データサイエンス
福島キャンプ2019

(新しい時系列計量分析の理論と応用) [1]

国友直人[2](編)

2019 年 8 月

[1]学術振興会・科学研究プロジェクト「新しい時系列計量分析の理論と応用」(2017 年度〜2020 年度) が主催した（2019 年 8 月 7 日, 福島大学) 研究集会 (明治大学先端数理科学インスティテュート (MIMS) 協賛）における講演集。

[2]明治大学政治経済学部

概要

経済や金融 (ファイナンス) などで観察されるデータなどを分析対象とする統計的時系列分析では幾つかの検討するべき基本的な問題がなお存在している。例えば経済において「通常の常識では起こりにくいとされる事象」についてのリスク解析や対策の重要性についての認識が高まっているが、実際に 2008 年に起きたリーマンショックや 2011 年ごろに発生したヨーロッパ諸国の金融危機などが顕著な例として挙げることができる。国際的に連動している現代の経済・金融市場では従来の計量分析ではほとんど考慮されてこなかった変動を経験しているのである。社会・経済を理解し、より良いものにしていくには、事前には予想が困難である自然災害や経済変動における稀ではあるが実際に起きると大きな影響のある不確実な事象を科学的に分析し、有効な対策を考察することが必要であり重要となっている。また近年に明らかになりつつあるように高頻度金融データやマクロ経済データの分析においては、経済変数間の関係の統計的分析が重要であるが、確率過程の統計分析や非定常多次元時系列分析にはなお解決すべき課題が少なくない。実は考察の範囲を広げて自然や生物を研究対象とする分野でも似たような現象が観察され、共通する問題と固有の統計的問題がある。

科学研究プロジェクト「新しい時系列計量分析の理論と応用」では主に経済・金融・社会における新しい動向を背景として、近年の日本経済・社会の理解の方法として重要な新しい計量分析の方法を開発、応用を検討している。特に金融現象やマクロ経済の統計分析では、時々起きる大きな経済変動は重要であるにもかかわらず、なお研究の蓄積が不十分であり、様々な研究の可能性がある。また、ミクロ金融の分析ではジャンプや確率過程の一般理論を踏まえた金融時系列分析はなお十分とは言えず、マイクロマーケット分析や確率過程の計量分析の方法を確立が望まれる。こうした経済・金融におけるマクロ分野、ミクロ分野と云う二つの時系列分析において新しい分析の枠組みを構築する必要がある。またマクロ経済データが象徴的であるように、経済・金融分野で観察されるデータは統計的には非定常性が見られるとともに本来的に多次元時系列であり、次元数も必ずしも小さくない場合も研究対象である。

本年度は研究プロジェクトの三年目であり、特にファイナンス市場の分析では重要な高次元データ分析、マイクロマーケットの実証分析、から計量経済分析の問題やデータサイエンスの問題をなど含めて活発な議論を行った。ここに収録した研究報告が統計学・データサイエンス、計量経済学、計量ファイナンス、経済時系列分析、などの理論と応用を展開する一助になることを期待する。

2019 年 9 月
編者

研究集会・プログラム

科学研究プロジェクト「新しい時系列計量分析の理論と応用」
日程：2019 年 8 月 7 日 (水)
会場：福島大学経済経営棟 5F
オーガナイザー：国友直人
プロジェクト参加者：国友直人 (明治大学)・大屋幸輔 (大阪大学)・佐藤整尚 (東京大学)・栗栖大輔 (東京工業大学)
会場責任者：井上健 (福島大学)
協賛：明治大学先端数理科学インスティテュート (MIMS)

< セッション I：高次元データ分析 >
Chair: 大屋幸輔
13:00〜13:50「高頻度データにおける高次元共分散行列の統計推測」
　　　　小池裕太 (東京大学)
13:50〜14:20「ノイズを含む高頻度高次元ボラティリティ分析」
　　　　国友直人

< 休憩 >

< セッション II：Financial Economics and Econometrics>
Chair: 三崎広海 (筑波大学)
14:30〜15:10「Estimation of risk aversion for Japanese stock market using implied moments」
　　　　大屋幸輔
15:10〜15:50 「指値注文市場における気配設定」
　　　　太田亘 (大阪大学)

< 休憩 >

< セッション III：Econometrics>
Chair:国友直人
16:00〜16:40「A two-sample alternative to using instruments in regressions with omitted variables」
　　　　蛭川雅之 (龍谷大学)
16:40〜17:10「Nonparametric estimation of density functions from repeated measurements」
　　　　栗栖大輔
17:10〜17:40「周波数回帰と経済データへの応用」
　　　　佐藤整尚

高頻度データにおける高次元共分散行列の統計推測

小池祐太

東京大学 MI センター・大学院数理科学研究科
CREST JST

2019 年 8 月 7 日

データサイエンス・福島キャンプ 2019
@ 福島大学

背景

- $Y_t = (Y_t^1, \dots, Y_t^d)^\top$ $(t \in [0,1])$: d 個の資産の対数価格過程
- $Y = (Y_t)_{t\in[0,1]}$ は確率基底 $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ 上で定義された連続 Itô セミマルチンゲールと仮定:

$$Y_t = Y_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dB_s, \qquad t \in [0,1].$$

  - $\mu = (\mu_s)_{s\in[0,1]}$: d 次元 $(\mathcal{F}_t)$-発展的可測過程
  - $\sigma = (\sigma_s)_{s\in[0,1]}$: $\mathbb{R}^{d\times r}$ 値 $(\mathcal{F}_t)$-発展的可測過程
  - $B = (B_s)_{s\in[0,1]}$: r 次元標準 $(\mathcal{F}_t)$-Brown 運動
  - 適当なモーメント条件を課す

背景

- Y は等間隔な時点 $t_h = t_h^n = h/n,\ h = 0, 1, \ldots, n$ で観測されているとする
- 目的 観測データ $(Y_{t_h})_{h=0}^n$ に基づいて, Y の**累積共分散行列**

$$[Y, Y]_1 = \int_0^1 \Sigma_t dt, \qquad \Sigma_t := \sigma_t \sigma_t^\top$$

を $n \to \infty$ のときに推定する (高頻度観測の設定)

- 最も自然な $[Y, Y]_1$ の推定量として, **実現共分散行列**

$$\widehat{[Y, Y]}_1^n := \sum_{h=1}^n (Y_{t_h} - Y_{t_{h-1}})(Y_{t_h} - Y_{t_{h-1}})^\top$$

が考えられる

背景

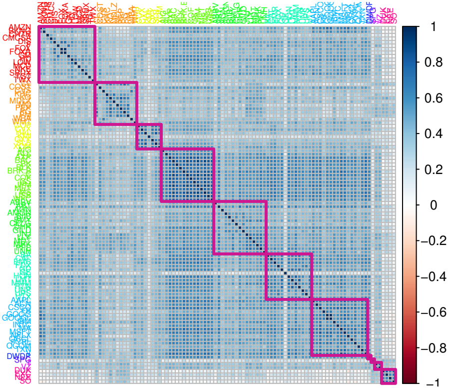
- 近年, 高頻度データ解析の分野でも, $n \to \infty$ のとき $d \to \infty$ なる高次元の設定で $[Y, Y]_1$ を推定する研究が活発に行われている
  - 応用: 最小分散ポートフォリオの重みの計算など
  - マーケットマイクロストラクチャー効果の影響で, サンプル数 n はせいぜい 400 程度しかとれないが, もし $d = 100$ ならば, 推定すべきパラメーター数は $d(d+1)/2 \approx 5000$ と n に比べて膨大となる!!
- 上の設定は特に d 変量正規確率変数を n 個独立に観測する状況を含み, その場合 $\widehat{[Y, Y]}_1^n$ は標本共分散行列に対応するから, 前述と同様の問題がやはり生じる
- 他方, 問題の解決方法も基本的には i.i.d. の設定のアナロジーで行われる: 累積共分散行列に何らかの意味でのスパース性を課す!!
- 次頁に代表的な先行研究を挙げる (他にもたくさんある)

代表的な先行研究

- **Wang and Zou (2010)**
  - $[Y, Y]_1$ の成分の多くが 0, もしくは非常に小さいというスパース性を課して, バンド化/閾値法によってナイーブな推定量を正則化する
  - Bickel and Levina (2008a,b) の高頻度版
- **Fan, Furger, and Xiu (2016)**
  - Y が連続時間のファクターモデルに従うと仮定し, その残差過程の共分散行列がスパースであると仮定して, ファクター部分は線形回帰, 残差部分は正則化で共分散行列を推定し, それを合算する
  - Fan et al. (2011) の高頻度版
- **Brownlees, Nualart, and Sun (2018)**
  - $[Y, Y]_1$ の逆行列にスパース性を課して, graphical Lasso によって $[Y, Y]_1$ の逆行列を推定する
  - Yuan and Lin (2007) の高頻度版 (高次元の理論は Ravikumar et al. (2011) に依拠)

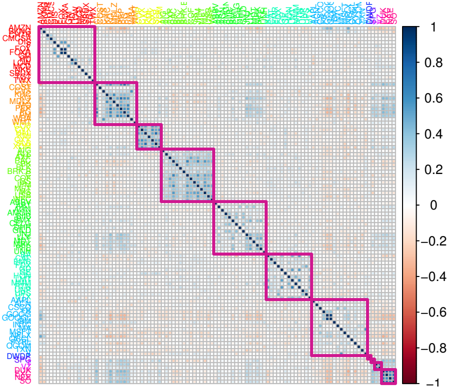
...しかし, そのようなスパース性の仮定の妥当性はどのように検証すればよいか?

図 1: S&P100 指数構成銘柄の実現相関行列 (2018 年 3 月)



15 分足リターンで計算. 長方形はセクターを表す

図 2: マーケットファクターで回帰後の残差過程の実現相関行列 (2018 年 3 月)



15 分足リターンで計算．長方形はセクターを表す．マーケットファクターは SPY で代理

連続時間ファクターモデルの残差スパース性検定

- ここでは Fan et al. (2016) の仮定に着目してみる (Wang and Zou (2010) の仮定は特殊ケースと考えられる)
- d 番目の資産 Y^d をファクター過程とみなし, 以下の連続時間 1 ファクターモデルを考える:

$$Y^j = \beta^j Y^d + R^j, \qquad j = 1, \dots, \underline{d} := d - 1. \tag{1}$$

  - β^j: 定数 (因子負荷量)
  - R^j: $[R^j, Y^d] \equiv 0$ を満たすセミマルチンゲール (残差過程)
- 各ペア (i, j) について, 以下の統計的仮説検定を考える:

$$H_0^{(i,j)} : [R^i, R^j]_1 = 0 \quad \text{a.s.} \quad \text{vs} \quad H_1^{(i,j)} : [R^i, R^j]_1 \neq 0 \quad \text{a.s.} \tag{2}$$

連続時間ファクターモデルの残差スパース性検定

- ある特定のペア (i,j) に着目した場合の検定 (2) は Bibinger and Mykland (2016) においてすでに考察されており, 次の検定統計量が提案されている:

$$T_n^{(i,j)} := \frac{\sqrt{n}\hat{\mathfrak{T}}_n^{ij}}{\sqrt{\hat{\mathfrak{V}}_n^{ij}}},$$

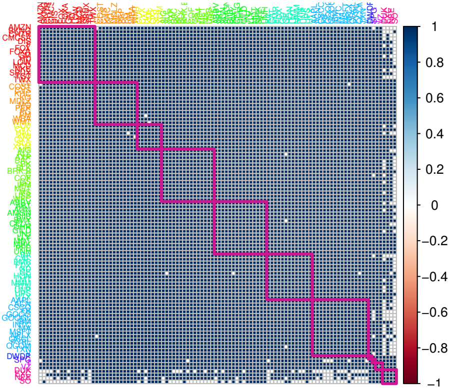
ここに,

$$\hat{\mathfrak{T}}_n^{ij} := \widehat{[Y^i, Y^d]}_1^n \widehat{[Y^j, Y^d]}_1^n - \widehat{[Y^i, Y^j]}_1^n \widehat{[Y^d, Y^d]}_1^n$$

で $\hat{\mathfrak{V}}_n^{ij}$ は $\hat{\mathfrak{T}}_n^{ij}$ の漸近分散の推定量

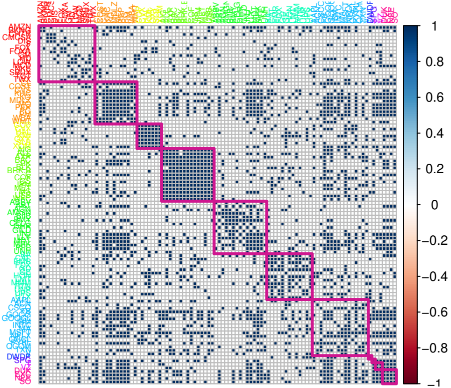
- 適当な正則条件を仮定すると, $H_0^{(i,j)}$ の下で $T_n^{(i,j)} \to^{\mathcal{L}} N(0,1)$ となり, $H_1^{(i,j)}$ の下で $|T_n^{(i,j)}| \to^p \infty$ となることが証明されている

図 3: S&P100 指数構成銘柄の実現相関行列 (2018 年 3 月): 5%水準で有意な成分=1, それ以外=0



15 分足リターンで計算. 長方形はセクターを表す

図 4: マーケットファクターで回帰後の残差過程の実現相関行列 (2018 年 3 月): 5%水準で有意な成分=1, それ以外=0



15 分足リターンで計算. 長方形はセクターを表す. マーケットファクターは SPY で代理

連続時間ファクターモデルの残差スパース性検定

- 残差過程の共分散行列の正則化推定のために必要なスパース性を検証するには, 仮説検定 (2) を $(i,j) \in \Lambda_n := \{(i,j) : 1 \leq i < j \leq d\}$ について同時に実行する必要がある
  - どの成分が有意かどうかも知りたい $\Longrightarrow$ 多重検定の問題
- 特に, この多重検定問題の family-wise error rate (FWER) をコントロールしたい
  - FWER = 正しい帰無仮説を 1 つでも誤って棄却してしまう確率
- FWER のコントロールを達成するために, Romano and Wolf (2005) で提案されたステップダウン法を利用したい
  - アルゴリズムの詳細については Romano and Wolf (2005) を参照

$\Rightarrow$ $\mathcal{L} \subset \Lambda_n$ が与えられたとき, 統計量 $\max_{(i,j)\in\mathcal{L}} |T_n^{(i,j)}|$ の分布を近似する方法が必要

- この目的は以下のようなタイプの CLT を証明すると達成できる (“**高次元漸近混合正規性**”):

$$\sup_{y \in \mathbb{R}^m} \left| P\left(\Xi_n S_n \leq y\right) - P\left(\Xi_n \widehat{\mathfrak{C}}_n^{1/2} \zeta_n \leq y\right) \right| \to 0 \qquad (3)$$

  - $S_n := \sqrt{n}\,\mathrm{vec}\left(\widehat{[Y,Y]}_1^n - [Y,Y]_1\right)$
  - $\widehat{\mathfrak{C}}_n$ は S_n の漸近共分散行列の推定量:

$$\widehat{\mathfrak{C}}_n := n \sum_{h=1}^{n} \chi_h \chi_h^\top - \frac{n}{2} \sum_{h=1}^{n-1} \left(\chi_h \chi_{h+1}^\top + \chi_{h+1} \chi_h^\top\right)$$

    ただし, $\chi_h := \mathrm{vec}\left[(Y_{t_h} - Y_{t_{h-1}})(Y_{t_h} - Y_{t_{h-1}})^\top\right]$
  - ζ_n は $\mathcal{F}$ と独立な d^2 次元標準正規確率ベクトル
  - Ξ_n は $m \times d^2$ ランダム行列で, 行ごとにスパースなもの (厳密な定式化は後述); m は n に依存しうる

Jacod's Paradise

- もし m, d ともに n に依存しないのであれば, (3) 式は **Jacod の安定収束理論**の帰結として得られる

定理 1 (Jacod and Protter (2012), Theorem 5.4.2)

確率 1 で $\int_0^1(\|\mu_s\|_{\ell_2}^2 + \|\Sigma_s\|_{\ell_2}^2)ds < \infty$ が成り立つならば, S_n は $n \to \infty$ のとき $\bar{\mathfrak{C}}^{1/2}\zeta$ に安定収束する. ただし, ζ は $\mathcal{F}$ と独立な d^2 次元標準正規確率ベクトル, $\bar{\mathfrak{C}} = (\bar{\mathfrak{C}}^{pq})_{1\leq p,q\leq d^2}$ は d^2 次非負定値対称行列に値をとるランダム行列で, その成分は次式で与えられる:

$$\bar{\mathfrak{C}}^{d(p-1)+q,d(p'-1)+q'} = \int_0^T \left(\Sigma_t^{pp'}\Sigma_t^{qq'} + \Sigma_t^{pq'}\Sigma_t^{p'q}\right) dt, \qquad p, q, p', q' = 1, \dots, d.$$

定理 1 の仮定は $\mu_s \equiv 0$ の場合は必要条件!!

Paradise Lost

- しかし, $n \to \infty$ のとき $d \to \infty$ となる高次元の設定下では, Jacod の理論は完全に破綻する
  - Jacod の定理の証明は, Skorohod 位相に関する緊密性の議論とあるマルチンゲール問題の解の一意性に強く依拠している
  - しかし, $d \to \infty$ のとき, S_n は通常の位相に関してはもはや緊密でない!
- 高次元設定下での CLT の証明は, 実質的には分布間の距離に対する定量的評価の問題となり, 一歩先の評価が必要
  - 次元に対してリーズナブルな上界が欲しい
  - 「分布間の距離としてどれを採用するか」も本質的
- 本研究の貢献 $d \to \infty$ のときに (3) 式のような収束を証明するための別の手段を構築

Chernozhukov-Chetverikov-Kato (CCK) 理論

- 最近, Chernozhukov, Chetverikov, and Kato (2013, 2015, 2017) の一連の研究で, (3) のような形の収束を証明するための新たな理論が開発された
  - Chernozhukov-Chetverikov-Kato 理論, 略して CCK 理論と呼ぶことにする
- オリジナルの CCK 理論は, $\Xi_n S_n$ が独立な確率ベクトルの和, もしくは Gauss 型で, かつ $\Xi_n \widehat{\mathfrak{C}}_n^{1/2}$ が非ランダムでなければ適用できない
- しかし, 証明のほとんどの部分では上述の数学的構造は使われない
  - 証明の一部を修正することで我々の状況にも適用できる!

CCK 理論: Gaussian comparison の場合

- 以下 $m \geq 3$ と仮定する

定理 2 (Chernozhukov et al. (2015), Theorem 2)

F, Z を平均 0 の m 次元 Gauss 型確率変数とし，ある定数 $\underline{\sigma} > 0$ が存在して $\min_j E[Z_j^2] \geq \underline{\sigma}^2$ が成り立つと仮定する．このとき，$\underline{\sigma}$ のみに依存する定数 $C > 0$ が存在して，

$$\sup_{x \in \mathbb{R}^m} |P(F \leq x) - P(Z \leq x)| \leq C \Delta^{1/3} \log^{2/3} m$$

が成り立つ．ただし，

$$\Delta = \max_{1 \leq i,j \leq m^2} |E[F_i F_j] - E[Z_i Z_j]|$$

CCK 理論: 問題の分割

- 集合 $A \subset \mathbb{R}$ と $\varepsilon > 0$ に対して以下のように定義:

$$A^\varepsilon := \{u \in \mathbb{R} : |u - v| \leq \varepsilon \text{ for some } v \in A\}$$

- 次の補題は初等的議論で示せる:

補題 1

F, Z を m 次元確率変数とし, ある定数 $\varepsilon, \eta > 0$ が存在して, 任意の $y \in \mathbb{R}^m$ と Borel 集合 $A \subset \mathbb{R}$ に対して

$$P\left(\max_{1\leq j\leq m}(F_j - y_j) \in A\right) \leq P\left(\max_{1\leq j\leq m}(Z_j - y_j) \in A^\varepsilon\right) + \eta$$

が成り立つと仮定する. このとき次が成り立つ:

$$\sup_{x\in\mathbb{R}^m}|P(F \leq x) - P(Z \leq x)| \leq \sup_{y\in\mathbb{R}^m} P\left(0 \leq \max_{1\leq j\leq m}(Z_j - y_j) \leq \varepsilon\right) + \eta.$$

CCK 理論: 反集中不等式

- Z が Gauss 型の場合, 補題 1 の最後の不等式の右辺第 1 項は次の形の**反集中不等式**を用いて次元に関してリーズナブルに抑えられる:

補題 2 (Nazarov の不等式 (Chernozhukov et al. (2017), Lemma A.1))

Z を m 次元 Gauss 型確率変数とし, ある定数 $\underline{\sigma} > 0$ が存在して $\min_j E[Z_j^2] \geq \underline{\sigma}^2$ が成り立つと仮定する. このとき, $\underline{\sigma}$ のみに依存する定数 $C > 0$ が存在して, 任意の $\varepsilon > 0$ に対して

$$\sup_{y \in \mathbb{R}^m} P\left(0 \leq \max_{1 \leq j \leq m}(Z_j - y_j) \leq \varepsilon\right) \leq C\varepsilon\sqrt{\log m}$$

が成り立つ.

CCK 理論: 最大値関数の近似

- $\beta > 0$ に対して, 関数 $\Phi_\beta : \mathbb{R}^m \to \mathbb{R}$ を

$$\Phi_\beta(x) = \beta^{-1} \log \left(\sum_{j=1}^{m} e^{\beta x_j} \right) \qquad (x = (x_1, \dots, x_m)^\top \in \mathbb{R}^m)$$

  で定義する

- 任意の $x \in \mathbb{R}^m$ に対して

$$0 \le \Phi_\beta(x) - \max_{1 \le j \le m} x_j \le \beta^{-1} \log m \tag{4}$$

  が成り立つ (cf. Chernozhukov et al. (2015), Eq.(1))

CCK 理論: 指示関数の近似

- 任意に $\varepsilon > 0$, $y \in \mathbb{R}^m$ と Borel 集合 $A \subset \mathbb{R}$ をとる
- $\beta = \varepsilon^{-1} \log m$ (従って $\varepsilon = \beta^{-1} \log m$) とおけば, (4) より

$$P\left(\max_j (F_j - y_j) \in A\right) \leq P(\Phi_\beta(F - y) \in A^\varepsilon) = E[1_{A^\varepsilon}(\Phi_\beta(F - y))]$$

が成り立つ

補題 3 (Chernozhukov et al. (2016), Lemma 5.1)

任意の $\varepsilon > 0$ と任意の Borel 集合 $A \subset \mathbb{R}$ に対して, 次の条件を満たす C^∞ 関数 $g : \mathbb{R} \to \mathbb{R}$ が存在する:

(i) ある普遍定数 $K > 0$ が存在して $\|g'\|_\infty \leq \varepsilon^{-1}$, $\|g''\|_\infty \leq K\varepsilon^{-2}$ および $\|g'''\|_\infty \leq K\varepsilon^{-3}$ が成り立つ.

(ii) 任意の $x \in \mathbb{R}$ に対して $1_A(x) \leq g(x) \leq 1_{A^{3\varepsilon}}(x)$ が成り立つ.

CCK 理論: 期待値の差の評価への帰着

- 補題 3 を $A = A^\varepsilon$ として適用すれば,

$$
\begin{aligned}
&E[1_{A^\varepsilon}(\Phi_\beta(F-y))] \le E[g(\Phi_\beta(F-y))], \\
&E[g(\Phi_\beta(Z-y))] \le E[1_{A^{4\varepsilon}}(\Phi_\beta(Z-y))] \\
&\le E\left[1_{A^{5\varepsilon}}\left(\max_j (Z_j - y_j)\right)\right] = P\left(\max_j (Z_j - y_j) \in A^{5\varepsilon}\right)
\end{aligned}
$$

  を得る

- 以上より, $\underline{\sigma}$ のみに依存する定数 $c_1 > 0$ が存在して,

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}^m} |P(F \le x) - P(Z \le x)| \\
&\le c_1 \varepsilon \sqrt{\log m} + \sup_{y\in\mathbb{R}^m} |E[g(\Phi_\beta(F-y))] - E[g(\Phi_\beta(Z-y))]|
\end{aligned}
$$

  が成り立つ

- 以下に注意:
  - 以上の議論は任意の確率変数 F に対して成立!
  - $g \circ \Phi_\beta$ は C^∞ 級関数!
- 更に, $g \circ \Phi_\beta$ の高次偏導関数の ℓ_1-ノルムは「次元に関して効率的に」評価できる:

補題 4 (Chernozhukov et al. (2014), Lemma 4.3)

$g : \mathbb{R} \to \mathbb{R}$ を C^3 級関数とする. このとき, 任意の $x \in \mathbb{R}^m$ に対して

$$\sum_{i,j=1}^{m} \left| \frac{\partial^2 (g \circ \Phi_\beta)}{\partial x_i \partial x_j}(x) \right| \leq \|g''\|_\infty + 2\beta \|g'\|_\infty,$$

$$\sum_{i,j,k=1}^{m} \left| \frac{\partial^3 (g \circ \Phi_\beta)}{\partial x_i \partial x_j \partial x_k}(x) \right| \leq \|g'''\|_\infty + 6\beta \|g''\|_\infty + 6\beta^2 \|g''\|_\infty$$

が成り立つ

CCK 理論: Stein の方法

- **Stein の方法**によって $|E[g(\Phi_\beta(F-y))]-E[g(\Phi_\beta(Z-y))]|$ を評価する

補題 5 (Stein の等式)

G を平均 0 の m 次元 Gauss 型確率変数とする．このとき，有界な偏導関数をもつ任意の C^1 級関数 $f:\mathbb{R}^m\to\mathbb{R}$ に対して

$$E[G_i f(G)] = \sum_{j=1}^{m} E[G_iG_j] E\left[\frac{\partial f}{\partial x_j}(G)\right] \quad (i=1,\dots,m)$$

CCK 理論: Slepian 補間 (スマート・パス法)

- 一般性を失わずに F と Z は互いに独立と仮定してよい
- 補題 4 より $\psi := g \circ \Phi_\beta$ の 2 階の偏導関数はすべて有界
- 関数 $\Psi : (0,1) \to \mathbb{R}$ を $\Psi(t) = E[\psi(\sqrt{t}F + \sqrt{1-t}Z - y)]$ $(t \in (0,1))$ で定義すると, Ψ は微分可能であり, 各 $t \in (0,1)$ について

$$\Psi'(t) = \frac{1}{2}\sum_{i=1}^{m} E\left[\left(\frac{F_i}{\sqrt{t}} - \frac{Z_i}{\sqrt{1-t}}\right)\frac{\partial\psi}{\partial x_i}(\sqrt{t}F + \sqrt{1-t}Z - y)\right]$$

- F が Gauss 型ならば, Stein の補題より

$$\Psi'(t) = \frac{1}{2}\sum_{i,j=1}^{m}\{E[F_iF_j] - E[Z_iZ_j]\}\, E\left[\frac{\partial^2\psi}{\partial x_i\partial x_j}(\sqrt{t}F + \sqrt{1-t}Z - y)\right]$$

CCK 理論: 証明の完成

- 従って, 微分積分学の基本定理と補題 4 より, ある普遍定数 $c_2 > 0$ が存在して,

$$
\begin{aligned}
|E[g(\Phi_\beta(F-y))] - E[g(\Phi_\beta(Z-y))]| &= |\Psi(1) - \Psi(0)| \\
&\le c_2(\varepsilon^{-2} + \varepsilon^{-1}\beta)\Delta \\
&\le 2c_2\varepsilon^{-2}(\log m)\Delta
\end{aligned}
$$

- 以上より

$$
\sup_{x \in \mathbb{R}^m} |P(F \le x) - P(Z \le x)| \le c_1 \varepsilon \sqrt{\log m} + 2c_2\varepsilon^{-2}(\log m)\Delta
$$

を得る

- $\varepsilon = \Delta^{1/3} \log^{1/6} m$ として定理 2 の結論を得る

Malliavin 解析

- 上記の議論からわかるように, F が Gauss 型であるという仮定は Stein の方法の適用にのみ必要
- 言い換えれば, F が Stein の方法を有効に適用できるクラスに属してさえいれば, 上の議論はそのまま成立する $\Rightarrow$ Malliavin 解析!!
  - Malliavin の部分積分公式は Stein の等式の無限次元版 (Sakamoto and Yoshida, 1999, 2004)
  - "Malliavin-Stein method" (Nourdin and Peccati, 2009)
- $F = \int_0^1 u_s dB_s$ と書ける場合を考える (u: $E[\int_0^1 u_s^2 ds] < \infty$ なる $\mathbb{R}^{m\times r}$ 値発展的可測過程)

補題 6

$G \in \mathbb{D}_{1,2}(\mathbb{R}^m)$ ならば, 有界な偏導関数をもつ任意の C^1 級関数 $f : \mathbb{R}^m \to \mathbb{R}$ に対して

$$E\left[F_i f(G)\right] = \sum_{j=1}^{m} E\left[\left(\int_0^1 u_s^{i\cdot} \cdot D_s G_j ds\right) \frac{\partial f}{\partial x_j}(G)\right] \quad (i = 1, \dots, m)$$

Nualart-Yoshida 補間

- 極限の混合正規性は問題となるか?
  - $Z = \mathfrak{C}^{1/2}\zeta$, $\zeta \sim N(0,1) \perp\!\!\!\perp \mathcal{F}$, $\mathfrak{C}$: m 次半正定値 $\mathcal{F}$-可測ランダム行列

  $\Longrightarrow$ 問題となる！
  - F と Z はもはや独立でない！
  - 上の議論をそのまま成立させるには, $\mathfrak{C}$ が対角行列でないとうまくいかない (cf. Nourdin et al., 2016)
- この問題は, $E[\psi(F-y)]$ と $E[\psi(Z-y)]$ の補間を周波数領域で行えば解決できる (Nualart and Yoshida, 2017)
  - F と Z の特性関数どうしを補間して, 反転公式を使う
- $\theta \in [0,1]$ と $\mathrm{z} \in \mathbb{R}^m$ に対して,

$$\lambda(\theta;\mathrm{z}) := \theta\sqrt{-1}\mathrm{z}\cdot F - (1-\theta^2)\frac{\mathrm{z}^\top\mathfrak{C}\mathrm{z}}{2}, \quad \varphi(\theta;\mathrm{z}) := E[e^{\lambda(\theta;\mathrm{z})}]$$

とおくと, $\varphi(1;\mathrm{z}), \varphi(0;\mathrm{z})$ がそれぞれ F, Z の特性関数となる

Nualart-Yoshida 補間

$$\frac{\partial \varphi}{\partial \theta}(\theta;\mathrm{z}) = E[e^{\lambda(\theta;\mathrm{z})}(\sqrt{-1}\mathrm{z}\cdot F + \theta \mathrm{z}^\top \mathfrak{C}\mathrm{z})]$$

であり, 右辺第 1 項に補題 6 を適用して整理すると,

$$\frac{\partial \varphi}{\partial \theta}(\theta;\mathrm{z}) = \theta E\left[e^{\lambda(\theta;\mathrm{z})}\sum_{i,j=1}^{m} \mathrm{z}_i\mathrm{z}_j\left((\sqrt{-1})^2\int_0^1 u_s^{i\cdot}\cdot D_sF_j ds + \mathfrak{C}_{ij}\right)\right]$$
$$-\sqrt{-1}\frac{1-\theta^2}{2}E\left[e^{\lambda(\theta;\mathrm{z})}\sum_{i,j,k=1}^{m} \mathrm{z}_i\mathrm{z}_j\mathrm{z}_k\left(\int_0^1 u_s^{i\cdot}\cdot D_s\mathfrak{C}_{jk} ds\right)\right] \quad (5)$$

($F_j, \mathfrak{C}_{jk} \in \mathbb{D}_{1,2}$ と仮定)

Nualart-Yoshida 補間

ψ が十分滑らかならば, Fourier の反転公式と微分積分学の基本定理より

$$
\begin{aligned}
&E[\psi(F-y)] - E[\psi(Z-y)] \\
&= \frac{1}{(2\pi)^m}\int_{\mathbb{R}^m} \hat{\psi}(z) e^{\sqrt{-1}z\cdot y}\{\varphi(1;z)-\varphi(0;z)\}dz \\
&= \frac{1}{(2\pi)^m}\int_{\mathbb{R}^m} \left(\int_0^1 \hat{\psi}(z) e^{\sqrt{-1}z\cdot y}\frac{\partial\varphi}{\partial\theta}(\theta;z)d\theta\right)dz
\end{aligned}
$$

($\hat{\psi}$ は ψ の Fourier 変換)

(5) 式を代入したあと再び反転公式を適用して以下を得る:

$$
\begin{aligned}
&E[\psi(F-y)] - E[\psi(Z-y)] \\
&= \int_0^1 \theta E\left[\sum_{i,j=1}^m \frac{\partial^2\psi}{\partial x_i \partial x_j}(\theta F + \sqrt{1-\theta^2}Z)\left(\int_0^1 u_s^{i\cdot}\cdot D_s F_j ds - \mathfrak{C}_{ij}\right)\right] d\theta \\
&+ \int_0^1 \frac{1-\theta^2}{2} E\left[\sum_{i,j,k=1}^m \frac{\partial^3\psi}{\partial x_i \partial x_j \partial x_k}(\theta F + \sqrt{1-\theta^2}Z)\left(\int_0^1 u_s^{i\cdot}\cdot D_s \mathfrak{C}_{jk} ds\right)\right] d\theta
\end{aligned}
\tag{6}
$$

従って, 補題 4 より, ある普遍定数 $c_3 > 0$ が存在して,

$$
\begin{aligned}
&|E[\psi(F-y)] - E[\psi(Z-y)]| \\
&\leq c_3 \varepsilon^{-2}(\log m) E\left[\max_{1\leq i,j\leq m}\left|\int_0^1 u_s^{i\cdot}\cdot D_s F_j ds - \mathfrak{C}_{ij}\right|\right] \\
&\quad + c_3 \varepsilon^{-3}(\log m)^2 E\left[\max_{1\leq i,j\leq m}\left|\int_0^1 u_s^{i\cdot}\cdot D_s \mathfrak{C}_{jk} ds\right|\right]
\end{aligned}
$$

CCK 理論: 混合正規極限の場合

まとめると以下を得る:

定理 3

$F, \mathfrak{C}, m$ はパラメーター $n \in \mathbb{N}$ に依存して, 以下の条件を満たすとする:

(i) $F \in \mathbb{D}_{1,2}(\mathbb{R}^m), \mathfrak{C} \in \mathbb{D}_{1,2}(\mathbb{R}^{m \times m})$

(ii) $(\log m)^2 E\left[\max_{1 \le i,j \le m}\left|\int_0^1 u_s^{i\cdot} \cdot D_s F_j ds - \mathfrak{C}_{ij}\right|\right] \to 0\ (n \to \infty)$

(iii) $(\log m)^{7/2} E\left[\max_{1 \le i,j \le m}\left|\int_0^1 u_s^{i\cdot} \cdot D_s \mathfrak{C}_{jk} ds\right|\right] \to 0\ (n \to \infty)$

(iv) $\lim_{b \downarrow 0} \limsup_{n \to \infty} P(\min_{1 \le j \le m} \mathfrak{C}_{jj} < b) = 0$

このとき, $\sup_{x \in \mathbb{R}^m} |P(F \le x) - P(Z \le x)| \to 0\ (n \to \infty)$ が成り立つ

CCK 理論: 混合正規極限の場合

定理 3 に関するコメント

- S_n は伊藤積分で書けるので定理 3 は適用可能だが, (3) を得るには不十分
  $\Longrightarrow$ S_n は実際には二重伊藤積分なので, 補題 6 を 2 回適用して評価を精密にする必要がある
  - 詳細は K. (2019a) 参照 (Ξ_n は上と同様の議論で対処できる)
  - もしくは, (6) 式を Nualart and Yoshida (2017) に従って整理すると, (3) を得るのに十分な結果を得られるかもしれない
- 定理 3 において, 実際には F は Skorohod 積分であれば十分 (anticipative でもよい)
  - 特に F のマルチンゲール性は使われていない (これは非効率的?)

