & Read's family

$$\rho^{\mathrm{CR}}(v) = \begin{cases} \dfrac{1}{\gamma+1}\{(1+\gamma v)^{(\gamma+1)/\gamma} - 1\}\,, & \gamma \neq -1, 0 \\ -\log(1-v)\,, & \gamma = -1 \\ \mathrm{e}^v - 1\,, & \gamma = 0 \end{cases}$$

の中では $\gamma = -1$ すなわち ELR 統計量のみがバートレット調整可能である.

ASYMPTOTIC PROPERTIES OF TIME SERIES NON-LIFE INSURANCE MODEL

KENTARO KOBAYASHI AND JUNICHI HIRUKAWA
NIIGATA UNIVERSITY

The object of main interest from the point of view of an insurance company is the total claim amount process or aggregate claim amount process $S(t)$, which is composed of the claim number process $N(t)$ and the sequence of claim size $\{X_i\}$. One would like to find sufficiently realistic, but simple, probabilistic models for $S(t)$, $N(t)$, and $\{X_i\}$. In the classical non-life insurance model, the sequence of claim size $\{X_i\}$ is assumed to be i.i.d. process (see Mikosch (2004)). However, this assumption is quite unrealistic. The claim size sequence $\{X_i\}$ should contain the time dependent structure. Here, we suggest to employ the locally stationary processes as the claim size sequence. We will derive the extension of central limit theorem (Anscombe' Theorem (see Gut (1988))) for the locally stationary non-life insurance model. We will also present the functional central limit theorem (Anscombe-Donsker Invariance Theorem (see Gut (1988), Billingsley (1968))) in our case.

Now, we can define the claim number process $N(t) = \#\{i \geq 1 : T_i \leq t\},\ t \geq 0$, i.e., $N = (N(t))_{t\geq 0}$ is a counting process on $[0, \infty)$. $N(t)$ is the number of the claims which occurred by time t. The object of our main interest is the total claim amount process $S(t) = \sum_{i=1}^{N(t)} X_i,\ t \geq 0$. The process $S = (S(t))_{t\geq 0}$ is random partial sum process which refers to the fact that the deterministic index n of the partial sums $S_n = X_1 + \cdots + X_n$ is replaced by the random variables $N(t)$: $S(t) = X_1 + \cdots + X_{N(t)},\ t \geq 0$. It is also often called a compound (sum) process. The following lemma is greatly useful.

Lemma 1 (Anscombe' Theorem)**.** Suppose that $Z_n \overset{d}{\Rightarrow} Z$, $n \to \infty$ and $N(t)/n(t) \overset{P}{\to} 1$ *as* $t \to \infty$, where $\{n(t)\}$ is a family of positive numbers tending to infinity, and that

$$(A)\quad \begin{cases} \text{Given } \epsilon > 0 \text{ and } \eta > 0 \text{ there exists } \delta > 0 \text{ and } n_0, \text{ such that} \\ P\{\max_{m:|m-n|<n\delta} |Z_m - Z_n| > \epsilon\} < \eta, \ \text{for all } n > n_0 \end{cases}$$

Then $Z_{N(t)}$ converges in distribution to Z as $t \to \infty$.

Now, we introduce locally stationary processes $X_{k,n}$ (see Dahlhaus (1996a,b)):

$$X_{k,n} = \mu\left(\frac{k}{n}\right) + u_{k,n}, \quad (\mu(u) \in C[0,1]),$$

where $\{u_{k,n}\}$ is generated by the following time varying MA (∞) model

$$u_{k,n} = \sum_{l=0}^{\infty} \alpha_l\left(\frac{k}{n}\right)\varepsilon_{k-l} = \sum_{l=0}^{\infty} \alpha_l\left(\frac{k}{n}\right) L^l \varepsilon_k = \alpha\left(\frac{k}{n}, L\right)\varepsilon_k$$

with $\alpha(u,L)=\sum_{l=0}^{\infty}\alpha_l(u)L^l$, $\alpha_l(u)\in C[0,1]$, $l=1,2,\ldots$, are time varying MA coefficients and $\{\varepsilon_s,\ s\geq 1\}$ is s set of independent random variables, such that $E[\varepsilon_s]=0$ and $E[\varepsilon_s]=\sigma^2$. Using this locally stationary processes, we define the total claim amount process $\{S_n\}$ as

$$S_n=\sum_{k=1}^{n}X_{k,n}.$$

Theorem 1. (Central Limit Theorem) For locally stationary stochastic processes $\{X_{k,n}\}$, we have the following central limit theorem:

$$\widetilde{Z}_n=\frac{S_n}{\sqrt{n}}\overset{d}{\Rightarrow}N\left(\sqrt{n}\mu_z,\sigma_z^2\right)\quad as\ n\to\infty,$$

where $\mu_z=\int_0^1\mu(u)du$ and $\sigma_z^2=\sum_{s=-\infty}^{\infty}\int_0^1\int_{-\pi}^{\pi}f(u,\lambda)e^{is\lambda}d\lambda du$.

Finally, the generalization of B-N decomposition for locally stationary processes is given by

$$u_{k,n}=\alpha\left(\frac{k}{n},1\right)\varepsilon_k-\tilde{\varepsilon}_{k,n},\qquad \tilde{\varepsilon}_{k,n}=\tilde{\alpha}\left(\frac{k}{n},L\right)(1-L)\varepsilon_j,$$

where $\tilde{\alpha}(u,L)=\sum_{l=0}^{\infty}\tilde{\alpha}_l(u)L^l$, $\tilde{\alpha}_l(u)=\sum_{j=l+1}^{\infty}\alpha_j(u)$. Let $\tilde{S}_n=S_n-n\mu$ and define partial sum process

$$X_n(t,\omega)=\frac{1}{\sigma\sqrt{n}}\tilde{S}_{[nt]}(\omega),\quad (0\leq t\leq 1).$$

Using B-N decomposition, we have $X_n(t)\Rightarrow\int_0^t\alpha(u,1)\,dW(u)$ (Functional Central Limit Theorem). Furthermore, we have the following result.

Theorem 2 (Anscombe' Theorem & Anscombe-Donsker Invariance Principle)**.** Suppose $\frac{N(n)}{a_n}\overset{P}{\to}\theta$, then

$Y_n(t,\omega)=\dfrac{1}{\sigma\sqrt{N(n)}}\tilde{S}_{[t\cdot N(n,\omega)]}(\omega)\overset{J_1}{\Rightarrow}W(t)$ (Anscombe-Donsker Invariance Principle),

$Z_{N(n)}=\dfrac{S_{N(n)}}{\sigma\sqrt{N(n)}}\overset{J_1}{\Rightarrow}W(1)\sim N(0,1)$ (Anscombe-Theorem).

References

[1] Gut, A. (1988) *Stopped Random Walks.* Springer, New York.

[2] Billingsley, P. (1968) *Convergence of Probability Measures.* Wiley, New York.

[3] Mikosch, T. (2004) *Non-Life Insurance Mathematics. An Introduction with Stochastic Processes*, Springer, Berlin.

[4] Dahlhaus, R. (1996a) On the Kullback Leibler information divergence of locally stationary processes, *Stochastic Process. Appl.* **62**, 139-168.

[5] Dahlhaus, R. (1996b) Maximam Likelihood Estimation and Model Selection for Locally Stationary Processes, *J. Nonparametr. Statist.* **6** 171-191.

Monitoring the intraday volatility pattern

Siegfried Hörmann
Département de Mathématique
Université Libre de Bruxelles

A functional time series consists of curves, typically one curve per day. The most important parameter of such a series is the mean curve. We propose two methods of detecting a change in the mean function of a functional time series. The change is detected on line, as new functional observations arrive. The general methodology is motivated by and applied to the detection of a change in the average intraday volatility pattern. The methodology is asymptotically justified by applying a new notion of weak dependence for functional time series. It is calibrated and validated by simulations based on real intraday volatility curves.

The talk is based on joint work with P. Kokoszka and R. Gabrys.

Which models to match?

(David Veredas, Roxana Halbleib and Matteo Barigozzi)

Most of the inference techniques based on matching (e.g. GMM, Indirect Inference, Efficient Method of Moments or the Method of Simulates Quantiles) depend on a subjective choice of the theoretical parametric functions that are used to estimate the parameters. In this article we develop a criteria for choosing them. The criteria is based on the likelihood ratio between the asymptotic distributions of estimators under two sets of theoretical functions. Both distributions are Gaussian, consistent but with different variance-covariance matrices. The criteria is suitable for choosing among nested functions, e.g. one set of moments versus a larger one, or one auxiliary model that is nested into another but also for non-nested sets of functions, such as two different estimation methods, two sets of moments, or two auxiliary models. A thorough Monte Carlo study based on two simple, yet important and illustrative, models shows the usefulness of the criteria.