主結果: 記号

- 確率変数 ξ と $p > 0$ に対して $\|\xi\|_p := (E[|\xi|^p])^{1/p}$
- ランダム配列 $F = (F_{i_1,\dots,i_q})_{1\le i_1,\dots,i_q\le r}$ と $p > 0$ に対して

$$\|F\|_{p,\ell_2} := \left\| \sqrt{\sum_{i_1,\dots,i_q=1}^{r} F_{i_1,\dots,i_q}^2} \right\|_p$$

- $m \times l$ 行列 A に対して $|||A|||_\infty := \max_{1\le i\le m} \sum_{j=1}^{l} |A^{ij}|$
- 正方行列 A の対角成分の最小値を $\min \operatorname{diag}(A)$ で表す
- 以下では B に関する (偏)Malliavin 解析を考える
- ある非ランダムな $m \times d^2$ 行列 Υ_n と $m \times d^2$ ランダム行列 $\boldsymbol{X}_n$ が存在して $\Xi_n = \Upsilon_n \circ \boldsymbol{X}_n$ と書けると仮定 ($\circ$ は成分ごとの積の意) (Ξ_n の「スパース性」に関する仮定)

定理 4 (K. (2019a), Theorem 4.1)

すべての $t \in [0,1]$ について $\mu_t \in \mathbb{D}_{1,\infty}(\mathbb{R}^d)$, $\sigma_t \in \mathbb{D}_{2,\infty}(\mathbb{R}^{d\times r})$ および $\boldsymbol{X}_n \in \mathbb{D}_{2,\infty}(\mathbb{R}^{m\times d^2})$ であると仮定する．また, $|||\Upsilon_n|||_\infty \geq 1$, $|||\Upsilon_n|||_\infty^5 = O(n^\varpi)$ と

$$\max_{1\leq i\leq m}\max_{1\leq j\leq d^2}\left(\|\boldsymbol{X}_n^{ij}\|_p + \sup_{0\leq t\leq 1}\|D_t\boldsymbol{X}_n^{ij}\|_{p,\ell_2} + \sup_{0\leq s,t\leq 1}\|D_{s,t}\boldsymbol{X}_n^{ij}\|_{p,\ell_2}\right) = O(1),$$

$$\max_{1\leq i\leq d}\sup_{0\leq t\leq 1}\left(\|\mu_t^i\|_p + \sup_{0\leq s,t\leq 1}\|D_s\mu_t^i\|_{p,\ell_2}\right) = O(1),$$

$$\max_{1\leq i\leq d}\sup_{0\leq t\leq 1}\left(\|\Sigma_t^{ii}\|_p + \sup_{0\leq s,t\leq 1}\|D_s\sigma_t^{i\cdot}\|_{p,\ell_2} + \sup_{0\leq s,t,u\leq 1}\|D_{s,t}\sigma_u^{i\cdot}\|_{p,\ell_2}\right) = O(1),$$

$$\sup_{0\leq t\leq 1-\frac{1}{n}}\left\|\max_{1\leq k,l\leq d}\left|\Sigma_{t+\frac{1}{n}}^{kl} - \Sigma_t^{kl}\right|\right\|_2 = O(n^{-\gamma})$$

がすべての $p \in [1,\infty)$ とある $\varpi \in (0,\frac{1}{2})$, $\gamma \in (0,\frac{1}{2}]$ について成り立つと仮定する．さらに, 以下を仮定する:

$$\lim_{b\downarrow 0}\limsup_{n\to\infty} P(\min\mathrm{diag}(\Xi_n\widehat{\mathfrak{C}}_n\Xi_n^\top) < b) = 0.$$

このとき, ある $\mathfrak{c} > 0$ について $d + m = O(n^{\mathfrak{c}})$ ならば, (3) 式が成立する

残差スパース性検定への応用

- 定理 4 より, $\mathcal{L} \subset \Lambda_n$ が与えられたとき, 適当な条件下で統計量 $\max_{(i,j)\in\mathcal{L}} |T_n^{(i,j)}|$ の分布はある混合 Gauss 型確率ベクトルの最大値の分布で近似できる
  $\Longrightarrow$ 後者の分布はシミュレーション (ブートストラップ法) で評価できる
- 標準的なブートストラップ法は非エルゴード性の問題によって機能しない (cf. Dovonon et al., 2013; Hounyo, 2017)
  $\Longrightarrow$ Gauss 型 MA(1) ワイルドブートストラップを用いる:

$$[\widehat{Y^i, Y^j}]_1^{n,*} := \sum_{h=1}^{n} (\eta_h^* - \eta_{h-1}^*)(Y^i_{t_h} - Y^i_{t_{h-1}})(Y^j_{t_h} - Y^j_{t_{h-1}}).$$

ここに, $\eta_h^* \overset{i.i.d.}{\sim} N(0, 0.5) \perp\!\!\!\perp \mathcal{F}$

  - 本質的に Hounyo (2017) の WBB ブートストラップの特殊ケース

残差スパース性検定への応用

$$\hat{\mathfrak{T}}^{ij}_{n,*} := \widehat{[Y^i, Y^d]}^{n,*}_1 \widehat{[Y^j, Y^d]}^{n}_1 + \widehat{[Y^i, Y^d]}^{n}_1 \widehat{[Y^j, Y^d]}^{n,*}_1 - \widehat{[Y^i, Y^j]}^{n,*}_1 \widehat{[Y^d, Y^d]}^{n}_1 - \widehat{[Y^i, Y^j]}^{n}_1 \widehat{[Y^d, Y^d]}^{n,*}_1$$

とし，

$$T^{(i,j)}_{n,*} := \frac{\sqrt{n}\mathfrak{T}^{ij}_{n,*}}{\sqrt{\hat{\mathfrak{V}}^{ij}_n}}$$

とおくと，$\max_{(i,j)\in\mathcal{L}} |T^{(i,j)}_n|$ の分布を $\max_{(i,j)\in\mathcal{L}} |T^{(i,j)}_{n,*}|$ の分布で近似できることが証明できる

- K. (2019a, Proposition 4.2) 参照

数値実験

- モデル (1) を以下の設定でシミュレーション (cf. Fan et al., 2016):
  - ファクター過程 Y^d は Heston モデル:

$$dY_t^d = \mu dt + \sqrt{v_t} dB_t^d,$$
$$dv_t = \kappa(\theta - v_t)dt + \eta\sqrt{v_t}\left(\rho dB_t^d + \sqrt{1-\rho^2} dB_t^{d+1}\right).$$

    パラメーターは $\mu = 0.05$, $\kappa = 3$, $\theta = 0.09$, $\eta = 0.3$, $\rho = -0.6$ とし, 初期値 v_0 は定常分布から生成
  - 残差過程: $dR_t^j = \gamma_j^\top d\underline{B}_t$ $(j = 1, \dots, \underline{d})$
    - ★ $\underline{B}_t = (B_t^1, \dots, B_t^{\underline{d}})$
    - ★ $\gamma_1, \dots, \gamma_{\underline{d}}$ は B と独立な $\underline{d}$ 次元確率ベクトルで, $\Gamma := (\gamma_i^\top \gamma_j)_{1 \le i,j \le \underline{d}}$ は同サイズのブロック 10 個をもつブロック行列. 各ブロックの対角成分は $[0.2, 0.5]$ 上の一様乱数で生成し, ブロック内相関は定数 ρ_γ とする
  - $\beta^1, \dots, \beta^{\underline{d}}$ は $[0.25, 2.25]$ 上の一様乱数で生成
- $\underline{d} = 100$ として, $\rho_\gamma \in \{0.25, 0.5, 0.75\}$ と変動させ FWER と平均検出力を比較

表 1: 有意水準 5%での FWER

ρ_γ	$n = 26$	$n = 39$	$n = 78$	$n = 130$	$n = 195$	$n = 390$
0.25	0.022	0.007	0.003	0.004	0.009	0.018
0.50	0.023	0.008	0.004	0.005	0.009	0.019
0.75	0.026	0.011	0.006	0.008	0.014	0.023

Romano-Wolf 法による多重検定の有意水準 5%での FWER(10,000 回のモンテカルロ実験の結果から計算). ブートストラップ複製は 999 回.

表 2: 有意水準 5%での平均検出力

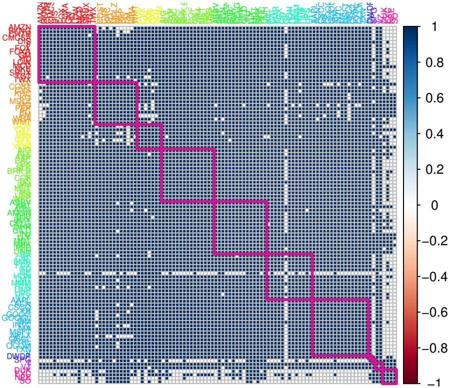
ρ_γ	$n = 26$	$n = 39$	$n = 78$	$n = 130$	$n = 195$	$n = 390$
0.25	0.000	0.000	0.000	0.006	0.048	0.567
0.50	0.000	0.001	0.037	0.458	0.956	1.000
0.75	0.004	0.017	0.393	0.977	1.000	1.000

Romano-Wolf 法による多重検定の有意水準 5%での平均検出力 (10,000 回のモンテカルロ実験の結果から計算). ブートストラップ複製は 999 回.

実データ解析例

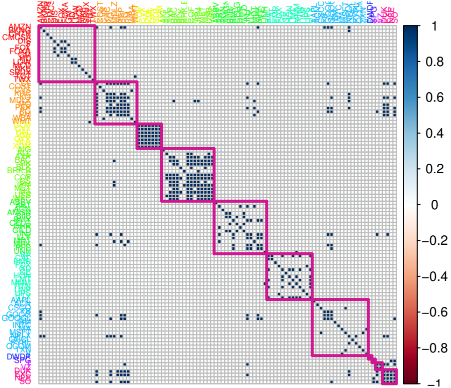
- 期間: 2018 年 3 月
  - この期間を区間 [0, 1] とみなす (オーバーナイトリターンは無視)
- 15 分足リターン
- 分析対象: S&P100 指数構成銘柄
- ファクター過程: SPDR S&P 500 ETF (SPY)
- データソース: ブルームバーグ
- 図 1–2 に対応する共分散行列のスパース性を検定するために, Romano-Wolf のステップダウン法を実行する
  - 棄却値の計算は 999 回のブートストラップ複製に基づく
- Fan et al. (2016) で報告されたブロック対角構造を確認するために, セクターによって銘柄をソートしておく

図 5: S&P100 指数構成銘柄の実現相関行列 (2018 年 3 月): FWER が 5%水準で有意な成分=1, それ以外=0



15 分足リターンで計算. 長方形はセクターを表す

図 6: マーケットファクターで回帰後の残差過程の実現相関行列 (2018 年 3 月): FWER が 5%水準で有意な成分=1, それ以外=0



15 分足リターンで計算．長方形はセクターを表す．
マーケットファクターは SPY で代理

まとめ

- サンプル数とともに次元が上昇するような高次元の設定下で実現共分散行列の漸近混合正規性を証明した
  - 証明は Malliavin 解析と Nualart-Yoshida 補間による Chernozhukov-Chetverikov-Kato 理論の変形
- 理論の応用例として, 連続時間 1 ファクターモデルの残差スパース性検定を説明
- 数値実験と実データ解析例によって理論の適用可能性を示した
- 論文
  - "Mixed-normal limit theorems for multiple Skorohod integrals in high-dimensions, with application to realized covariance", *Electronic Journal of Statistics*, 13 (2019), no.1, 1443–1522.

今後の課題

- より複雑な状況への適用
  - 非同期観測の場合 (Hayashi-Yoshida 推定量は二重伊藤積分で書ける!)
  - マイクロストラクチャーノイズ
  - ジャンプ
- ファクターを PCA 等で構成した場合の残差スパース性検定
- FWER ではなく False Discovery Rate (FDR) や False Discovery Proportion (FDP) のコントロールを考える
- 高次元精度行列の統計推測への応用
  - De-biased graphical Lasso を使うと, インプットの共分散行列推定量の高次元漸近混合正規性からほぼ自動的に精度行列推定量の高次元漸近混合正規性が出てくる (Koike (2019b) 参照)
  - 最適収束レートを達成する精度行列推定法と組み合わせることは可能か？ (例えば Sun and Zhang (2013) の scaled Lasso)

References I

M. Bibinger and P. A. Mykland. Inference for multi-dimensional high-frequency data with an application to conditional independence testing. *Scand. J. Stat.*, 43:1078–1102, 2016.

P. J. Bickel and E. Levina. Regularized estimation of large covariance matrices. *Ann. Statist.*, 36(1):199–227, 2008a.

P. J. Bickel and E. Levina. Covariance regularization by thresholding. *Ann. Statist.*, 36(6):2577–2604, 2008b.

C. Brownlees, E. Nualart, and Y. Sun. Realized networks. *J. Appl. Econometrics*, 33(7):986–1006, 2018.

V. Chernozhukov, D. Chetverikov, and K. Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.*, 41(6):2786–2819, 2013.

V. Chernozhukov, D. Chetverikov, and K. Kato. Gaussian approximation of suprema of empirical processes. *Ann. Statist.*, 42(4):1564–1597, 2014.

References II

V. Chernozhukov, D. Chetverikov, and K. Kato. Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields*, 162:47–70, 2015.

V. Chernozhukov, D. Chetverikov, and K. Kato. Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. *Stochastic Process. Appl.*, 126:3632–3651, 2016.

V. Chernozhukov, D. Chetverikov, and K. Kato. Central limit theorems and bootstrap in high dimensions. *Ann. Probab.*, 45(4):2309–2353, 2017.

P. Dovonon, S. Gonçalves, and N. Meddahi. Bootstrapping realized multivariate volatility measures. *J. Econometrics*, 172:49–65, 2013.

J. Fan, Y. Liao, and M. Mincheva. High-dimensional covariance matrix estimation in approximate factor models. *Ann. Statist.*, 39(6):3320–3356, 2011.

J. Fan, A. Furger, and D. Xiu. Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high-frequency data. *J. Bus. Econom. Statist.*, 34(4): 489–503, 2016.

References III

U. Hounyo. Bootstrapping integrated covariance matrix estimators in noisy jump-diffusion models with non-synchronous trading. *J. Econometrics*, 197: 130–152, 2017.

J. Jacod and P. Protter. *Discretization of Processes*. Springer, 2012.

Y. Koike. Mixed-normal limit theorems for multiple Skorohod integrals in high-dimensions, with application to realized covariance. *Electron. J. Stat.*, 13 (1):1443–1522, 2019a.

Y. Koike. De-biased graphical Lasso for high-frequency data. Working paper. arXiv:1905.01494, 2019b.

I. Nourdin and G. Peccati. Stein's method on Wiener chaos. *Probab. Theory Related Fields*, 145:75–118, 2009.

I. Nourdin, D. Nualart, and G. Peccati. Quantitative stable limit theorems on the Wiener space. *Ann. Probab.*, 44(1):1–41, 2016.

D. Nualart and N. Yoshida. Asymptotic expansion of Skorohod integrals. Working paper. Available at arXiv: https://arxiv.org/abs/1801.00120, 2017.

References IV

P. Ravikumar, M. J. Wainwright, G. Raskutti, and B. Yu. High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electron. J. Stat.*, 5:935–980, 2011.

J. P. Romano and M. Wolf. Exact and approximate stepdown methods for multiple hypothesis testing. *J. Amer. Statist. Assoc.*, 100(469):94–108, 2005.

Y. Sakamoto and N. Yoshida. Malliavin calculus and statistical asymptotic theory (in Japanese). *Proc. Inst. Statist. Math.*, 47(1):175–199, 1999.

Y. Sakamoto and N. Yoshida. Asymptotic expansion formulas for functionals of ϵ-Markov processes with a mixing property. *Ann. Inst. Statist. Math.*, 56(3): 545–597, 2004.

T. Sun and C.-H. Zhang. Sparse matrix inversion with scaled Lasso. *J. Mach. Learn. Res.*, 14:3385–3418, 2013.

Y. Wang and J. Zou. Vast volatility matrix estimation for high-frequency financial data. *Ann. Statist.*, 38(2):943–978, 2010.

M. Yuan and Y. Lin. Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.

Detecting the number of factors of quadratic variation in the presence of microstructure noise : High-Dimension Case

Naoto Kunitomo

Meiji University

August 7, 2019

This is a joint work with Daisuke Kurisu.

Outline

- Introduction
- Estimation of QV
- Representing the problem as a Random Field model
- Characteristic Roots in High-Dimension
- A Mathematical Interpretation and Discussions
- Simulation Results
- Further Problems

Introduction

- One of important observations on the asset price movements has been that many financial asset prices move in similar way in their trends, volatilities and jumps.
- There is a question how to cope with many asset prices when the number of factors of volatilities or quadratic variation of asset prices is less than the dimension of observed prices. Kunitomo and Kurisu (2019) is developping a statistical method for detecting the number of factors.
- The problem is related to the reduced rank regression problem, which has been well-known in statistical multivariate analysis (Anderson (1984, AS) and Anderson (2003)).
- There is an important question when the dimension of financial asset prices is large (high-dimension) because we have a large number of assets traded in financial markets. This is an extension of Kunitomo and Kurisu (2019).

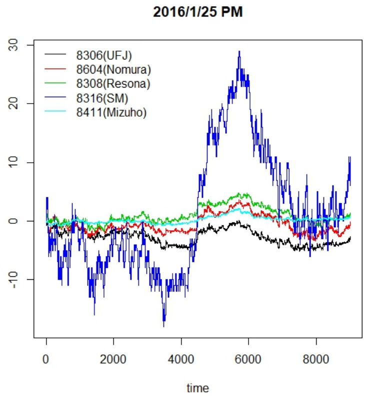


Figure: Intra-day movements of 5 Stock Prices at TSE (afternoon session of January 25, 2016).

We set the same values for the starting process since we focus on the volatility structure (of quadratic variation) of asset prices.

Introduction

- The new feature in our formulation is the fact that we are dealing with the continuous stochastic process as the hidden process with diffusions and jump processes while we have discrete observations with measurement errors.
- We develop a new way to determine the number of factors of quadratic covariation or the integrated volatility of asset prices based on the separating information maximum likelihood (SIML) method when the dimension of observed vectors is large.
- The SIML method is originally developed by Kunitomo, Sato and Kurisu (2018).
- Kunitomo, N. and Kurisu, D. (2019). Detecting factors of quadratic variation in the presence of microstructure noise. MIMS-RBP SDS-10.

Estimation of QV

We adopt the the construction of the whole filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,1]}, P)$ where both the hidden process $\mathbf{X}$ and the noise $\mathbf{V}$. Let $\mathbf{Y}(t_i^n) = (Y_j(t_i^n))$ be the (p-dimensional) observed (log-)prices at $t_i \in [0,1]$ and $i = 1, \cdots, n$, which satisfyies

$$\mathbf{Y}(t_i^n) = \mathbf{X}(t_i^n) + \mathbf{v}(t_i^n) \ \ (i = 1, \cdots, n) \ ,$$

- $\mathbf{X}(t_i^n)$ $(= (X_j(t_i^n))$: the $p \times 1$ hidden stochastic vector process.
- $\mathbf{v}(t_i^n)$ $(= (v_j(t_i^n)))$: a sequence of (mutually) independent market microstructure noises with $\mathcal{E}[\mathbf{v}(t_i^n)] = \mathbf{0}$ and $\mathcal{E}[\mathbf{v}(t_i^n)\mathbf{v}(t_i^n)^{'}] = \mathbf{\Sigma}_v$ (a positive definite matrix).
- We also assume that $\mathbf{v}$ is independent of $\mathbf{X}$.

Estimation of QV

We also assume that $\mathbf{X}$ is a p-dimensional continuous-time stochastic process which is given by

$$\mathbf{X}(t) = \mathbf{X}(0) + \int_0^t \mathbf{b}(s)ds + \int_0^t \boldsymbol{\sigma}(s)d\mathbf{W}(s)$$
$$+ \int_0^t \int_{\|\mathbf{x}\|<1} \boldsymbol{\Delta}(s,x)(\boldsymbol{\mu} - \boldsymbol{\nu})(ds, d\mathbf{x}) + \int_0^t \int_{\|\mathbf{x}\|\geq 1} \boldsymbol{\Delta}(s,\mathbf{x})\boldsymbol{\mu}(ds, d\mathbf{x}),$$

- $\mathbf{b}(s)$: the p-dimensional adapted drift process, $\boldsymbol{\sigma}(s)$: the $p \times q_1$ instantaneous predictable volatility process, $\mathbf{c}(s) = \boldsymbol{\sigma}(s)\boldsymbol{\sigma}'(s)$,
- $\mathbf{W}(s) = (W_j(s))$: the $q_1 \times 1$ standard Brownian motions.
- $\boldsymbol{\Delta}(\omega, s, \mathbf{x})$ is a $\mathbf{R}^p$-valued predictable function on $\Omega \times [0, \infty) \times \mathbf{R}^{q_2}$,
- $\boldsymbol{\mu}(\cdot)$ is a Poisson random measure on $[0, \infty) \times \mathbf{R}^{q_2}$ and $\boldsymbol{\nu}(ds, d\mathbf{x}) = ds \otimes \lambda(d\mathbf{x})$ is the predictable compensator of $\boldsymbol{\mu}$ with a σ-finite measure λ on $(\mathbf{R}^p, \mathcal{B}^p)$.

Estimation of QV

The jump terms are denoted as $\Delta\mathbf{X}_s = (\Delta X_j(s))$
$(\Delta X_j(s) = X_j(s) - X_j(s-)$ and $X_j(s-) = \lim_{u\uparrow s} X_j(u)$ at any $s \in [0,1]$.
Here $\|\cdot\|$ is the Euclidean norm on $\mathbf{R}^p$.

Assumption 2.1

(a) The path $t \mapsto \mathbf{b}(t,\omega)$ is locally bounded.

(b) The process $\boldsymbol{\sigma}$ is continuous and $\int_t^{t+u} ||\boldsymbol{\sigma}(s)||ds > 0$ a.s. for all $t, u > 0$.

(c) For some positive K, the measurable functions $\mathbf{b}(s,\mathbf{x})$, $\boldsymbol{\sigma}(s,\mathbf{x})$, and $\boldsymbol{\Delta}(s,\mathbf{x})$ $(0 \le s \le 1)$ satisfy

$$\|\mathbf{b}(s,\mathbf{x})\|^2 + \|\boldsymbol{\sigma}(s,\mathbf{x})\|^2 + \int_{U_0} \|\boldsymbol{\Delta}(s,u,\mathbf{x})\|^2 \boldsymbol{\lambda}(du) \le K\,[\,1 + \|x\|^2\,]\ ,\ \mathbf{x} \in \mathbf{R}^p\ ,$$

where $U_0 = \{\|\mathbf{u}\| < 1\}$.

(d) The noise terms $\mathbf{v}(t_i^n)(= (v_j(t_i^n))', 1 \le i \le n, 1 \le j \le p)$ are a sequence of i.i.d. random variables with $\mathcal{E}[v_j(t_i^n)] = 0$, $\mathcal{E}[\mathbf{v}(t_i^n)\mathbf{v}'(t_i^n)] = \boldsymbol{\Sigma}_v$ (p.d.), $\mathcal{E}[v_j^4(t_i^n)] < \infty$, $\mathbf{v} \perp\!\!\!\perp \mathbf{X}$.

Estimation of QV

The fundamental quantity for the continuous-time Itô semimartingale with $p \geq 1$ is the quadratic variation (QV) matrix, which is given by

$$\mathbf{\Sigma}_x = \int_0^1 \mathbf{c}(s)ds + \sum_{0 \leq s \leq 1} (\Delta \mathbf{X}_s)(\Delta \mathbf{X}_s)^{'} = (\sigma_{gh}^{(x)}) \ .$$

The SIML estimator of $\hat{\mathbf{\Sigma}}_x$ (Kunitomo, Sato and Kurisu (2018)) for the integrated volatility is defined by

$$\hat{\mathbf{\Sigma}}_x = \frac{1}{m_n} \sum_{k=1}^{m_n} \mathbf{z}_k \mathbf{z}_k^{'},$$

where $\mathbf{z}_k = (z_{gk})$, $j = 1, \cdots, p$; $k = 1, \cdots, m_n$, $m_n = O(n^{\alpha})$, $0 < \alpha < 1$, which are constructed by the transformation from $\mathbf{Y}_n = (\mathbf{y}_k^{'})$ $(n \times p)$ to $\mathbf{Z}_n$ $(= (\mathbf{z}_k^{'}))$ by

$$\mathbf{Z}_n = \mathbf{K}_n \left(\mathbf{Y}_n - \bar{\mathbf{Y}}_0 \right)$$

Estimation of QV

where $\mathbf{K}_n = \sqrt{n}\mathbf{P}_n\mathbf{C}_n^{-1}$,

$$\mathbf{C}_n^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}_{n\times n} ,$$

and

$$\mathbf{P}_n = (p_{jk}^{(n)}) \ , \ p_{jk}^{(n)} = \sqrt{\frac{2}{n+\frac{1}{2}}} \cos\left[\frac{2\pi}{2n+1}(k-\frac{1}{2})(j-\frac{1}{2})\right] .$$

Intuitively, the SIML estimator $\hat{\Sigma}_x$ is an estimator of the spectral density matrix of X at the origin, i.e. $\Sigma_x = f_X(0)$. (Indeed, this is true if we apply the SIML method to state space models.)

Representing the problem as a Random Field

We consider the inference probelm when the dimension p is large. For $j = 1, \cdots, p$ and at $s = j/p$, we set $X_s^*(t) = X_j(t) \ \ (0 \leq t \leq 1)$, where $X_j(t)$ is the j-th component of $\mathbf{X}(t)$. We construct the random field $X_s(t)$ on $s \in [0,1], t \in [0,1]$ by

$$X_s^*(t) = X_{j-1}(t) + p[s - \frac{j-1}{p}][X_j(t) - X_{j-1}(t)] \ \ ,$$

for $\frac{j-1}{p} \leq s \leq \frac{j}{p}$ and $j = 1, \cdots, p$. Then we have the continuous time process

$$\begin{aligned} X_s^*(t) &= X_0^*(t) + \int_0^t b_s^*(u)du + \int_0^t \boldsymbol{\sigma}_s^*(u)d\mathbf{B}_u \\ &+ \int_u \int_{\|\mathbf{x}\|<1} \boldsymbol{\Delta}_s^*(x,u)(\boldsymbol{\mu}(u) - \boldsymbol{\nu}(u))(du, d\mathbf{x}, s) \\ &+ \int_u \int_{\|\mathbf{x}\|\geq 1} \boldsymbol{\Delta}_s^*(\mathbf{x},u)\mu(du, d\mathbf{x}, s). \end{aligned}$$

We write the quadratic variation as

$$\sigma_x(s,t) = \int_0^1 \boldsymbol{\sigma}_s^*(u)\boldsymbol{\sigma}_t^{*'}(u)du + \sum_{0\le u\le 1} \Delta X_s(u)\Delta X_t(u)\ ,$$

which is a non-negative definite (random or non-random) matrix. Then, by using Mercer Theorem,

$$\sigma_x(s,t) = \sum_{i\ge 1} \mu_i\beta_i(s)\beta_i(t)\ \ ,$$

where $0 \le \mu_1 \le \mu_2 \le \cdots$ are the eigenvalues and $\beta_i(u)$ are the associated eigenfunction.

In the same way, we construct the noise process $V_s^*(t)$ $(0 \le s \le 1, 0 \le t \le 1)$ and $Y_s^*(t)$ $(0 \le s \le 1, 0 \le t \le 1)$ by

$$V_s^*(t) = V_{j-1}(t) + p[s - \frac{j-1}{p}][V_j(t) - V_{j-1}(t)]$$

and

$$Y_s^*(t) = Y_{j-1}(t) + p[s - \frac{j-1}{p}][Y_j(t) - Y_{j-1}(t)]$$

for $s \in [0,1]$ and $t \in [0,1]$

Then we have

$$Y_s^*(t) = X_s^*(t) + V_s^*(t) \quad (0 \le s \le 1, 0 \le t \le 1).$$

Characteristic Roots in High-Dimension

Let the random field $g_m(s,t)$ $(0 \le s,t \le 1)$ from $g_m(\frac{i}{p},\frac{j}{p}) = [G_m]_{ij}$, $\sigma_x(\frac{i}{p},\frac{j}{p}) = [\mathbf{\Sigma}_x]_{ij}$ and $\sigma_v(\frac{i}{p},\frac{j}{p}) = [\mathbf{\Sigma}_v]_{ij}$ where $\mathbf{G}_m = ((G_m)_{ij})$ $(i,j=1,\cdots,p)$, $\mathbf{\Sigma}_x = ((\Sigma_x)_{ij})$ $(i,j=1,\cdots,p)$ and $\mathbf{\Sigma}_v = ((\Sigma_v)_{ij})$ $(i,j=1,\cdots,p)$. We denote

$$\sigma_m(\frac{i}{p},\frac{j}{p}) = \sigma_x(\frac{i}{p},\frac{j}{p}) + a_m\sigma_v(\frac{i}{p},\frac{j}{p}) .$$

Let

$$\mathbf{D}_m(\lambda) = \begin{vmatrix} g_m(\frac{1}{p},\frac{1}{p}) - \lambda\sigma_v(\frac{1}{p},\frac{1}{p}) & g_m(\frac{1}{p},\frac{2}{p}) - \lambda\sigma_v(\frac{1}{p},\frac{2}{p}) & \cdots & g_m(\frac{1}{p},\frac{p}{p}) - \lambda\sigma_v(\frac{1}{p},\frac{p}{p}) \\ g_m(\frac{2}{p},\frac{1}{p}) - \lambda\sigma_v(\frac{2}{p},\frac{1}{p}) & g_m(\frac{2}{p},\frac{2}{p}) - \lambda\sigma_v(\frac{2}{p},\frac{2}{p}) & \cdots & g_m(\frac{2}{p},\frac{p}{p}) - \lambda\sigma_v(\frac{2}{p},\frac{p}{p}) \\ \vdots & \vdots & \vdots & \vdots \\ g_m(\frac{p}{p},\frac{1}{p}) - \lambda\sigma_v(\frac{p}{p},\frac{1}{p}) & g_m(\frac{p}{p},\frac{2}{p}) - \lambda\sigma_v(\frac{p}{p},\frac{2}{p}) & \cdots & g_m(\frac{p}{p},\frac{p}{p}) - \lambda\sigma_v(\frac{p}{p},\frac{p}{p}) \end{vmatrix} .$$

Then the characteristic roots when we take $\mathbf{H} = \mathbf{\Sigma}_v = \sigma_v^2 \mathbf{I}_p$ as the leading case. The underlying characteristic equation can be written as the solutions of

$$|\mathbf{D}_m(\lambda)| = 0 \ .$$

We transform the present problem into the stochastic integral equation and random field. For $\frac{i-1}{p} \le t \le \frac{i}{p}$ and $\frac{j-1}{p} \le s \le \frac{j}{p}$ $(i, j = 1, \cdots, p)$, define

$$\begin{aligned} g_m(t,s) &= g_m(\frac{i-1}{p}, \frac{j-1}{p}) \\ &+ c_p(t)[g_m(\frac{i}{p}, \frac{j-1}{p}) - g_m(\frac{i-1}{p}, \frac{j-1}{p})] \\ &+ c_p(s)[g_m(\frac{i-1}{p}, \frac{j}{p}) - g_m(\frac{i-1}{p}, \frac{j-1}{p})] \\ &+ c_p(t)c_p(s)[g_m(\frac{i}{p}, \frac{j}{p}) - g_m(\frac{i}{p}, \frac{j-1}{p}) - g_m(\frac{i-1}{p}, \frac{j}{p}) + g_m(\frac{i-1}{p}, \frac{j-1}{p})] \ , \end{aligned}$$

where $c_p(t) = p[t - (i-1)/p]$ and $c_p(s) = p[s - (j-1)/p]$.

Under some conditions such as the boundedness of fourth corder moments. we have

$$\mathcal{E}\left[\left(\int_0^1\int_0^1 g_m(t,s)dtds\right)^2\right] < +\infty \ .$$

Hence we can use the standard arguments on L^2. For a fixed p, we write

$$|(g_m(\frac{i}{p},\frac{j}{p}) - a_m\sigma_v^2\delta(i,j)) - (\lambda - a_m)\sigma_v^2\mathbf{I}_p| = 0 \ .$$

If we take $n, m \to \infty$, then

$$|\mathrm{plim}_{m\to\infty}(g_m(\frac{i}{p},\frac{j}{p}) - a_m\sigma_v^2\delta(i,j)) - (\lambda - a_m)\sigma_v^2\mathbf{I}_p| = 0$$

implies

$$\lambda_i - a_m \xrightarrow{p} 0 \ \ (i = 1, \cdots, r_x) \ .$$

For the case when p is large, we may set $p = p_m$. In order to develop the asymptotic theory, we impose a condition that the number of eigenvalues is less than m. Also we assume that the volatility and co-volatility functions are well-defined.

Condition II : $\frac{p_m}{m} \longrightarrow 0$ as $n, m \longrightarrow +\infty$.
It may be possible to extend the following results when $p_m/m \to c$ and $c > 0$ as $m \to \infty$.
Miyahara (1982) has pointed out the fundamental difficulty on L^2 arguments in the present situation (See Section 4). The rank condition should be replaced by the following condition of continuous integral equation :
Condition III : There exist eigenfunctions $\beta_k(t)$ $(0 \le t \le 1, k \ge 1)$ such that

$$\int_0^1 \sigma_x(t,s)\beta_k(t)dt = 0 \ \ (\text{at } s = \frac{k}{p}, k = 1, \cdots, r_x) \ .$$

Because $g_m(t,s)$ is a symmetric non-negative definite operator on $[0,1] \times [0,1]$, the set of eigenvalues is at most denumerable and eigenfunctions corresponding to the different eigenvalues are orthogonal by the (standard) integral equation theory.

We construct $\sigma_m(t,s), \sigma_x(t,s), \sigma_v(t,s)$ and $b_k(t)$ from $\mathbf{\Sigma}_m, \mathbf{\Sigma}_x, \mathbf{\Sigma}_v$ and b_{ij} for $g_m(t,s)$. We take b_{jk} $(j,k=1,\cdots,p)$ such that and construct $b_k(t)$ from b_{jk} for $0 \leq t \leq 1$.

We consider the stochastic integral equation as

$$\int_0^1 g_m(s,t)b_k(t)dt = \lambda_k \int_0^1 \sigma_v(t,s)b_k(t)dt \; (\text{at } s = \frac{k}{p}, k = 1,\cdots,p) ,$$

where λ_k $(0 \le \lambda_1 \le \lambda_2 \le \cdots, 1 \le k \le \infty)$. This equation corresponds to the continuous version. By using $\sigma_m(t,s) = \sigma_x(t,s) + a_m\sigma_v(t,s)$, we re-write

$$\int_0^1 [g_m(t,s) - a_m\sigma_v(t,s)]\, b_k(t)dt = [\lambda_k - a_m] \int_0^1 \sigma_v(t,s)b_k(t)dt \; .$$

For a fixed p, we take the probability limit as $m \longrightarrow \infty$.
Since we have the rank condition (R), $\mathcal{E}[g_m(t,s)] = \sigma_x(t,s) + a_m\sigma_v(t,s)$,

$$\int_0^1 \text{plim}_m[g_m(t,s) - a_m\sigma_v(t,s)]\beta_k(t)dt = \text{plim}_m[\lambda_k - a_m] \int_0^1 \sigma_v(t,s)\beta_k(t)dt$$

we find that as $m \longrightarrow \infty$,

$$\lambda_j - a_m \xrightarrow{p} 0 \; (j = 1,\cdots,r_x)$$

and $0 < r_x = p - q_x$.