Classification in segmented regression problems

Cathy WS Chen

Department of Statistics,
Graduate Institute of Statistics & Actuarial Science
Feng Chia University
Taichung 407, Taiwan

Heterogeneity in many datasets stems from the different behaviors of several underlying groups or subpopulations. The aim of this paper is to classify observations in such a dataset into these latent groups when each group's behavior is piecewise linearly related to a set of covariates. We assume that each group can be represented by a segmented regression model, but the group membership for each observation is unobserved or lost. A full Bayesian approach is proposed to simultaneously classify observations and estimate segmented regression parameters. The estimated marginal likelihood and the Deviance Information Criterion are used to select the number of mixture groups. We demonstrate the accuracy and performance of the proposed MCMC estimators in a simulation study and illustrate the methodology in an empirical study.

Local Linear Regression on Manifolds and its Geometric Interpretation

Ming-Yen Cheng
Department of Mathematics, National Taiwan University, Taipei 106, Taiwan.

Hau-TiengWu
Department of Mathematics, Princeton University, Fine Hall, Washington Road, Princeton NJ 08544-1000, USA.

High-dimensional data arise frequently in many fields of contemporary science. For example, medical images, genetic microarray data, or functional data are observed over time and with different candidate predictors. In addition, it is common that the sample size is small compared to the dimensionality of the data. In the linear regression setting, variable selection is one of the fundamental problems and has been extensively studied in the literature. When the dimension p is large, it becomes a difficult task to estimate the finite number of parameters because significant variables can be highly correlated with some of the unimportant variables. The rationale for variable selection is that only some of the regression coefficients are nonzero, so the principle is to regularize the parameter estimation problem by penalizing the spurious regression coefficients. Parametric regression is often too restrictive and results in large bias. There has been a vast amount of literature on nonparametric regression because it relaxes the restrictive model assumptions in parametric regression, thus can capture the underlying structure in a flexible way and avoid the excessive bias induced by model mis-specification in parametric regression. However, when the dimensionality is high, nonparametric models suffer from the curse of dimensionality problem i.e. the estimation accuracy deteriorates rapidly as p increases. Variable selection and dimension reduction become even more difficult in this setting. In fact, almost all the existing dimension reduction methods are based on the assumption that a (linear) central dimension reduction space, denoted as CS, exists and are focused on estimation of that. A a consequence, models built on results of these methods are closely related to some special cases in semiparametric regression such as multiple-index models. Semiparametric regression is an emerging area because it enjoys both the modeling flexibility of nonparametric regression and the modeling stability of parametric regression discussed in the above. Owing to the above considerations, there has been an increasing trend in studying variable selection in the semiparametric regression setting. On the other hand, the key motivations here are existence of CS and sparsity of nonzero constant coefficients or functions, called global assumptions, which may not be easily validated in practice. This is one of the motivations of our approach to model the regression function on the underlying manifold of the explanatory variables.

When the dimension of the predictors is high, it has been observed that the data usually lie on a lower dimensional manifold (Hall, Marron, and Neeman:2005; Bickel and Li, 2007; Jung, Foskey, and Marron, 2011; Aswani, Bickel, and Tomlin, 2011). We study nonparametric local linear regression when the predictors in this case. Recently Aswani, Bickel, and Tomlin (2011) suggested to regularize the local linear regression problem, performed in the ambient space, using information obtained by learning the manifold. By contrast, our approach is to construct local linear regression directly on an approximation to the manifold, thus is faster to compute when the dimension of the ambient space is large. Under mild conditions, asymptotic expressions for the conditional bias and variance of the proposed estimator are derived for both interior and boundary cases. One implication of this is the optimal convergence rate depends only on the intrinsic dimension d of the manifold, like in the usual multivariate nonparametric regression, but not on the dimension of the ambient space p. Another implication is that, in the diffusion map framework, the proposed method can be used to estimate Laplace Beltrami operator when in the interior and can be used to estimate linear combinations of the second order covariant derivatives when close to the boundary provided that the boundary is smooth. Further, the bias and variance expressions are used to construct a simple and effective bandwidth selection rule. An extensive simulation study and an example are used to compare the computational speed and estimation accuracy of our method with existing ones for various combinations of sample size, p, and d.

Keywords. bandwidth selection, classification, diffusion map, dimension reduction, manifold learning, nonparametric regression.

References

Aswani, A., Bickel, P., and Tomlin, C. (2011). Regression on manifolds: Estimation of the exterior derivative. *Annals of Statistics*, , 48-81.

Bickel, P. J., and Li, B (2007). Local Polynomial Regression on Unknown Manifolds. *IMS Lecture Notes-Monograph Series*, **54**, 177-186.

Jung, S., Foskey, M., and Marron, J. S (2011). Principal arc analysis on direct product manifolds *Annals of Applied Statistics*, **5**, 578–603.

Hall, P., Marron, J. S., and Neeman, A. (2005). Geometric representation of high dimension, low sample size data. *ournal of the Royal Statistical Society Series B*, **67**, 427–444.

Statistical Arbitrage and Fractional Cointegration [1]

Ngai Hang Chan
Department of Statistics
Chinese University of Hong Kong
Shatin, NT, Hong Kong
nhchan@sta.cuhk.edu.hk

Abstract

This talk discusses some of the recent developments of statistical arbitrage and fractional cointegration. By virtue of some of the asymptotic results about fractional co-integration tests, a pair-trading strategy is constructed from which statistical arbitrage can be developed. The talk concludes with some of the applications to financial data.

[1]Wakayama Time Series Symposium, Wakayama University, Japan. Dec 2–4, 2011. Research supported in part by grants from HKSAR-RGC-GRF.

Skew-symmetric distributions and Fisher information

Marc Hallin

E.C.A.R.E.S., Université Libre de Bruxelles, 50 Av. F.D. Roosevelt, 1050 Brussels, Belgium and

O.R.F.E., Princeton University, Sherrerd Hall, Princeton, NJ 08544, USA

Christophe Ley

E.C.A.R.E.S. and Département de Mathématique, Université Libre de Bruxelles, Boulevard du Triomphe, CP210, 1050 Brussels, Belgium

Keywords: Skewing function, Skew-normal distributions, Skew-symmetric distributions, Singular Fisher information, Symmetric kernel

Skew-symmetric densities recently received much attention in the literature, giving rise to increasingly general families of univariate and multivariate skewed densities. Most of those families, however, suffer from the major drawback of a potentially singular Fisher information in the vicinity of symmetry. All existing results indicate that Gaussian densities (possibly after restriction to some linear subspace) play a very special and somewhat mysterious role in that context. In this talk, we totally dispel that widespread opinion by providing a full characterization of the information singularity phenomenon, highlighting its relation to a possible link between symmetric kernels and skewing functions–a link that can be interpreted as the mismatch of two densities.

References

[1] Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, **12**, 171–178.

[2] Azzalini, A. and Genton, M. G. (2008). Robust likelihood methods based on the skew-t and related distributions, *International Statistical Review*, **76**, 106–129.

[3] Hallin, M. and Ley, C. (2011). Skew-symmetric distributions and Fisher information – a tale of two densities, *Bernoulli*, to appear.

[4] Ley, C. and Paindaveine, D. (2010a). On the singularity of multivariate skew-symmetric models, *Journal of Multivariate Analysis*, **101**, 1434–1444.

[5] Wang, J., Boyer, J. and Genton, M. G. (2004). A skew-symmetric representation of multivariate distribution, *Statistica Sinica*, **14**, 1259–1270.

Covariance Tapering for Prediction of Large Spatial Data Sets in Transformed Random Fields

Toshihiro Hirano and Yoshihiro Yajima
University of Tokyo and University of Tokyo

Interpolation of a spatially correlated random process is widely used in mining, hydrology, forestry and other fields. This method is often called kriging in geostatistical literature and requires the inverse of the covariance matrix of observations in different spatial points. The operation count for computing the inverse is of order n^3 with sample size n. Hence as the sample size is larger, the computation becomes a more formidable one in practice.

To deal with this problem Furrer et al. (2006) proposed covariance tapering. A basic idea of covariance tapering is to reduce a spatial covariance function to zero beyond some range by multiplying the true spatial covariance function by a positive definite but compactly supported function. Then the resulting covariance matrix is so sparse that it is much easier and faster to obtain its inverse matrix. Furrer et al. (2006) proved the asymptotic efficiency of the BLUP using covariance tapering which we call the tapered BLUP for the original BLUP. Zhu and Wu (2010) investigated properties of covariance tapering for convolution-based nonstationary models and proved that the BLUP is asymptotically efficient in specific assumptions.

An alternative approach to reduce the computational time of the inverse matrix is to calculate a spatial prediction based on a small and manageable number of observations that are close to a prediction point. This approach often shows good performance. However, it is not clear how we may choose samples in a neighborhood of the prediction point and theoretical properties are not derived completely. On the other hand, in covariance tapering it is shown that the MSE ratio of the tapered predictor and the true predictor converges to 1 as the sample size goes to infinity regardless of the selection of the taper range (Furrer et al. (2006)).

Covariance tapering is also used for the estimation of parameters of a covariance function. The log-likelihood function of Gaussian random fields includes the determinant and the inverse of the covariance matrix between the observations in different spatial points, which it is difficult to calculate for large data sets. Kaufman et al. (2008) applied covariance tapering to the log-likelihood function and showed that the estimators maximizing the tapered approximation of the log-likelihood are strongly consistent. Du et al. (2009) proved that this tapered MLE has the asymptotic normality in one dimensional case. Recently, Wang and Loh (2011) showed the asymptotic normality of the tapered MLE in multidimensional case by letting the taper range converge to 0 when the sample size goes to infinity.