We summarize our heuristic arguments.

Theorem 3.1 : Assume some regularity conditions and take $m = m_n = [n^{\alpha}]$ $(0 < \alpha < 1)$, $l_n = [n^{\beta}]$ with $0 < \alpha < \beta < 1$. Also assume that the market micro-structure noises are mutually independent Gaussian random variables. For a fiexd p, as $m \to \infty$,

$$\left[\frac{m}{a_m(2)}\frac{1}{2r_x}[\sum_{i=1}^{r_x}(\lambda_i - a_m)]^2\right] \xrightarrow{w} \chi^2(1)$$

where $r_x = p - q_x$ and $1 \leq r_x < \infty$.

Since Theorem 3.1 does not depend on p and r_x, we obtain the following result, which has some important meanings in applications.

Corollary 3.2 : In Theorem 3.1 as $p \to \infty$,

$$\left[\frac{m}{a_m(2)}\frac{1}{2r_x}[\sum_{i=1}^{r_x}(\lambda_i - a_m)]^2\right] \xrightarrow{w} \chi^2(1)$$

where $r_x = p - q_x$ and $1 \leq r_x \leq \infty$.

A Mathematical Interpretation and Discussions

There is a fundamental problem of our formulation and derivations in the previous sections. We have treated the high-dimension case when $p \longrightarrow \infty$ and $q_x = q_x^{(1)} + q_x^{(2)} \longrightarrow \infty$. However, when $p = +\infty$ there can be an infinite number of Brownian motions as the limit, and it is not clear that we have the proper mathematical meaning of our formulation. Fortunately, however, there is an earlier work on the cylindrical Brownian motion on a Banach space by Miyahara (1982) and we can use his arguments. We are reviewing the first part of Miyahara (1982). For the probability space $(\mathbf{\Omega}, \mathcal{F}, P)$ and an increasing family of $\sigma-$field $\mathcal{F}_t$ $(t \geq 0)$, let $\mathbf{H}$ be a real separable Hilbert space with norm denoted by $\| \cdot \|$. A mapping $B_t(h, \omega) : [0, \infty) \times \mathbf{H} \times \mathbf{\Omega} \to \mathbf{R}^{'}$ is called a cylindrical Brownian motion (c.s.m.) on $\mathbf{H}$ if if satisfies the following conditions. (i) $B_0(h, \cdot) = 0$ and $B_t(h, \cdot)$ is $\mathcal{F}-$adapted, (ii) For any $h \in \mathbf{H}$, $h \neq 0$, $B_t(h . \cdot)/\|h\|$ is a one dimensional Brownian motion, (iii) For any $t \in [0, \infty)$ and $\alpha, \beta \in \mathbf{R}^1$ and $h, k \in \mathbf{H}$, $B_t(\alpha h + \beta k) = \alpha B_t(h) + \beta B(k)$ $(P - a.s.)$.

Let $\mathbf{V}$ be a Banach space which is a dense set of $\mathbf{H}$, and let $\mathbf{V}^{'}$ be the dual Banach space of $\mathbf{V}$. Then we have $\mathbf{V} \subset \mathbf{H} \subset \mathbf{V}^{'}$ and we denote the canonical bilinar norm on $\mathbf{V} \times \mathbf{V}^{'}$ by $<,>_{V \times V'}$ or simply by $<,>$. A $\mathbf{V}^{'} -$ *valued* process $\tilde{B}_t$ is called an H-Brownian motion on $\mathbf{V}^{'}$ (H-B.M.) if it satisfies (i) $\tilde{B}_{\cdot}(\omega) : \mathbf{R}_{+} \rightarrow \mathbf{V}^{'}$ is almost surely continuous and $\tilde{B}_0(\omega) = 0$, (ii) For each $y \in \mathbf{V}, y \neq 0$, $< y, \tilde{B}_t > / \|y\|$ is a one-dimensional Brownian motion.

Then Miyahara (1982) gives the following fundamental result for us.

Theorem M (Miyahara 1982) : Let $\tilde{B}_t$ be a $H-$B.M. on $\mathbf{V}^{'}$. Then there exists a unique cylindrical Brownian motion B_t on $\mathbf{H}$ which satisfies the following equality for any $y \in \mathbf{V}$ and $t \in \mathbf{R}_{+}$

$$< y, \tilde{B}_t >= B_t(y) \;\; (P - a.s..)$$

We can interpret any real-valued Brownian motions as the bilinear norm of the cylindrical Brownian motions on H. (We have used such Brownian motions in (2.2).) Miyahara (1982) has also shown that even when we have $p = \infty$, we have the corresponding Banach-valued Brownian motions and the solutions of stochastic differential equations.

We denote $X_s^*(t) = (X_j(t))$ at $s = j/p$ $(j = 1, \cdots, p)$ with $\mathbf{E}[\|X(t)\|^2] < \infty$.

In a similar framework a Corollary of Araujo (1978) gives the following CLT for a sequence of Banach-valued random variables in $\mathbf{H}_1$.

Let $\mathbf{W}_u(s) = (W_{i/n}(j/p))$ at $u = i/n, s = j/p$, $i = 1, \cdots, n; j = 1, \cdots, p$ be i.i.d. random vectors with $\mathcal{E}[W_u(j/p)] = 0$ and $\mathcal{W}[U_u(j/p)^2] = \sigma_u^2$ (> 0).

Proposition 4.1 : Let $W_u(s)$ are i.i,d. random variables at $u = i/n$ $(i = 1, \cdots, n)$ and

$$W_u(s) = W_u((j-1)/p) + [s - \frac{j-1}{p}][W_u(i/p) - W_u((i-1)/p)] .$$

Assume

We construct the random field $W_u(s,t)$ $(0 \leq u \leq 1, 0 \leq s,t \leq 1)$ from the discrete process $U_u(i/p.j/p)$ for $0 \leq u \leq 1; i,j = 1, \cdot, p$. It can be done such as that for $\frac{j-1}{p} \leq r \leq \frac{i}{p}$ and $\frac{k-1}{p} \leq s \leq \frac{k}{p}$ $(j,k = 1, \cdots, p)$. Define

$$
\begin{aligned}
W_u(s,t) &= W(\frac{j-1}{p}, \frac{k-1}{p}) \\
&+ c_p(s)[W(\frac{j}{p}, \frac{k-1}{p}) - W(\frac{j-1}{n}, \frac{k-1}{p})] \\
&+ c_p(t)[W(\frac{j-1}{n}, \frac{k}{p}) - W(\frac{j-1}{p}, \frac{k-1}{p})] \\
&+ c_p(s)c_p(t)[W(\frac{j}{p}, \frac{k}{p}) - W(\frac{j}{p}, \frac{k-1}{p}) - W(\frac{j-1}{p}, \frac{k}{p}) + W(\frac{j-1}{p}, \frac{k-1}{p})] \,,
\end{aligned}
$$

where $c_p(r) = p[s - (i-1)/p]$ and $c_p(t) = p[s - (j-1)/p]$.

Then a Corollary of Araujo (1978) implies the following CLT for a sequence of Banach-valued random variables in $\mathbf{H}_2$, which would be useful for our analysis.

Proposition 4.2 : Define $W_u(s,t)$ be i.i,d. random variables at $u = i/n \ (i = 1, \cdots, n)$ as above. Assume

$$\sup_{1 \leq i \leq n} \mathcal{E}[|W_u(s,t)|^{2+\delta}] < \infty$$

for some $\delta > 0$. Then $(1/\sqrt{n}) \sum_{i=1}^{n} W_{i/n}(s,t)$ converges weakly to a Gaussian random variable.

Simulation Results

Let $\mathbf{X} = (\mathbf{X}(t))_{t\geq 0}$ $(= (X^{(1)}(t), \cdots, X^{(p)}(t))')_{t\geq 0}$ be the vector of Itô semimartingale satisfying

$$dX^{(j)}(t) = \sigma_t^{(j)} dB_t^{(j)},\ j = 1, \cdots, q_{1,x}$$
$$dX^{(3)}(t) = Z_t^{(j)} dN_t^{(j)}\ j = 1, \cdots, q_{2,x},$$

where $\mathbf{B} = (B^{(1)}, \cdots, B^{(q_{1,x})})$ is the two dimensional (standard) Brownian motion vector, $\mathbf{N} = (N^{(1)}, \cdots, N^{(q_{2,x})}$ is the Poisson process with intensity λ_j $(j = 1, \cdots, q_{2,x})$ as 10. We assume that $\mathbf{N}$ is independent of $\mathbf{W}$ $(q_x = q_{1,x} + q_{2,x})$ and $Z = (Z_t)_{t\geq 0}$ is the jump sizes with $Z_t \sim \mathrm{N}(0, 5^{-2})$. (See Cont and Tankov (2004) for the generation of jump processes.) For the volatility process σ of the diffusion part, we set

$$d(\sigma_t^{(j)})^2 = a_j(\mu_j - (\sigma_t^{(j)})^2)dt + \kappa_j \sigma_t^{(j)} dW_t^{\sigma^{(j)}}, \quad j = 1, \cdots, q_{1,x},$$

where $\sigma^{(1)}$ and $\sigma^{(2)}$ are independent.

When $q_{1,x} = 2$ and $q_{1,x} = 1$, we set $a_1 = 2$, $a_2 = 3$, $\mu_1 = 0.8$, $\mu_2 = 0.7$, $\kappa_1 = \kappa_2 = 0.5$, $\mathcal{E}[dB_t^{\sigma^{(j)}} dB_t^{(j)}] = \rho_j dt$, $\rho_1 = \rho_2 = -0.5$. (We also have $q_{1,x} = 5$ and $q_{1,x} = 5$,)0 In our simulation, we consider the following two models :

$$\text{Model 1} \;:\; \mathbf{Y}(t) = \mathbf{\Gamma}_1 (X^{(1)}(t), X^{(2)}(t))' + \mathbf{v}(t)$$

and

$$\text{Model 2} \;:\; \mathbf{Y}(t) = \mathbf{\Gamma}_1 (X^{(1)}(t), X^{(2)}(t))' + \mathbf{\Gamma}_2 X^{(3)}(t) + \mathbf{v}(t).$$

Here we denote the coefficients matrices ($p \times q_{1,x}$ and $p \times q_{2,x}$, respectively) as $\mathbf{\Gamma}_1 = (\gamma_1^{(1)}, \ldots, \gamma_1^{(p)})'$, $\mathbf{\Gamma}_2 = (\gamma_2^{(1)}, \ldots, \gamma_2^{(p)})'$and they are sampled as $\gamma_1^{(j)} \sim [0.25, 1.75]$ and $\gamma_2^{(j)} \sim [0.25, 1.75]$, $j = 1, \ldots, p$.

The observation vectors are

$$\mathbf{Y}(i/n) = (Y^{(1)}(i/n), \ldots, Y^{(p)}(i/n))', \; i = 1, \ldots, n$$

and we set $\Delta = \Delta_n = 1/n$. As the market microstructure noise vectors, we set $\mathbf{v}_{i/n} = (v^{(1)}(i/n), \ldots, v^{(p)}(i/n))'$, and use independent Gaussian noises for each component, that is,

$$(v^{(1)}(i/n), \ldots, v^{(p)}(i/n))' \sim i.i.d. N_p(\mathbf{0}, c\mathbf{I}_p) \; (i = 1, \ldots, n) \, ,$$

with a pre-specified value c. In all simulations, we set $p = 100$.

Simulation Results

Let N be the number of Monte Carlo iterations. We plotted the mean value of the eigenvalues of the SIML estimator for the quadratic variation in Figures . To compute the SIML estimators, we set $m_n/n = 0.06$ and $l_n/n = 0.09$.

$$\hat{\mathbf{\Sigma}}_x = \frac{1}{m_n}\sum_{j=1}^{m_n}\mathbf{z}_j\mathbf{z}_j^{'}\ ,\ \ \hat{\mathbf{\Sigma}}_v = \frac{1}{l_n}\sum_{j=n-l_n+1}^{n} a_{jn}^{-1}\mathbf{z}_j\mathbf{z}_j^{'},$$

where $a_{kn} = 4n\sin^2\left[\frac{\pi}{2}\left(\frac{2k-1}{2n+1}\right)\right]$.

In the following Figures, we set $0 \leq \hat{\lambda}_1 \leq \ldots \leq \hat{\lambda}_p$ are eigenvalues of $\hat{\mathbf{\Sigma}}_v^{-1}\hat{\mathbf{\Sigma}}_x$.

In Model-1 and Molde-2 we have p=100 dimensions observation vectors ($p = 100$). Model-1 has two factor of diffusion type ($q_x = 2, r_x = 98$) while Model-2 has two diffusion type factors and one jumps factor ($q_x = 3, r_x = 97$). Figures show that the estimated characteristic roots reflect the true rank of hidden stochastic process. They show the distributions of the test statistic we are developed in this paper. It seems that our method of evaluating the rank condition of hidden volatility factors based on the characteristic roots and the SIML estimation detects the number of factors properly in these two numerical simulations.

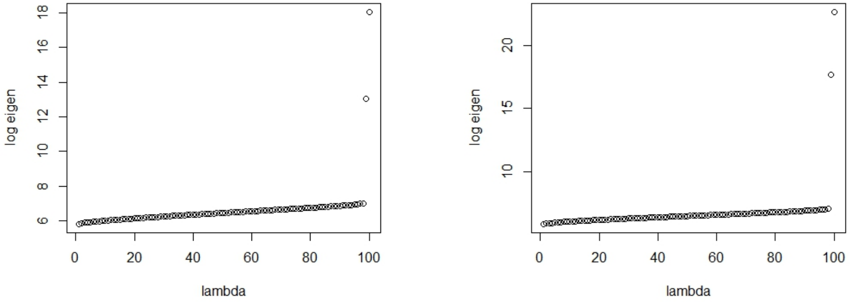


Figure 5.1 : Mean of estimated log characteristic roots (log eigen values) of Model 1 when $c = 10^{-6}$(left) and $c = 10^{-8}$(right).
We set $\Delta = 1/20000$, $m_n/n = 0.06$ and $l_n/n = 0.09$. The number of Monte Carlo iteration is 300.

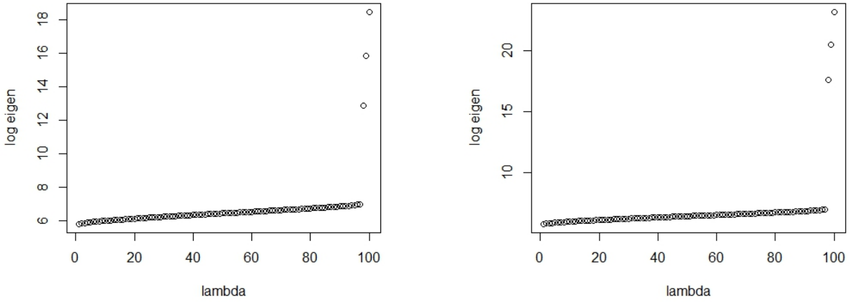


Figure 5.2 : Mean of estimated log characteristic roots (log eigen values) of Model 2 when $c = 10^{-6}$(left) and $c = 10^{-8}$(right).
We set $\Delta = 1/20000$, $m_n/n = 0.06$ and $l_n/n = 0.09$. The number of Monte Carlo iteration is 300.

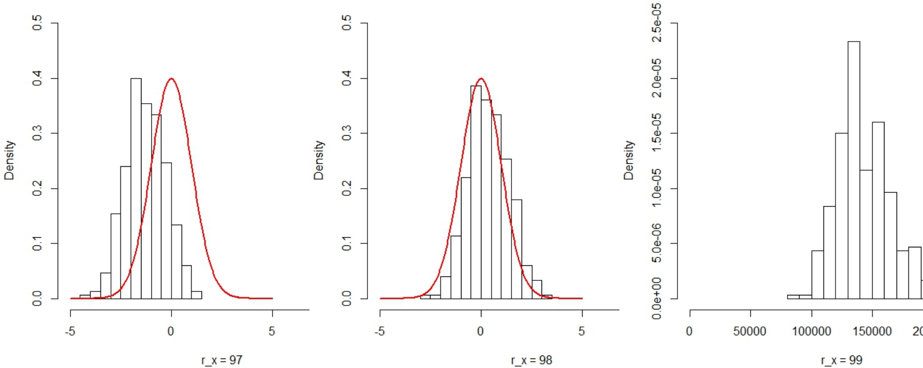


Figure 5.3 : Empirical distributions of test statistic T_n of Model 1 when $r_x = 97$(left), $r_x = 98$(center) and $r_x = 99$(right).
We set $\Delta = 1/20000$, $c = 10^{-8}$, $m_n/n = 0.06$, and $l_n/n = 0.09$.
The number of Monte Carlo iteration is 300. The red line is the density of the chi square distribution with 1 degree of freedom.

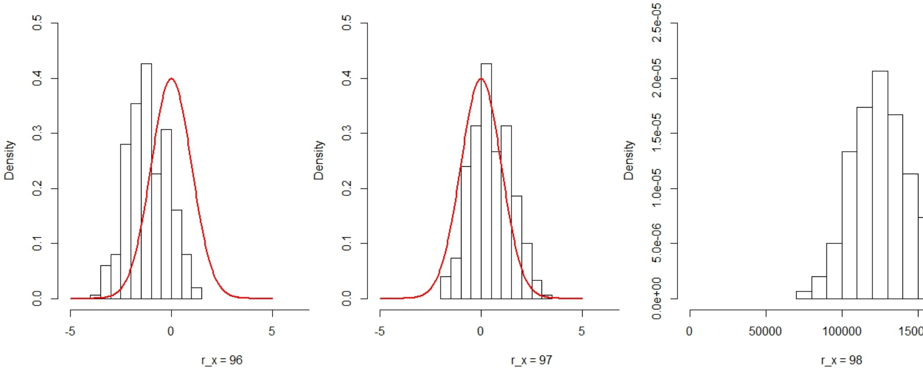


Figure 5.4 : Empirical distributions of test statistic T_n of Model 2 when $r_x = 96$(left), $r_x = 97$(center) and $r_x = 98$(right).
We set $\Delta = 1/20000$, $c = 10^{-8}$, $m_n/n = 0.06$, and $l_n/n = 0.09$.
The number of Monte Carlo iteration is 300. The red line is the density of the chi square distribution with 1 degree of freedom.

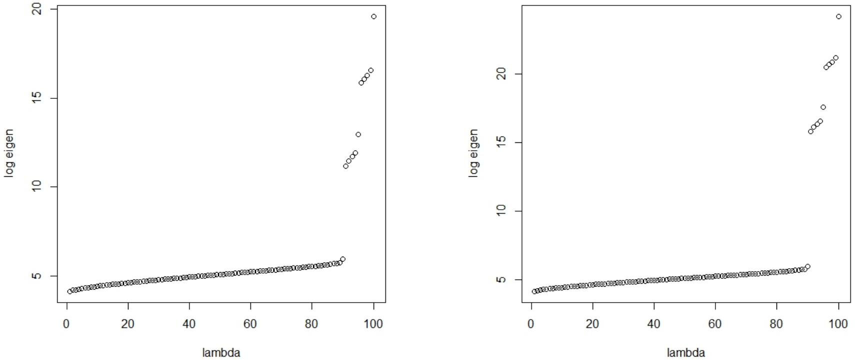


Figure 5.5 : Mean of estimated log characteristic roots (log eigen values) of Model 1 when $c = 10^{-6}$(left) and $c = 10^{-8}$(right).
We set $\Delta = 1/20000$, $m_n/n = 0.06$ and $l_n/n = 0.09$.
The number of Monte Carlo iteration is 300.

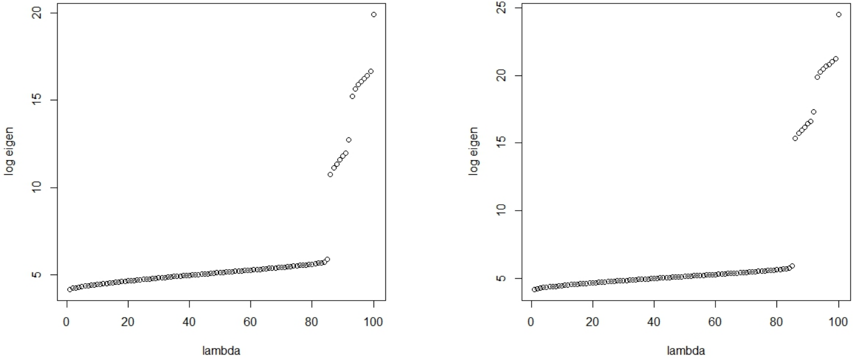


Figure 5.6 : Mean of estimated log characteristic roots (log eigen values) of Model 2 when $c = 10^{-6}$(left) and $c = 10^{-8}$(right).
We set $\Delta = 1/20000$, $m_n/n = 0.06$ and $l_n/n = 0.09$.
The number of Monte Carlo iteration is 300.

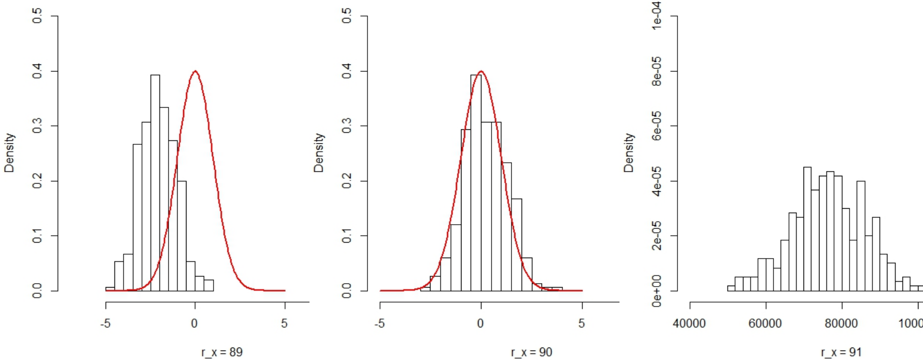


Figure 5.7 : Empirical distributions of test statistic T_n of Model 1 when $r_x = 89$(left), $r_x = 90$(center) and $r_x = 91$(right). We set $\Delta = 1/20000$, $c = 10^{-8}$, $m_n/n = 0.06$, and $l_n/n = 0.09$.
The number of Monte Carlo iteration is 300.
The red line is the density of the chi square distribution with 1 degree of

Further Problems

- We can develo a detecting procedure for the number of hidden factors by using the SIML method when the true hidden stochastic process is a class of Itô semimartingales, there can be market microstructure noises and the dimension is large at the same time.
- Our procedure is partly based on the Banach-valued random variables and CLT(central limit theorems).
- We have derived the asymptotic distributions of characteristic roots.
- From our limited simulations and an empirical application, our approach works well in practical situations.
- The results in Kunitomo and Kurisu (2019) can be generalized when the dimension is high. (Kunitomo, N. and Kurisu, D. (2019). Detecting factors of quadratic variation in the presence of microstructure noise. MIMS-RBP SDS-10.)
- There are several related problems of Ito semi-martingales with market microstructure noises when the dimension is large.

References

- Anderson, T.W. (1984), Estimating Linear Statistical Relationships. *Ann. Statist.* **12** 1-45.
- Christensen, K., M. Podolskij and M. Vetter (2013), On Covariation Estimation for Multivariate Continuous Itô Semimartingales with Noise in Non-Synchronous Observation Schemes. *J. Multivariate Anal.* **120** 59-84.
- Kunitomo, N., Sato and D. Kurisu (2018), *Separating Information Maximum Likelihood Estimation for High Frequency Financial Data*. Springer.
- Araujo, A. P. (1978), "On the Central Limit Theorem in Banach Spaces," *Journal of Multivariate Analysis*, 8, 598-613.
- Ito, K. and Nishio, M. (1968), "On the Convergence of Sum of Independent Banach Space Valued Random Variables," *Osaka Journal of Mathematics*, 35-48.
- Miyahara, Yoshio (1982), *Stochastic Evolution Equations and White Noise Analysis*, Carleton Mathematical Lecture Notes No.42.

- Mandrekar and Büdinger (2015), *Stochastic Integration in Banach Space*, Springer.
- Malliavin, P. and M. E. Mancino (2009), "A Fourier Transform Method for Nonparametric Estimation of Multivariate Volatility," *The Annals of Statistics*, 37-4, 1993-2010.

Estimation of risk aversion for Japanese stock market using implied moments

Kosuke Oya
Graduate School of Economics & MMDS, Osaka Univsersity
「データサイエンス・福島キャンプ 2019」
2019/08/07, 福島大学

Outline

What is risk neutral probablity density?

- Two probability measures $\mathcal{P}$ and $\mathcal{Q}$
- Volatility index, variance risk premium, volatility spread

Moment under $\mathcal{P}$

- Sample moment & realized moment

Estimation for relative risk aversion

Summary and future studies

What is risk neutral probablity density?

Risk neutral valuation

- S: a finite set of state
- $\pi_{\mathcal{P}}(s)$: probability of state s
- $pc(s)$: price of contingent claim that pays ¥1 in state s
- $x(s)$: asset whose payoff for state s

$p(x)$: price of asset x

$$p(x) = \sum_{s \in S} pc(s)\, x(s) = \sum_{s \in S} \pi_{\mathcal{P}}(s) \underbrace{\left(\frac{pc(s)}{\pi_{\mathcal{P}}(s)} \right)}_{m(s)} x(s) = \mathrm{E}_{\mathcal{P}}[\, m(s)\, x(s)\,]$$

R^f: risk free rate, $1/R^f = \sum_{s \in S} pc(s) = \sum_{s \in S} \pi_{\mathcal{P}}(s) m(s) = \mathrm{E}_{\mathcal{P}}[m(s)]$

$$p(x) = \frac{1}{R^f} \sum_{s \in S} \underbrace{\left(\frac{m(s)\pi_{\mathcal{P}}(s)}{\mathrm{E}_{\mathcal{P}}[m(s)]} \right)}_{\pi_{\mathcal{Q}}(s)} x(s) = \frac{1}{R^f} \sum_{s \in S} \pi_{\mathcal{Q}}(s) x(s) = \frac{1}{R^f} \mathrm{E}_{\mathcal{Q}}[\, x(s)\,]$$

What is risk neutral pdf?

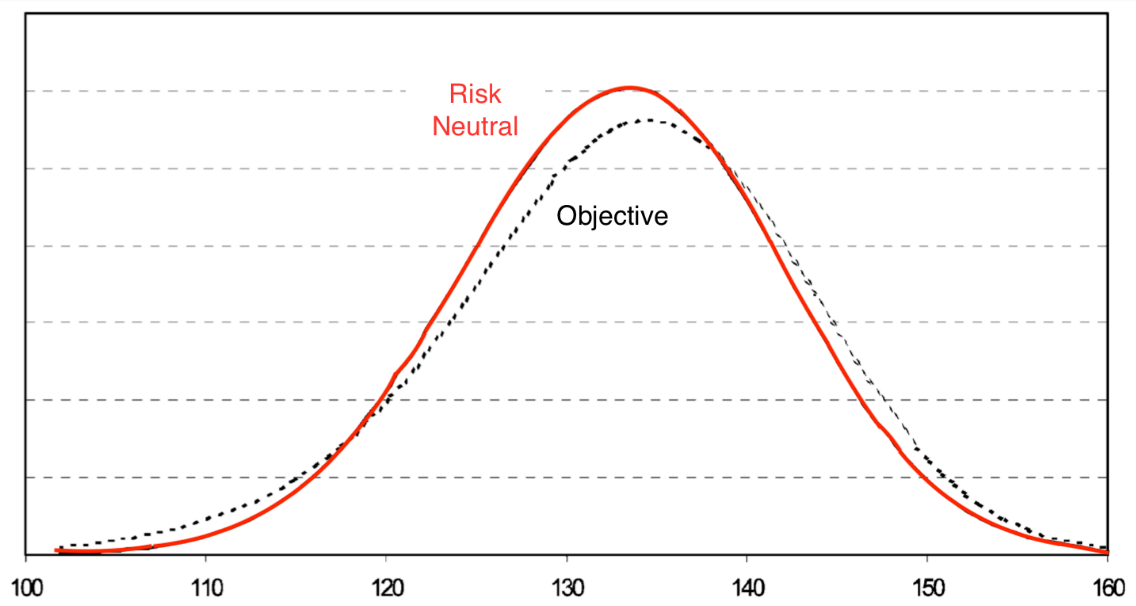
- $S(t)$: stock index price at time t,
- $R(t) = \ln S(t+1)/S(t)$: one period return,
- $\mathcal{P}$: physical probability measure, $\mathcal{Q}$: risk neutral prob. measure,
- $p(r)$: (physical) prob. density func. of $R(t)$, $q(r)$: risk neutral pdf,

Bakshi et al. (2003)

- power utility with relative risk aversion γ suppl.
- pricing kernel $m(r) = e^{-\gamma r}$

$$q(r) = \frac{m(r)\,p(r)}{\int m(r)p(r)\mathrm{d}r} = \frac{e^{-\gamma r}\,p(r)}{\int e^{-\gamma r}p(r)\mathrm{d}r} \quad (1)$$

  - risk neutral $\gamma = 0 \;\Leftrightarrow\; q(r) = p(r)$
  - risk averse $\gamma > 0$ ($q(r) = m(r)p(r)$ for simplicity),
    $q(r) > p(r)$ for $r < 0$, and vice versa



Italian Government Bond Futures: Fig.6 in Fornari and Mele (2001)

This is not example of $q(r) = e^{-\gamma r}p(r)$

Statistical devices

Estimation for risk neutral probablity density

- under the assumption of (specific parametric model)
  specific distribution $p(r)$ and functional form $m(r)$
- Aït-Sahalia and Lo (1998)：Nonparametric estimation

moment (provides useful & enough information)

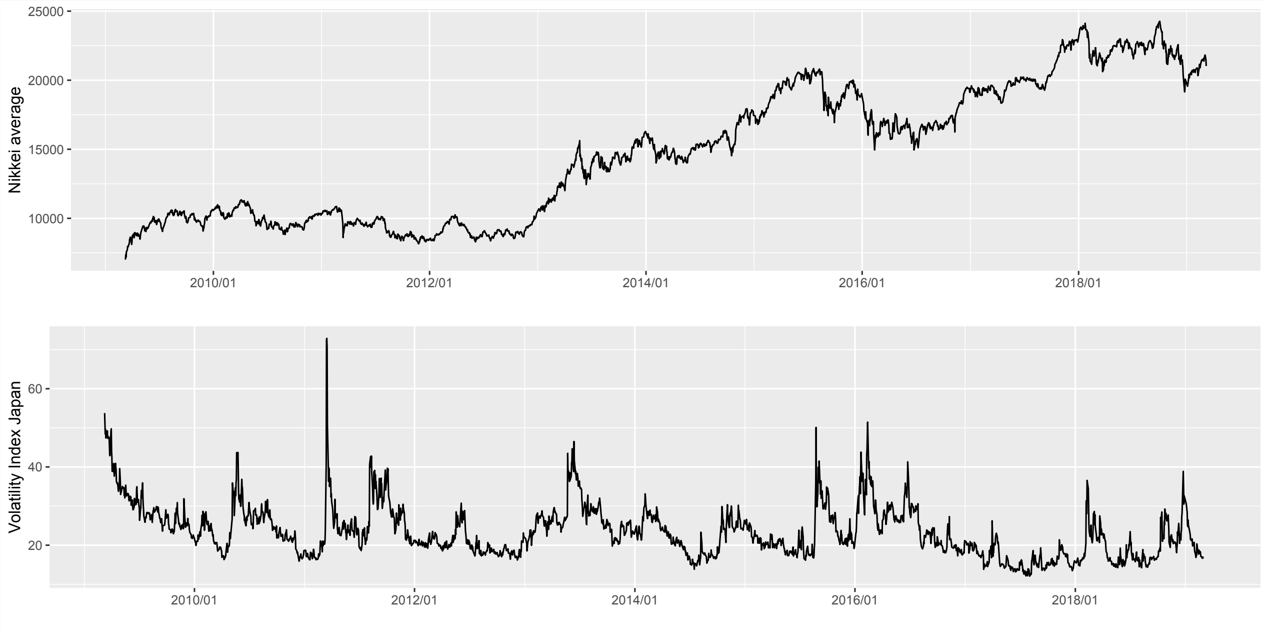
- 1st order moment: risk neutral valuation
- 2nd order moment: (variance, volatility) risk reference
  - $\sqrt{\text{Model free implied variance}}$: volatility index (VIX, VXJ, VI)
  - “implied” means ... impled from observed option price

Hereafter, k th order moments under $\mathcal{P}$ and $\mathcal{Q}$

$$\mu_{\mathcal{P}}^{(k)} = \mathrm{E}_{\mathcal{P}}[R(t)^k] \text{ and } \mu_{\mathcal{Q}}^{(k)} = \mathrm{E}_{\mathcal{Q}}[R(t)^k], \text{ respectively}$$

Implied moment：Volatility index

- VIX は Cboe Global Markets が開発した index で，S&P500 の将来の収益率の変動の大きさの期待値として算出されている
- 我が国では大阪大学 CSFI（現：MMDS）が 2008 年 7 月に学術研究目的として VXJ (Volatility Index Japan) を公表，2010 年 11 月に日本経済新聞社が日経平均 VI の公表開始し，2012 年 2 月に大阪取引所において日経平均 VI 先物の取引が開始
- vol. index は株価指数との間に負の相関をもつことから恐怖指数（fear gauge）としても知られる



出所： 大阪大学 mmds

Implied moment：Volatility index

- MFIV は，put と call option の価格 $P(K)$ と $C(K)$ に関する項を権利行使価格 K に関して積分する以下の式から

$$\mathrm{E}_Q[(R(t)-\mu_Q^{(1)})^2] = 2\int_0^F \frac{P(K)}{K^2}\mathrm{d}K + 2\int_F^\infty \frac{C(K)}{K^2}\mathrm{d}K$$

- 実際には K は離散的で，$P(K)$ と $C(K)$ も対応する K のすべての領域で観測できるわけではないため近似計算が必要
- この積分を離散近似によって算出しているのが VIX
- 大阪大学が公開している VXJ はこの積分を Gatheral (2006) の方法で適切に計算したもの (Fukasawa et al., 2011)
- Carr and Madan (2001) や Fukasawa (2012) は 2 階連続微分可能な $S(t)$ の関数の Q のもとでの期待値を導出
    - $\mu_Q^{(3)}$，$\mu_Q^{(4)}$，そして歪度と尖度も option 価格を使って同様に

Implied moment ：分散リスク・プレミアム（VRP）

それぞれの測度のもとでの分散：$\sigma^2_{\mathcal{P}} = \mu^{(2)}_{\mathcal{P}} - \left(\mu^{(1)}_{\mathcal{P}}\right)^2$,　$\sigma^2_{\mathcal{Q}} = \mu^{(2)}_{\mathcal{Q}} - \left(\mu^{(1)}_{\mathcal{Q}}\right)^2$

分散リスク・プレミアム (VRP)：　$VRP = \sigma^2_{\mathcal{Q}} - \sigma^2_{\mathcal{P}}$

VRP を金融市場に関連する予測変数とする研究

- Bollerslev et al. (2009)：市場収益率に対する予測力
  long-run risk model の拡張
  （追加的な不確実性の源泉：vol. の変動の risk）
- Bollerslev et al. (2014)：
  US を含む主要 8 カ国の VRP と共通する global VRP
- 大屋 (2011)：景気指標に対する予測力
- Ubukata and Watanabe (2014)：
  予測力あり（credit spread，景気指標），予測力なし（市場収益率）
- 渡部 (2016)：実現分散に Heterogenous AR モデルとその拡張モデル

Implied moment：$\mathcal{P}$ のもとでの moment との関係

Pricing kernel $m(R) = e^{-\gamma R}$ の仮定のもと　suppl.