The BLUP is identical with the conditional expectation if an underlying random field is Gaussian and consequently is the optimal predictor in the MSE sense whereas if an original data takes a nonnegative value or has a skewed distribution, we frequently apply a nonlinear transformation to it to get a data which is nearer Gaussian. Typical ones are a chi-squared process and a lognormal process (Cressie (1993)). For example a precipitation data is approximately regarded as a chi-squared process because the standardized square root values known as anomalies are closer to a Gaussian distribution (Johns et al. (2003) and Furrer et al. (2006)). On the other hand the variable such as topsoil concentrations of cobalt and copper takes a positive value and has a right skewed sampling distribution. This kind of spatial data is often obtained in large numbers and modeled by the lognormal distribution (Moyeed and Papritz (2002) and De Oliveira (2006)). However the optimality of the BLUP and the tapered BLUP for an original data is not clear because it is non-Gaussian.

We consider the class of the transformation $Z(\boldsymbol{s}) = T(Y(\boldsymbol{s}))$ where $\{Y(\boldsymbol{s})\}$ is a Gaussian random field with zero-mean and unit variance. If we assume $E(T(Y(\boldsymbol{s}))^2) < \infty$, $T(\cdot)$ can be expressed by the Hermite polynomial expansion. Granger and Newbold (1976) considered this class of the transformed models in a time series context and calculated the mean, the covariance function and the mean squared error of predictors. Our work can be also regarded as an extension of their results to spatial processes. Moreover since the conditional mean of $Z(\boldsymbol{s}_0)$ given $\{Y(\boldsymbol{s})\}$ requires the inverse of the covariance matrix of $\{Y(\boldsymbol{s})\}$, covariance tapering is useful to reduce the computational difficulty.

Finally we show that the BLUP, the BLUP using covariance tapering and the optimal predictor are asymptotically equivalent in the MSE sense if the covariance function of the underlying Gaussian random field is Matérn type. This is an extension of Furrer et al. (2006). Monte Carlo simulations support theoretical results.

References

[1] Cressie, N. (1993). *Statistics for Spatial Data*, revised ed. Wiley.

[2] De Oliveira, V. (2006). On optimal point and block prediction in log-Gaussian random fields. *Scandinavian Journal of Statistics*, 33:523-540.

[3] Du, J., Zhang, H. and Mandrekar, V. S. (2009). Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *Annals of Statistics*, 37:3330-3361.

[4] Furrer, R., Genton, M. G. and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15:502-523.

[5] Granger, C.W.J. and Newbold, P. (1976). Forecasting transformed series. *Journal of the Royal Statistical Society: Series B*, 38:189-203.

[6] Johns, C., Nychka, D., Kittel, T. and Daly, C. (2003). Infilling sparse records of spatial fields. *Journal of the American Statistical Association*, 98:796-806.

[7] Kaufman, C., Schervish, M. and Nychka, D. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103:1545-1555.

[8] Moyeed, R. A. and Papritz, A. (2002). An empirical comparison of kriging methods for nonlinear spatial point prediction. *Mathematical Geology*, 34:365-386.

[9] Wang, D. and Loh, W.-L. (2011). On fixed-domain asymptotics and covariance tapering in Gaussian random field models. *Electronic Journal of Statistics*, 5:238-269.

[10] Zhu, Z. and Wu, Y. (2010). Estimation and prediction of a class of convolution-based spatial nonstationary models for large spatial data. *Journal of Computational and Graphical Statistics*, 19:74-95.

Asymptotics of L-statistics with dependent data and their applications to risk measure estimation

Hideatsu Tsukahara
Department of Economics, Seijo University, Tokyo
e-mail address: tsukahar@seijo.ac.jp

Let $(X_n)_{n\in\mathbb{N}}$ be a strictly stationary process with a stationary distribution function (df) F, and denote by $\mathbb{F}_n$ the empirical df based on the sample $X_1, \ldots, X_n$. Consider a general L-statistic of the following form:

$$T_n = \frac{1}{n}\sum_{i=1}^{n} c_{ni} h(X_{n:i}),$$

where c_{ni}'s are constants, and $X_{n:1} \quad X_{n:2} \quad \cdots \quad X_{n:n}$ are the order statistics based on the sample $X_1, \ldots, X_n$. Let $g := h \circ F \quad ^1$, and de ne the centering constants

$$\mu_n := \int_0^1 g(u) J_n(u)\,\mathrm{d}u = \int_{[0,1]} g(u)\,\mathrm{d}\Psi_n(u).$$

Consistency is a basic desirable property of statistical estimators. We can show that under very weak integrability conditions, we have $T_n \quad \mu_n \to 0$, a.s. This result was stated and proved in van Zwet (1980) for the i.i.d. case, but his proof remains to be valid for the ergodic case.

We also prove the asymptotic normality of T_n with conditions which are slightly weaker than the ones in the existing literature. We require that the observations $(X_n)_{n\in\mathbb{N}}$ be strongly mixing with a certain decaying rate of the mixing coefficient. We also assume the bounded growth of g and J_n, and smoothness of J_n. Let $C_k(u, v) := \mathrm{P}(\xi_1 \quad u, \xi_k \quad v)$ and put

$$(u, v) := u \wedge v \quad uv + \sum_{k=2}^{\infty}[C_k(u, v) \quad uv] + \sum_{k=2}^{\infty}[C_k(v, u) \quad uv].$$

Then we show that the asymptotic normality of the general L-statistic:

$$\sqrt{n}(T_n \quad \mu_n) \quad \xrightarrow{\mathcal{L}} \quad N(0, \quad ^2)$$

where

$$^2 := \int_0^1 \int_0^1 \quad (u, v) J(u) J(v)\,\mathrm{d}g(u)\mathrm{d}g(v) < \infty$$

When we try to construct approximate con dence intervals for L-statistics, we need to estimate the asymptotic variance. This asymptotic variance can be represented as the sum of the autocovariances of a certain stationary sequence of random variables Y_n de ned by

$$Y_n = \int_{\infty}^{\infty} [\mathbf{1}_{\{X_n \le x\}} \quad F(x)] J(F(x))\,\mathrm{d}h(x).$$

And it is in turn equal to 2 times the value of spectral density at frequency 0. The only difficulty is that we cannot observe Y_n because it involves the unknown df F. Then it is natural to replace it with the empirical df $\mathbb{F}_n$ and use instead

$$Y_{i,n} := \int_{\infty}^{\infty} [\mathbf{1}_{\{X_i \le x\}} \quad \mathbb{F}_n(x)] J(\mathbb{F}_n(x))\,\mathrm{d}h(x), \quad i = 1, \ldots, n.$$

Let

$$\widetilde{\gamma}_n(k) := \frac{1}{n}\sum_{i=1}^{n\ \ k} Y_{i,n}Y_{i+k,n} \quad \text{and} \quad \widetilde{f}_n(0) := \frac{1}{2\ }\sum_{|k|<K_n} w(k/K_n)\widetilde{\gamma}_n(k).$$

and

$$\widehat{f}_n(0) = \frac{1}{2\ }\sum_{|k|<K_n} w(k/K_n)\widehat{\gamma}_n(k) \tag{1}$$

where w is a so-called lag window. Then 2 $\widetilde{f}_n(0)$ give a consistent estimator of the asymptotic variance $\ ^2$ under certain regularity conditions.

We apply the above results to distortion risk measures of the following form

$$_D(X) := \int_{[0,1]} F^{\ \ 1}(u)\,\mathrm{d}D(u) = \int_{\mathbb{R}} x\,\mathrm{d}D\circ F(x), \tag{2}$$

where D is a convex distortion function, which is simply a df D on $[0,1]$ A natural estimator of $\ (X)$ is

$$\widehat{\ }_n = \int_0^1 \mathbb{F}_n^{\ 1}(u)\,\mathrm{d}D(u) = \sum_{i=1}^n c_{ni}X_{n:i}, \tag{3}$$

where $c_{ni} := D(i/n) \quad D((i \quad 1)/n)$, and this is a simple L-statistic.

It is easy to see that

$$\begin{aligned} \mathrm{E}\ \int_{[0,1]} \mathbb{F}_n^{\ 1}(u)\,\mathrm{d}D(u) \ &= \int_{\ \infty}^0 \mathrm{E}(\ \ D(\mathbb{F}_n(x)))\,\mathrm{d}x + \int_0^\infty \mathrm{E}[1 \quad D(\mathbb{F}_n(x))]\,\mathrm{d}x \\ \int_{\ \infty}^0 \ \ D(\mathrm{E}(\mathbb{F}_n(x)))\,\mathrm{d}x &+ \int_0^\infty [1 \quad D(\mathrm{E}(\mathbb{F}_n(x)))]\,\mathrm{d}x = \int_{[0,1]} F^{\ \ 1}(u)\,\mathrm{d}D(u) \end{aligned}$$

Therefore $\mathrm{E}(\widehat{\ }_n) \quad (X) \quad 0$, i.e., $\widehat{\ }_n$ has a negative bias. We show that the moving block bootstrap can be used to correct the bias with weakly dependent data.