- Bakshi and Madan (2006)：$\mu^{(k)}_{\mathcal{P}}, (k = 2, 3, 4)$ と γ で $\mu^{(1)}_{\mathcal{Q}}$，$\mu^{(2)}_{\mathcal{Q}}$ を表現

$$\mu^{(1)}_{\mathcal{Q}} = -\gamma\mu^{(2)}_{\mathcal{P}} + \frac{\gamma^2}{2}\mu^{(3)}_{\mathcal{P}} + o(\gamma^2)$$

$$\mu^{(2)}_{\mathcal{Q}} = \mu^{(2)}_{\mathcal{P}} - \gamma\mu^{(3)}_{\mathcal{P}} + \frac{\gamma^2}{2}(\mu^{(4)}_{\mathcal{P}} - (\mu^{(2)}_{\mathcal{P}})^2) + o(\gamma^2)$$

- Bakshi et al. (2003)：$\mathcal{P}$ のもとでの歪度 $\theta_{\mathcal{P}}$，尖度 $\kappa_{\mathcal{P}}$ で $\theta_{\mathcal{Q}}$ を表現

$$\theta_{\mathcal{Q}} \approx \theta_{\mathcal{P}} - \gamma\,(\kappa_{\mathcal{P}} - 3)\,(\,\sigma^2_{\mathcal{P}}\,)^{1/2}$$

  - risk neutral pdf $q(r)$ は，リスク回避度 $\gamma > 0$，$\kappa_{\mathcal{P}} > 3$ のとき，現実の確率密度関数 $p(r)$ よりも負の方向に歪む
  - R の（$\mathcal{P}$ で）正規分布に従っているなら，$q(r)$ も歪度=0

Implied moment：ボラティリティ・スプレッド

Theorem (Bakshi and Madan (2006): Volatility spread)

$$\frac{\sigma_Q^2 - \sigma_{\mathcal{P}}^2}{\sigma_{\mathcal{P}}^2} \approx -\gamma \underbrace{(\sigma_{\mathcal{P}}^2)^{1/2}}_{2nd\ order} \times \underbrace{\theta_{\mathcal{P}}}_{3rd} + \gamma^2 \times \underbrace{\frac{\sigma_{\mathcal{P}}^2}{2}}_{2nd} \times (\underbrace{\kappa_{\mathcal{P}}}_{4th} - 3)$$

vol. スプレッドが拡大するのは

- 分布の裾が厚い ($\kappa_{\mathcal{P}} > 3$)， 負の歪度 ($\theta_{\mathcal{P}} < 0$)
- リスク回避度 ($\gamma > 0$)

スプレッドがゼロとなるのは

- リスク回避度 $\gamma = 0$
- $R \sim N(\mu_{\mathcal{P}}^{(1)}, \sigma_{\mathcal{P}}^2) \Rightarrow \theta_{\mathcal{P}} = 0, \kappa_{\mathcal{P}} = 3$

Moment under $\mathcal{P}$

確率測度 $\mathcal{P}$ のもとでの moment

低頻度観測データ

- 月次の分散：日次収益率の利用（月内の営業日数は小標本）
- 日次の分散：日次収益率系列に対する rolling window での標本分散

高頻度観測データ

- 実現（realized）moment：tick data によるモーメント推定
- 実現分散（Realized Variance: RV）：$\sigma^2_{\mathcal{P}}$ の推定量
- 高頻度観測の収益率の期待値はゼロと仮定 $\mu^{(1)}_{\mathcal{P}} = 0$

実現 moment（Realized moment: RM）

$$RM^{(k)} = \sum_{i=1}^{n} r(t_i)^k$$

実現 moment の漸近的性質

- 期間 $[0, T]$ において，価格 $S(t_i)$，$0 = t_0 < t_1 < \cdots < t_n = T$ が観測
- $W(s)$ は std Wiener 過程，$\sigma(s)$ は実数値適合過程，$J(t)$ は jump 過程
- 収益率 $r(t_i) = p(t_i) - p(t_{i-1})$，時点 s での jump $\Delta p(s) = p(s) - p(s-)$

対数価格 $p(t) = \ln S(t)$ が以下のように表されていると考える

$$\ln S(t) = p(t) = p(0) + \int_0^t \sigma(s) \mathrm{d}W(s) + J(t)$$

$n \to \infty$ のとき

$$\text{実現分散 :} RM^{(2)} \xrightarrow{p} \int_0^T \sigma(s)^2 \mathrm{d}s + \sum_{0<s\leq T} (\Delta p(s))^2$$

$$RM^{(k)} \xrightarrow{p} \sum_{0<s\leq T} (\Delta p(s))^k, \quad k = 3, 4$$

実現 moment の漸近的性質

- $RM^{(k)}$, $(k = 3, 4)$ は $R(t)$ が従っている確率過程の連続部分に起因する
  - 3 次の場合は収益率と分散変動の共分散であるレバレッジ効果
  - 4 次の場合は分散変動の分散

  を捉えることができていない
- ただし，実現 moment の漸近的な性質は $t_i - t_{i-1} \to 0$, $(n \to \infty)$ だが，実際には microstructure noise の影響を考慮して，5 分間隔といった時間間隔が選ばれることが多い (生方・渡部, 2011)
  - microstructure noise に関しては
    Ubukata and Oya (2009); Aït-Sahalia and Jacod (2014)
- 5 分間隔の $RM^{(k)}$, $(k = 3, 4)$ では，$R(t)$ の jump 部分に起因する項が支配的になっていると考えられる
- 次節での実証では，$RM^{(k)}$ を使った相対的リスク回避度 γ の推定と長期の低頻度観測データによる推定の両方を行い，推定結果を総合的に検討する

Estimation for relative risk aversion

Previous studies

- Aït-Sahalia and Lo (1998) and Aıt-Sahalia and Lo (2000)
  - オプション価格を利用した状態価格密度の nonparametric estimators とリスク回避度の推定
  - リスク回避度で調整された nonparametric VaR
- Bliss and Panigirtzoglou (2004)
  - オプション価格を利用した $q(r)$ の推定とリスク回避度の推定
- Bakshi and Madan (2006)
  - ボラティリティ・スプレッドとリスク回避度 γ の推定
- Yoon and Byun (2012)
  - BM06 による S&P 500, 日経 225 and KOSPI 200 での γ の推定
  - $\hat{\gamma}$： S&P 500 > 日経 225 > KOSPI 200
- 大屋 (2017)
  - BM06 の vol. スプレッドの紹介と実現 moment を使っての γ の推定

Volatility spread, skewness, kurtosis

Monthly volatility spread

- σ_Q^2 には VXJ の 2 乗値の月中平均，$\sigma_{\mathcal{P}}^2$ には月次実現分散

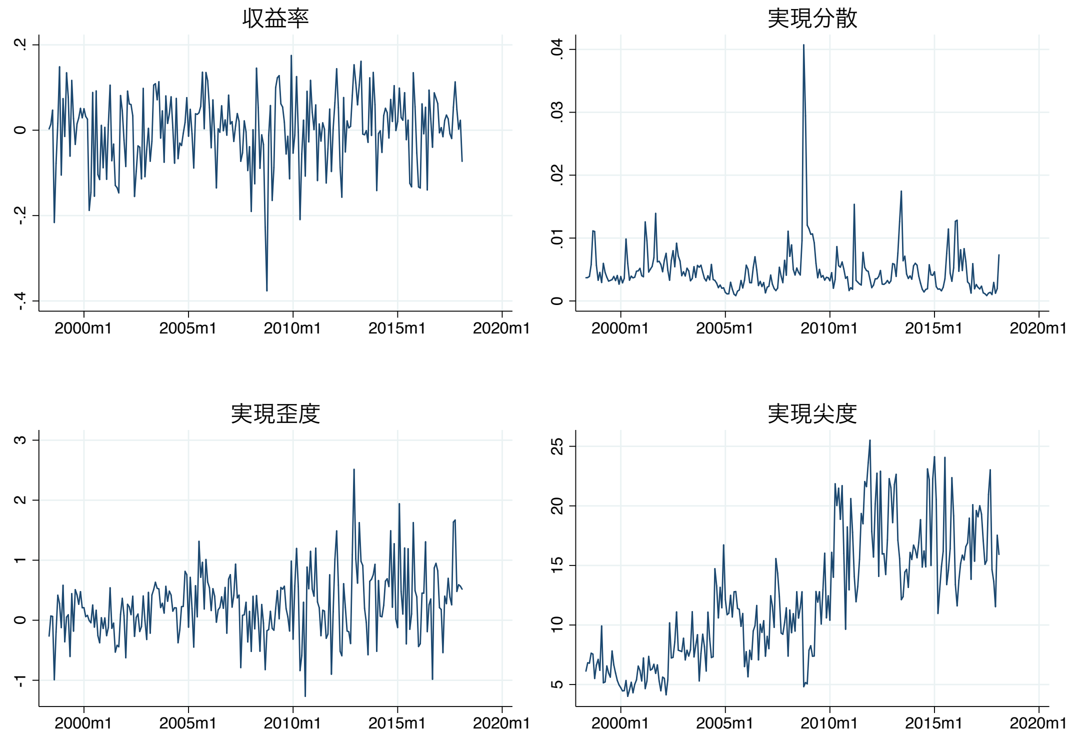
実現分散，歪度，尖度 (Amaya et al., 2015)

- t 月の第 i 日の第 j 番目の収益率を $r_i(j)$, $j = 1, \ldots, n_i$
- $DRV_i = \sum_{j=1}^{n_i} r_i(j)^2, \quad DRS_i = \dfrac{\sqrt{n_i}\sum_{j=1}^{n_i} r_i(j)^3}{DRV_i^{3/2}}, \quad DRS_i = \dfrac{n_i \sum_{j=1}^{n_i} r_i(j)^4}{DRV_i^2}$
- $RV_t = \sum_{i=1}^{N_i} DRV_i, \quad RS_i = \dfrac{1}{N_i}\sum_{i=1}^{N_i} DRS_i, \quad RK_i = \dfrac{1}{N_i}\sum_{i=1}^{N_i} DRK_i$

月次標本歪度・標本尖度

- rolling window による日次収益率を使った標本歪度・標本尖度

日経平均株価：月次収益率，RV，RS，RK



日経平均株価：標本歪度，標本尖度

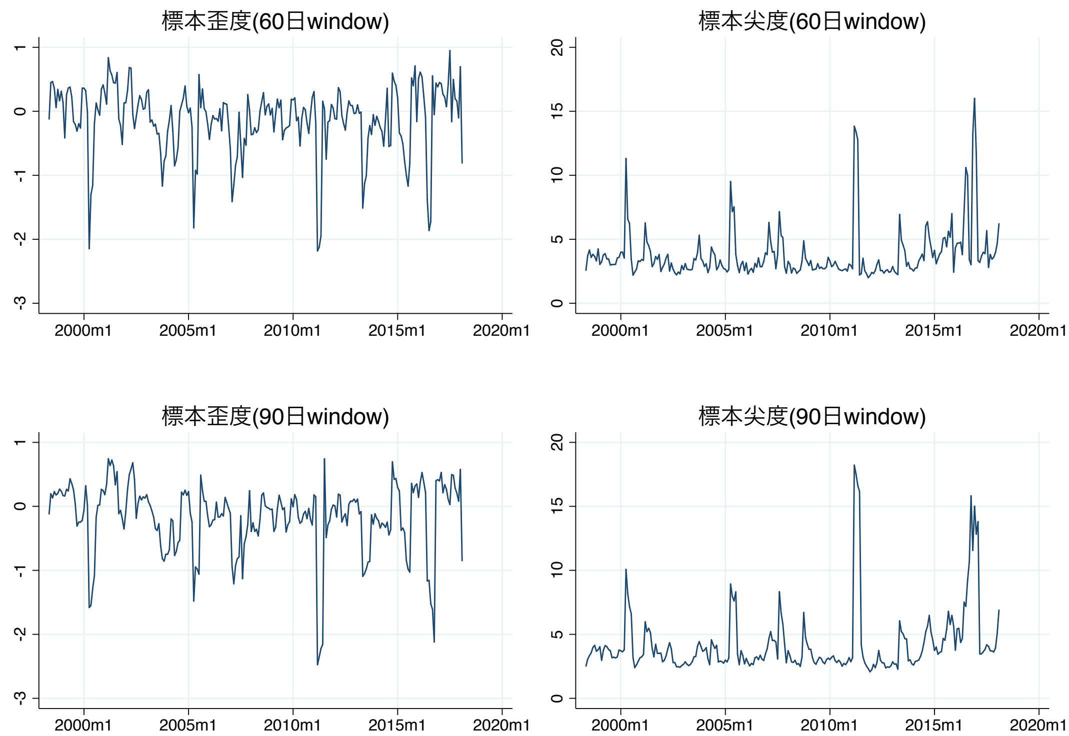


表 1：歪度・尖度の推定値（期間別）

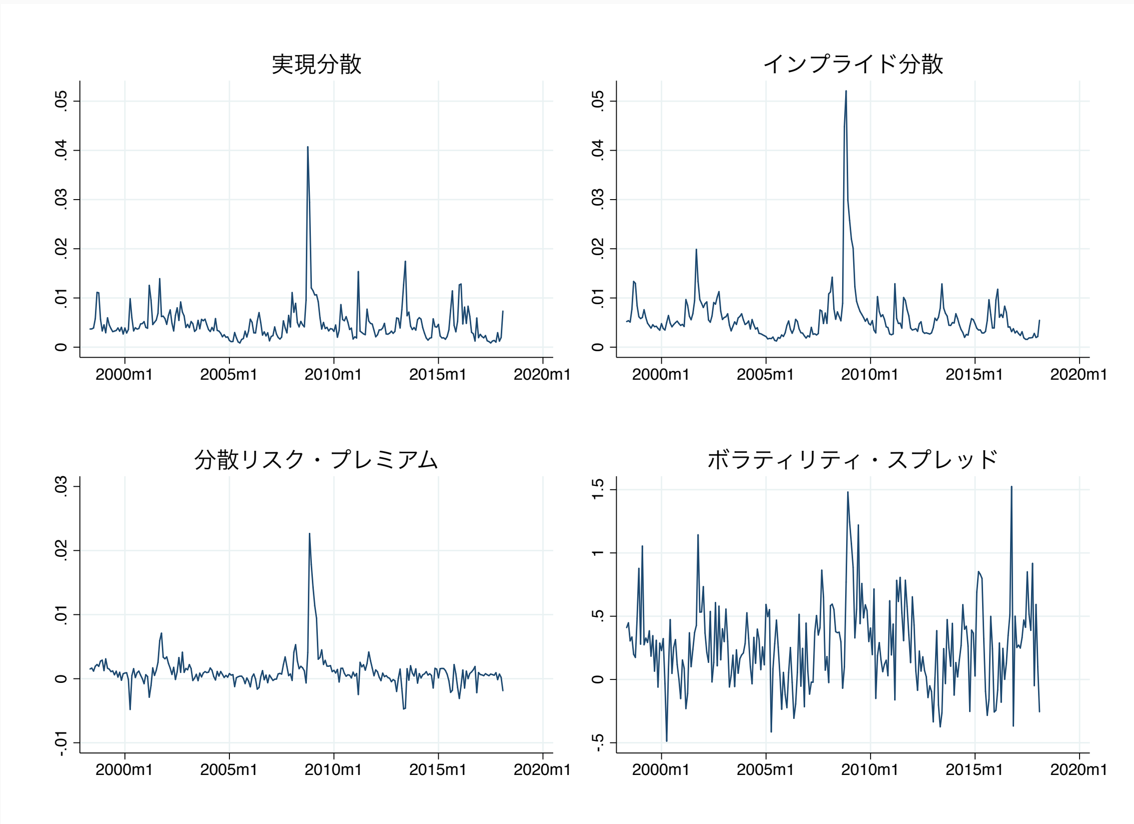
		RS	SS_{60}	SS_{90}	RK	SK_{60}	SK_{90}
期間 1	mean	0.077	-0.032	-0.057	7.022	3.502	3.728
98/Mar – 04/Oct	se	0.046	0.087	0.096	0.388	0.197	0.225
期間 2	mean	0.254	-0.248	-0.316	11.679	3.707	4.336
04/Nov – 11/Jun	se	0.066	0.084	0.102	0.723	0.372	0.597
期間 3	mean	0.452	-0.102	-0.14	17.474	4.235	4.732
11/Jul – 18/Feb	se	0.081	0.090	0.097	0.496	0.412	0.577

RS，RK は実現歪度，実現尖度，SS は標本歪度，SK は標本尖度で，添字はウィンドウの大きさを表す。

- 実現歪度は期間 2，3 でその標本平均が有意に正
- 標本歪度はいずれの期間でも負だが，有意なのは期間 2 のみ
- 尖度はどちらも，期間 1 から 3 になるにつれて，その標本平均の値を増加させている
- 実現尖度は標本尖度よりも大きな値をとっている

検討結果へ

分散リスク・プレミアム，ボラティリティ・スプレッド



リスク回避度の推定

GMM 推定

$$\varepsilon(t) = \frac{\sigma^2_{\mathcal{Q}}(t) - \sigma^2_{\mathcal{P}}(t)}{\sigma^2_{\mathcal{P}}(t)} + \gamma \left(\sigma^2_{\mathcal{P}}(t)\right)^{1/2} \theta_{\mathcal{P}}(t) - \frac{\gamma^2}{2} \sigma^2_{\mathcal{P}}(t) \left(\kappa_{\mathcal{P}}(t) - 3\right)$$

- $\theta_{\mathcal{P}}(t)$ と $\kappa_{\mathcal{P}}(t)$
    - 推定 HF：実現歪度，実現尖度
    - 推定 LF60：60 日 rolling window による標本歪度，標本尖度
    - 推定 LF90：90 日 rolling window による標本歪度，標本尖度
- 操作変数ベクトル $Z(t)$

  set 1： $Z(t) = (1, \sigma^2_{\mathcal{Q}}(t))'$，　set 2： $Z(t) = (1, \sigma^2_{\mathcal{Q}}(t), \sigma^2_{\mathcal{Q}}(t-1))'$
- $J_T(\gamma) = T g_T(\gamma)' W_T g_T(\gamma)$ を最小化する γ
    - W_T は Newey-West の HAC ウエイト行列
    - $g_T(\gamma) \equiv \frac{1}{T} \sum_{t=1}^{T} Z(t-1)\varepsilon(t)$

表 2a：推定 HF

期間	推定	操作変数	γ	HAC 標準誤差	Hansen's J 統計量	p 値
1: 98/Mar–04/Oct			5.498	0.348	1.293	0.256
2: 04/Nov–11/Jun	HF	set1	4.352	0.358	1.435	0.231
3: 11/Jul–18/Feb			2.825	0.312	1.330	0.249
1: 98/Mar–04/Oct			5.311	0.288	2.099	0.350
2: 04/Nov–11/Jun	HF	set2	4.439	0.332	1.650	0.438
3: 11/Jul–18/Feb			2.823	0.312	1.336	0.513

- 操作変数 set 1，set 2 どちらの推定結果もほぼ同じ
- 期間 1 から 3 に向けて，低下傾向

表 2b：推定 LF

期間	推定	操作変数	γ	HAC 標準誤差	Hansen's J 統計量	p 値
1: 98/Mar–04/Oct			7.243	3.129	2.467	0.116
2: 04/Nov–11/Jun	LF60	set1	9.153	2.868	5.081	0.024
3: 11/Jul–18/Feb			7.008	1.856	1.469	0.226
1: 98/Mar–04/Oct			6.079	2.407	5.747	0.056
2: 04/Nov–11/Jun	LF60	set2	6.730	1.473	1.499	0.473
3: 11/Jul–18/Feb			7.067	1.857	1.528	0.466
1: 98/Mar–04/Oct			7.604	2.564	2.504	0.114
2: 04/Nov–11/Jun	LF90	set1	6.830	1.736	2.132	0.144
3: 11/Jul–18/Feb			6.373	1.601	1.435	0.231
1: 98/Mar–04/Oct			5.792	2.335	2.762	0.251
2: 04/Nov–11/Jun	LF90	set2	5.872	1.084	2.389	0.303
3: 11/Jul–18/Feb			6.370	1.597	1.444	0.486

- 標準誤差と比べると各期間の推定値の差は顕著ではない
- 操作変数 set 2，LF90 の方が標準誤差が小さい

表 3：制約付き推定（操作変数 set 2）

制約	期間	推定	γ	HAC 標準誤差	Hansen's J 統計量	p 値
	1		4.817	0.367	3.616	0.164
$\theta_{\varphi} = 0$	2	HF	3.972	0.344	1.841	0.398
	3		2.508	0.331	0.859	0.651
	1		10.042	2.182	5.707	0.058
$\theta_{\varphi} = 0$	2	LF90	7.711	0.771	1.705	0.426
	3		7.919	1.565	1.515	0.469
	1		-85.169	64.583	0.711	0.701
$\kappa_{\varphi} = 3$	2	HF	-13.970	6.526	1.427	0.490
	3		-10.383	3.179	0.556	0.757
	1		4.111	4.589	3.869	0.145
$\kappa_{\varphi} = 3$	2	LF90	11.392	4.111	1.679	0.432
	3		9.968	4.903	2.054	0.358

- 制約 $\theta_{\varphi} = 0$ については，推定 HF では制約なしの結果とほぼ同じで，推定 LF でも差は小さい
- 制約 $\kappa_{\varphi} = 3$ は明らかに間違っている

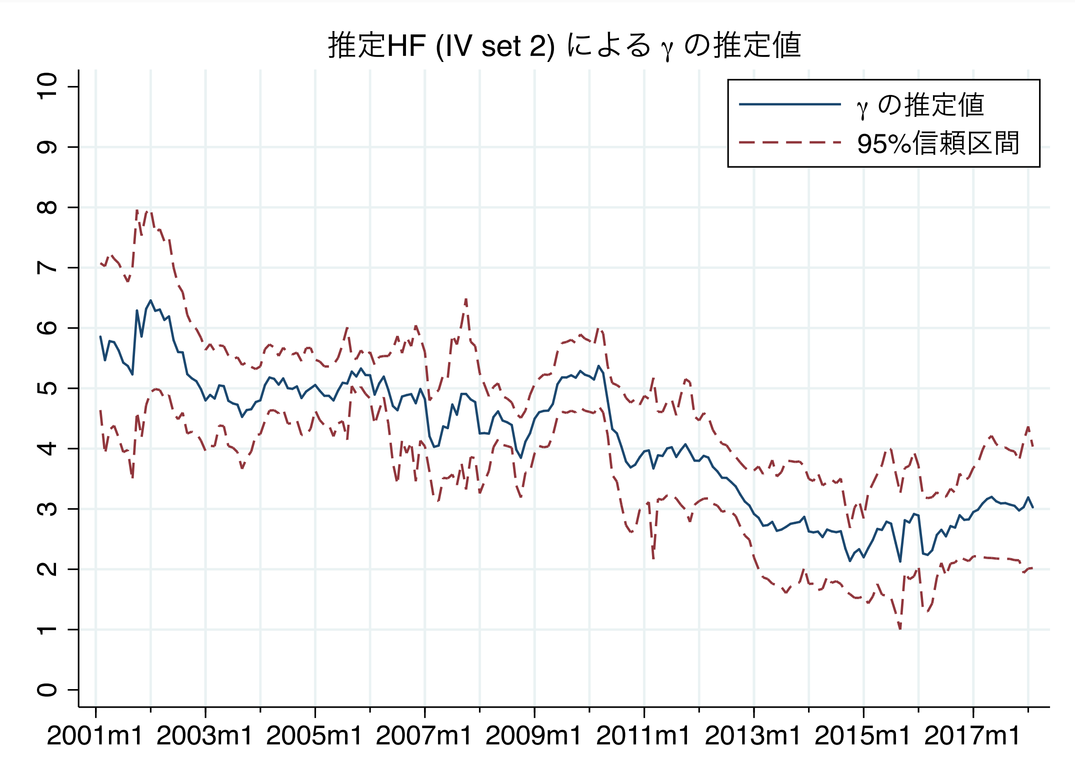
検討結果

期間 1 表 1 より歪度は RS でも SS_{60}，SS_{90} でもゼロからの乖離は有意ではないので，推定結果は推定 HF(IV set2) を採用

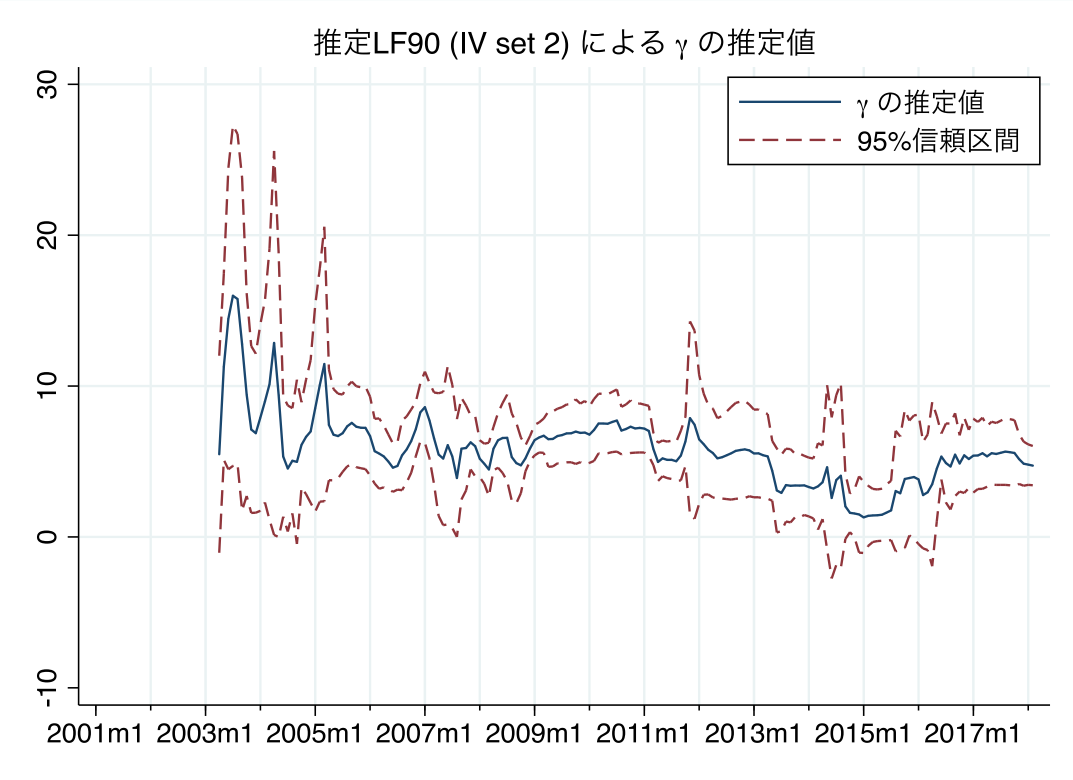
期間 2 表 1 より歪度は，RS は有意に正，SS_{60}，SS_{90} ともに有意に負を示しているが，表 3 での歪度=0 の制約を課した推定結果と制約を課さない推定結果に大きな差はないため，s.e. の小さい推定 HF(IV set2) を採用

期間 3 表 1 より RS は有意に正となっているが，表 3 の結果より，推定 HF(IV set2) を採用

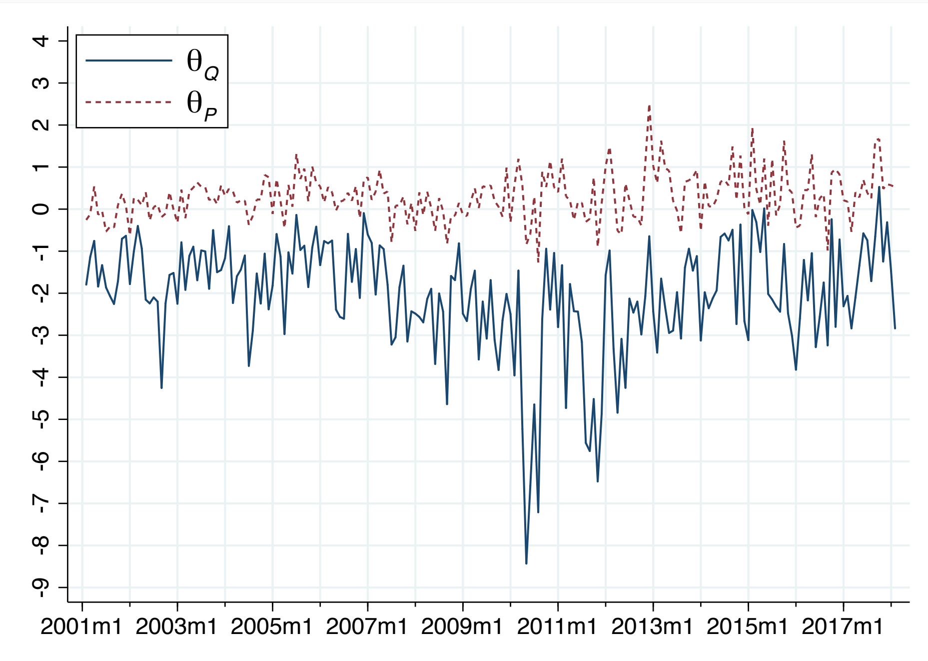
ローリング・ウィンドウによる推定結果: 推定 HF



ローリング・ウィンドウによる推定結果: 推定 LF



ローリング・ウィンドウによる推定結果: $\mathcal{P}$ と Q のもとでの歪度



Bakshi et al. (2003) : $\theta_Q \approx \theta_{\mathcal{P}} - \gamma\,(\kappa_{\mathcal{P}} - 3)\,(\,\sigma^2_{\mathcal{P}}\,)^{1/2}$

Summary and future studies

まとめと今後の検討課題

まとめ

1. リスク中立確率測度から市場参加者のリスクに対するスタンスに関して，リスク回避度をひきだす Bakshi and Madan (2006) の方法を日本市場の日経平均株価に適用
2. 金融危機以前はリスク回避度は低下傾向を示していたが，金融危機を機に一旦，リスク回避度の上昇がみられ，その後，再び低下傾向を示していることが明らかになった

今後の検討課題

- $\theta_{\mathcal{P}}$，$\kappa_{\mathcal{P}}$ の推定
- pricing kernel(utility function) の特定化

参考文献

References

Aït-Sahalia, Yacine and Jean Jacod (2014) *High-frequency financial econometrics*: Princeton University Press.

Aït-Sahalia, Yacine and Andrew W Lo (1998) "Nonparametric estimation of state-price densities implicit in financial asset prices," *Journal of Finance*, Vol. 53, No. 2, pp. 499–547.

Aıt-Sahalia, Yacine and Andrew W Lo (2000) "Nonparametric risk management and implied risk aversion," *Journal of econometrics*, Vol. 94, No. 1-2, pp. 9–51.

Amaya, Diego, Peter Christoffersen, Kris Jacobs, and Aurelio Vasquez (2015) "Does realized skewness predict the cross-section of equity returns?" *Journal of Financial Economics*, Vol. 118, No. 1, pp. 135–167.

Bakshi, Gurdip and Dilip Madan (2006) "A theory of volatility spreads," *Management Science*, Vol. 52, No. 12, pp. 1945–1956.

Bakshi, Gurdip, Nikunj Kapadia, and Dilip Madan (2003) "Stock return characteristics, skew laws, and the differential pricing of individual equity options," *Review of Financial Studies*, Vol. 16, No. 1, pp. 101–143.

Bliss, Robert R and Nikolaos Panigirtzoglou (2004) "Option-implied risk aversion estimates," *Journal of Finance*, Vol. 59, No. 1, pp. 407–446.

Bollerslev, Tim, George Tauchen, and Hao Zhou (2009) "Expected stock returns and variance risk premia," *Review of Financial Studies*, Vol. 22, No. 11, pp. 4463–4492.

Bollerslev, Tim, James Marrone, Lai Xu, and Hao Zhou (2014) "Stock return predictability and variance risk premia: statistical inference and international evidence," *Journal of Financial and Quantitative Analysis*, Vol. 49, No. 3, pp. 633–661.

Carr, Peter and Dilip Madan (2001) "Optimal positioning in derivative securities."

Fornari, Fabio and Antonio Mele (2001) "Recovering the probability density function of asset prices using GARCH as diffusion approximations," *Journal of Empirical Finance*, Vol. 8, No. 1, pp. 83–110.

Fukasawa, Masaaki (2012) "The normalizing transformation of the implied volatility smile," *Mathematical Finance*, Vol. 22, No. 4, pp. 753–762.

Fukasawa, Masaaki, Isao Ishida, Nabil Maghrebi, Kosuke Oya, Masato Ubukata, and Kazutoshi Yamazaki (2011) "Model-free implied volatility: from surface to index," *International Journal of Theoretical and Applied Finance*, Vol. 14, No. 4, pp. 433–463.

Gatheral, Jim (2006) *The volatility surface: a practitioner's guide*: John Wiley & Sons.

Ubukata, Masato and Kosuke Oya (2009) "Estimation and testing for dependence in market microstructure noise," *Journal of Financial Econometrics*, Vol. 7, No. 2, pp. 106–151.

Ubukata, Masato and Toshiaki Watanabe (2014) "Market variance risk premiums in Japan for asset predictability," *Empirical Economics*, Vol. 47, No. 1, pp. 169–198.

Yoon, Sun-Joong and Suk Joon Byun (2012) "Implied risk aversion and volatility risk premiums," *Applied Financial Economics*, Vol. 22, No. 1, pp. 59–70.

生方雅人・渡部敏明 (2011) 「実現ボラティリティ–ボラティリティの計測方法の発展とリスクマネジメントへの応用可能性 (特集ボラティリティと証券投資)」,『証券アナリストジャーナル』, 第 49 巻, 第 8 号, 16–26 頁.

大屋幸輔 (2011) 「ボラティリティの景気予測力 -バリアンス・リスクプレミアムの検証から-」,『浅子和美・飯塚信夫・宮川努編『世界同時不況と景気循環分析』, 第 7 章』, 141–157 頁.

——— (2017) 「ボラティリティ・スプレッド」,『先物・オプションレポート』, 第 29 巻, 第 12 号.

渡部敏明 (2016) 「日経 225 分散リスク・プレミアムの予測力」,『先物・オプションレポート』, 第 28 巻, 第 11 号.