In simulation study, we introduce the following simple stochastic volatility model to evaluate the bias and root mean squared error (RMSE). Let $X_t = \ _tZ_t$ and suppose that $V_t := 1/\ ^2_t$ follows the rst-order autoregressive gamma process introduced in Gaver and Lewis (1980):

$$V_t = \ \ V_{t\ \ 1} + \varepsilon_t,$$

where V_t has a gamma distribution with shape parameter and inverse-scale parameter β for each t, (ε_t) is a sequence of i.i.d. random variables, and $0 \quad < 1$. It is known that the distribution of ε_t is compound Poisson. Let (Z_t) be a sequence of independent random variables with standard normal distribution, which are also independent of (ε_t). Then it is well known that X_t has a scaled t-distribution with 2 degrees of freedom and scale paprameter $\beta/\ $. This allows us to calculate the true values of VaR, expected shortfall, and proportional odds risk measure.

The bias is not of serious size, and the moving block bootstrap seems to be working reasonably well. The estimated RMSEs are large probably reflecting the heavy tail of the t-distribution with four degrees of freedom. Although RMSE is slightly smaller for every risk measure in the i.i.d. case, there does not seem to be a big difference in the behavior of the estimates between in the stochastic volatility case and i.i.d. case, reflecting the quite weak dependence in this stochastic volatility model.

和歌山研究集会報告書

三浦良造　一橋大学名誉教授

発表題目。

「Asymptotic Normality of Estimators derived from Rank Statistics for Generalized Lehmann's Alternative Models when the Observations are a sequence of weakly dependent random variables: a General Model and special cases including Skew Symmetric models.」

発表概要。

　レーマン対立仮説は、順位検定統計量の確率分布を対立仮説においても計算できるモデルとして提案された(1953 年)。このモデルを一般化し、その中にあるパラメターを推定する方式を、観測列が iid,つまり独立で同一分布に従う確率変数列である場合について、提案したのが三浦論文(1985 年、1993 年)である。今回の講演では、この iid の仮定を緩め、観測列が弱い依存性を持つ場合に拡張する試みを紹介した。モデルはすべて 1 変量の場合である。観測値が独立で同一分布に従うという仮定の下で導かれたこれまでの成果を紹介しながら、その数学的証明が、弱い依存性を持つ観測列の経験分布関数の場合にもその収束が保証されるという近年の成果(1996，2000，2011 など)に基づき、独立同一分布性に基づくこれまでの成果を弱い依存性を持つ場合に拡張できることを説明した。

　和歌山におけるこの研究集会に先立つ京都での研究集会では変換パラメターの推定だけに関して発表したが、和歌山では、二つのパラメター、変換パラメターと位置パラメターの同時推定に関して発表した。さらに、単純線形回帰モデルの偶然誤差項が一般化されたレーマン対立仮説モデルに従う場合について、順位統計量に基づく統計的推論（推定と検定）が可能であるようにするための統計量の構成を紹介した。この統計量の漸近正規性の証明は現在検討中であるがこれまでの数学的方法をそのまま使って証明できる方向であることを報告した。

　さらに付け加えると、一般化されたレーマン対立仮説モデルは、現在よく研究されている skew symmetric distribution をその特殊形として含むので、その点の解説も行った。その視点からも一般化されたレーマン対立仮説モデルの研究が有意義であろうと思われる。

　　この問題分野に詳しい Marc Hallin 教授が出席されていたので、発表者の考え方と証明の方法を聴いていただき大変有意義であった。海外からの適切な参加者を得て発表者自身はこの研究集会への参加が大変有意義であったことを報告したい。

参考文献

：Miura.三浦（1985）. “Hodges-Lehmann　型推定と Lehmann 対立仮説 “　日本数学会春季予稿集　特別講演

：Tsukahara, H. & Miura, R.(1993). “One sample estimation for generalized Lehmann’s

alternative models."   Statistica Sinica. Vol.3. 83-101.
==
：Tsukahara, H. (2011). " Estimation of Distortion Risk Measure."
Working Paper. Department of Economics Seijo University.
: Shao,Q.M.&Yu,H. (1996).
Weak convergence for weighted empirical processes of dependent sequences,
Ann. Probab. 24 (4) (1996) 2052–2078.
: Louhichi,S.(2000). Weak convergence for empirical processes of associated sequences.
Ann. Inst. Henri Poincaré, Probabilités et Statistiques 36, 5 (2000) 547–567
==
：Azzalini, A.(1985) ."A Class of Distributions which includes the Normal Ones."
Scandinavian Journal of Statistics. Vol.12. 171-178.
：Azzalini, A.(2005)."The skew normal distribution and related multivariate families."
Scandinavian Journal of Statistics. Vol.32. 159-188.
==
：Lehmann, E.L.(1953) . "The power of rank tests." Annals of Mathematical Statistics. Vol.24. 23-43.
: Hallin,M. and his colleagues: "Serial and Nonserial sign-and-rank statistics: asymptotic representation and asymptotic normality."(2006) Ann.Statist

Empirically Effective Bond Pricing Model and Analysis on Term Structures of Implied Interest Rates in Financial Crisis

Takeaki KARIYA(GSB, Meiji University), kariya@kisc.meiji.ac.jp
Jingsui WANG (GSB, Meiji University), ouk@kisc.meiji.ac.jp
Zhu WANG (ICI)
Eiichi DOI (SST, Meiji University)
Yoshiro YAMAMURA (GSB, Meiji University), yyama@kisc.meiji.ac.jp

In his book(1993) Kariya proposed a government bond (GB) pricing model that simultaneously values individual fixed-coupon (non-defaultable) bonds of different coupon rates and maturities via a discount function approach, and Kariya and Tsuda (1994) verified its empirical effectiveness of the model as a pricing model for Japanese Government bonds (JGBs) though the empirical setting was limited to a simple case. In this paper we first clarify the theoretical relation between our stochastic discount function approach and the spot rate or forward rate approach in mathematical finance. Then we make a comprehensive empirical study on the capacity of the model in view of its pricing capability for individual GBs with different attributes and in view of its capacity of describing the movements of term structures of interest rates that JGBs imply as yield curves. Based on various tests of validity in a GLS (Generalized Least Squares) framework we propose a specific formulation with a polynomial of order 6 for the mean discount function that depends on maturity and coupon as attributes and a specific covariance structure. It is shown that even in the middle of the Financial Crisis, the cross-sectional model we propose is shown to be very effective for simultaneously pricing all the existing JGBs and deriving and describing zero yields.

This paper has two distinct objectives. One is to show that our stochastic discount function approach has a theoretical legitimacy in pricing non-defaultable bonds or equivalently government bonds (GBs) and deriving a term structure of interest rates in comparison with the interest rate approach in mathematical finance, which is rather dominant in pricing bonds and interest derivatives. Another objective is to show an empirical effectiveness of our cross-sectional model in pricing Japanese government bonds (JGBs) and describing variations of term structures of interest rates that JGBs imply. We also make a comprehensive empirical analysis on time series variations of implied interest rates and swap rates in the Financial Crisis Period.

The first objective is treated in Sections 2 and 3. A main feature of this model is that a stochastic realization of each individual GB price at present time 0 is viewed as equivalent to a realization at 0 of the whole stochastic process of the random cash-flow discount rate, which is defined on its term to maturity and depends on attributes of each individual bond such as coupon rate, term to maturity, etc. Another feature associated with the first one is that the cross-sectional correlation structure of all the GB prices at 0 is obtained through that of the corresponding random discount functions. Though this model is a cross-sectional model as it stands, it can be extended to certain types of dynamic models, which will be discussed elsewhere.

On the other hand, in time-continuous mathematical finance with arbitrage-free paradigm, pricing non-defaultable bonds is made to correspond to specifying a spot interest rate such as CIR (Cox-Ingersoll-Ross) model or forward interest rate model such as HJM (Heath-Jarrow-Morton) model. However, the model is inevitably diffusion (Markovian) model though actual interest rate processes in practice move with business cycles, which are not Markovian. It is remarked here that a diffusion interest rate model typically converges to a Martingale process with a constant mean if it converges. (see Chung and William (1990) p.97 for Vasicek model). In addition, such a single interest rate model will be unable to value all the existing GBs at 0 simultaneously with the correlations being taken into account. This is because it cannot describe differences of bond prices due to such attributes as coupon rate and maturity. In fact, it is often observed empirically that bond prices formed in the market are affected by the difference of these attributes, which is difficult to be taken into account in terms of interest rates as they are rather static. In modeling a swap rate process Collin-Dufrense and Solnik (2001) and Feldhutter and Lando (2007) take the dependence of swap rates on credit attributes into account and specify a swap rate process as the sum of an abstract risk-free rate process and a convenience yield process where they are assumed to be independent. Here the convenience yield is supposed to represent such attributes as liquidity premium, credit premium (collateral condition), etc. Recent developments in the area of interest rate models in mathematical finance are found in Brigo and Mercurio (2006) and Filipovic (2009). While, in Anderson, et al (1996) the derivation and estimation of yield curves in a traditional or practical approach is well exposited from a recent perspective.