補足：効用関数と pricing kernel

Find p_t (the value at time t) of a payoff x_{t+1}

- Assume that the investor can freely buy or sell as much of the payoff x_{t+1} (which is $r.v.$) as he wishes, at the price p_t.
  - Let e_t be the original consumption level at time t and ξ be the amount of the asset he choose to buy.
  - the consumptions at t and $t+1$ are $c_t = e_t - p_t\xi$ and $c_{t+1} = e_{t+1} + x_{t+1}\xi$
- Investor's utility function $U(c_t, c_{t+1}) = u(c_t) + \beta\,\mathrm{E}_t[\,u(c_{t+1})\,]$
  - $u(\cdot)$ is increasing, reflecting a desire for more consumption, and concave, reflecting the declining marginal value of additional consumption
  - β captures impatience, called the subjective discount factor

$$\max_{\xi} u(c_t) + \beta\,\mathrm{E}[\,u(c_{t+1})\,], \quad s.t.\;\; c_t = e_t - p_t\xi \;\; and \;\; c_{t+1} = e_{t+1} + x_{t+1}\xi$$

The first order condition

$$p_t = \mathrm{E}_t\left[\beta\,\frac{u'(c_{t+1})}{u'(c_t)}\,x_{t+1}\right] = \mathrm{E}_t[\,m_{t+1}(x_{t+1})\,x_{t+1}\,], \quad m_{t+1}(x_{t+1}) = \beta\,\frac{u'(c_{t+1}(x_{t+1}))}{u'(c_t)}$$

Under the assumption of

- time-additive utility func. $u(S_t) + \beta\,u(S_{t+1})$, β: subjective discount factor
- power utility $u(S) = \dfrac{S^{1-\gamma}-1}{1-\gamma}$, γ is the relative risk-aversion

Pricing kernel (stochastic discount factor) is given as

$$m_{t+1} = \beta\,\frac{u'(S_{t+1})}{u'(S_t)} = \beta\,e^{-\gamma\,(\log S_{t+1} - \log S_t)} = \beta\,e^{-\gamma\,R_{t+1}}$$

where $R_{t+1} = \log S_{t+1} - \log S_t$. Then the risk-neutral density $q(R)$ is

$$q(R) = \frac{m(R)\,p(R)}{\int m(R)\,p(R)\mathrm{d}R} = p(R) \times \frac{e^{-\gamma R}}{\int e^{-\gamma R}\,p(R)\mathrm{d}R}$$

- risk-averse investors pay more attention to unpleasant states in comparison to the physical counterpart.

slide 4

Representaion $\mu_{\mathcal{Q}}^{(1)}$ in terms of $\mu_{\mathcal{P}}^{(k)}$

- Denote m.g.f. of R under $p(R)$ and $q(R)$ as $\mathcal{M}_{\mathcal{P}}(\lambda)$ and $\mathcal{M}_{\mathcal{Q}}(\lambda)$.
- Assume $\mu_{\mathcal{P}} = 0$

$$\begin{aligned}\mathcal{M}_{\mathcal{P}}(\lambda) &= \int e^{\lambda r} p(r) \mathrm{d}r \\ &= \int \left\{ 1 + \lambda r + \frac{1}{2}\lambda^2 r^2 + \frac{1}{6}\lambda^3 r^3 + \frac{1}{24}\lambda^4 r^4 \right\} p(r)\mathrm{d}r + o(\lambda^4) \\ &= 1 + \frac{\lambda^2}{2}\mu_{\mathcal{P}}^{(2)} + \frac{\lambda^3}{6}\mu_{\mathcal{P}}^{(3)} + \frac{\lambda^4}{24}\mu_{\mathcal{P}}^{(4)} + o(\lambda^4)\end{aligned}$$

$$\mathcal{M}_{\mathcal{Q}}(\lambda) = \int e^{\lambda r} q(r)\mathrm{d}r = \frac{\int e^{\lambda r} e^{-\gamma r} p(r)\,\mathrm{d}r}{\int e^{-\gamma r} p(r)\mathrm{d}r} = \frac{\mathcal{M}_{\mathcal{P}}(\lambda - \gamma)}{\mathcal{M}_{\mathcal{P}}(-\gamma)}$$

$$\mu_{\mathcal{Q}}^{(1)} = \int r\, q(r)\mathrm{r} = \frac{\mathcal{M}_{\mathcal{P}}(\lambda - \gamma)'\big|_{\lambda=0}}{\mathcal{M}_{\mathcal{P}}(-\gamma)}$$

slide 11

Theorem 1 in BM(2006): Variance spread

For $m(R)$ satisfying $m(R) = 1 - \mathcal{A}_1 R + (1/2)\mathcal{A}_2 R^2 + O(R^3)$ where

$$\mathcal{A}_1 = -\,\partial m/\partial R\big|_{R=0} \text{ and } \mathcal{A}_2 = \partial^2 m/\partial R^2\big|_{R=0},$$

$$\frac{\sigma_{\mathcal{Q}}^2 - \sigma_{\mathcal{P}}^2}{\sigma_{\mathcal{P}}^2} \approx -\mathcal{A}_1 (\sigma_{\mathcal{P}}^2)^{1/2} \times \theta_{\mathcal{P}} + \frac{\mathcal{A}_2 \sigma_{\mathcal{P}}^2}{2} \times \left(\kappa_{\mathcal{P}} - 1 - \frac{2\mathcal{A}_1^2}{\mathcal{A}_2} \right).$$

- Suppose that $m(R) = \int_0^\infty e^{-zR} \nu(\mathrm{d}z)$ for some measure ν on $\mathcal{R}^+$.
- This $m(R)$ encompasses HARA marginal utility class and loss-aversion utilities.
- For example $z \sim G(\gamma, 1/b)$, then $\mathrm{E}[z] = \gamma b$, $\mathrm{E}[z^2] = \gamma(\gamma+1)b^2$ m.g.f. of z is $\mathcal{M}(\lambda) = (1 - b\lambda)^{-\gamma}$.
- Then we have $m(R) = \mathcal{M}(-R) = (1 + bR)^{-\gamma}$ which represents HARA marginal utility.

Quotes competition in limit order markets

Wataru Ohta

Graduate School of Economics
and Center for Mathematical Modeling and Data Science
Osaka University

Outline

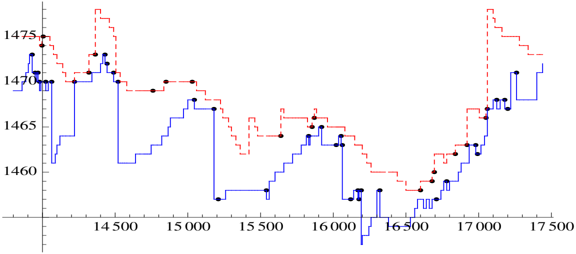
1. Purpose

- Order submission strategies in limit order markets
  - no designated market makers
  - matching orders submitted by public traders
- How transactions take place in limit order markets
  - Is there a systematic pattern to the ask quote and bid quote?
  - What is the optimal order submission strategy under a certain market condition?
    - a limit order undercutting the best quote by one tick
    - a limit order undercutting the best quote by more than one tick
    - a limit order at the best quote
    - a limit order submitted behind the best quote
  - What affects the size of holes?
  - How does the tick size affect quote competition of limit orders?

2-1. Empirical studies

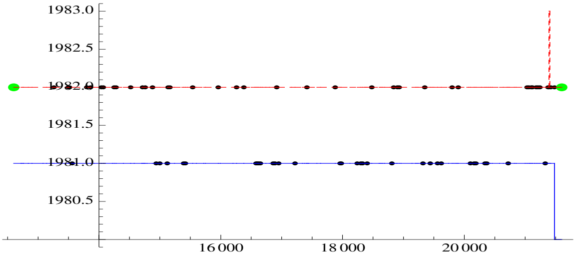
- Higher volatility can relate with lower liquidity.
- Egginton, Van Ness, Van Ness (2016)
  - Quote stuffing relates with decreased liquidity and increased short-term volatility.
- Hasbrouck (2018)
  - Higher short-term volatility relates with recurrent cycles of undercutting similar to the Edgeworth cycles.

2-2. June 24, 2002, Code 7269 in Tokyo



- Horizontal axis denotes time in seconds from the opening of the day, and vertical axis denotes price.
- Red line is the ask, blue line is the bid, black dots indicate transactions.

2-3. June 4, 2002, Code 8273 in Tokyo

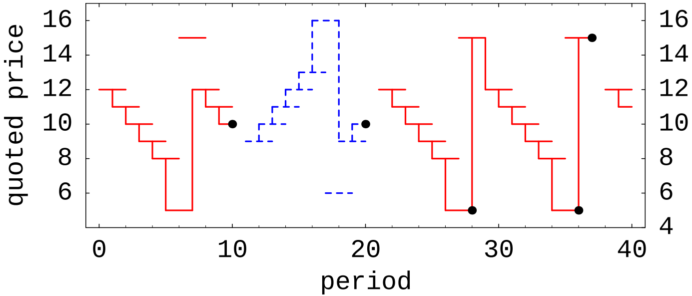


- Traders queue at the same quote.
- There is no quote competition.
- Tick size is the binding constraint in choosing quotes of limit orders.

3-1. Theoretical studies

- There can be too many states of the limit order book.
- Foucault (1999, Journal of Financial Markets)
  - assumes that limit orders are automatically canceled in one period after their submission in a discrete time setting.
  - This paper assumes that limit orders expire in two periods after their submission.
- Foucault, Kadan, and Kandel (2005, Review of Financial Studies)
  - Sellers and buyers arrive at the market alternately.
  - Traders cannot submit limit orders at or behind the best quotes.
  - This paper assume sellers and buyers arrive at the market randomly and traders can choose any price under the tick size restriction.
- Parlour (1998, Review of Financial Studies)
  - assumes that the bid-ask spread is one tick
- Goettler, Parlour, and Rajan (2005, Journal of Finance; 2009, Journal of Financial Economics)
  - numerical solutions
- Rosu (2009, Review of Financial Studies; 2009, Mimeo)
  - continuous time setting with zero tick size

3-2. Theoretical results of this paper



- An example of quote dynamics on the equilibrium path when limit orders expire in two periods
- Red lines denote ask quotes, blue lines denote bid quotes, dots indicate transactions.
- Cyclic quote dynamics
- There is possibility of queuing under the large tick size.

4-1. Model

- Discrete and infinite periods
- In each period, one potential trader arrives at the exchange and submits an order.
- The probability of a trader arriving is $\alpha \in (0, 1]$, and the probability of a seller conditional on a trader arriving is $\beta \in (0, 1)$.
  - α and β are exogenous.
  - α is the trader arrival rate.
  - β represents the order imbalance.
  - A trader is a seller with the probability of $\pi_s = \alpha\beta$, and a buyer with the probability of $\pi_b = \alpha(1 - \beta)$.
- No asymmetric information

4-2. Model

- A trader can submit an order with a volume of one share only when he arrives at the market.
- The order is either a limit order or a market order.
  - A limit order: price contingent
    - chooses the limit price under the restriction of the tick size k.
    - cannot cancel or modify his order once he submits it.
    - a limit order automatically expires in two periods after their submission.
  - A market order: not price contingent
    - immediately executed at the best available price on the market

4-3. Model

- A seller holds one share of the asset and evaluates it as $v_L \geq 0$.
  - A payoff for a seller is $P - v_L$ if he sells a share at a price P.
- A buyer holds no share, and evaluates a share as v_H $(v_H > v_L)$.
  - The payoff for a buyer is $v_H - P$ if he buys a share at a price P.
- A trader receives zero payoff if he does not trade.
- The discount rate is assumed to be zero.
- Traders maximize expected payoffs.
- The market is transparent: traders observe the book before submitting orders.

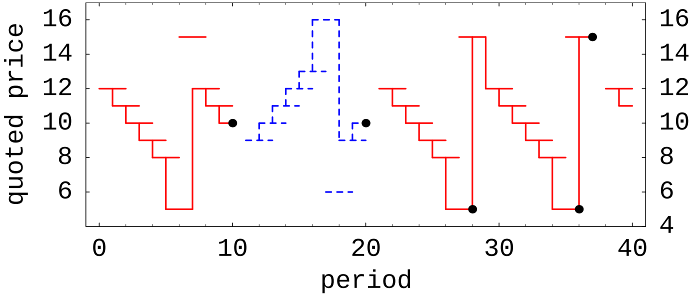
4-4. Trading rule

- The price-time precedence rule
  - Among limit buy (sell) orders, the highest (lowest)-priced orders have priority.
  - Among limit orders at the same price, priority is given to orders submitted earlier.
- The discriminatory pricing rule
  - a transaction price is the price of the matched limit order.
- The order choice depends on the state of the book and the future order flow.
  - Trade-off between price and execution probability in submitting an order.
  - a market order: executed immediately at a low price
  - a limit order: execution is uncertain though the price is favorable once the order is executed.

4-5. Pure-strategy Markov Perfect Equilibrium

- Markov strategy which depends only on the type of trader and on the book.
- Pure strategy.

5-1. An example of equilibrium quote dynamics



- $v_L = 0$, $v_H = 21$, $k = 1$, the trader arrival rate $\alpha = 1$, and the order imbalance $\beta = 1/2$.
- At first, one-tick quote-cutting
- At a certain level, the best quote jumps more than one tick.
- A limit sell order behind the best ask is submitted.

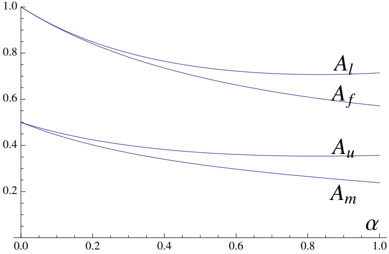
5-2. Edgeworth cycle

- An equilibrium with cyclic quote dynamics exists when the tick size k is small enough.
- Edgeworth cycle
    - Edgeworth (1925)
    - Maskin and Tirole (1988, Econometrica)
    - Noel (2007, Journal of Industrial Economics) in Canadian retail gasoline markets

5-3. Edgeworth cycle

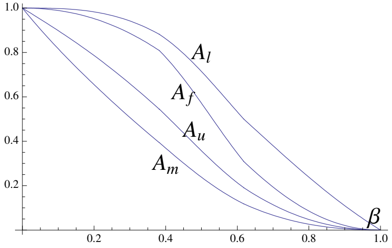
- A seller who arrives at an empty book submits a limit sell order at a relatively high ask, and allows the next seller to undercut his ask.
  - Hindering quote-cutting by the aggressive ask is costly.
- The next seller undercuts the best ask by only one tick.
  - expects further quote-cutting.
  - one-tick quote-cutting minimizes the cost in price to get higher priority against the limit orders in the book.
- When the best ask reaches a certain level after one-tick quote-cutting, a seller submits a very aggressive ask (quote jump)
  - Jumping quote is reasonable because hindering further quote cutting raises the execution probability.
- The next seller submits a limit sell order far behind the best ask.
  - gives up further quote-cutting because the price has already become too low.
  - The execution probability of a limit order behind the best ask is small.
  - The high price compensates for the large loss in the execution probability.

5-4. The trader arrival rate α and the critical asks



- Horizontal axis is the trader arrival rate α and vertical axis is price.
- A_f is the first ask submitted to an empty book.
- A_u is the end of the range of one-tick quote-cutting.
- A_m is the most aggressive ask.
- A_l is the least aggressive ask.

5-5. The order imbalance β and the critical asks



- When sellers outnumber buyers (when β is greater),
  - Sellers submit more aggressive limit sell orders
    - narrow the size of holes on the ask side.
    - shorten the steps of one-tick quote-cutting
  - Holes created by the jumping quote on the ask side are smaller.
  - Cyclic quote dynamics on the ask side are more frequently observed

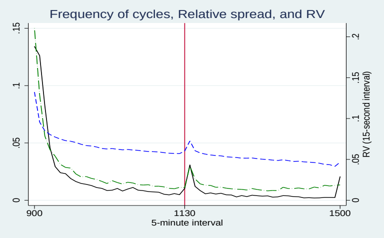
5-6. Multiple equilibria: Queuing

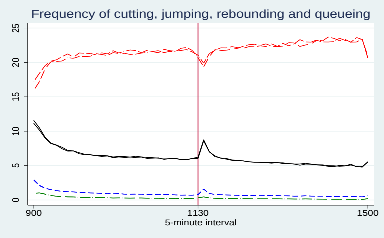
- When the tick size is large, traders queue at the same quote.
  - The cost of quote-cutting to obtain price priority is large.
  - Queuing does not occur when the tick size is small.
- The higher the arrival rate of sellers (buyers) is, the more likely sellers (buyers) queue at the same quote.
  - The benefit of quote-cutting is small when the number of traders on the opposite side of the market is small.
  - Order imbalance caused by liquidity needs increases the depth.

6-1. Empirical study in the Tokyo Stock Exchange

- Data: Nikkei Needs tick data
- Period: from August 2014 to December 2015
- Stocks in the TOPIX100
  - Tick size reduction after July 22, 2014
  - Price range from 500 yen to 1,000 yen (the tick size is 0.1 yen) and from 5000 yen and 10,000 yen (the tick size is 1 yen)
  - 26 stocks with more than 200 trading days satisfying the above price range condition
- Divide the trading period into 5-minute intervals
  - Proportion of the number of specific types of order submission to the total number of quote changes for each interval
  - Cycle: two consecutive quote cutting followed by quote rebounding
  - OIB: (number of market and limit buy order - number of market and limit sell order)/(number of market and limit buy order + number of market and limit sell order).
  - Winsorize variables at the 2nd and 98th percentiles

6-2. Frequency of cycles





- Black: proportion of cycles to the total number of quote changes
- Blue: relative bid-ask spread
- Green: realized variance

- Black: proportion of quote cutting
- Blue: proportion of quote jumping
- Green: proportion of quote rebounding
- Red: proportion of queuing

6-3. Quote cycles in the 5-min interval from 9:05 to 9:09

- Total: 8268 stock*days
- 4913 stock*days have no cycle, 3355 stock*days have cycles

	N of cycles	Tick spread	Relative spread	Realized variance	Realized Kurtosis
	Intervals without cycles (4913 stock*days)				
Mean	0	4.20	0.06	0.09	4.21
SD	0	1.98	0.03	0.08	1.87
Min	0	1.41	0.19	0.006	1.86
Max	0	17.46	0.26	0.82	15.63
	Intervals with cycles (3355 stock*days)				
Mean	3.87	5.81	0.08	0.18	4.13
SD	6.69	2.79	0.04	0.13	1.86
Min	1	1.46	0.02	0.006	1.86
Max	138	17.46	0.26	0.82	15.63

6-4. Bid-ask spread and moments of quote-midpoint returns

- Bid-ask spread and moments of quote-midpoint returns classified by quintiles of OIB for the 5-minute intervals from 9:05 to 9:09

	Tick spread	Relative spread	Realized variance	Realized skewness	Realized kurtosis
Quintile 1 (low OIB)	4.90	0.069	0.129	-1.031	4.266
Quintile 2	4.81	0.068	0.124	-0.449	4.141
Quintile 3	4.77	0.067	0.127	0.024	4.032
Quintile 4	4.90	0.069	0.129	0.498	4.175
Quintile 5 (high OIB)	4.89	0.068	0.135	1.039	4.286
Total	4.85	0.068	0.129	0.019	4.179
F-value	1.00	0.66	2.29	1091.1	4.87
p-value	0.41	0.62	0.06	0.00	0.00

- F-value and p-value are for F-test of equality of quintile means

6-5. Order submission

- Proportion (percent) of specific sell orders to the total number of quote changes classified by quintiles of OIB for the 5-minute intervals from 9:05 to 9:09

	Cycle on the ask side	One-tick cutting	Quote jumping	Quote rebounding	Queuing
Quintile 1 (low OIB)	0.12	11.06	2.49	1.54	23.41
Quintile 2	0.07	9.19	2.19	1.17	19.75
Quintile 3	0.04	7.96	2.04	0.96	17.18
Quintile 4	0.04	6.89	1.97	0.85	14.70
Quintile 5 (high OIB)	0.02	5.36	1.71	0.67	11.17
Total	0.06	8.09	2.07	1.04	17.23
F-value	90.82	831.6	89.21	158.6	2645
p-value	0.00	0.00	0.00	0.00	0.00

- F-value and p-value are for F-test of equality of quintile means

7. Summary

- In limit order markets, cyclic quotes dynamics can be observed, and holes emerge in the book during cycles.
- When the trader arrival rate is high, cyclic quote dynamics are more frequently observed and the holes are larger.
- When sellers outnumber buyers, cycles on the ask side are more frequently observed and the holes on the ask side are smaller. However, sellers queue at the same ask if the tick size is relatively large.

References

Egginton, J. F., B. F. Van Ness, and R. A. Van Ness, 2016, Quote Stuffing, Financial Management 45(3): 583-608.

Foucault, Thierry, 1999, Order flow composition and trading costs in a dynamic limit order market, Journal of Financial Markets 2, 99-134.

Foucault, Thierry, Ohad Kadan, and Eugene Kandel, 2005, Limit order book as a market for liquidity, Review of Financial Studies 18, 1171-1217.

Goettler, Ronald L., Christine A. Parlour, and Uday Rajan, 2005, Equilibrium in a dynamic limit order market, Journal of Finance 60, 2149-92.

Goettler, Ronald L., Christine A. Parlour, and Uday Rajan, 2009, Informed traders and limit order markets, Journal of Financial Economics 93, 67-87.

Hasbrouck, J., 2018, High-Frequency Quoting: Short-Term Volatility in Bids and Offers, Journal of Financial and Quantitative Analysis 53(2): 613-641.

Maskin, Eric, and Jean Tirole, 1988, A theory of dynamic oligopoly, ii: Price competition, kinked demand curves, and Edgeworth cycles, Econometrica 56, 571-99.

Noel, Michael D., 2007, Edgeworth price cycles: Evidence from the Toronto retail gasoline market, Journal of Industrial Economics 55, 69-92.

Parlour, Christine A., 1998, Price dynamics in limit order markets, Review of Financial Studies 11, 789-816.

Rosu, Ioanid, 2008, Liquidity and information in order driven markets, Mimeo.

Rosu, Ioanid, 2009, A dynamic model of the limit order book, Review of Financial Studies 22, 4601-41.

A Two-Sample Alternative to Using Instruments in Regressions with Omitted Variables

Masayuki Hirukawa *Ryukoku University*
Irina Murtazashvili *Drexel University*
Artem Prokhorov *University of Sydney*

7 August 2019

Fukushima University

1 Motivation

- Relevant variables that are omitted lead to a challenging problem.

- Examples of the missing regressor:

  1. An **ability measure** in Mincer's (1974) wage regression.

     – Card (1995) argues that the estimation result suffers from the "ability bias" unless the regression includes a variable representing ability as a regressor.

     – A test score is typically considered as a variable representing ability.

     – Test scores are unavailable in the Current Population Survey (CPS) and the Panel Study of Income Dynamics (PSID).

2. **Work experience** in gender wage gap (e.g., Zabalza and Arrufat, 1985; Black, Trainor, and Spencer, 1999).
   - No actual work experience is included in the General Household Survey (GHS) and CPS.
   - Actual work experience is available in the National Longitudinal Survey (NLS) and PSID.

- In the absence of a proxy, the most common solution is to look for an instrument, but...
  1. Weak instruments?
  2. Finite-sample properties of IV estimates?

1.1 Our Proposal

- Why not turn to a two-sample approach using a proxy?

- Our two-sample estimation procedure is to run ordinary least squares (OLS) after replacing the missing regressor with a nonparametric estimate of its conditional mean.

  – This estimator does not rely on the combined sample constructed via the nearest-neighbor matching (NNM).

  – Hirukawa and Prokhorov (2018) have already proposed a NNM-based two-sample estimation (to be discussed shortly).

1.2 Contributions

1. Nonparametric imputation of the missing regressor:

    - Imposing a parametric form on the conditional mean of the missing regressor (e.g., Fang, Keane and Silverman, 2008; Flavin and Nakagawa, 2008) has at least two problems:

    (a) Bias due to misspecification of the conditional mean.

    (b) Lack of adjustments in standard errors under the assumption that imputed predictors are free of estimation errors.

2. Nonparametrically generated regressors:

    - Our approach is analogous to Pagan (1984).

    - The regressor is not imputed within the same sample in our estimation procedure.

Plan of Talk

1. Motivation
2. Setup
3. Strategy 1: IV Estimation
4. Strategy 2: Two-Sample Estimation
5. Finite-Sample Performance
6. Application to Estimation of Return to Schooling
7. Concluding Remarks

2 Setup

True Model.

$$Y = \beta_0 + X_1'\beta_1 + X_2'\beta_2 + X_{3I}'\beta_3 + u,$$

where $X_1 \in \mathbb{R}^{d_1}$, $X_2 \in \mathbb{R}^{d_2}$ and $X_{3I} \in \mathbb{R}^{d_{3I}}$.

- Either β_1 or β_3 is the parameter of interest.

Standard Situation. A single complete sample (Y, X_1, X_2, X_{3I}) is available.

- If $E(u | X_1, X_2, X_{3I}) = 0$, then OLS for this regression is consistent.

Our Case. X_2 is missing (in the primary sample).

Two Datasets.

	Y	X_1	X_2	X_3
$\mathcal{S}_1$	✓	✓		✓
$\mathcal{S}_2$			✓	✓

Note: $X_3 = (X_{3I}, X_{3E})$

Estimation Strategies.

1. **IV estimation** with X_2 left omitted, where only $\mathcal{S}_1$ is used.

2. **Two-sample estimation** with (a proxy of) X_2 imputed from a different data source, where both $\mathcal{S}_1$ and $\mathcal{S}_2$ are used.

3 Strategy 1: IV Estimation

- The true regression

$$Y = \beta_0 + X_1'\beta_1 + X_2'\beta_2 + X_{3I}'\beta_3 + u$$

can be rewritten as a "short" regression

$$Y = \beta_0 + X_1'\beta_1 + X_{3I}'\beta_3 + \left(u + X_2'\beta_2\right) = X_S'\beta_S + v,$$

where $X_S := \left(1, X_1', X_{3I}'\right)'$.

  - X_2 is treated as the vector of omitted variables.
  - (X_1, X_{3I}) and X_2 may be correlated.

- It is still possible to estimate β_S (and thus β_1) consistently if the vector of instruments $Z \in \mathbb{R}^{d_Z}$ is available.

- The estimation is based on the moment restriction

$$E(Z_S v) := E\left\{\begin{bmatrix} 1 \\ Z \end{bmatrix}\left(u + X_2'\beta_2\right)\right\} = 0_{(d_Z+1)\times 1}.$$

  - Even if $d_Z \geq d_1 + d_{3I}$ is satisfied, consistency of this estimation strategy is built on somewhat restrictive assumptions such as

$$Z \perp\!\!\!\perp X_2 \text{ and } E(X_2)'\beta_2 = 0.$$

  - A simple IV estimator in the just-identified case $(d_Z = d_1 + d_{3I})$ is

$$\hat{\beta}_{IV,S} := \hat{Q}_{ZX}^{-1}\hat{R}_{ZY} := \left(\frac{1}{n}\sum_{i=1}^{n} Z_{S,i}X_{S,i}'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} Z_{S,i}Y_i\right).$$

4 Strategy 2: Two-Sample Estimation

- What if instruments may be unavailable or there may be other reasons for avoiding Strategy 1 (e.g., weak instruments)?

- Suppose that econometricians have access to the auxiliary data set

$$\mathcal{S}_2 = \mathcal{S}_{2m} = \left\{ \left(X_{2j}, X_{3Ij}, X_{3Ej} \right) \right\}_{j=1}^{m}.$$

- $\mathcal{S}_1$ and $\mathcal{S}_2$ contain common variables $X_3 = (X_{3I}, X_{3E})$.

  - X_{3I} and X_{3E} are included in and excluded from the regression

$$Y = \beta_0 + X_1'\beta_1 + X_2'\beta_2 + X_{3I}'\beta_3 + u.$$

4.1 Additional Notations

- The vector of matching variables X_3 consists of continuous ("C") and discrete ("D") variables so that

$$d_3 := \dim(X_3) = \dim(X_{3C}) + \dim(X_{3D}) =: d_{3C} + d_{3D}.$$

  - Assume that $d_{3C} > 0$, i.e., $\mathcal{S}_1$ and $\mathcal{S}_2$ have at least one continuous variable in common.

- Let $\mathbb{X}_3 := \mathbb{X}_{3C} \times \mathbb{X}_{3D}$, where $\mathbb{X}_{3C} := \text{supp}(X_{3C})$ and $\mathbb{X}_{3D} := \text{supp}(X_{3D})$.

- Define

$$g(X_3) = \begin{bmatrix} g_1(X_3) \\ g_2(X_3) \end{bmatrix} := \begin{bmatrix} E(X_1|X_3) \\ E(X_2|X_3) \end{bmatrix} \text{ and}$$
$$\eta = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} := \begin{bmatrix} X_1 - g_1(X_3) \\ X_2 - g_2(X_3) \end{bmatrix}.$$

4.2 The Estimator

- Reformulate the true "long" regression

$$Y = \beta_0 + X_1'\beta_1 + X_2'\beta_2 + X_{3I}'\beta_3 + u$$

as

$$Y = X'\beta + \epsilon,$$

where

$$X := \left(1, X_1', g_2\left(X_3\right)', X_{3I}'\right)' \text{ and } \epsilon := u + \eta_2'\beta_2.$$

- The estimator fundamentally takes the form of an OLS-type regression of Y on X.
  - Actually $g_2\left(\cdot\right)$ is unknown and thus must be estimated nonparametrically.

- Given $\hat{g}_2(\cdot)$, some consistent nonparametric estimator of $g_2(\cdot)$, the entire estimation procedure takes the following two steps:

  1. Regard $\left\{\left(X_{2j}, X_{3j}\right)\right\}_{j=1}^{m}$ in $\mathcal{S}_2$ and $\{X_{3i}\}_{i=1}^{n}$ in $\mathcal{S}_1$ as m data and n design points, respectively, and obtain n nonparametric estimates $\{\hat{g}_2(X_{3i})\}_{i=1}^{n}$.

  2. Run OLS for the regression of Y_i on $\hat{X}_i := \left(1, X'_{1i}, \hat{g}_2(X_{3i})', X'_{3Ii}\right)'$.

- The estimator of β in the final form is

$$\hat{\beta}_{PILS} := \hat{Q}_{\hat{X}\hat{X}}^{-1}\hat{R}_{\hat{X}Y} := \left(\frac{1}{n}\sum_{i=1}^{n}\hat{X}_i\hat{X}_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\hat{X}_iY_i.$$

  – Because this estimator is the OLS with $\hat{g}_2(X_{3i})$ imputed in place of the missing regressor X_{2i}, we call it the **plug-in least squares (PILS)** estimator hereinafter.

4.3 Kernel Regression Smoother for Mixed Data

- From among many nonparametric methods, we adopt the kernel regression smoother for mixed continuous and categorical data by Racine and Li (2004) to estimate $g_2(\cdot)$.

- The kernel is $\mathcal{W}(t_3; x_3, \mathbf{h}, \boldsymbol{\lambda}) := \mathcal{K}(t_{3C}; x_{3C}, \mathbf{h}) \mathcal{L}(t_{3D}; x_{3D}, \boldsymbol{\lambda})$.

$\mathcal{K}$:	product kernel for the continuous component of X_3.
$\mathcal{L}$:	product kernel for the discrete component of X_3.
t_3:	data point.
x_3:	design point.
$\mathbf{h}, \boldsymbol{\lambda}$:	bandwidths.

- The nonparametric regression estimator of $g_2(\cdot)$ is defined as

$$\hat{g}_2(X_{3i}) := \frac{\sum_{j=1}^{m} X_{2j} \mathcal{W}\left(X_{3j}; X_{3i}, \mathbf{h}, \boldsymbol{\lambda}\right)}{\sum_{j=1}^{m} \mathcal{W}\left(X_{3j}; X_{3i}, \mathbf{h}, \boldsymbol{\lambda}\right)}, \ i = 1, \ldots, n.$$

- Which univariate kernel to choose for the continuous component?
  - A standard symmetric kernel such as the Epanechnikov kernel; or
  - The **beta kernel** by Chen (1999), which takes the form of

$$K_{B(x,b)}(t) = \frac{t^{x/b}(1-t)^{(1-x)/b}}{B\{x/b+1,(1-x)/b+1\}}\mathbf{1}\{0 \le t \le 1\}$$

for the design point $x \in [0,1]$ and the smoothing parameter b.

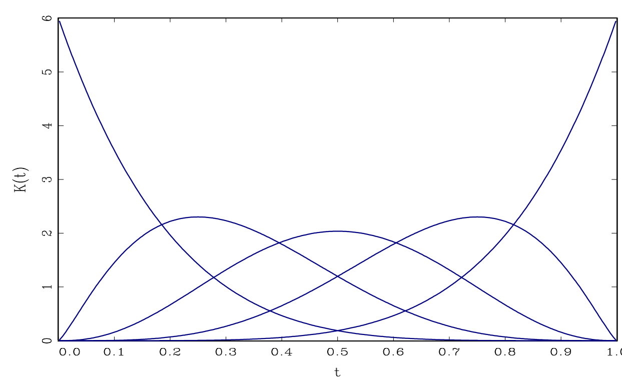


Figure: Shapes of the Beta Kernel for $b = 0.2$ and $x = 0.00, 0.25, 0.50, 0.75, 1.00$

4.4 Regularity Conditions

Assumption 1. The two random samples $(\mathcal{S}_1, \mathcal{S}_2) = (\mathcal{S}_{1n}, \mathcal{S}_{2m})$ are drawn independently from the joint distribution of (Y, X_1, X_2, X_3) with finite fourth-order moments.

Assumption 2. X_{3C} is continuously distributed with a convex and compact support $\mathbb{X}_{3C}$, and its density is bounded and bounded away from zero on $\mathbb{X}_{3C}$.

Assumption 3. (i) $E\left(u | X_1, X_3\right) = 0$ and $\sigma_u^2\left(X_1, X_3\right) := E\left(u^2 \middle| X_1, X_3\right) \in (0, \infty)$. **(ii)** $E\left(\eta_1 \eta_2'\right) = 0$. **(iii)** $g_2\left(\cdot\right)$ is non-constant on $\mathbb{X}_{3C}$ if X_{3E} contains at least one continuous variable, and $g_2\left(\cdot\right)$ is strictly nonlinear on $\mathbb{X}_{3C}$ otherwise.

Assumption 4. **(i)** Let $f(\cdot)$ be the marginal pdf of X_{3C}. Then, the second-order derivatives of $f(\cdot)$ and $f(\cdot)\,g_2(\cdot)$ with respect to X_{3C} are continuous and bounded uniformly on $\mathbb{X}_{3C}$. **(ii)** There exist some constants $\gamma \in (0, \infty)$ and $C \in [1, \infty)$ so that $\sup_{x_3 \in \mathbb{X}_3} E\left(|Y|^{2+\gamma} \middle| X_3 = x_3 \right) \leq C$.

Assumption 5. The univariate continuous kernel is either (a) a symmetric and bounded pdf that satisfies the first-order Lipschitz condition or (b) the beta kernel.

Assumption 6. Sequences of the smoothing parameters $h_p \left(= h_p(m) > 0\right)$, $b_p \left(= b_p(m) > 0\right)$ and $\lambda_q \left(= \lambda_q(m) \in (0,1)\right)$ and the boundary parameter $\eta_p \left(= \eta_p(m) > 0\right)$ for $p = 1, \ldots, d_{3C}$ and $q = 1, \ldots, d_{3D}$ satisfy one of the following conditions as $m \to \infty$: (a) for a symmetric kernel, $h_p, \lambda_q \to 0$ and $\log m / \left(m \prod_{p=1}^{d_{3C}} h_p \right) \to 0$; and (b) for the beta kernel, $b_p, \lambda_q, \eta_p \to 0$, $b_p/\eta_p \to 0$ and $\log m / \left(m \sqrt{\prod_{p=1}^{d_{3C}} b_p \eta_p} \right) \to 0$.

4.5 General Asymptotic Results of PILS

Theorem 1. *If Assumptions 1-6 hold, then* $\hat{\beta}_{PILS} \xrightarrow{p} \beta$ *as* $n, m \to \infty$.

Theorem 2. *If Assumptions 1-6 hold, then* $\sqrt{n}\left(\hat{\beta}_{PILS} - \beta - B_{g_2}\right) \xrightarrow{d} N\left(0_{(d+1)\times 1}, V_X\right) := N\left(0_{(d+1)\times 1}, Q_{XX}^{-1}\Omega_{\psi\psi}Q_{XX}^{-1}\right)$ *as* $n, m \to \infty$, *where*

$$B_{g_2} = \hat{Q}_{\hat{X}\hat{X}}^{-1}\frac{1}{n}\sum_{i=1}^{n}\left[\hat{X}_i\left\{g_2\left(X_{3i}\right) - \hat{g}_2\left(X_{3i}\right)\right\}' - E(\hat{X})\left\{\hat{g}_2\left(X_{3i}\right) - E\left(\hat{g}_2\left(X_{3i}\right)\middle| X_{3i}\right)\right\}'\right]\beta_2,$$

$$Q_{XX} = E\left(XX'\right),\ \hat{Q}_{\hat{X}\hat{X}} = \frac{1}{n}\sum_{i=1}^{n}\hat{X}_i\hat{X}_i',$$

$$\Omega_{\psi\psi} = \lim_{n,m\to\infty}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_i\psi_i', \text{ and}$$

$$\psi_i = \hat{X}_i\epsilon_i - E(\hat{X})\left[\hat{g}_2\left(X_{3i}\right) - E\left\{\hat{g}_2\left(X_{3i}\right)\middle| X_{3i}\right\}\right]'\beta_2.$$

In addition, the asymptotic variance V_X can be consistently estimated by

$$\hat{V}_X := \hat{Q}_{\hat{X}\hat{X}}^{-1} \hat{\Omega}_{\hat{\psi}\hat{\psi}} \hat{Q}_{\hat{X}\hat{X}}^{-1},$$

where

$$\begin{aligned}
\hat{\Omega}_{\hat{\psi}\hat{\psi}} &= \frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_i \hat{\psi}_i', \\
\hat{\psi}_i &= \hat{X}_i \hat{\epsilon}_i + \left(\hat{\delta}_i - \frac{1}{n} \sum_{i=1}^{n} \hat{\delta}_i \right), \\
\hat{\epsilon}_i &= Y_i - \hat{X}_i' \hat{\beta}_{PILS}, \\
\hat{\delta}_i &= - \left(\frac{1}{n} \sum_{k=1}^{n} \hat{X}_k \right) \hat{g}_2 \left(X_{3i} \right)' \hat{\beta}_{PILS,2},
\end{aligned}$$

and $\hat{\beta}_{PILS,2}$ is the PILS estimate of β_2.

Remark. PILS is a partial mean estimator of Newey (1994), and thus $\hat{\Omega}_{\hat{\psi}\hat{\psi}}$ can be obtained straightforwardly by his approach.

4.6 $\sqrt{n}$-Asymptotic Normality vs Curse of Dimensionality

Assumption 6'. Sequences of the smoothing and boundary parameters satisfy one of the followings as $m \to \infty$: (a) for a symmetric kernel, there is a bandwidth $h\,(= h\,(m) > 0)$ so that $h_1, \ldots, h_{d_{3C}} \propto h$, $\lambda_1, \ldots, \lambda_{d_{3D}} \propto h^2$, $h \to 0$, and $\log m / \left(m h^{d_{3C}}\right) \to 0$; and (b) for the beta kernel, there are a smoothing parameter $b\,(= b\,(m) > 0)$ and a boundary parameter $\eta\,(= \eta\,(m) > 0)$ so that $b_1, \ldots, b_{d_{3C}} \propto b$, $\lambda_1, \ldots, \lambda_{d_{3D}} \propto b$, $\eta_1, \ldots, \eta_{d_{3C}} \propto \eta$, $b, \eta \to 0$, $b/\eta \to 0$, and $\log m / \left\{m\,(b\eta)^{d_{3C}/2}\right\} \to 0$.

Corollary 1. *Let $h \propto (\log m/m)^{\alpha}$ and $b \propto (\log m/m)^{2\alpha}$ for some constant $\alpha > 0$. Also suppose that one of the following divergence patterns in (n, m) is true: (i) $n/m \to \kappa \in (0, \infty)$; (ii) $n/m \to 0$; or (iii) $n/m \to \infty$ and $n/m^{4\alpha} \to 0$. Then, under Assumptions 1-5 and 6',$\sqrt{n}\left(\hat{\beta}_{PILS} - \beta\right) \xrightarrow{d} N\left(0_{(d+1)\times 1}, V_X\right)$ holds for $d_{3C} \leq 3$ as $n, m \to \infty$.*

4.7 A Comparison with the MSII Estimator

NNM.

- The **matched-sample indirect inference (MSII)** estimation by Hirukawa and Prokhorov (2018) starts from obtaining a combined sample by means of NNM on the observed variables.

- For each observation of X_3 in $\mathcal{S}_1$, **K-NNM** picks out K closest matches of X_2 from $\mathcal{S}_2$ through finding first to Kth closest matches of X_3 in S_2 with respect to some distance function such as the Mahalanobis distance.

  – The resulting combined sample can be written as

$$\mathcal{S} = \left\{ \left(Y_i, X_{1i}, X_{2j_1(i)}, \ldots, X_{2j_K(i)}, X_{3Ii}, X_{3Ei} \right) \right\}_{i=1}^{n}.$$

Matched-Sample OLS (MSOLS).

- The OLS estimation (**MSOLS**) for the regression of Y_i on

  $$X_{i,j(i)} := \left(1, X'_{1i}, X'_{2j(i)}, X'_{3Ii}\right)',$$

  where

  $$X_{2j(i)} := \frac{1}{K}\sum_{k=1}^{K} X_{2j_k(i)},$$

  generates a non-vanishing, classical measurement error bias.

  – The bias is attributed to using $X_{2j(i)}$ as a proxy for X_{2i}.

  – The source of attenuation (= bias toward zero) is $\mathbf{\Sigma}_2 = E\left(\eta_2\eta'_2\right)$.

Bias-Corrected Estimation.

- As is well known in the literature on errors-in-variables models, the bias cannot be corrected in general without imposing additional identification conditions.
  - It can be corrected analytically with no such extra conditions because $\mathcal{S}_2$ serves as repeated measurements.

- The MSII is a bias-corrected estimator defined as

$$\hat{\beta}_{II} := \left(\frac{1}{n} \sum_{i=1}^{n} X_{i,j(i)} X'_{i,j(i)} - \frac{1}{K} \hat{\Sigma} \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} X_{i,j(i)} Y_i,$$

where

$$\hat{\Sigma} := \text{diag}\left\{ 0_{(d_1+1)\times(d_1+1)}, \hat{\Sigma}_2, 0_{d_{3I}\times d_{3I}} \right\},$$

and $\hat{\Sigma}_2$ is the **difference-based variance estimator** that can be obtained by reordering $\mathcal{S}_2$ with respect to X_3.

Two-Step Bias Correction for $\sqrt{n}$-Asymptotic Normality.

- MSII can attain $\sqrt{n}$-asymptotic normality when the number of matching variables is only one.

- Hirukawa and Prokhorov (2018) propose a two-step estimator called the **fully-modified MSII (MSII-FM)** estimator.
  - In its second step, MSII-FM eliminates the second-order bias due to the so-called **matching discrepancy** (Abadie and Imbens, 2006) asymptotically by means of a polynomial approximation.
  - The estimator can achieve parametric convergence when the number of matching variables is four or less.

Novel Features of PILS.

1. PILS does not rely on NNM.
   - It requires no bias correction that is a key ingredient in MSII(-FM).
   - The asymptotic variance of PILS does not much exceed that of OLS.
2. The asymptotic analysis for PILS explicitly incorporates discrete matching variables.
3. The role of excluded continuous matching variables for identification is clarified.
   - Hirukawa and Prokhorov (2018) maintain the assumption that all common variables enter the regression and are used for both estimation and matching.

5 Finite-Sample Performance

Regressions.

$$[L]: Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u, \text{ and}$$
$$[S]: Y = \beta_0 + \beta_1 X_1 + (u + \beta_2 X_2) := \beta_0 + \beta_1 X_1 + v.$$

where

$$\beta_0 = \beta_1 = \beta_2 = 1.$$

Data Generation.

1. $Z, X_{3E_C} \overset{iid}{\sim} U[-2, 2]$.

2. $X_{3E_D} \overset{iid}{\sim} Bernoulli\,(1/2) - 1/2 = \begin{cases} -1/2 & \text{wp } 1/2 \\ 1/2 & \text{wp } 1/2 \end{cases}.$

3. $X_1 = \left(1 + X_{3E_C}^2 + X_{3E_D}\right) + (\pi Z + \xi)$, where $\xi \overset{iid}{\sim} N(0,1)$ and $\pi \in \{1.00, 0.70, 0.15\}$.

$$corr(X_1, Z) = \sqrt{\frac{240\pi^2}{240\pi^2 + 481}} \approx \begin{cases} 0.58 & \text{for } \pi = 1.00 \\ 0.44 & \text{for } \pi = 0.70 \\ 0.11 & \text{for } \pi = 0.15 \end{cases}.$$

4. $X_2 = \left\{h\left(X_{3E_C}\right) + X_{3E_D}\right\} + \eta_2$, where $\eta_2 \overset{iid}{\sim} N(0,1)$ and

$$h(x) = \begin{cases} x & \text{[Model A]} \\ \{\exp(x) - \exp(-x)\} / \{\exp(x) + \exp(-x)\} & \text{[Model B]} \\ (3/4)\sqrt{x+2} - 1 & \text{[Model C]} \\ x^3/2 - x^2/2 - 2x - 2/3 & \text{[Model D]} \\ |x| - 1 & \text{[Model E]} \\ x + (5/\tau)\phi(x/\tau) - (5/2)\{\Phi(2/\tau) - 1/2\},\ \tau = 3/4 & \text{[Model F]} \end{cases}.$$

5. $Y = 1 + X_1 + X_2 + u$, where $u \overset{iid}{\sim} N(0,1)$.

Samples.

1. An unobserved complete sample:

$$\mathcal{S}^* = \{(Y_i, X_{1i}, X_{2i}, X_{3i}, Z_i)\}_{i=1}^{n}.$$

2. Two observable samples:

$$\mathcal{S}_1 = \{(Y_i, X_{1i}, X_{3i}, Z_i)\}_{i=1}^{n} \text{ and } \mathcal{S}_2 = \left\{\left(X_{2j}, X_{3j}\right)\right\}_{j=1}^{m}.$$

3. The matched sample via NNM with respect to X_3:

$$\mathcal{S} = \left\{\left(Y_i, X_{1i}, X_{2j_1(i)}, \ldots, X_{2j_K(i)}, X_{3i}, X_{3j_1(i)}, \ldots, X_{3j_K(i)}, Z_i\right)\right\}_{i=1}^{n},$$

where the number of matches $K \in \{1, 2, 4, 8, 16, 32, 64, 128\}$.

Remark. Sample sizes are $n = m \in \{500, 1000\}$, and the number of replications is 1000.

Estimators.

1. OLS*: Infeasible OLS estimator for $[L]$ using $\mathcal{S}^*$.
2. OLS-S: (Inconsistent) OLS estimator for $[S]$ using $\mathcal{S}_1$.
3. IV-S: IV estimator for $[S]$ using $\mathcal{S}_1$.
4. MSOLS: (Inconsistent) MSOLS estimator for $[L]$ using $\mathcal{S}$.
5. MSII: MSII estimator for $[L]$ using $\mathcal{S}$.
6. PILS-E: PILS estimator for $[L]$ using $\mathcal{S}_1$, $\mathcal{S}_2$ and the Epanechnikov kernel.
7. PILS-B: PILS estimator for $[L]$ using $\mathcal{S}_1$, $\mathcal{S}_2$ and the beta kernel.
8. PARA A parametric analogue to PILS for $[L]$ using $\mathcal{S}_1$ and $\mathcal{S}_2$ by Fang, Keane and Silverman (2008).

Bandwidth Selection.

$\hat{h} = \hat{\sigma}_X (\log m/m)^{0.3}$ for the Epanechnikov kernel.
$\hat{b} = \hat{\sigma}_U (\log m/m)^{0.6}$ for the beta kernel, where $U := (X + 2)/4 \in [0, 1]$.
$\hat{\lambda} = (\log m/m)^{0.6}$ for the discrete kernel.

A Few Words on PARA.

- PARA takes the following two steps:

  1. Run OLS for the regression of X_{2j} on $\left(1, X_{3E_C,j}, X_{3E_D,j}\right)$ using $\mathcal{S}_2$ and obtain n parametric predictors
  $$\left\{\hat{X}_{2i} = \hat{\alpha}_0 + \hat{\alpha}_1 X_{3E_C,i} + \hat{\alpha}_2 X_{3E_D,i}\right\}_{i=1}^{n}$$
  using the OLS estimates $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2)$.

  2. Run OLS for the regression of Y_i on $\left(1, X_{1i}, \hat{X}_{2i}\right)$.

- Under our Monte Carlo design, PARA is consistent if and only if $h(\cdot)$ is odd.

  – Consistency holds for
  $$h(x) = \begin{cases} x & \text{[Model A]} \\ \{\exp(x) - \exp(-x)\} / \{\exp(x) + \exp(-x)\} & \text{[Model B]} \end{cases}.$$

Model A: $h(x) = x$ ($n = m = 500$).

π	Estimator	OLS*	OLS-S	IV-S	MSOLS	MSII	PILS-E	PILS-B	PARA
1.00	$Mean$	0.9984	1.0611	0.9956	1.0227	0.9975	1.0069	1.0050	0.9989
	SD	0.0224	0.0425	0.0737	0.0373	0.0427	0.0347	0.0345	0.0318
	$RMSE$	0.0225	0.0744	0.0738	0.0436	0.0428	0.0354	0.0349	0.0318
	$\overline{SE}$	0.0223	0.0434	0.0738	0.0391	0.0395	0.0318	0.0318	0.0317
	CR	95%	–	95%	–	93%	93%	94%	95%
0.70	$Mean$	0.9985	1.0740	0.9934	1.0278	0.9975	1.0088	1.0065	0.9992
	SD	0.0245	0.0470	0.1063	0.0413	0.0476	0.0389	0.0386	0.0351
	$RMSE$	0.0246	0.0876	0.1065	0.0498	0.0477	0.0399	0.0392	0.0351
	$\overline{SE}$	0.0245	0.0479	0.1062	0.0429	0.0435	0.0350	0.0349	0.0348
	CR	95%	–	95%	–	92%	92%	92%	95%
0.15	$Mean$	0.9992	1.0920	0.8116	1.0353	0.9979	1.0118	1.0091	1.0000
	SD	0.0270	0.0529	2.7255	0.0468	0.0542	0.0445	0.0443	0.0394
	$RMSE$	0.0271	0.1061	2.7320	0.0586	0.0543	0.0461	0.0452	0.0394
	$\overline{SE}$	0.0272	0.0536	5.5917	0.0476	0.0483	0.0388	0.0388	0.0386
	CR	95%	–	99%	–	92%	90%	91%	94%

Notes: $Mean$ = simulation average of the parameter estimate; SD = simulation standard deviation of the parameter estimate; $RMSE$ = root mean-squared error of the parameter estimate; $\overline{SE}$ = simulation average of the standard error; and CR = coverage rate for the nominal 95% confidence interval.

Model E: $h(x) = |x| - 1$ ($n = m = 500$).

π	Estimator	OLS*	OLS-S	IV-S	MSOLS	MSII	PILS-E	PILS-B	PARA
1.00	$Mean$	0.9981	1.2282	0.9939	1.1647	0.9666	1.0116	1.0061	1.1770
	SD	0.0237	0.0332	0.0620	0.0362	0.1397	0.0398	0.0397	0.0334
	$RMSE$	0.0238	0.2306	0.0623	0.1686	0.1436	0.0414	0.0402	0.1801
	$\overline{SE}$	0.0239	0.0337	0.0628	0.0477	0.1173	0.0395	0.0396	0.0337
	CR	96%	–	95%	–	97%	94%	94%	–
0.70	$Mean$	0.9981	1.2753	0.9905	1.2055	0.9308	1.0164	1.0091	1.2166
	SD	0.0264	0.0365	0.0896	0.0405	0.4623	0.0467	0.0467	0.0370
	$RMSE$	0.0264	0.2777	0.0901	0.2094	0.4675	0.0495	0.0476	0.2198
	$\overline{SE}$	0.0266	0.0368	0.0906	0.0579	0.3786	0.0459	0.0461	0.0371
	CR	96%	–	96%	–	96%	93%	94%	–
0.15	$Mean$	0.9987	1.3397	0.9013	1.2658	0.9728	1.0262	1.0153	1.2726
	SD	0.0298	0.0408	1.3240	0.0463	1.1735	0.0582	0.0586	0.0416
	$RMSE$	0.0299	0.3421	1.3276	0.2698	1.1738	0.0638	0.0606	0.2758
	$\overline{SE}$	0.0302	0.0405	1.9154	0.0751	1.9124	0.0558	0.0562	0.0413
	CR	96%	–	98%	–	95%	91%	94%	–

Notes: $Mean$ = simulation average of the parameter estimate; SD = simulation standard deviation of the parameter estimate; $RMSE$ = root mean-squared error of the parameter estimate; $\overline{SE}$ = simulation average of the standard error; and CR = coverage rate for the nominal 95% confidence interval.

6 Application to Estimation of Return to Schooling

- We estimate the earnings equation

$$\begin{aligned}\log(earnings) = {} & \beta_0 + \beta_1 education + \beta_2 ability \\ & + \beta_3 experience + \beta_4 experience^2 \\ & + \beta_5 married + \beta_6 black + \beta_7 south + \beta_8 urban + u\end{aligned}$$

  using the following two samples:

  1. $\mathcal{S}_1$ from the 1972 wave of PSID ($n = 2430$).
     - An ability measure (*IQ score*) is included (but this may not be a good ability proxy).
  2. $\mathcal{S}_2$ from the `CARD` dataset provided by Wooldridge (2013) ($m = 1102$).
     - Another ability measure ("Knowledge of the World of Work" or *KWW score*) is included.

IV Estimation.

- Father's education is used as the IV for individual's education in the short regression with $ability$ omitted.

- The sample correlation between these variables is 0.434.

Two-Sample Estimation.

- List of common variables:

  included: *education* (C), *married* (D), *black* (D), *south* (D), *urban* (D)
  excluded: *age* (C), *south while growing up* (D)

- A single match ($K = 1$) and the Mahalanobis metric are adopted for NNM.

Estimation Results.

	(1) OLS*	(2) OLS-S	(3) IV-S	(4) MSOLS	(5) MSII-FM	(6) PILS	(7) PARA
Education	0.0635	0.0718	0.0953	0.0727	0.0685	0.0598	−0.2850
	(0.0056)	(0.0053)	(0.0049)	(0.0059)	(0.0080)	(0.0070)	(0.0891)
Experience	0.0809	0.0818	0.0830	0.0826	0.0766	0.0825	0.1170
	(0.0043)	(0.0043)	(0.0044)	(0.0049)	(0.0065)	(0.0043)	(0.0094)
$Experience^2$	−0.0017	−0.0017	−0.0017	−0.0017	−0.0016	−0.0017	−0.0017
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Ability	0.0313	–	–	−0.0012	0.0029	0.0075	0.2444
	(0.0078)	(–)	(–)	(0.0027)	(0.0081)	(0.0029)	(0.0609)
Married	0.3717	0.3793	0.3844	0.3799	0.3777	0.3954	0.2226
	(0.0539)	(0.0536)	(0.0535)	(0.0535)	(0.0535)	(0.0541)	(0.0605)
Black	−0.1302	−0.1741	−0.1249	−0.1849	−0.1504	−0.1749	−0.1819
	(0.0323)	(0.0316)	(0.0304)	(0.0393)	(0.0806)	(0.0316)	(0.0309)
South	−0.0921	−0.0983	−0.0814	−0.0979	−0.0989	−0.1036	−0.1372
	(0.0288)	(0.0287)	(0.0293)	(0.0286)	(0.0289)	(0.0287)	(0.0299)
Urban	0.1363	0.1499	0.1278	0.1538	0.1404	0.1541	0.0138
	(0.0282)	(0.0284)	(0.0288)	(0.0297)	(0.0390)	(0.0281)	(0.0421)
Data Combination?	No	No	No	Yes	Yes	Yes	Yes
Sample Size: n	2430	2430	2430	2430	2430	2430	2430
m	–	–	–	1102	1102	1102	1102

7 Concluding Remarks

- The PILS estimation procedure is developed for models where endogenous regressors enter the model due to an omitted variable.
  - Convergence properties of PILS are explored.

- Monte Carlo study confirms attractive finite-sample properties of PILS.
  - It is numerically demonstrated that PILS is generally more efficient than IV and dominates other competing estimators (e.g., MSII) in terms of mean squared error.

- The curse of dimensionality in continuous matching variables is still a concern.
  - We may adopt propensity score matching as a means of dimension reduction using multiple matching variables.

References

[1] Abadie, A., and G. W. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica*, 74, 235-267.

[2] Black, B., M. Trainor, and J. E. Spencer (1999): "Wage Protection Systems, Segregation and Gender Pay Inequalities: West Germany, the Netherlands and Great Britain," *Cambridge Journal of Economics*, 23, 449-464.

[3] Card, D. (1995): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in L. N. Christophides, E. K. Grant, and R. Swidinsky (eds.), *Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp*. Toronto: University of Toronto Press, 201-222.

[4] Chen, S. X. (1999): "Beta Kernel Estimators for Density Functions," *Computational Statistics and Data Analysis*, 31, 131 - 145.

[5] Fang, H., M. P. Keane, and D. Silverman (2008): "Sources of Advantageous Selection: Evidence from the Medigap Insurance Market," *Journal of Political Economy*, 116, 303 - 350.

[6] Flavin, M., and S. Nakagawa (2008): "A Model of Housing in the Presence of Adjustment Costs: A Structural Interpretation of Habit Persistence," *American Economic Review*, 98, 474 - 495.

[7] Hirukawa, M., and A. Prokhorov (2018): "Consistent Estimation of Linear Regression Models Using Matched Data," *Journal of Econometrics*, 203, 344 - 358.

[8] Mincer, J. A. (1974): *Schooling, Experience and Earnings*. New York: National Bureau of Economic Research.

[9] Newey, W. K. (1994): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233 - 253.

[10] Pagan, A. (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25, 221 - 247.

[11] Racine, J. S., and Q. Li (2004): "Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Data," *Journal of Econometrics*, 119, 99 - 130.

[12] Wooldridge, J. M. (2013): *Introductory Econometrics: A Modern Approach*, 5th Edition. Mason, OH: South-Western Cengage Learning.

[13] Zabalza, A., and J. L. Arrufat (1985): "The Extent of Sex Discrimination in Great Britain," in A. Zabalza and Z. Tzannatos (eds.), *Women and Equal Pay: The Effects of Legislation on Female Employment and Wages in Britain*, Cambridge, U.K.: Cambridge University Press, 70 - 96.

Nonparametric estimation of density functions from repeated measurements

Daisuke Kurisu

School of Engineering,
Tokyo Institute of Technology

Joint works with
Karun Adusumilli (University of Pennsylvania)
Taisuke Otsu (London School of Economics)
Yoon-Jae Whang (Seoul National University)

August 7, 2019 @福島大学

Outline

- Introduction
- Estimation of density functions
- Estimation of distribution functions under symmetric error
- Estimation of distribution functions under asymmetric error
- Conclusion

Introduction

- Kurisu, D and Otsu, T. (2019). On the uniform convergence of deconvolution estimators from repeated measurements. *submitted*.
  STICERD Econometrics Paper Series EM604.

- Adusumilli, K., Kurisu, D., Otsu, T. and Whang, Y.-J. (2019). Inference on distribution functions under measurement error.
  R&R for *J. Econometrics*.
  STICERD Econometrics Paper Series EM594.

Estimation of density functions

We consider the following repeated measurements model.

- Consider a bivariate i.i.d. sample $\{Y_{j,1}, Y_{j,2}\}_{j=1}^{n}$ of (Y_1, Y_2), which is generated by

$$Y_1 = X + \epsilon_1, \;\; Y_2 = X + \epsilon_2,$$

  where $(X, \epsilon_1, \epsilon_2)$ are unobservables.
- X: an error-free variable of interest, (ϵ_1, ϵ_2) are measurement error for X.
- We are interested in estimating the densities of X, ϵ_1, and ϵ_2.
- Applications:
  (1) Household income and expenditure survey data (AKOW(2019))
  $\rightarrow$ reinterview studies (to assess the credibility of the survey data).
  (2) Medical data
  $\rightarrow$ measurements of patient's status by different methods (to assess the compatibility of measurements).

Estimation of density functions

Assumption M

- (ϵ_1, ϵ_2) are independent copies of a random variable ϵ.
- X is independent of (ϵ_1, ϵ_2).
- X and ϵ have square integrable Lebesgue densities f_X and f_ϵ, respectively.
- The characteristic functions $\varphi_X(\cdot) = E[e^{\mathrm{i}\cdot X}]$ and $\varphi_\epsilon(\cdot) = E[e^{\mathrm{i}\cdot\epsilon}]$ vanish nowhere, and $E[\epsilon] = 0$.

Estimation of density functions

Define

$$\psi(u_1, u_2) = E[e^{\mathrm{i}(u_1 Y_1 + u_2 Y_2)}] = \varphi_X(u_1 + u_2)\varphi_\epsilon(u_1)\varphi_\epsilon(u_2).$$

Under the condition $E[|Y_1|] < \infty$, Kotlarski's indentity gives us a explicit identification formula of φ_X, that is

$$\varphi_X(u) = \exp\left(\int_0^u \frac{\partial\psi(0, u_2)/\partial u_1}{\psi(0, u_2)} du_2\right).$$

The Li and Vuong's (LV) estimator (Li and Vuong('98, JMA)) of φ_X is defined by taking its sample counterpart.

Estimation of density functions

- (Definition of LV estimator):

$$\hat{\varphi}_X(u) = \exp \int_0^u \frac{\partial \hat{\psi}(0, u_2)/\partial u_1}{\hat{\psi}(0, u_2)} du_2,$$

where $\hat{\psi}(u_1, u_2) = \frac{1}{n}\sum_{j=1}^n e^{\mathrm{i}(u_1 Y_{j,1} + u_2 Y_{j,2})}$ and $\frac{\partial \hat{\psi}(u_1,u_2)}{\partial u_1} = \frac{1}{n}\sum_{j=1}^n \mathrm{i}\, Y_{j,1} e^{\mathrm{i}(u_1 Y_{j,1} + u_2 Y_{j,2})}$.

- Based on this estimator, the density f_X of X can be estimated by

$$\hat{f}_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}ux} \hat{\varphi}_X(u) \varphi_K(hu) du,$$

where $\varphi_K(u) = \int_{\mathbb{R}} e^{\mathrm{i}ux} K(x) dx$ is the Fourier transform of a kernel function K and $h = h_n$ is a sequence of positive numbers (bandwidths) such that $h_n \to 0$ as $n \to \infty$.

Estimation of density functions

- Based on the expression $\varphi_\epsilon(u) = \psi(0,u)/\varphi_X(u)$, the characteristic function φ_ϵ of ϵ can also be estimated by

$$\hat{\varphi}_\epsilon(u) = \frac{\hat{\psi}(0,u)}{\hat{\varphi}_X(u)}.$$

- The estimator $\hat{f}_\epsilon$ of the density f_ϵ is given by replacing $\hat{\varphi}_X$ in the definition of $\hat{f}_X$ with $\hat{\varphi}_\epsilon$.

Estimation of density functions (CK)

- We also consider a regularized version of the LV estimator proposed in Comte and Kappus('15, JMA).
- (Definition of CK estimator): Their main idea is to regularize $\hat{\varphi}_X$ as

$$\tilde{\varphi}_X(u) = \frac{\tilde{\varphi}_X^{mod}(u)}{\max\{1, |\tilde{\varphi}_X^{mod}(u)|\}},$$

where

$$\tilde{\varphi}_X^{mod}(u) = \exp \int_0^u \frac{\partial \hat{\psi}(0, u_2)/\partial u_1}{\tilde{\psi}(0, u_2)} du_2$$

with

$$\tilde{\psi}(0, u_2) = \frac{\hat{\psi}(0, u_2)}{\min\{1, \sqrt{n}|\hat{\psi}(0, u_2)|\}}.$$

Estimation of density functions (CK)

- The additional term, $\min\{1, \sqrt{n}|\hat{\psi}(0, u_2)|\}$ circumvents unfavorable effects caused by small values of the denominator.
- The denominator of $\tilde{\varphi}_X$, $\max\{1, |\tilde{\varphi}_X^{mod}(u)|\}$ is introduced to improve the quality of the estimator by imposing that the estimand is a characteristic function, which should not take values lager than one.
- The CK estimator $\tilde{f}_X$ of the density f_X is defined by replacing $\hat{\varphi}_X$ with $\tilde{\varphi}_X$.
- Also, φ_ϵ can be estimated by

$$\tilde{\varphi}_\epsilon(u) = \frac{\hat{\psi}(0, u)}{\bar{\varphi}_X(u)}, \text{ where } \bar{\varphi}_X(u) = \frac{\tilde{\varphi}_X(u)}{\min\{1, \sqrt{n}|\tilde{\varphi}_X(u)|\}}.$$

- The CK estimator $\tilde{f}_\epsilon$ of f_ϵ is given by replacing $\hat{\varphi}_X$ in the definition of $\hat{f}_X$ with $\tilde{\varphi}_\epsilon$.

Estimation of density functions

To estimate densities, we need to choose the kernel function K, and impose the following conditions.

Assumption K

- The kernel function K satisfies $\int_{\mathbb{R}} K(x)dx = 1$, $\int_{\mathbb{R}} x^{\ell} K(x)dx = 0$ for $\ell = 1, \ldots, p-1$, and $\int_{\mathbb{R}} |x|^p K(x)dx < \infty$ with a positive even integer p.
- Also, $\varphi_K(u) = 0$ for any $|u| > 1$.

This assumption says that K is a p-th order kernel function.

Estimation of density functions

(Examples of kernel functions)
Kernel functions satisfying Assumption K is typically constructed by specifying its Fourier transform φ_K. Examples of φ_K include

- (Finite order kernel):
  $\varphi_K(u) = (1-u^2)^k \mathbb{I}\{u \in [-1,1]\}$, $k \geq p+2$.
- (Infinite order kernel 1):

$$\varphi_K(u) = \begin{cases} 1 & \text{if } |u| \leq c_0, \\ \exp\left\{ \frac{-b\exp(-b/(|u|-c_0)^2)}{(|u|-1)^2} \right\} & \text{if } c_0 < |u| < 1, \\ 0 & \text{if } 1 \leq |u|, \end{cases}$$

  for $0 < c_0 < 1$ and $b > 0$.
- (Infinite order kernel 2): The sinc kernel function.

$$\varphi_K(u) = \mathbb{I}\{u \in [-1,1]\}, \ K(x) = \sin(x)/x.$$

Estimation of density functions

We consider the following two scenarios for the characteristic functions φ_X and φ_ϵ.

- Assumption OS (ordinary smooth): For some positive constants $\beta_x > 1$, $C_x \geq c_x$, ω_x, $\beta_\epsilon > 1$, $C_\epsilon \geq c_\epsilon$, and ω_ϵ, it holds

$$c_x|u|^{-\beta_x} \leq |\varphi_X(u)| \leq C_x|u|^{-\beta_x}, \text{ for all } |u| \geq \omega_x,$$
$$c_\epsilon|u|^{-\beta_\epsilon} \leq |\varphi_\epsilon(u)| \leq C_\epsilon|u|^{-\beta_\epsilon}, \text{ for all } |u| \geq \omega_\epsilon.$$

- The conditions $\beta_x, \beta_\epsilon > 1$ are introduced to guarantee the consistency of the density functions.
- We need to use the lower and upper bounds of the characteristic functions to obtain suitable bounds of the stochastic and deterministic bias terms of the estimators.
- Examples: Laplace, Gamma.

Estimation of density functions

- Assumption SS (super smooth): For some positive constants ρ_x, $C_x \geq c_x$, ω_x, ρ_ϵ, $C_\epsilon \geq c_\epsilon$, ω_ϵ, and some constants $\beta_x, \beta_\epsilon \in \mathbb{R}$, it holds

$$
\begin{aligned}
c_x|u|^{\beta_x}\exp(-|u|^{\rho_x}/\mu_x) \leq |\varphi_X(u)| \leq C_x|u|^{\beta_x}\exp(-|u|^{\rho_x}/\mu_x), \\
\text{for all } |u| \geq \omega_x, \\
c_\epsilon|u|^{\beta_\epsilon}\exp(-|u|^{\rho_\epsilon}/\mu_\epsilon) \leq |\varphi_\epsilon(u)| \leq C_\epsilon|u|^{\beta_\epsilon}\exp(-|u|^{\rho_\epsilon}/\mu_\epsilon), \\
\text{for all } |u| \geq \omega_\epsilon.
\end{aligned}
$$

- We use lower and upper bounds to control estimation errors of estimators.
- Examples: Normal, Mixture Normal, Cauchy.

Estimation of density functions

Theorem (KO(2019))

Suppose that Assumption M holds true, and $E[|Y_1|^{2+\eta}] < \infty$ for some $\eta > 0$.

(i) : Under Assumption OS and $n^{-1/2} T_n^{3\beta_x + 2\beta_\epsilon + 1} \log T_n \to 0$ as $n \to \infty$, it holds

$$\sup_{u \in [-T_n, T_n]} |\hat{\varphi}_X(u) - \varphi_X(u)| = O_p\left(n^{-1/2} T_n^{2\beta_x + 2\beta_\epsilon + 1} \log T_n\right),$$

$$\sup_{u \in [-T_n, T_n]} |\hat{\varphi}_\epsilon(u) - \varphi_\epsilon(u)| = O_p\left(n^{-1/2} T_n^{3\beta_x + 2\beta_\epsilon + 1} \log T_n\right).$$

Estimation of density functions

Theorem (conti.)

(i) Additionally, suppose that Assumption K holds for $p \geq \max\{\beta_x, \beta_\epsilon\}$. Set $T_n = h^{-1}$. Then

$$\sup_{|u| \leq h^{-1}} |\hat{f}_X(x) - f_X(u)| = O_p\left(n^{-1/2} h^{-2\beta_x - 2\beta_\epsilon - 2} \log h^{-1} + h^{\beta_x - 1}\right),$$

$$\sup_{|u| \leq h^{-1}} |\hat{f}_\epsilon(x) - f_\epsilon(u)| = O_p\left(n^{-1/2} h^{-3\beta_x - 2\beta_\epsilon - 2} \log h^{-1} + h^{\beta_\epsilon - 1}\right).$$

Estimation of density functions

Theorem (conti.)

(ii) Under Assumption SS and
$n^{-1/2}T_n^{1-3\beta_x-2\beta_\epsilon}(\log T_n)\exp\left(3T_n^{\rho_x}/\mu_x + 2T_n^{\rho_\epsilon}/\mu_\epsilon\right) \to 0$ as $n \to \infty$,
it holds

$$\sup_{u\in[-T_n,T_n]} |\hat{\varphi}_X(u) - \varphi_X(u)|$$
$$= O_p\left(n^{-1/2}T_n^{1-2\beta_x-2\beta_\epsilon}(\log T_n)\exp\left(\frac{2T_n^{\rho_x}}{\mu_x} + \frac{2T_n^{\rho_\epsilon}}{\mu_\epsilon}\right)\right),$$
$$\sup_{u\in[-T_n,T_n]} |\hat{\varphi}_\epsilon(u) - \varphi_\epsilon(u)|$$
$$= O_p\left(n^{-1/2}T_n^{1-3\beta_x-2\beta_\epsilon}(\log T_n)\exp\left(\frac{3T_n^{\rho_x}}{\mu_x} + \frac{2T_n^{\rho_\epsilon}}{\mu_\epsilon}\right)\right).$$

Theorem (conti.)

(ii) Additionally, suppose that Assumption K holds and that there exists $0 < c \leq 1$ such that $\varphi_K(x) = 1$ for $|x| \leq c$. Set $T_n = h^{-1}$. Then

$$\sup_{|u| \leq h^{-1}} |\hat{f}_X(u) - f_X(u)|$$
$$= O_p\left(n^{-1/2} h^{2\beta_x + 2\beta_\epsilon - 2} (\log h^{-1}) e^{\frac{2h^{-\rho_X}}{\mu_X} + \frac{2h^{-\rho_\epsilon}}{\mu_\epsilon}} + h^{\rho_X/q - \beta_x - 1} e^{-\frac{c^{\rho_X} h^{-\rho_X}}{\mu_X}}\right),$$
$$\sup_{|u| \leq h^{-1}} |\hat{f}_\epsilon(u) - f_\epsilon(u)|$$
$$= O_p\left(n^{-1/2} h^{3\beta_x + 2\beta_\epsilon - 2} (\log h^{-1}) e^{\frac{2h^{-\rho_X}}{\mu_X} + \frac{2h^{-\rho_\epsilon}}{\mu_\epsilon}} + h^{\rho_X/q - \beta_x - 1} e^{-\frac{c^{\rho_X} h^{-\rho_X}}{\mu_X}}\right),$$

where $q = 1$ when $\beta_x, \beta_\epsilon > 0$, and $q > 1$ when $\beta_x, \beta_\epsilon \leq 0$.

Estimation of density functions

- The same uniform convergence results in (i) and (ii) hold true even if we replace the LV estimator $(\hat{\varphi}_X, \hat{\varphi}_\epsilon, \hat{f}_X, \hat{f}_\epsilon)$ with the CK estimator $(\tilde{\varphi}_X, \tilde{\varphi}_\epsilon, \tilde{f}_X, \tilde{f}_\epsilon)$.
- To obtain the uniform consistency of $\hat{\varphi}$ (or $\tilde{\varphi}$) on $[-T_n, T_n]$, it is sufficient to assume

$$n^{-1/2} T_n^{2\beta_x + 2\beta_\epsilon + 1} \log T_n \to 0 \text{ under Assumption OS},$$
$$n^{-1/2} T_n^{1-2\beta_x - 2\beta_\epsilon} (\log T_n) e^{\frac{2T_n^{\rho_x}}{\mu_x} + \frac{2T_n^{\rho_\epsilon}}{\mu_\epsilon}} \to 0 \text{ under Assumption SS},$$

  as $n \to \infty$.
- We can take $c = 1$ in Theorem (ii) if we use the sinc kernel function.
- We can also derive uniform convergence rates when X is ordinary smooth and ϵ is super smooth and vise versa since the decay rates of the characteristic functions are not essential in our proofs.

Estimation of density functions

(Comparison with Li and Vuong ('98, JMA))

- We note that LV established the uniform convergence rates of their estimators under the assumption that both X and ϵ have bounded support. Our theorem does not require such boundedness.
- Also, the convergence rates obtained in our theorem are typically faster than those obtained in LV. For example, if we set $T_n = O\left((n/\log\log n)^{\alpha/2(1+\beta_x+\beta_\epsilon)}\right)$ with $0 < \alpha < 1/2$ as in Lemma 3.1 of LV, our Theorem (i) implies that

$$\sup_{u \in [-T_n, T_n]} |\hat{\varphi}_X(u) - \varphi_X(u)| = O_p\left(\left(\frac{n}{\log\log n}\right)^{-\frac{1}{2}+\alpha-\frac{\alpha}{2(1+\beta_X+\beta_\epsilon)}}\right),$$

and this convergence rate is faster than that given in LV, i.e., $\left(\frac{n}{\log\log n}\right)^{-\frac{1}{2}+\alpha}$. Similar comments apply to other cases.

Estimation of density functions

(Comparison with Bonhomme and Robin('10, Rev. Econ. Stud.))

- The convergence rates in our theorem are also faster than those given in Bonhomme and Robin ('10). For example, under Assumption OS, Bonhomme and Robin ('10, Theorem 1) implies that

$$\begin{aligned}\sup_{u\in[-T_n,T_n]}|\hat{\varphi}_X(u)-\varphi_X(u)| &= O_p\left(n^{-1/2}T_n^{3\beta_x+3\beta_\epsilon+2}\log T_n\right),\\ \sup_{u\in[-T_n,T_n]}|\hat{\varphi}_\epsilon(u)-\varphi_\epsilon(u)| &= O_p\left(n^{-1/2}T_n^{3\beta_x+3\beta_\epsilon+2}\log T_n\right).\end{aligned}$$

- It also should be noted that our assumption on (Y_1, Y_2) is weaker than an assumption in Bonhomme and Robin ('10) since we do not need the existence of moment generating functions of Y_1^2 and Y_1Y_2.
- More precisely, the same convergence rate given in Lemma 1 of their paper can be obtained under weaker condition by proving a multivariate version of Neumann and Reiss ('09, Theorem 4.1).

Estimation of density functions

- We obtained the same uniform convergence rates for the LV estimator and the CK estimator.
- It is open whether the LV estimator can achieve the L_2 convergence rate in Comte and Kappus('15, JMA).
- To control the L_2 risk of the LV-type estimators which are defined by the ratio of the (regularized) empirical averages, it seems crucial to introduce some regularization as in Comte and Kappus('15).

Estimation of density functions

Multivariate version of Theorem 4.1 in Neumann and Reiß('09, Bernoulli)

- The relaxation of assumptions in the previous works can be achieved by a multivariate version of theorem 4.1 in Neumann and Reiß('09) on the normalized empirical characteristic function process.
- Let $\{\boldsymbol{Y}_j = (Y_{j,1}, \ldots, Y_{j,d})'\}_{j=1}^n$ be $\mathbb{R}^d$-valued i.i.d. random variables. For $\boldsymbol{t} = (t_1, \ldots, t_d)' \in \mathbb{R}^d$, define

$$\psi(\boldsymbol{t}) = E[e^{i\boldsymbol{t}\cdot\boldsymbol{Y}}],\ \hat{\psi}(\boldsymbol{t}) = \frac{1}{n}\sum_{j=1}^{n} e^{i\boldsymbol{t}\cdot\boldsymbol{Y}_j},$$

Estimation of density functions

and for $\boldsymbol{k} = (k_1, \ldots, k_d)' \in \mathbb{N}^d$, $|\boldsymbol{k}| = \sum_{j=1}^d k_j$,

$$C_n(\boldsymbol{t}) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (e^{i\boldsymbol{t}\cdot \boldsymbol{Y}_j} - E[e^{i\boldsymbol{t}\cdot \boldsymbol{Y}_1}]) = \sqrt{n}(\hat{\psi}(\boldsymbol{t}) - \psi(\boldsymbol{t})),$$

$$C_n^{(\boldsymbol{k})}(\boldsymbol{t}) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\partial^{|\boldsymbol{k}|}}{\partial t_1^{k_1} \cdots \partial t_d^{k_d}} (e^{i\boldsymbol{t}\cdot \boldsymbol{Y}_j} - E[e^{i\boldsymbol{t}\cdot \boldsymbol{Y}}]),$$

$$E[\| C_n^{(\boldsymbol{k})} \|_{L_\infty(w)}] = E\left[\sup_{\boldsymbol{t}\in\mathbb{R}^d} \left(w(\|\boldsymbol{t}\|) |C_n^{(\boldsymbol{k})}(\boldsymbol{t})| \right) \right],$$

where $w(t) = (\log(e + |t|))^{-1/2-\delta}$ for some $\delta > 0$ and $\| \cdot \|$ is the Euclidean norm on $\mathbb{R}^d$.

Estimation of density functions

For example,

$$C_n^{\boldsymbol{k}}(\boldsymbol{t}) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\partial}{\partial t_1}(e^{\mathrm{i}\boldsymbol{t}\cdot\boldsymbol{Y}_j} - E[e^{\mathrm{i}\boldsymbol{t}\cdot\boldsymbol{Y}}]),\ \boldsymbol{k} = (0,1)',$$

$$= \sqrt{n}\left\{\frac{1}{n}\sum_{j=1}^{n}(g(\boldsymbol{Y}_j) - E[g(\boldsymbol{Y}_1)])\right\},\ g \in \mathbb{G}_{1,k},$$

$$\mathbb{G}_{1,k} := \left\{\boldsymbol{y} \mapsto \frac{\partial}{\partial t_1}\left(\cos(\boldsymbol{t}\cdot\boldsymbol{y}) + \mathrm{i}\sin(\boldsymbol{t}\cdot\boldsymbol{y})\right) : \boldsymbol{t} \in \mathbb{R}^d\right\}.$$

From Corollary 19.35 in van der Vaart (1998), we have that

$$\sup_{n\geq 1} E[\|C_n^{(\boldsymbol{k})}\|_{L_\infty(w)}] \leq CJ_{[\cdot]}\left(\sqrt{E[Y_1^{2k}]}, \mathbb{G}_{1,k}\right)$$

for a universal constant C and where $J_{[\cdot]}(\delta, G) = \int_0^\delta \sqrt{\log(N_{[\cdot]}(\epsilon, G))}d\epsilon$ is the L^2-bracketing number of $\mathbb{G}_{1,k}$.

Estimation of density functions

Lemma 3(KO(2019))

Assume $E\left[\left(\prod_{j=1}^{d}|Y_j|^{(k_j \vee 1/2)}\right)^{2+\eta}\right] < \infty$ for some $\eta > 0$. Then

$$\sup_{n \geq 1} E[\| C_n^{(\boldsymbol{k})} \|_{L_\infty(w)}] < \infty.$$

We can also apply this result to examples in Bonhomme and Robin('10) to weaken their assumptions on the existence of m.g.f.s.

Estimation of density functions

Since

$$\|C_n^{(\boldsymbol{k})}\|_{L_\infty(w)} \geq \sqrt{n} \sup_{\|\boldsymbol{u}\| \leq T_n} \left| \frac{\partial^{|\boldsymbol{k}|}}{\partial u_1^{k_1} \cdots \partial u_d^{k_d}} \left(\hat{\psi}(\boldsymbol{u}) - \psi(\boldsymbol{u}) \right) \right| \inf_{\|\boldsymbol{u}\| \leq T_n} w(\|\boldsymbol{u}\|),$$

Lemma 3 implies that

$$E\left[\sup_{\|\boldsymbol{u}\| \leq T_n} \left| \frac{\partial^{|\boldsymbol{k}|}}{\partial u_1^{k_1} \cdots \partial u_d^{k_d}} \left(\hat{\psi}(\boldsymbol{u}) - \psi(\boldsymbol{u}) \right) \right| \right] \leq \frac{\sup_{n \geq 1} E[\|C_n^{(\boldsymbol{k})}\|_{L_\infty(w)}]}{\sqrt{n} \inf_{\|\boldsymbol{u}\| \leq T_n} w(\|\boldsymbol{u}\|)}$$
$$= O\left(n^{-1/2} \log T_n \right).$$

Then we can show

$$\sup_{\|\boldsymbol{u}\| \leq T_n} \left| \frac{\partial^{|\boldsymbol{k}|}}{\partial u_1^{k_1} \cdots \partial u_d^{k_d}} \left(\hat{\psi}(\boldsymbol{u}) - \psi(\boldsymbol{u}) \right) \right| = O_p\left(n^{-1/2} \log T_n \right)$$

from Markov's inequality.

Estimator of distribution function under symmetric error

(We use the notations in AKOW(2019))

Suppose we observe a random sample $\{X_j\}_{j=1}^n$ generated from

$$X = X^* + \epsilon,$$

where X^* is an unobservable variable of interest and ϵ is its measurement error. We assume that ϵ is independent of X^*.
More specifically, suppose that we observe

$$X_{j,k} = X_j^* + \epsilon_{j,k} \text{ for } k = 1, 2 \text{ and } j = 1, \ldots, n.$$

where $\{X_j^*\}_{j=1}^n$ and $\{\epsilon_{j,k}\}_{j,k}$ are two independent sets of i.i.d. random variables.

Estimator of distribution function under symmetric error

If the pdf f_ϵ of ϵ is "symmetric", its Fourier transform f_ϵ^{ft} can be estimated by (Delaigle, Hall and Meister ('08, AoS))

$$\hat{f}_\epsilon^{\mathrm{ft}}(\omega) = \left| \frac{1}{n} \sum_{j=1}^{n} \cos\{\omega(X_{j,1} - X_{j,2})\} \right|^{1/2}$$

We can estimate the cdf of F_{X^*} by

$$\hat{F}_{X^*}(t) = \frac{1}{2} - \frac{1}{2n} \sum_{j=1}^{n} \sum_{k=1}^{2} \hat{\mathbb{L}}\left(\frac{t - X_{j,k}}{h} \right), \ \hat{\mathbb{L}}(u) = \frac{1}{2\pi} \int_{-1}^{1} \frac{\sin(\omega u)}{\omega} \frac{K^{\mathrm{ft}}(\omega)}{\hat{f}_\epsilon^{\mathrm{ft}}(\omega/h)} d\omega,$$

where K^{ft} is the Fourier transform of a kernel function K.

Estimator of distribution function under symmetric error

- We are concerned with approximation for the distribution of the maximal deviation

$$T_n = \sup_{t \in T} |\hat{F}_{X^*}(t) - F_{X^*}(t)|,$$

  where T is a compact interval specified by the researcher.
- A direct use of such approximation is construction of the confidence band for F_{X^*} over T.

Estimator of distribution function under symmetric error

- Why distribution function?
- Let $\{X_j\}_{j=1}^{n_1}$ and $\{Y_j\}_{j=1}^{n_2}$ be two independent samples of X and Y. X and Y are generated as

$$X = X^* + \epsilon,\ \ Y = Y^* + \delta,$$

- To test the hypothesis of the (first-order) stochastic dominance

$$H_0 : F_{X^*}(t) \leq F_{Y^*}(t) \text{ for all } t \in T,$$

against the negation of H_0, density-based methods are cumbersome to be handled.

- In AKOW, we applied the stochastic order test to the Korea Household Income and Expenditure Survey data to investigate welfare changes of different population subgroups between 2006 and 2012.

Estimator of distribution function under symmetric error

- Bootstrap approximation: the bootstrap version of $\hat{F}_{X^*}$ is given by

$$\hat{F}^{\sharp}_{X^*}(t) = \frac{1}{2} - \frac{1}{2n}\sum_{j=1}^{n}\sum_{k=1}^{2}\hat{\mathbb{L}}\left(\frac{t - X^{\sharp}_{j,k}}{h}\right),$$

where $X^{\sharp}_{j,k}$ is randomly drawn from the pooled observations $\{X_{j,k}\}$ with equal weights. The bootstrap counterpart of T_n is obtained as

$$T^{\sharp}_n = \sup_{t \in T}|\hat{F}^{\sharp}_{X^*}(t) - \hat{F}_{X^*}(t)|.$$

- Let $\hat{c}_\alpha$ denote the $(1-\alpha)$-th quantile of the bootstrap statistic $T^{\sharp}_n$.

Theorem 1(AKOW(2019))

Under some regularity conditions and suppose either Assumption OS or SS holds. Then,

$$P(T_n \leq \hat{c}_\alpha) \geq 1 - \alpha + o(1).$$

Estimator of distribution function under symmetric error

- Bootstrap based implementation of stochastic dominance test: to test the (first-order) stochastic dominance

$$H_0 : F_{X^*}(t) \leq F_{Y^*}(t) \text{ for all } t \in T,$$

against the negation of H_0, we can use a Kolmogorov-type test.

- The test statistics and its bootstrap counterpart are given by

$$D_{n_1,n_2} = \sup_{t \in \mathcal{T}} \{\hat{F}_{X^*}(t) - \hat{F}_{Y^*}(t)\},$$
$$D^{\sharp}_{n_1,n_2} = \sup_{t \in \mathcal{T}} \left\{ \hat{F}^{\sharp}_{X^*}(t) - \hat{F}^{\sharp}_{Y^*}(t) - \{\hat{F}_{X^*}(t) - \hat{F}_{Y^*}(t)\} \right\},$$

where $\hat{F}^{\sharp}_{X^*}$ and $\hat{F}^{\sharp}_{Y^*}$ are computed using bootstrap samples $\{X^{\sharp}_i\}_{i=1}^{n_1}$ and $\{Y^{\sharp}_i\}_{i=1}^{n_2}$ from $\{X_i\}_{i=1}^{n_1}$ and $\{Y_i\}_{i=1}^{n_2}$, respectively.

Estimator of distribution function under symmetric error

Let $\hat{c}_\alpha^D$ denotes the $(1-\alpha)$-th quantile of the bootstrap statistics $D^\sharp_{n_1,n_2}$.

Theorem 6(AKOW(2019))

Suppose either Assumption OS or SS holds. Then, under some regularity conditions, we have the following results.

- Under H_0,
$$P(D_{n_1,n_2} > \hat{c}_\alpha^D) \leq \alpha + o(1).$$
- Under the alternative H_1(i.e. H_0 is false),
$$P(D_{n_1,n_2} > \hat{c}_\alpha^D) \to 1, \text{ as } n \to \infty.$$

Estimator of distribution function under asymmetric error

Let $m < n$. We can estimate F_{X^*} under possibly asymmetric error distribution by

$$\check{F}^A_{X^*}(t) = \frac{1}{2} - \frac{1}{2m}\sum_{j=1}^{m}\sum_{k=1}^{2}\check{\mathbb{L}}\left(\frac{t - X_{j,k}}{h}\right),$$

$$\check{\mathbb{L}}(u) = \frac{1}{\pi}\int_0^1 \frac{1}{\omega}\mathrm{Im}\left[\frac{e^{\mathrm{i}\omega u}}{\check{f}^{\mathrm{ft}}_{\epsilon}(\omega/h)}\right] K^{\mathrm{ft}}(\omega)d\omega,$$

where $\check{f}^{\mathrm{ft}}_{\epsilon}$ is the Comte and Kappus' estimator (computed from "n" observations).
Let $T^A_n = \sup_{t\in T}|\check{F}^A_{X^*}(t) - F_{X^*}(t)|$, and $\hat{c}^A_{\alpha}$ denote the $(1-\alpha)$-th quantile of the bootstrap statistic $T^{A\sharp}_n = \sup_{t\in T}|\check{F}^{A\sharp}_{X^*}(t) - \check{F}^A_{X^*}(t)|$, where

$$\check{F}^{A\sharp}_{X^*}(t) = \frac{1}{2} - \frac{1}{2m}\sum_{j=1}^{m}\sum_{k=1}^{2}\check{\mathbb{L}}\left(\frac{t - X^{\sharp}_{j,k}}{h}\right).$$

Estimator of distribution function under asymmetric error

Theorem 2(AKOW(2019))

Under some regularity conditions and suppose

(i) Assumption OS holds true, and $mn^{\zeta_0 - 1}h^{2\beta_\epsilon - 2\beta_x - 1} \to 0$ as $n \to \infty$ for some $0 < \zeta_0 < 1$, or

(ii) Assumption SS holds true, and $mn^{\zeta_1 - 1}(\log n)^{2\beta_x + 1}(\log \log n)^2 \to 0$ as $n \to \infty$ for some $0 < \zeta_1 < 1$.

Then

$$P(T_n^A \leq \hat{c}_\alpha^A) \geq 1 - \alpha - o(1).$$

- We can construct an asymptotic confidence band for F_{X^*} over T with level α as $[\hat{F}_{X^*}(t) \pm \hat{c}_\alpha]$ (Theorem 1) or $[\check{F}_{X^*}^A(t) \pm \hat{c}_\alpha^A]$ (Theorem 2).
- Other applications and real data analysis are given in AKOW(2019).

Conclusion

- We investigated the uniform convergence of deconvolution estimators from repeated measurements.
- We achieved faster uniform convergence rates than those obtained in existing results under weaker assumptions.
- Our results can be applied to uniform inference on distribution functions under possibly asymmetric measurement error.

(i) Kurisu, D and Otsu, T. (2019). On the uniform convergence of deconvolution estimators from repeated measurements. STICERD DP EM604. *submitted*.

(ii) Adusumilli, K., Kurisu, D., Otsu, T. and Whang, Y.-J. (2019). Inference on distribution functions under measurement error. R&R for *J. Econometrics*. STICERD DP EM594.

References

- Bonhomme, S. and Robin, J.-M. (2010). Generalized non-parametric deconvolution with an application to earnings dynamics. *Rev. Econ. Stud.* **77**, 491-533.
- Comte, F. and Kappus, J. (2015). Density deconvolution from repeated measurement without symmetry assumption on the errors. *J. Multivariate Anal.* **140**, 31-46.
- Delaigle, A., Hall, P. and A. Meister (2008). On deconvolution with repeated measurements. *Ann. Statist.* **36**, 665-685.
- Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *J. Multivariate Anal.* **65**, 139-165.
- Neumann, M. and Reiß, M. (2009). Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli*, 15, 223-248.

データサイエンス・サマーキャンプ福島

周波数回帰と経済データへの応用

佐藤　整尚　（東京大学・経済）

(データサイエンスの業界でよく使われているJupyter Notebook を使って発表したいと思います。）

SIMLでの変換

$$z = P_n C_n^{-1}(x - X_0)$$

where with $x = (x_1, x_2, \cdots, x_n)^\top$, $X_0 = (x_0, x_0, \cdots, x_0)^\top$, $z = (z_1, z_2, \cdots, z_n)^\top$,
C_n^{-1} is an $n \times n$-matrix representing the first-order difference, and P_n is an $n \times n$-orthogonal matrix such that $C_n^{-1} C_n^{\top -1} = P_n D_n P_n^\top$ with a diagonal matrix D_n. That is,

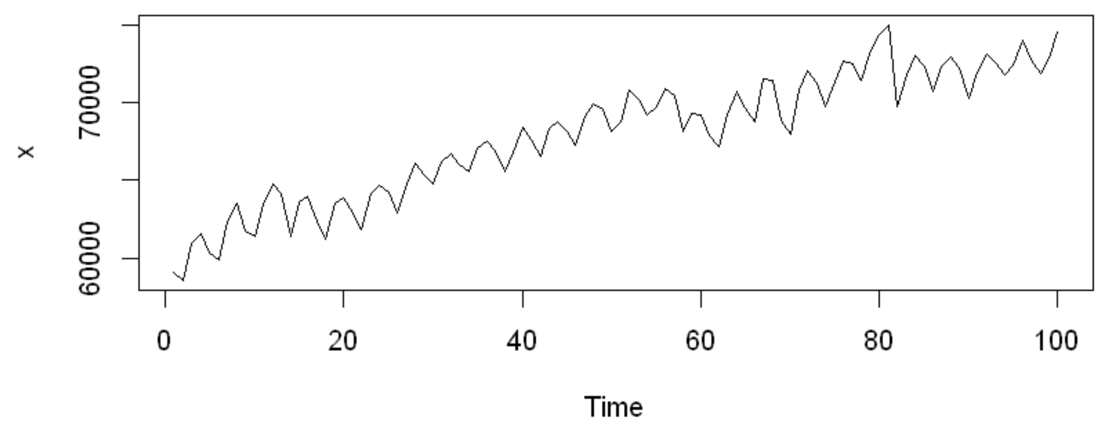
$$C_n^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \vdots \\ 0 & 0 & -1 & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix} \}n$$

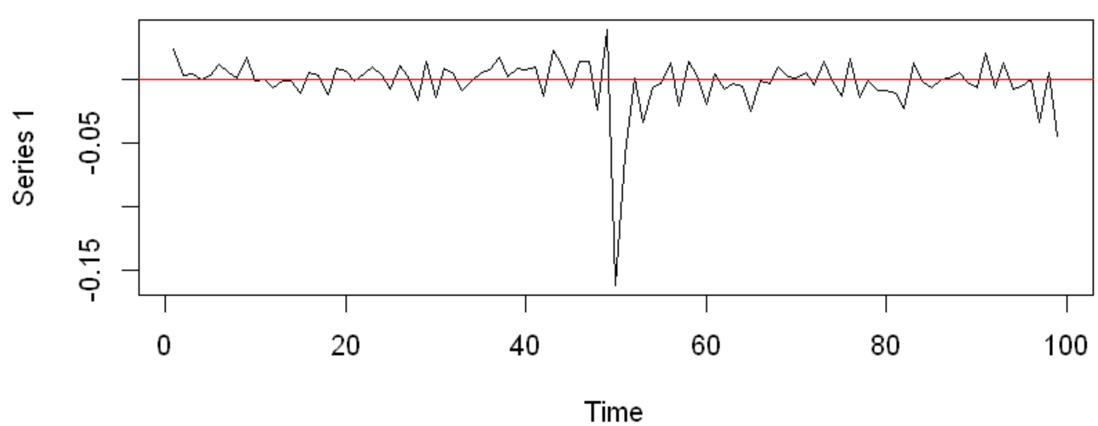
$$P_n = (p_{ij}^{(n)})_{1 \le i,j \le n}$$

$$p_{ij}^{(n)} = \sqrt{\frac{2}{n + 1/2}} \cos\left[\frac{2\pi}{2n+1}(i - \frac{1}{2})(j - \frac{1}{2})\right].$$

To detect a trend component in x, we use z_i only with $i = 1, \cdots, m(< n)$ for an estimation, because z_i with smaller i contains information about a longer cyclic component.

In [33]:
```
n <- length(shouhi)-1
Pmat <- sqrt(2/(n+0.5))*
            cos(outer(seq(n)-0.5,2*pi/(2*(n)+1)*(seq(n)-0.5)))
z <- Pmat %*% diff(log(shouhi))
parm(2,1)
tsplot(shouhi)
#abline(v=82)
tsplot(z)
abline(h=0,col=2)
```



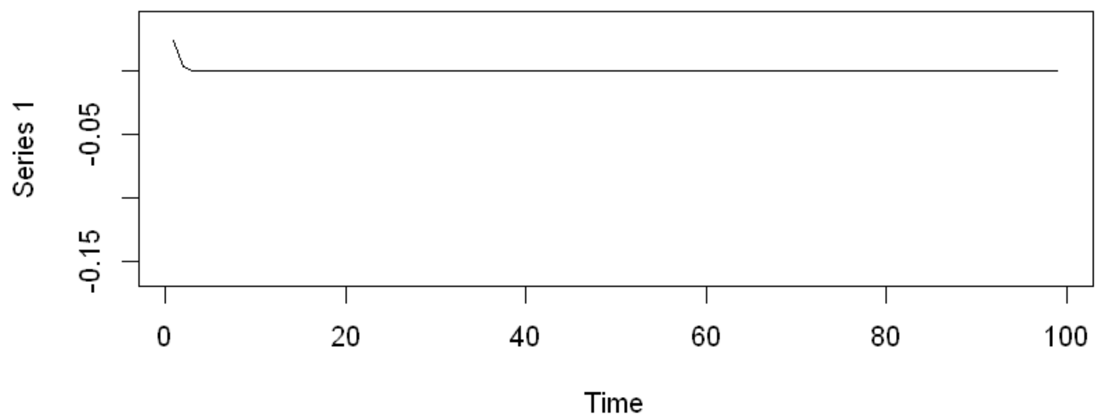


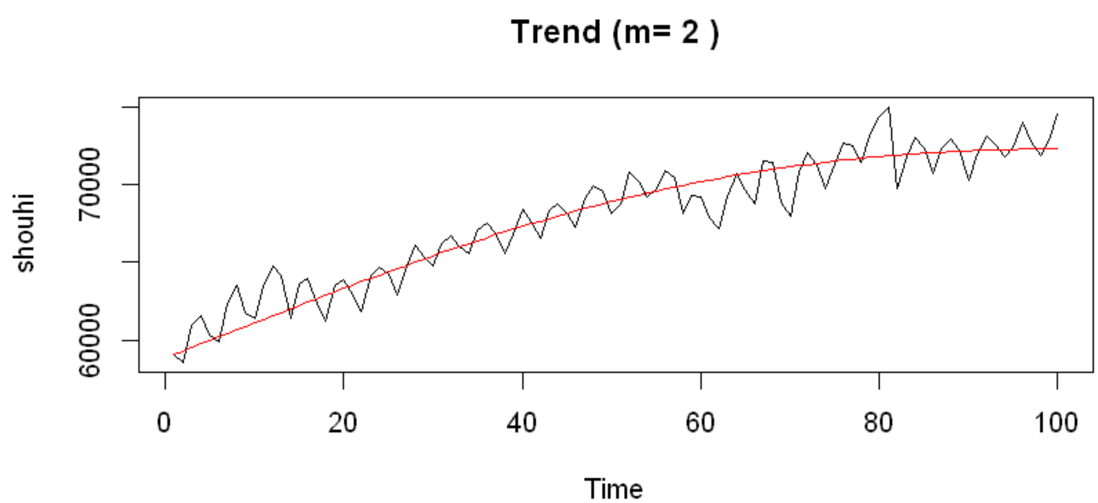
Trend 推定

trend series: T_x for x:

$$P_{m,n}^{(n)} = (p_{ij}^{(n)})_{1\le i\le m, 1\le j\le n}$$
$$T_x = C_n (P_{m,n}^{(n)})^\top P_{m,n}^{(n)} C_n^{-1} x + X_0,$$

In [35]:
```
parm(2,1)
m <- 2
zT <- z
zT[(m+1):n] <- 0
plot.ts(z,type="n")
lines(zT)
#abline(h=0,col=2)
plot.ts(shouhi)
Trend <- exp(cumsum(c(log(shouhi[1]),t(Pmat) %*% zT)))
lines(Trend,col=2)
title(paste("Trend (m=",m,")"))
```





季節成分の推定

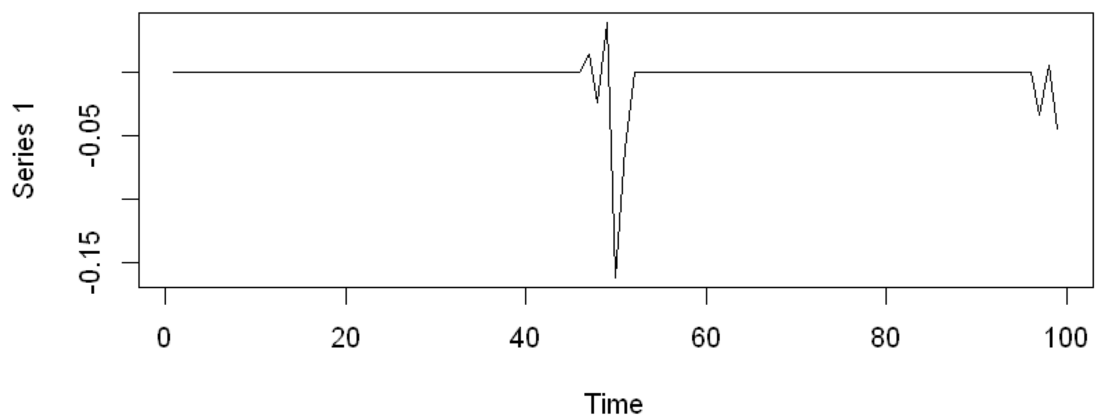
季節周波数付近の z のみを逆変換すればよい

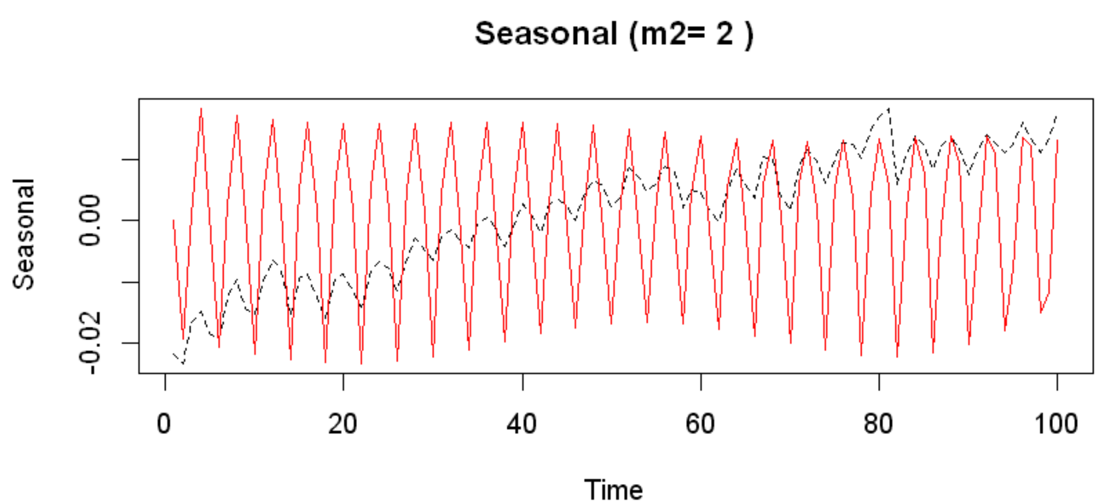
$$z = P^{(n)}_{n,n} C_n^{-1} x$$

$$z_s = (0, \cdots, 0, z_{n/2-2}, \cdots, z_{n/2+2}, 0, \cdots, 0, z_{n-2}, \cdots, z_n)'$$

$$S_x = C_n (P^{(n)}_{n,n})^\top z_s + X_0,$$

In [36]:
```
parm(2,1)
m2 <- 2
n2 <- as.integer(n/2)
zS <- z*0
zS[(n2-m2):(n2+m2)] <- z[(n2-m2):(n2+m2)]
zS[(n-m2):(n)] <- z[(n-m2):(n)]
plot.ts(z,type="n")
lines(zS)
#abline(h=0,col=2)
plot.ts(log(shouhi),lty=2,ylab="",yaxt="n")
par(new=T)
Seasonal <- cumsum(c(0,t(Pmat) %*% zS))
plot.ts(Seasonal,col=2)
title(paste("Seasonal (m2=",2,")"))
```



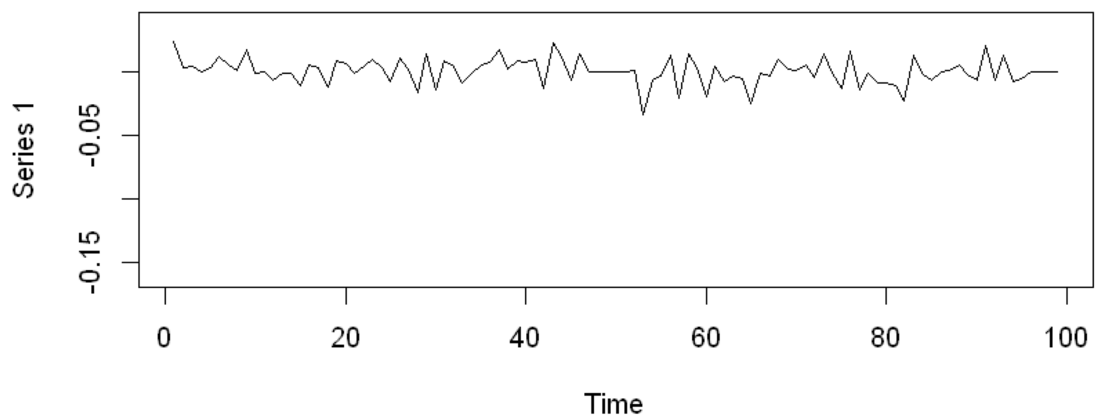


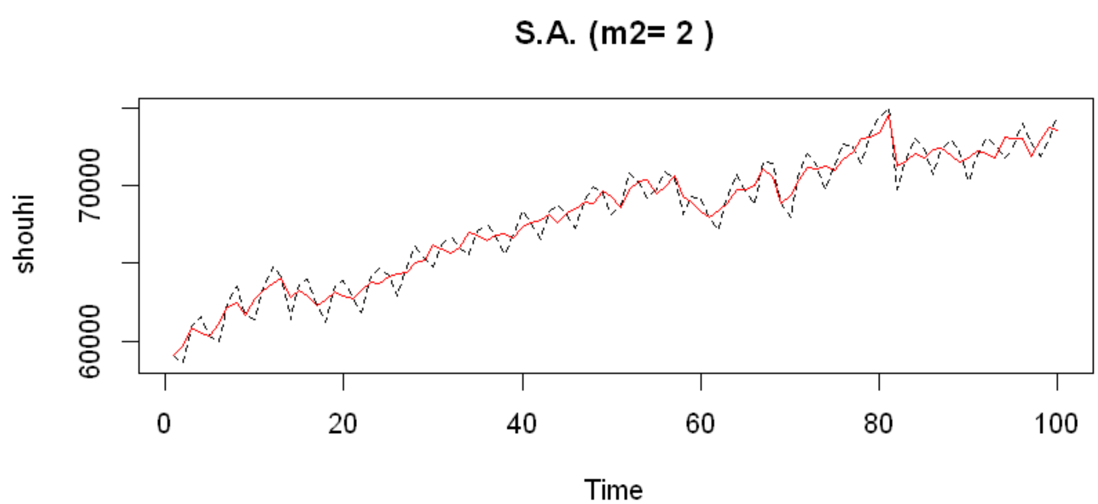
季節調整値

z において季節性に対応する部分を 0 にすれば良い。

$$z_a = (z_1, z_2, \cdots, z_{n/2-3}, 0, 0, 0, 0, 0, z_{n/2+3}, \cdots, z_{n-3}, 0, 0, 0)'$$
$$A_x = C_n (P_{n,n}^{(n)})^\top z_a + X_0,$$

In [37]:
```
parm(2,1)
m2 <- 2
n2 <- as.integer(n/2)
zAD <- z
zAD[c((n2-m2):(n2+m2),(n-m2):(n))] <- 0
plot.ts(z,type="n")
lines(zAD)
plot.ts(shouhi,lty=2)
SA <- cumsum(c(log(shouhi)[1],t(Pmat) %*% zAD))
lines(exp(SA),col=2)
title(paste("S.A. (m2=",m2,")"))
```

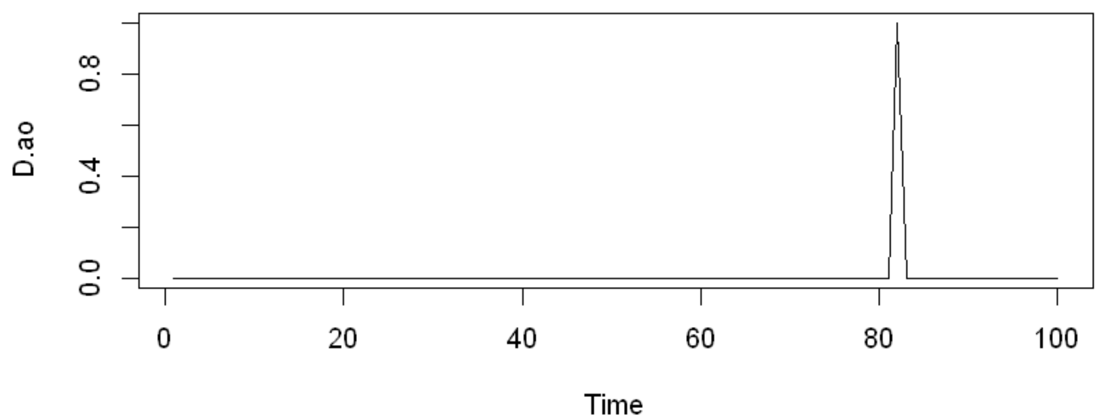


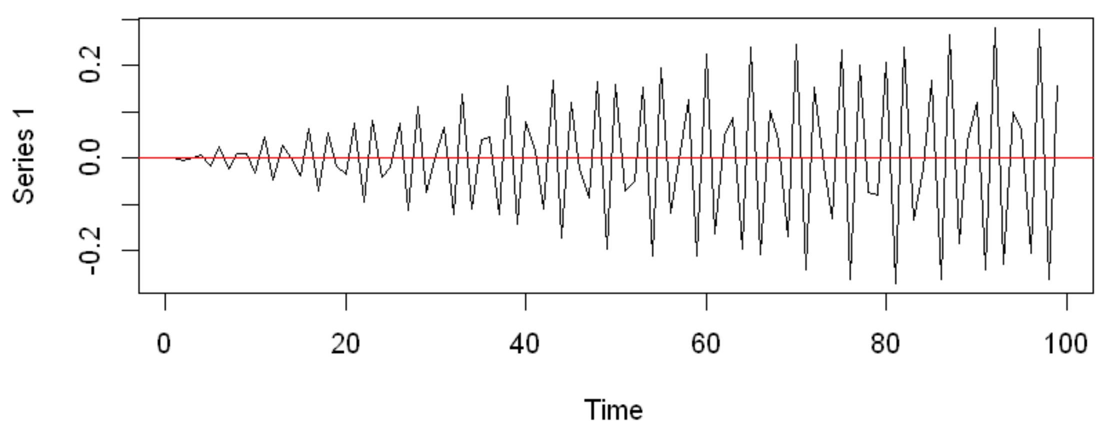


異常値処理

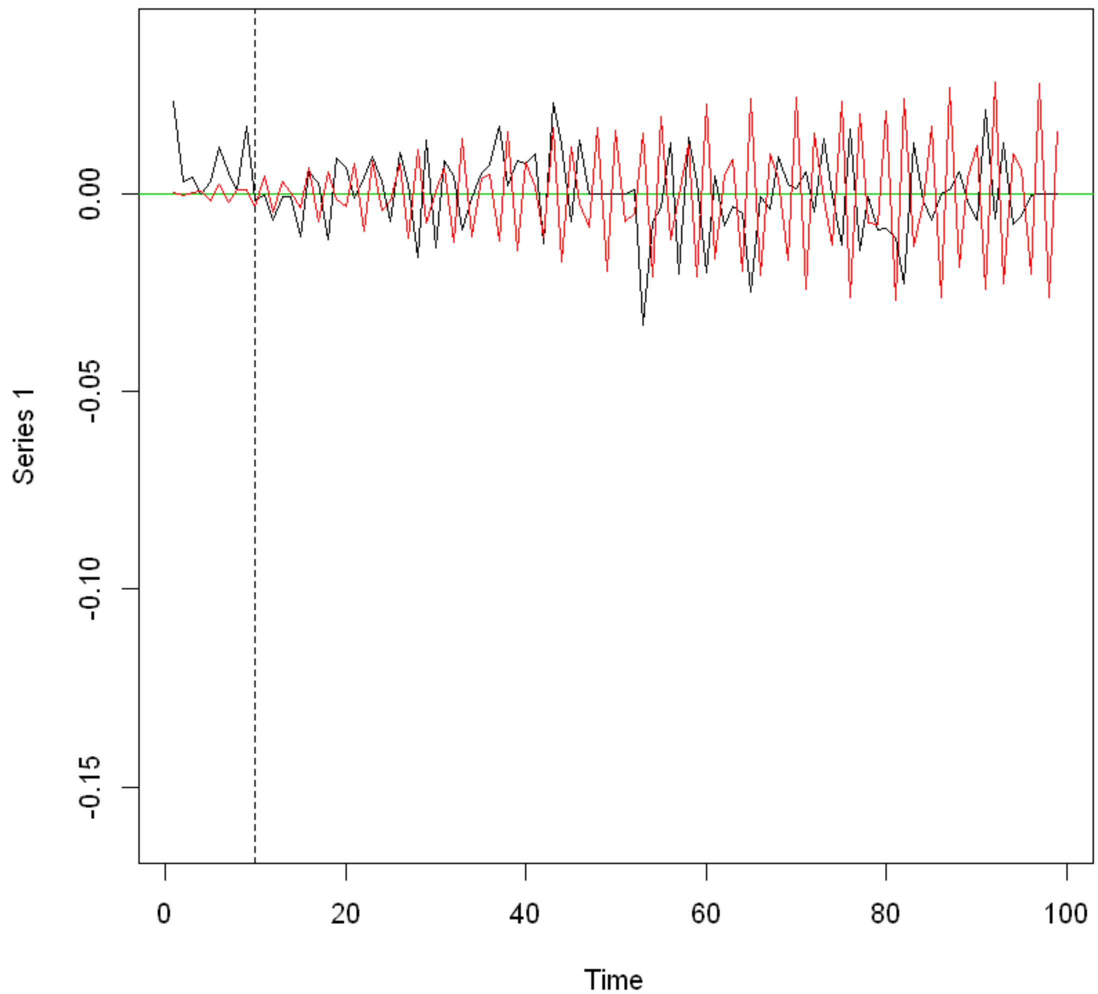
ダミー変数をPで変換して回帰するが、トレンド成分か季節成分以外の周波数において推定する。

In [7]:
```
#Additive Outlier
D.ao <- rep(0,n+1)
D.ao[82] <- 1
parm(2,1)
plot.ts(D.ao)
z.ao <- Pmat %*% diff(D.ao)
plot.ts(z.ao)
abline(h=0,col=2)
```



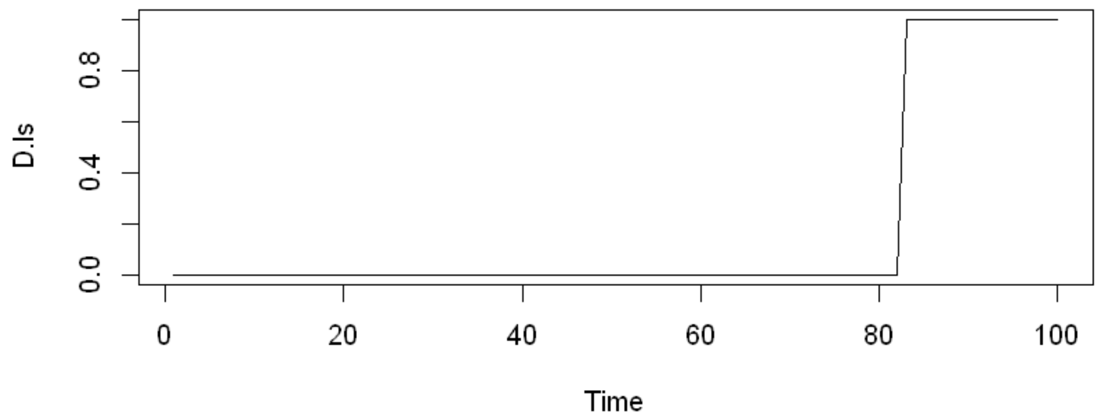


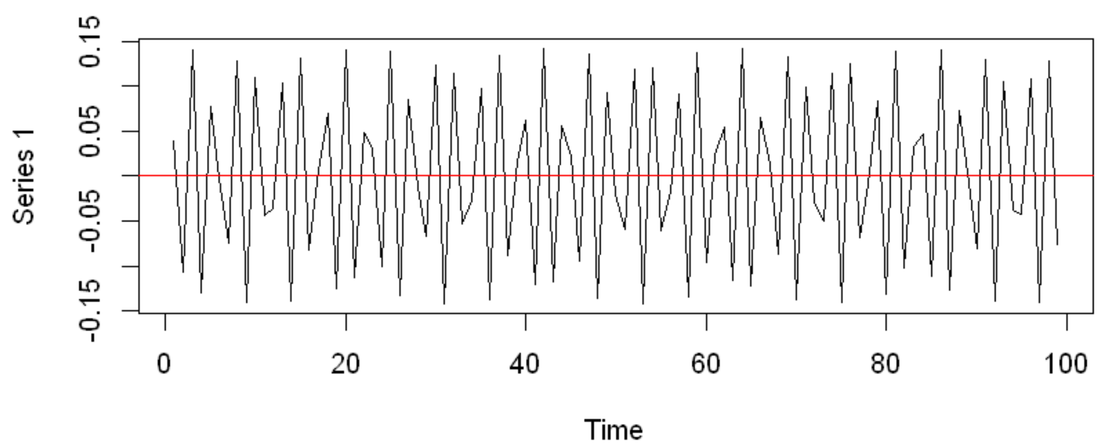
In [10]:
```
plot.ts(z,type="n")
abline(h=0,col=3)
lines(zAD)
lines(z.ao/10,col=2)
abline(v=10,lty=2)
```



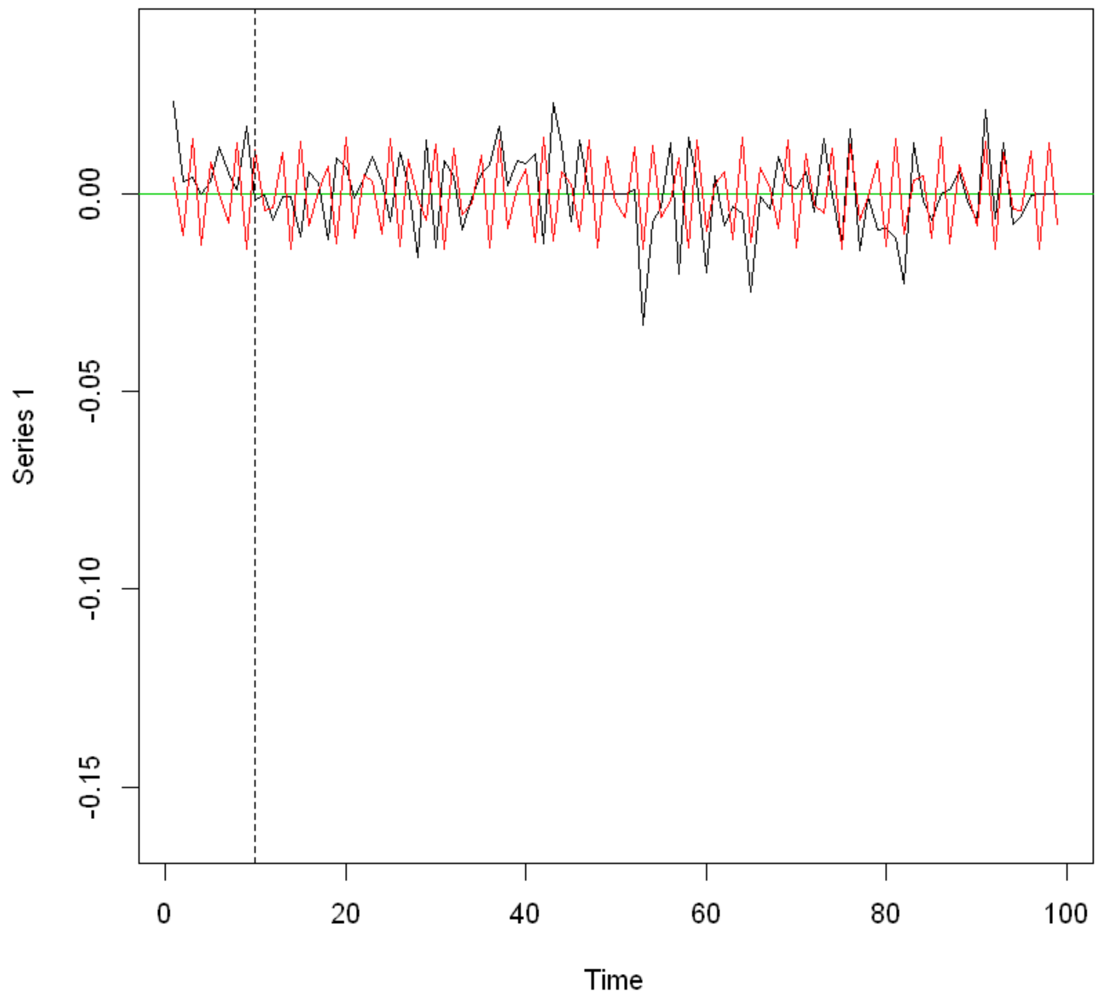
In [14]:

```
#Level Shift
D.ls <- rep(0,n+1)
D.ls[83:(n+1)] <- 1
parm(2,1)
plot.ts(D.ls)
z.ls <- Pmat %*% diff(D.ls)
plot.ts(z.ls)
abline(h=0,col=2)
```





In [15]:
```
plot.ts(z,type="n")
abline(h=0,col=3)
lines(zAD)
lines(z.ls/10,col=2)
abline(v=10,lty=2)
```

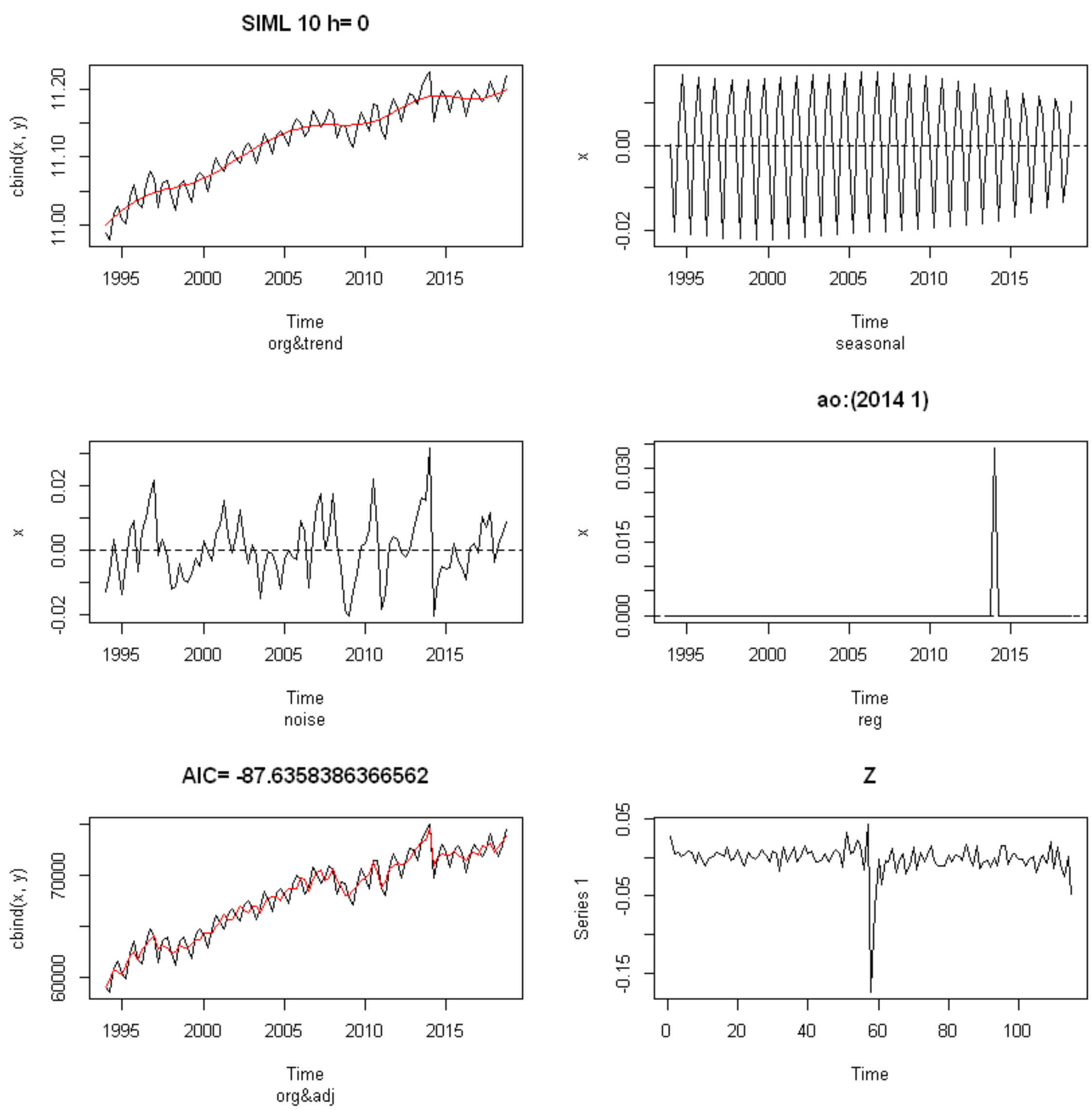


In [20]:
```
#ver.2012.3.21
DDAIC <- 4
TCRATE <- 0.8 #^(12/frequency)
len <- length
#dyn.load("decomp6r.dll")
len <- length
parm <- function(m,n) {
par(mfrow=c(m,n))
}
tsplot <- function(x,y=NULL,...)
{
 if(is.null(y)) {
 plot.ts(x,...)
}
else {
 plot.ts(cbind(x,y),plot.type="single",...,col=1:4)
}
}
"regsiml"<-
function(data, reg = NULL, m1 = 2, period = 4, sorder = 1, log = 0,pb=
2,pa=2,hh=-20)
{

    pb <- pb*(sorder > 0)
    pa <- pa*(sorder > 0)
    odata <- data
    if(log > 0) data <- log(data)
        n0 <- len(data)
    if(sorder > 0) {
        dx <- diff(data)

    dx <- c(0,rep(dx[1:period],pb),dx,rep(dx[(n0-period):(n0-1)],pa))
    x.tmp <- cumsum(dx)
    x.tmp <- x.tmp + (data[1]-x.tmp[period*pb+1])
    data <- x.tmp
    }
    n <- len(data)
    n1 <- n-1
    if(!is.null(reg)) {
        reg <- cbind(reg)
        k <- ncol(reg)
    }
    dx <- diff(data)
    if(hh < 0) {
        dy <- dx
        iddd1 <- NULL
        if(!is.null(reg)) {if(any(dimnames(reg)[[2]]=="ao")) {
                                iddd <- (cbind(reg[,dimnames(re
g)[[2]]=="ao"]) == 1)
                                if(pb > 0) {
                                for(ii in seq(pb)) {
                                    iddd1 <- rbind(iddd1,iddd[(1:perio
d)*(ii > 0),,drop=F])
                                }
                                }
                                iddd1 <- rbind(iddd1,iddd)
```

In [39]:

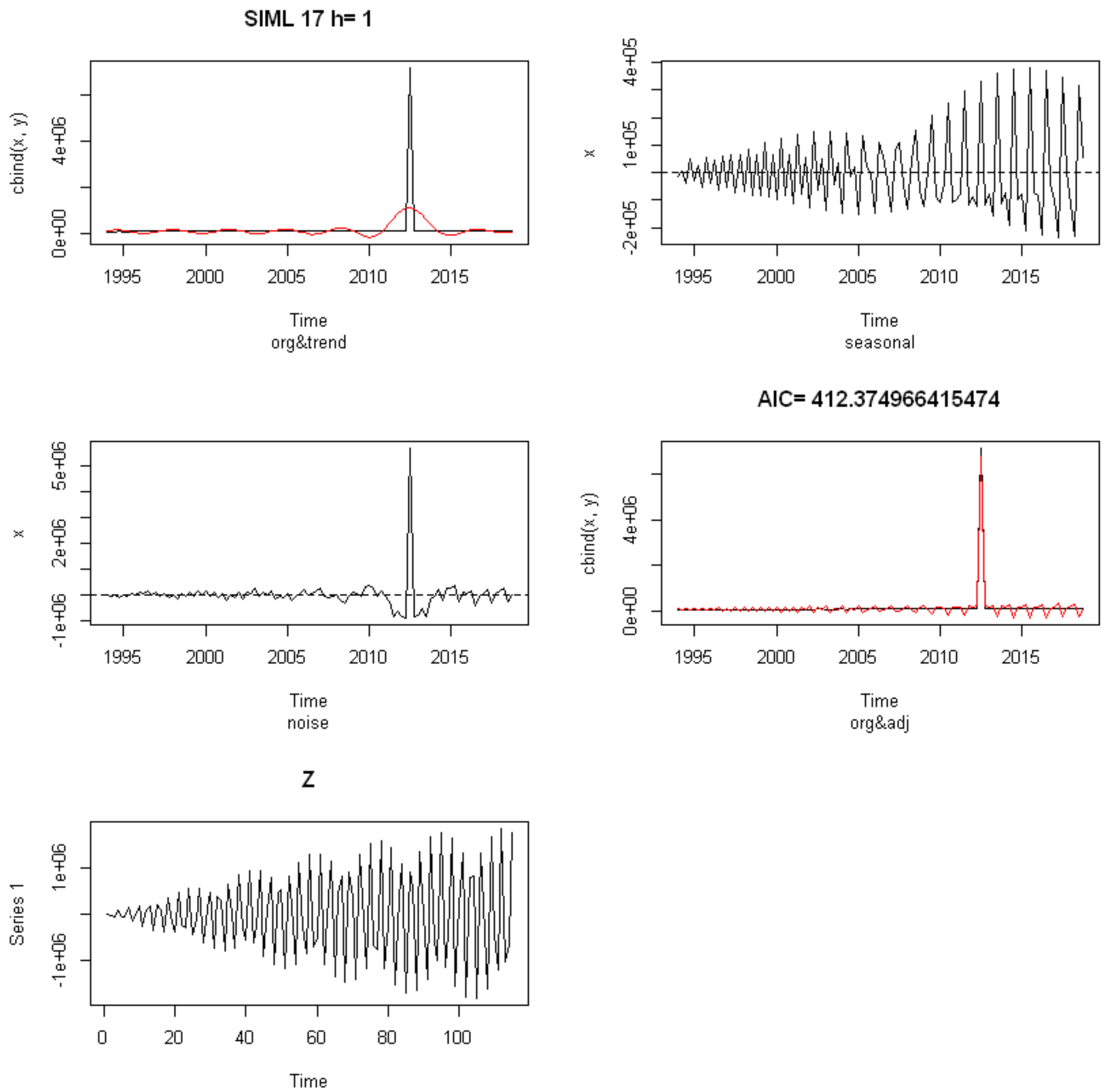
```
#length(shouhi)
#ts(shouhi,start=c(1994,1),frequency=4)
zz <- x12siml(shouhi, reg = NULL, trend = 10, ar=0, ilog = 1, frequency
= 4, start =
 c(1994, 1), iplot = T, sorder = 1,hh=0,ao=c(2014,1))
```

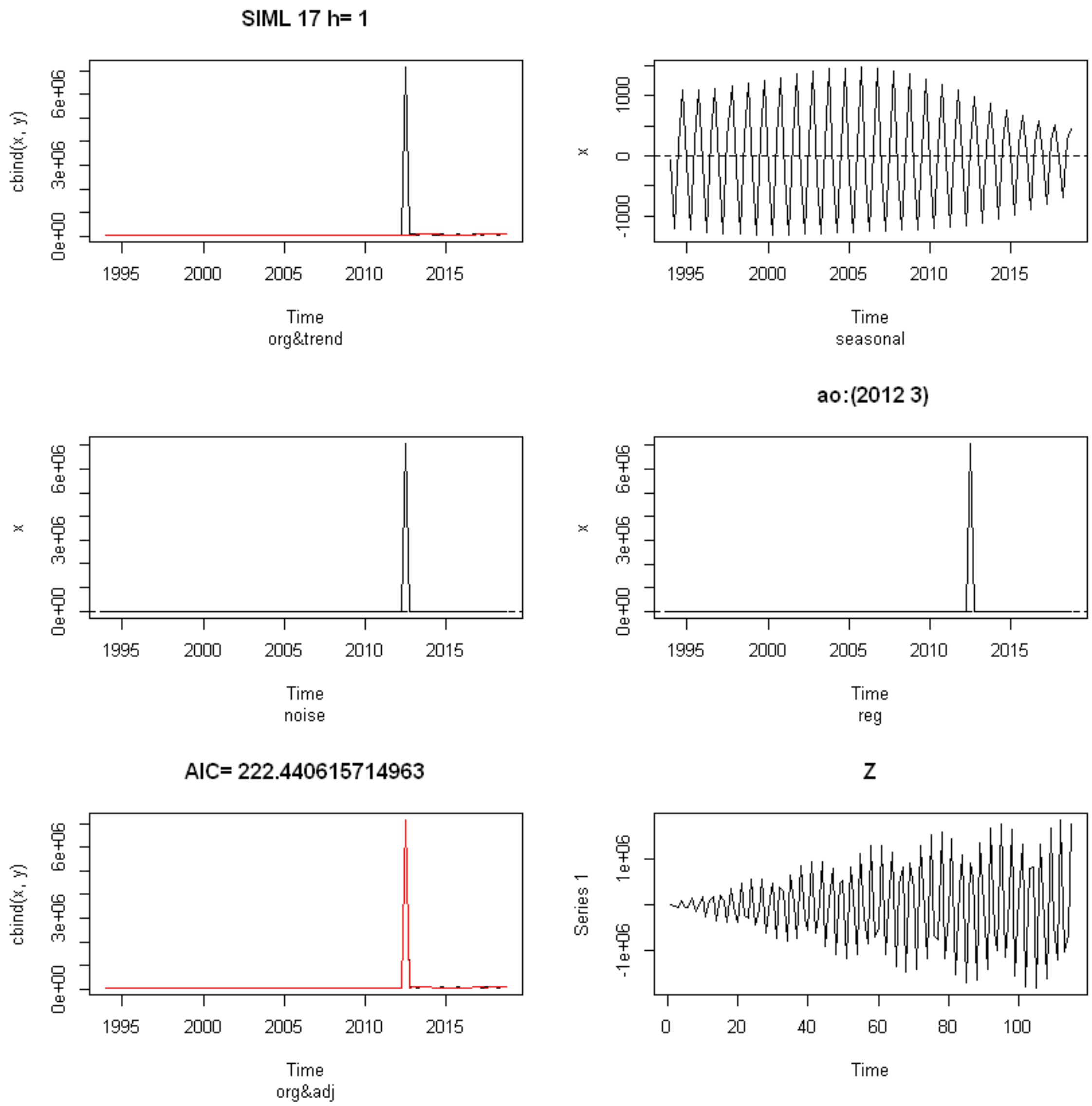
```
[1] 0.3541929 0.3930816
     Mean Sum Sq R Squared F-value  Df 1 Df 2 Pr(>F)
[1,] "0.0103"    "0.1115"  "1.1298" "1"  "9"  "0.3155"
     Mean Sum Sq R Squared F-value   Df 1 Df 2  Pr(>F)
[1,] "0.0099"    "0.1657"  "21.4431" "1"  "108" "0"
```

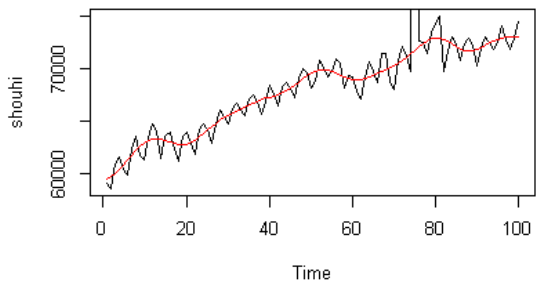


In [40]:
```
shouhi.o <- shouhi
shouhi.o[75] <- shouhi[75]*100
zz <- outlier(shouhi.o,c(2008,1,2015,4),type="ao",trend=17,ilog=0)
plot.ts(shouhi,type="n")
lines(shouhi.o)
lines(zz$trend,col=2)
```

```
[1] "make mat"
[1] "make mat"
[1] "make mat"
[1] 5243278 6336719
     Mean Sum Sq R Squared F-value  Df 1 Df 2 Pr(>F)
[1,] "166397.1"  "0.1445"  "2.7024" "1"  "16" "0.1197"
     Mean Sum Sq R Squared F-value  Df 1 Df 2  Pr(>F)
[1,] "914164.6"  "1e-04"   "0.0083" "1"  "109" "0.9276"
[1] "411.721884555715" "2008"             "1"
[1] 5243278 4842078
[1] "410.156568702371" "2008"             "2"
[1] 5243278 4764030
[1] 5243278 5171799
[1] 5243278 5765884
     Mean Sum Sq R Squared F-value  Df 1 Df 2 Pr(>F)
[1,] "175171.4"  "0.0519"  "0.8757" "1"  "16" "0.3633"
     Mean Sum Sq R Squared F-value  Df 1 Df 2  Pr(>F)
[1,] "914098.2"  "2e-04"   "0.0241" "1"  "109" "0.8768"
[1] 5243278 5245184
     Mean Sum Sq R Squared F-value Df 1 Df 2 Pr(>F)
[1,] "179900.2"  "0"       "2e-04" "1"  "16" "0.9884"
     Mean Sum Sq R Squared F-value  Df 1 Df 2  Pr(>F)
[1,] "913666.2"  "0.0012"  "0.1273" "1"  "109" "0.722"
[1] 5243278 5875371
     Mean Sum Sq R Squared F-value  Df 1 Df 2 Pr(>F)
[1,] "172549.8"  "0.0801"  "1.3924" "1"  "16" "0.2552"
     Mean Sum Sq R Squared F-value  Df 1 Df 2  Pr(>F)
[1,] "912168.8"  "0.0044"  "0.4858" "1"  "109" "0.4873"
[1] 5243278 4967688
[1] "408.454530909937" "2009"             "4"
[1] 5243278 3945295
[1] "402.189922619381" "2010"             "1"
[1] 5243278 5266835
     Mean Sum Sq R Squared F-value   Df 1 Df 2 Pr(>F)
[1,] "111472.9"  "0.6161"  "25.6726" "1"  "16" "1e-04"
     Mean Sum Sq R Squared F-value  Df 1 Df 2  Pr(>F)
[1,] "913306.9"  "0.002"   "0.2131" "1"  "109" "0.6452"
[1] "398.101638787292" "2010"             "2"
[1] 5243278 7607983
     Mean Sum Sq R Squared F-value   Df 1 Df 2 Pr(>F)
[1,] "116852.2"  "0.5781"  "21.9241" "1"  "16" "2e-04"
     Mean Sum Sq R Squared F-value  Df 1 Df 2  Pr(>F)
[1,] "911475.3"  "0.006"   "0.6525" "1"  "109" "0.421"
[1] 5243278 4504819
[1] 5243278 4714534
[1] 5243278 5106308
[1] 5243278 4000143
[1] 5243278 6285875
     Mean Sum Sq R Squared F-value Df 1 Df 2 Pr(>F)
[1,] "145489.3"  "0.346"   "8.464" "1"  "16" "0.0102"
     Mean Sum Sq R Squared F-value  Df 1 Df 2  Pr(>F)
[1,] "912557.3"  "0.0036"  "0.3926" "1"  "109" "0.5322"
[1] 5243278 5772231
     Mean Sum Sq R Squared F-value   Df 1 Df 2 Pr(>F)
[1,] "105169.8"  "0.6582"  "30.8174" "1"  "16" "0"
     Mean Sum Sq R Squared F-value Df 1 Df 2  Pr(>F)
```


SIML 17 h= 1
cbind(x, y)
Time
org&trend
x
Time
seasonal
x
Time
noise
AIC= 412.374966415474
cbind(x, y)
Time
org&adj
Z
Series 1
Time


SIML 17 h= 1
cbind(x, y)
0e+00
3e+06
6e+06
1995
2000
2005
2010
2015
Time
org&trend
x
-1000
0
1000
Time
seasonal
x
Time
noise
ao:(2012 3)
x
Time
reg
AIC= 222.440615714963
cbind(x, y)
Time
org&adj
Z
Series 1
-1e+06
1e+06
0
20
40
60
80
100
Time



In [ ]: