

ファイナンス計量分析の新展開 2010¹

国友直人² & 川崎能典³
(共編)

2010年4月

¹この報告書は文部科学省・科学研究費プロジェクト「ファイナンス計量分析の新展開と日本の金融市場」(2009年度～2011年度)が開催した研究集会で行われた連続講義、報告された研究内容をまとめたものである。

²東京大学大学院経済学研究科

³統計数理研究所

概要

この研究報告書は2009年度に開催した研究会「ファイナンス計量分析の新展開と日本の金融市場」での特別講義、研究報告・議論などの一部分をまとめたものである。2010年3月29日-30日に開催した同研究会では、関係者の熱心な協力もあり、特にこれまであまり日本語では利用可能でない金融分野における統計的手法の様々な側面について議論された。西山陽一氏（統計数理研究所）による「点過程における統計分析の基礎」に関する特別講義、が計画されたが、研究集会では計量ファイナンス分野において発展がめざましい「信用リスクの理論と応用」、「高頻度金融データの理論と応用」などを特に重視した。これらの話題は直接に研究会に参加した関係者以外の研究者、大学院生、業界関係者にとっても有用な情報を含んだものである、と研究会幹事であった国友直人（東京大学経済学研究科）及び川崎能典（統計数理研究所）は判断したので記録にとどめることとした。この報告書に収録した内容は、これまで様々な応用の場所で断片的にしか利用可能でなかった確率論、統計学的議論が多く含まれているが、近年ではいずれの話題も計量ファイナンスなどと呼ばれている、金融に係わる統計的リスク管理の分野で重要な役割を演じることが多い。

統計学・統計科学分野の関係者、金融分野において実際に応用されている確率論的方法や統計的方法の理解と発展に関心がある研究者、大学院生、業界の研究開発に携わる実務家にとって、ここに収録された内容が貴重な資料として勉学の参考になれば幸いである。この報告書が特に今後の研究の一助となることを期待したい。

2010年4月
編者

目次

第 I 部：連続講義録

「点過程における統計分析の基礎」 西山陽一

第 II 部：研究報告会 (コンファレンス)

< 統計的リスク管理 >

1. 「歪みリスク尺度の応用」 塚原英敦
2. 「Weighted Nadaraya-Watson estimation of conditional expected shortfall」 加藤健吾

< 信用リスクの計量 >

3. 「t 分布 2 ファクターモデルを用いた中小企業 CLO のデフォルト依存関係の分析」 吉規寿郎・中川秀敏
4. 「粒子フィルタリングと社債・信用リスク分析」 三崎広海
5. 「Self-exciting 性をもつイベント発生強度モデルによる社債ポートフォリオのリスク解析」 山中卓・杉原正顕・中川秀敏

< マイクロ市場の計量-I >

6. 「株式市場における終値形成について」 太田亘
7. 「高頻度データ分析：取引時間と価格の相互依存性について」 林高樹

< マイクロ市場の計量-II >

8. 「高頻度金融データの分離情報最尤推定の頑健性」 国友直人・佐藤整尚
9. 「高頻度データの日内季節性調整について」 川崎能典
10. 「バイアス補正実現分散推定量の構築」 大屋幸輔

第I部：連続講義録

「点過程における統計分析の基礎」 西山陽一

点過程による統計分析の基礎
研究集会「桜の季節に計量ファイナンス2010」
における講義資料

西山陽一
(統計数理研究所)
nisiyama@ism.ac.jp

2010年3月23日版

第1章 最も簡単な「確率過程の統計解析」 入門

1.1 条件付き期待値

(Ω, \mathcal{F}, P) を確率空間とする。 \mathcal{F} の部分 σ -加法族 \mathcal{G} , すなわち

$$\{\phi, \Omega\} \subset \mathcal{G} \subset \mathcal{F}$$

が与えられたとする。

条件付き期待値について、まずは直感的に説明しよう。条件付き期待値とは、 \mathcal{F} -可測な実数値確率変数 X が与えられたときに、それをより貧弱な \mathcal{G} に関して可測な実数値確率変数 $E[X|\mathcal{G}]$ で近似しようという操作のことである。特に \mathcal{G} が自明な σ -加法族 $\{\phi, \Omega\}$ である場合には、それは普通の期待値 $E[X]$ という定数で近似しているということに過ぎない。条件付き期待値とは、少し情報をもっている場合にはそれを使ったランダムな近似にしよう…ということをやっているのである。

ではきちんと定義しよう。

定理 1.1.1 X が \mathcal{F} -可測な実数値可積分確率変数であるとき、ある \mathcal{G} -可測な実数値可積分確率変数 $E[X|\mathcal{G}]$ であって

$$\int_G E[X|\mathcal{G}] dP = \int_G X dP \quad \forall G \in \mathcal{G}$$

をみたすものが *a.s.* の意味で一意的に存在する。この $E[X|\mathcal{G}]$ のことを \mathcal{G} が与えられたもとでの X の条件付き期待値とよぶ。

- 特に $\mathcal{G} = \{\phi, \Omega\}$ であるとき $E[X|\mathcal{G}] = E[X]$ である。
- 特に $\mathcal{G} = \mathcal{F}$ であるとき $E[X|\mathcal{G}] = X$ である。
- 一般に中間的に $\{\phi, \Omega\} \subset \mathcal{G} \subset \mathcal{F}$ であるときには、 \mathcal{F} -可測であった X を \mathcal{G} -可測でよりシンプルな確率変数に近似したものが $E[X|\mathcal{G}]$ であると解釈すればよ

い. \mathcal{G} が貧弱な σ -加法族であるときには $E[X|\mathcal{G}]$ は階段関数に近いようなシンプルな形をしており, \mathcal{G} が豊富な σ -加法族であるときには $E[X|\mathcal{G}]$ は情報をたくさんもった複雑な形をしている.

このことをもう少し詳しく説明することを試みよう. 読者は「 σ -加法族 \mathcal{G} は『情報』である」といった説明を受けたご経験があると思われるが, その真意は, 「任意の $G \in \mathcal{G}$ に対し $\omega \in G$ であるか否かが観測者にわかっている」ということであると解釈されたい.

さて, 例えば $\mathcal{G} = \{\phi, A, A^c, \Omega\}$ である場合に, A および A^c のことを便宜的に根元事象とよぶことにしよう. もう少し一般に, $\Omega = A_1 \cup \dots \cup A_p$ という非交和分割があり, $\mathcal{G} = \{\phi, A_1, \dots, A_p \text{ のいくつかの和の形で表される集合}\}$ であるとき, 空集合を除いてより小さな分割が不可能な集合 (すなわち A_1, \dots, A_p) のことを根元事象とよぶことにしよう. このとき, 「可測」という概念の定義を思い出すと, \mathcal{G} -可測関数というのは各々の根元事象の上で定数になっている関数のことである. (なぜなら, Y が \mathcal{G} -可測になるためには任意の $a \in \mathbb{R}$ に対して $\{\omega; Y(\omega) \geq a\} \in \mathcal{G}$ とならねばならないが, ひとつの A_i の上で $Y(\cdot)$ が異なる値 y_1, y_2 をとる場合には $a = (y_1 + y_2)/2$ に対して上記の条件が成り立たないからである.) さて, 条件付き期待値 $E[X|\mathcal{G}]$ というのは \mathcal{G} -可測でないとならないので, 各々の根元事象の上で定数になっていなければならない. しかも, 「期待値の意味で」 X を「近似」していなければならない. つまり, 条件付き期待値とは

$$E[X|\mathcal{G}](\omega) = x_i, \quad \forall \omega \in A_i, \quad i = 1, \dots, p$$

の形をしていて

$$\sum_{i=1}^p x_i P(G \cap A_i) = \sum_{i=1}^p \int_{G \cap A_i} X(\tilde{\omega}) P(d\tilde{\omega})$$

が全ての $G \in \mathcal{G}$ について成り立つようにする値 x_1, \dots, x_p から定められる確率変数のことである. 特に $G = A_i$ の場合も成り立たなければならないので,

$$x_i P(A_i) = \int_{A_i} X(\tilde{\omega}) P(d\tilde{\omega})$$

とならねばならず, $P(A_i) > 0$ の場合には

$$x_i = \frac{\int_{A_i} X(\tilde{\omega}) P(d\tilde{\omega})}{P(A_i)}$$

である。 $P(A_i) = 0$ の場合には x_i の値ははっきり定まらないが、これが「a.s. の意味で一意的に存在」という断り書きがつくゆえんである。結局、

$$E[X|G](\omega) = \begin{cases} \frac{\int_{A_i} X(\bar{\omega})P(d\bar{\omega})}{P(A_i)}, & \forall \omega \in A_i, \quad P(A_i) > 0 \text{ のとき}, \quad i = 1, \dots, p \\ \text{任意の定数 } x_i, & \forall \omega \in A_i, \quad P(A_i) = 0 \text{ のとき}, \end{cases}$$

というのが条件付き期待値の正体である。これは各 A_i 上では定数になっているような階段関数であるが、 G がきめの細かい σ -加法族である場合には $X(\omega)$ の比較的滑らかな（詳しい）近似になる。ランダムに ω をピックアップするとき、与えられた情報が G であると仮定された場合には、観測者は $X(\omega)$ の値ははっきり知らないが、少なくとも ω がどの A_i に入っているかは知っているのだから、その情報に基づく $X(\omega)$ の近似値 x_i を計算できる。これを条件付き期待値と呼んで $E[X|G](\omega)$ と書くのである。直観的な説明が可能な例として有限事象の場合を挙げたが、一般の σ -加法族に関する条件付き期待値の定義も、この直観の延長線上にあるとご理解いただきたい。

1.2 最も簡単な「計数過程の統計解析」入門

1.2.1 強度の導入まで

計数過程の統計的モデリングは、実際には「強度」のモデリングである。本小節の最終目標は、計数過程の「強度」を定義することである。

T を固定された正の定数とする。確率過程 $X = \{X_t; t \in [0, T]\}$ とは、ある確率空間 (Ω, \mathcal{F}, P) 上で定義された実数値（ときには \mathbb{R}^d -値）確率変数 X_t たちが $t \in [0, T]$ によって添え字づけられて集まったものである。（本書の後の方では $[0, T]$ の代わりに $[0, \infty)$ や $[0, \infty]$ をも考える立場をとるが、今のところ有限区間としておこう。）さて、セットアップがこれだけでは何も面白いことはできないので、徐々に縛りを入れて面白い具体形にしていく。最初の縛りは適合性である。 \mathcal{F} の部分 σ -加法族の増大列 $(\mathcal{F}_t)_{t \in [0, T]}$ が導入されているものとし、「各 t に対し X_t が \mathcal{F}_t -可測」であるときに、確率過程 X は $(\mathcal{F}_t)_{t \in [0, T]}$ に適合しているという。本章を通じ、以下で現れる確率過程は全て適合過程であるとする。

確率過程 X が càdlàg であるとは、殆ど全ての ω に対し、実現 $t \mapsto X_t(\omega)$ が右連続かつ左極限をもつときにいう。確率過程 M がマルチンゲールであるとは、それが càdlàg 過程であって

$$E[M_t | \mathcal{F}_s] = M_s \quad \text{a.s. for } s \leq t$$

が成り立つときにいう。

さて、いよいよお目当ての計数過程を登場させよう。確率過程 $N = \{N_t; t \in [0, T]\}$ が計数過程であるとは、 $N_0 = 0$ かつ $t \rightsquigarrow N_t$ がジャンプが 1 の単調増加 càdlàg 過程であるときにいう。本節を通じ、 $E[N_T] < \infty$ を仮定する。このとき、確率 1 で有限個であるようなランダム時刻の列 $0 < \tau_1 < \dots < \tau_{N_T} \leq T$ が存在して

$$N_t = \sum_k 1\{\tau_k \leq t\}$$

という表現ができる。

さて、ドゥーブ-メイヤー分解定理によると、ある単調増加予測可能過程 A が一意的に存在して $N - A$ をマルチンゲールすることができる。つまり

$$\begin{aligned} N_t &= A_t + M_t \\ \text{データ} &= \text{トレンド} + \text{ノイズ} \end{aligned}$$

という分解が一意的にできる。一見するとこの事実のどこがありがたいのかわからないかもしれないが、その真価は分解の一意性にある。つまり、トレンド + マルチンゲールへの分解自体は無数に存在する。例えば安直に

$$\begin{aligned} N_t &= N_t + 0 \\ \text{データ} &= \text{トレンド} + \text{ノイズ} \end{aligned}$$

という分解もアリのような気がするかもしれない。ドゥーブ-メイヤー分解定理の真価は、トレンドの項に「予測可能かつ有界変動」であることを仮定した瞬間、このような安直な方法が排除されて分解が一意になる点にある（ちなみに N は予測可能ではない）。つまり、同定理は「モデリングがびしっと出来るぞ」と言っているのである。なお我々はまだ「予測可能」の定義をしていないが、とりあえず「適合かつ左連続ならば予測可能」とのみご記憶いただきたい。また、上の例の場合は N が単調増加過程を扱っているので A も自動的に単調増加であり、従って有界変動である。なお、上の A のことをカンペンセイターというが、モデリングするときには通常それが t に関し絶対連続であることを仮定して

$$A_t = \int_0^t \lambda_s ds$$

の形で表して、この λ のことを強度と呼んで、これに関してモデリングを行う。そうしてできたモデルにおいて、例えば最尤推定量 $\hat{\theta}_n$ のようなものを考えると、不思議

なことに、ある予測可能過程 $\{H_s; s \in [0, T]\}$ が存在して

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \int_0^T H_s(dN_s - \lambda_s ds) + \text{誤差項}$$

のように書けることが多い。実は右辺第1項は確率積分と呼ばれ、それ自身ひとつのマルチンゲールの終点値なのであるが、これに対する中心極限定理を経由して最尤推定量の漸近正規性は証明される。そういうわけで、確率積分を学習すること、およびそれに対する中心極限定理をマスターすることは重要なのである。なお、強度の統計的モデリングの例については1.5.1節を参照されたい。論理構造をクリアにするために引き続き理論中心の記述を続けるが、読者におかれては一休みして1.5.1節を先に読まれ、強度のイメージを養ってから再度ここに戻ってきて頂きたい。

1.2.2 二次変分と確率積分

二次変分とは、「条件付き分散」のようなものである。また、確率積分は、平均がゼロの独立確率変数の和を一般化したようなものである。平均がゼロの独立確率変数の和に関しては中心極限定理が

「分散の収束」 + 「リンデベルグ条件」

のもとで成り立つことはよくご存知と思われるが、これを確率積分に一般化する際には、条件は

「二次変分の収束」 + 「リンデベルグ条件」

に置き換えられる。そういうわけで、二次変分の導入は中心極限定理の理解に不可欠である。

いったん計数過程から離れて一般のマルチンゲールの文脈に戻ろう。 M が2乗可積分マルチンゲールであるとは、それが

$$\sup_{t \in [0, T]} E[M_t^2] < \infty$$

を満たすマルチンゲールであるときにいう。このとき、再びドゥーブ-メイヤー分解定理により、 $M^2 - \langle M \rangle$ がマルチンゲールとなるような単調増加予測可能過程 $\langle M \rangle$ が一意的に存在する。これを二次変分と呼ぶ。二次変分は M^2 をより貧弱な可測性をもつもので近似したものである。すなわち、究極の貧弱な可測性への近似は単なる期待値

$E[M_t^2]$ であるが、そこまではせず、それより少し手前の「条件付き期待値」で止めたものが $\langle M \rangle$ である。

次に、予測可能過程 H が与えられたときに、

$$t \rightsquigarrow \int_0^t H_s dM_s$$

という確率過程を次のような雰囲気 で定義する (国田-渡辺の理論)。固定された時間区間 $[0, T]$ の刻み $0 = t_0^n < t_1^n < \dots < t_n^n = T$ をとって、任意の $t \in [0, T]$ に対し

$$\sum_{k: t_{k-1}^n \leq t} H_{t_{k-1}^n} (M_{t_k^n} - M_{t_{k-1}^n})$$

をとりあえず計算し、刻みをどんどん細かくとったときの極限を $\int_0^t H_s dM_s$ と「定義」する。実はこのような定義は

$$E \left[\int_0^T |H_s|^2 d\langle M \rangle_s \right] < \infty$$

という仮定があれば厳密に well-defined である。できあがったものはそれ自身あたらしい 2 乗可積分マルチンゲールになり、その二次変分は

$$\left\langle \int_0^\cdot H_s dM_s \right\rangle_t = \int_0^t H_s^2 d\langle M \rangle_s$$

となる。ちなみに、確率積分 $\int_0^t H_s dM_s$ は、特に $s \rightsquigarrow M_s$ が有界変動である場合には、実現 ω を固定するごとの通常のルベーク-スティルチェス積分に一致する (計数過程の場合はそうである)。一方 $s \rightsquigarrow M_s$ が連続である場合には $t \rightsquigarrow \int_0^t H_s dM_s$ も連続になる。

では話を計数過程に戻そう。以下、強度の存在を仮定する (つまり $A_t = \int_0^t \lambda_s ds$)。 $N_t - A_t$ はマルチンゲールであったが、実は $E[N_T] = E[A_T] < \infty$ を仮定しただけで 2 乗可積分マルチンゲールとなり、その二次変分は

$$\langle N - A \rangle_t = A_t = \int_0^t \lambda_s ds$$

となる。よって、確率積分に関していえば

$$\left\langle \int_0^\cdot H_s d(N_s - A_s) \right\rangle_t = \int_0^t H_s^2 \lambda_s ds$$

となる。

1.2.3 マルチンゲール中心極限定理

この小節では、確率空間の列 $(\Omega^n, \mathcal{F}^n, (\mathcal{F}_t^n)_{t \in [0, T_n]}, P^n)$ が与えられたとし、マルチンゲールの列 $M^n = \{M_t^n; t \in [0, T_n]\}$ を考え、それに対して中心極限定理が成立するための条件を紹介しよう。結果だけ紹介することに満足すれば、話はとても単純である。

定理 1.2.1 原点を出発する 2 乗可積分マルチンゲールの列 $M^n = \{M_t^n; t \in [0, T_n]\}$ に対し

(i) $(M^n)_{T_n} \rightarrow^p C$ (極限は定数)

(ii) リンデベルグ条件

が満たされるならば

$$M_{T_n}^n \rightarrow^d \mathcal{N}(0, C)$$

が成り立つ。

リンデベルグ条件は確率過程のジャンプに関する仮定であるから、特に連続マルチンゲールの場合には自動的に満たされている。よって上記の定理においては (i) のみチェックすればよい。計数過程の場合には次のような形になる。

定理 1.2.2 計数過程の列 $\{N_t^n; t \in [0, T_n]\}$ 、その強度の列 $\{\lambda_s^n; s \in [0, T_n]\}$ 、および予測可能過程の列 $\{H_s^n; s \in [0, T_n]\}$ であって

$$E^n \left[\int_0^{T_n} |H_s^n|^2 \lambda_s^n ds \right] < \infty$$

を満たすものが与えられたとする。もしも

(i) $\int_0^{T_n} |H_s^n|^2 \lambda_s^n ds \rightarrow^p C$ (極限は定数)

(ii) $\int_0^{T_n} |H_s^n|^2 1_{\{|H_s^n| > \varepsilon\}} \lambda_s^n ds \rightarrow^p 0, \quad \forall \varepsilon > 0$

が満たされるならば

$$\int_0^{T_n} H_s^n (dN_s^n - \lambda_s^n ds) \rightarrow^d \mathcal{N}(0, C)$$

が成り立つ。

1.2.4 最尤推定量の漸近正規性

この小節では、計数過程の強度のパラメトリックモデルの最もシンプルなものにおいて、未知パラメータの最尤推定量の漸近正規性の証明をしよう。もちろん、これはより複雑なモデルへ向けてのイラストレーションに過ぎない。

$N = \{N_t; t \in [0, \infty)\}$ は (n に依存しない) ひとつの計数過程であるとし, これが時間区間 $[0, T_n]$ (ただし $T_n \rightarrow \infty$) において観測されているとする. 強度が

$$\lambda_s = \alpha(Z_s; \theta)$$

が与えられるパラメトリックモデルを考えよう. ただし Z は可測空間 $(\mathcal{X}, \mathcal{A})$ に値をとる確率過程であってエルゴード性をもつものとする. つまり, 真値 θ_0 のもとで, ある確率測度 $\mu = \mu_{\theta_0}$ が存在して任意の μ -可積分関数 f に対し

$$\frac{1}{T_n} \int_0^{T_n} f(Z_s) ds \xrightarrow{p} \int_{\mathcal{X}} f(z) \mu(dz)$$

が成立することを仮定する.

観測 $\{N_t, Z_t; t \in [0, T_n]\}$ に基づく対数尤度は

$$l_n(\theta) = \int_0^{T_n} \log \alpha(Z_s; \theta) dN_s - \int_0^{T_n} \alpha(Z_s; \theta) ds$$

となることが知られている. 以下, 簡単のため θ は 1 次元であるとし, θ に関する 1 回微分を ', 2 回微分を '' などと表すと,

$$l'_n(\theta) = \int_0^{T_n} \frac{\alpha'(Z_s; \theta)}{\alpha(Z_s; \theta)} dN_s - \int_0^{T_n} \alpha'(Z_s; \theta) ds$$

となる. 最尤推定量は $l'_n(\hat{\theta}_n) = 0$ を満たす $\hat{\theta}_n$ として定義されるが, そのこととテイラー展開により

$$(1.1) \quad 0 = \frac{1}{\sqrt{T_n}} l'_n(\hat{\theta}_n) = \frac{1}{\sqrt{T_n}} l'_n(\theta_0) + \frac{1}{T_n} l''_n(\tilde{\theta}_n) \sqrt{T_n} (\hat{\theta}_n - \theta_0)$$

という展開式を得る. ただし $\tilde{\theta}_n$ は $\hat{\theta}_n$ と θ_0 を結ぶ線分上の点である.

ここで極めて重要な事実が二つある. ひとつめは, 右辺第 1 項, すなわち $l'_n(\theta)$ に $\theta = \theta_0$ を代入したものはマルチンゲールの終点値となることである. 実際,

$$\begin{aligned} l'_n(\theta_0) &= \int_0^{T_n} \frac{\alpha'(Z_s; \theta_0)}{\alpha(Z_s; \theta_0)} dN_s - \int_0^{T_n} \alpha'(Z_s; \theta_0) ds \\ &= \int_0^{T_n} \frac{\alpha'(Z_s; \theta_0)}{\alpha(Z_s; \theta_0)} (dN_s - \alpha(Z_s; \theta_0) ds). \end{aligned}$$

したがって (1.1) の右辺第 1 項 $\frac{1}{\sqrt{T_n}} l'_n(\theta_0)$ にマルチンゲール中心極限定理が適用できて

$$\frac{1}{\sqrt{T_n}} \int_0^{T_n} \frac{\alpha'(Z_s; \theta_0)}{\alpha(Z_s; \theta_0)} (dN_s - \alpha(Z_s; \theta_0) ds) \xrightarrow{d} \mathcal{N}(0, I(\theta_0)).$$

実際、二次変分は

$$\begin{aligned}
 & \left\langle \frac{1}{\sqrt{T_n}} \int_0^{T_n} \frac{\alpha'(Z_s; \theta_0)}{\alpha(Z_s; \theta_0)} (dN_s - \alpha(Z_s; \theta_0) ds) \right\rangle_{T_n} \\
 &= \frac{1}{T_n} \int_0^{T_n} \left| \frac{\alpha'(Z_s; \theta_0)}{\alpha(Z_s; \theta_0)} \right|^2 \alpha(Z_s; \theta_0) ds \\
 &\xrightarrow{P} \int_{\mathcal{X}} \frac{|\alpha'(z; \theta_0)|^2}{\alpha(z; \theta_0)} \mu(dz) \\
 &=: I(\theta_0).
 \end{aligned}$$

リンデベルグ条件をチェックするのも容易である。重要な事実の二つ目は、(1.1)の右辺第2項の係数 $\frac{1}{T_n} l'_n(\hat{\theta}_n)$ が、一様対数の法則も用いることにより、 $-I(\theta_0)$ に確率収束することである。この事実の証明は退屈な印象を与えるかもしれないので今は省略する。これら二つの事実を併せると、(1.1)は

$$\frac{1}{\sqrt{T_n}} l'_n(\theta_0) - I(\theta_0) \sqrt{T_n} (\hat{\theta}_n - \theta_0) = o_P(1)$$

となり、結局

$$\begin{aligned}
 \sqrt{T_n} (\hat{\theta}_n - \theta_0) &= I(\theta_0)^{-1} \frac{1}{\sqrt{T_n}} l'_n(\theta_0) + o_P(1) \\
 &\xrightarrow{d} \mathcal{N}(0, I(\theta_0)^{-1})
 \end{aligned}$$

を得る。

1.3 最も簡単な「拡散過程の統計解析」入門

1.3.1 伊藤積分およびマルチンゲール中心極限定理

本節の最終目標は

$$X_t = X_0 + \int_0^t S(Z_s; \theta) ds + \int_0^t \sigma(Z_s) dW_s$$

の形をした、いわゆる拡散過程モデルにおいて最尤推定量の漸近正規性の証明の概略を与えることである。ただし $s \rightsquigarrow W_s$ は標準ウィナー過程であるとし、 $s \rightsquigarrow Z_s$ は予測可能過程であるとする（よく知られた例では $Z_t = X_t$ である）。当然ながらモデル

の右辺最終項の定義をはっきりさせなければならない。さらには、計数過程の場合と同様、最尤推定量 $\hat{\theta}_n$ は、多くの場合ある予測可能過程 $\{H_s; s \in [0, T]\}$ を用いて

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \int_0^T H_s dW_s + \text{誤差項}$$

という確率積分の形で表現されることもあり、いわゆる伊藤積分 $\int_0^t H_s dW_s$ を学習する必要に迫られる訳であるが、準備は 1.2.2 節で見た（より一般的な）国田-渡辺理論の紹介の部分ですでにできている。すなわち、刻み $0 = t_0^n < t_1^n < \dots < t_n^n = T$ をとって、任意の $t \in [0, T]$ に対し

$$\sum_{k: t_{k-1}^n \leq t} H_{t_{k-1}^n} (W_{t_k^n} - W_{t_{k-1}^n})$$

をとりあえず計算し、刻みをどんどん細かくとったときの極限を $\int_0^t H_s dW_s$ と「定義」する。このような定義は

$$E \left[\int_0^T H_s^2 ds \right] < \infty$$

という仮定があれば well-defined である。できあがったものはそれ自身あたらしい 2 乗可積分マルチンゲールになり、その二次変分は

$$\left\langle \int_0^{\cdot} H_s dW_s \right\rangle_t = \int_0^t H_s^2 ds$$

となる。

なお、拡散過程の具体例については 1.5.2 節を参照されたい。論理構造をクリアにするために引き続き理論中心の記述を続けるが、読者におかれては 1.5.2 節を先に読まれ、拡散過程のイメージを養ってから再度ここに戻ってきて頂きたい。

さて、本小節を終えるにあたって定理 1.2.1 の特別な場合として次が得られることに注意しておこう。

定理 1.3.1 標準ウィナー過程の列 $\{W_s^n; s \in [0, T_n]\}$ および予測可能過程の列 $\{H_s^n; s \in [0, T_n]\}$ であって

$$E^n \left[\int_0^{T_n} |H_s^n|^2 ds \right] < \infty$$

を満たすものが与えられたとする。もしも

$$\int_0^{T_n} |H_s^n|^2 ds \rightarrow^p C \quad (\text{極限は定数})$$

が満たされるならば

$$\int_0^{T_n} H_s^n dW_s^n \rightarrow^d \mathcal{N}(0, C)$$

が成り立つ.

1.3.2 最尤推定量の漸近正規性

この小節では, 前小節で予告した最もシンプルな1次元の拡散過程モデル

$$X_t = X_0 + \int_0^t S(Z_s; \theta) ds + \int_0^t \sigma(Z_s) dW_s$$

において, 未知パラメータ θ の最尤推定量の漸近正規性の証明をしよう. ただし $s \rightsquigarrow W_s$ は標準ウィナー過程であるとし, $s \rightsquigarrow Z_s$ は可測空間 $(\mathcal{X}, \mathcal{A})$ に値をとる予測可能過程であってエルゴード性をもつものとする. つまり, 真値 θ_0 のもとで, ある確率測度 $\mu = \mu_{\theta_0}$ が存在して任意の μ -可積分関数 f に対し

$$\frac{1}{T_n} \int_0^{T_n} f(Z_s) ds \rightarrow^p \int_{\mathcal{X}} f(z) \mu(dz)$$

が成立することを仮定する. (よく知られた例では $Z_t = X_t$ である.) X, Z が時間区間 $[0, T_n]$ (ただし $T_n \rightarrow \infty$) において観測されたときの対数尤度は

$$l_n(\theta) = \int_0^{T_n} \frac{S(Z_s; \theta)}{\sigma(Z_s)^2} dX_s - \frac{1}{2} \int_0^{T_n} \frac{S(Z_s; \theta)^2}{\sigma(Z_s)^2} ds$$

となることが知られている. 以下, 簡単のため θ は1次元であるとし, θ に関する1回微分を', 2回微分を''などと表すと,

$$l_n'(\theta) = \int_0^{T_n} \frac{S'(Z_s; \theta)}{\sigma(Z_s)^2} dX_s - \int_0^{T_n} \frac{S'(Z_s; \theta) S(Z_s; \theta)}{\sigma(Z_s)^2} ds$$

となる. 計数過程のときと同様, 最尤推定量は $l_n'(\hat{\theta}_n) = 0$ を満たす $\hat{\theta}_n$ として定義されるが, そのこととテイラー展開により

$$(1.2) \quad 0 = \frac{1}{\sqrt{T_n}} l_n'(\hat{\theta}_n) = \frac{1}{\sqrt{T_n}} l_n'(\theta_0) + \frac{1}{T_n} l_n''(\tilde{\theta}_n) \sqrt{T_n} (\hat{\theta}_n - \theta_0)$$

という展開式を得る. ただし $\tilde{\theta}_n$ は $\hat{\theta}_n$ と θ_0 を結ぶ線分上の点である.

ここから先のストーリーは計数過程の場合と全く同じである。すなわち、極めて重要な二つの事実に注目しよう。ひとつめは、右辺第1項、すなわち $l'_n(\theta)$ に $\theta = \theta_0$ を代入したものはマルチンゲールの終点値となることである。実際、

$$\begin{aligned} l'_n(\theta) &= \int_0^{T_n} \frac{S'(Z_s; \theta_0)}{\sigma(Z_s)^2} dX_s - \int_0^{T_n} S'(Z_s; \theta_0) S(Z_s; \theta_0) ds \\ &= \int_0^{T_n} \frac{S'(Z_s; \theta_0)}{\sigma(Z_s)} dW_s. \end{aligned}$$

したがって (1.2) の右辺第1項 $\frac{1}{\sqrt{T_n}} l'_n(\theta_0)$ にマルチンゲール中心極限定理が適用できて

$$\frac{1}{\sqrt{T_n}} \int_0^{T_n} \frac{S'(Z_s; \theta_0)}{\sigma(Z_s)} dW_s \rightarrow^d \mathcal{N}(0, I(\theta_0)).$$

実際、二次変分は

$$\begin{aligned} &\left\langle \frac{1}{\sqrt{T_n}} \int_0^{T_n} \frac{S'(Z_s; \theta_0)}{\sigma(Z_s)} dW_s \right\rangle_{T_n} \\ &= \frac{1}{T_n} \int_0^{T_n} \left| \frac{S'(Z_s; \theta_0)}{\sigma(Z_s)} \right|^2 ds \\ &\rightarrow^p \int_{\mathcal{X}} \left| \frac{S'(z; \theta_0)}{\sigma(z)} \right|^2 \mu(dz) \\ &=: I(\theta_0). \end{aligned}$$

重要な事実の二つ目は、(1.2) の右辺第2項の係数 $\frac{1}{T_n} l''_n(\hat{\theta}_n)$ が、一様対数の法則も用いることにより、 $-I(\theta_0)$ に確率収束することである。この証明は今は省略する。これら二つの事実を併せると、(1.2) は

$$\frac{1}{\sqrt{T_n}} l'_n(\theta_0) - I(\theta_0) \sqrt{T_n} (\hat{\theta}_n - \theta_0) = o_P(1)$$

となり、結局

$$\begin{aligned} \sqrt{T_n} (\hat{\theta}_n - \theta_0) &= I(\theta_0)^{-1} \frac{1}{\sqrt{T_n}} l'_n(\theta_0) + o_P(1) \\ &\rightarrow^d \mathcal{N}(0, I(\theta_0)^{-1}) \end{aligned}$$

を得る。

1.4 アプローチのまとめ

以上で見たように、多くのパラメトリックモデルにおいて、対数尤度 $l_n(\theta)$ に対して

$$0 = \frac{1}{\sqrt{n}} l'_n(\hat{\theta}_n) = \frac{1}{\sqrt{n}} l'_n(\theta_0) + \frac{1}{n} l''_n(\tilde{\theta}_n) \sqrt{n}(\hat{\theta}_n - \theta_0)$$

という展開式が成立する。右辺の第2項の係数は $-I(\theta_0)$ に確率収束する。よって右辺第1項が $\mathcal{N}(0, I(\theta_0))$ に分布収束することを示せば

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta_0) &= I(\theta_0)^{-1} \frac{1}{\sqrt{n}} l'_n(\theta_0) + o_P(1) \\ &\rightarrow^d \mathcal{N}(0, I(\theta_0)^{-1}) \end{aligned}$$

が得られる。この必勝パターンを可能にするのが $\frac{1}{\sqrt{n}} l'_n(\theta_0)$ に対するマルチンゲール中心極限定理である。

1.5 例

1.5.1 計数過程モデルの例

$0 < \tau_1 < \tau_2 < \dots$ をランダムな点列とする。

$$N_t = \sum_k 1\{\tau_k \leq t\}$$

とおき、そのカンペンセイター A_t が t に関して絶対連続であるとき、すなわち

$$A_t(\omega) = \int_0^t \lambda_s(\omega) ds$$

と書けるとき、 $s \rightsquigarrow \lambda_s$ を強度と呼ぶのであった。その例をいくつか挙げよう。

例 1.5.1 (均質ポアソン過程) $\lambda_t(\omega) \equiv \lambda$ (定数) であるとき、 N は強度 λ の均質ポアソン過程であるという。

例 1.5.2 (不均質ポアソン過程) $\lambda_t(\omega) = \alpha(t)$ (ω に依存しない関数) であるとき、不均質ポアソン過程であるという。

例 1.5.3 (one-point process) 正のランダム点が 1 個だけ発生する場合を考える： $\tau_1 \sim F(t)$. ただし F は $(0, \infty)$ 上の確率分布とする. その密度関数を $f(t)$ とし, 危険関数 (hazard function) を

$$\alpha(t) = \frac{f(t)}{1 - F(t-)}$$

と定義する. このときの強度は

$$\lambda_t(\omega) = \alpha(t)1\{t \leq \tau_1(\omega)\}$$

となる. これが強度がランダムとなる最も簡単な例である.

例 1.5.4 (更新過程) 独立同一分布な正の確率変数列 Y_1, Y_2, \dots が与えられたとする. これを用いて

$$\begin{aligned}\tau_1 &= Y_1 \\ \tau_2 &= Y_1 + Y_2 \\ \tau_3 &= Y_1 + Y_2 + Y_3\end{aligned}$$

などと定義する. この $0 < \tau_1 < \tau_2 < \dots$ によって上述のように定義される N を更新過程であるという. その強度は

$$\lambda_t(\omega) = \alpha(t - \tau_{k-1}(\omega)), \quad t \in (\tau_{k-1}(\omega), \tau_k(\omega)]$$

となる.

単純な例はこれぐらいにして, 以下でもう少し統計的に役に立つ例を述べよう.

X_1, X_2, \dots, X_n は独立に同一分布 $F(t)$ に従う正の確率変数であるとする. これらを小さいものから並べなおしたものを

$$0 < \tau_1 < \tau_2 < \dots < \tau_n$$

とおく。この τ_k ($k = 1, \dots, n$) に対して計数過程 N を上述のように定義する。その強度は

$$\begin{aligned} \lambda_t &= \alpha(t) \sum_{i=1}^n 1_{[t, \infty)}(X_i) \\ &= \alpha(t) \sum_{k=1}^n 1_{[t, \infty)}(\tau_k) \\ &= \alpha(t) \left\{ n - \sum_{k=1}^n 1_{[0, t)}(\tau_k) \right\} \\ &= \alpha(t) \{n - N_{t-}\} \\ &= \alpha(t) \{ \text{時刻 } t- \text{ に生き残っている個体数} \} \end{aligned}$$

となる。

我々の目標は分布関数 F を推定することである。この例の場合、そのためには $0 < \tau_1 < \tau_2 < \dots < \tau_n$ など持ち出さなくても X_i ($i = 1, \dots, n$) のままでよい。なぜなら、

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n 1_{[0, t)}(X_i)$$

とおくと、一様大数の法則から

$$\sup_{t \in [0, \infty)} |\hat{F}_n(t) - F(t)| \rightarrow 0 \quad \text{a.s.}$$

がいえるからである。だから、このような状況ではマルチンゲール理論を使う必要はない。では次の例はどうであろうか？

例 1.5.5 (センサーを受けたデータ) 正の確率変数列

$$\begin{aligned} X_1, X_2, \dots, X_n &\sim_{i.i.d.} F(t), \quad \text{死亡時刻} \\ C_1, C_2, \dots, C_n &\sim_{i.i.d.} G(t), \quad \text{センサリング時刻} \end{aligned}$$

を考える。これは、個体 i は時刻 X_i にある病気で死亡する（はず）であるが、時刻 C_i に、その病気とは関係ない理由で観測できなくなる、というモデルである。我々の目標は F を推定することである。 G には興味がないとする。与えられるデータは次の $\{T_i, \Delta_i\}$ ($i = 1, \dots, n$) であるとする：

$$\begin{aligned} T_i &= X_i \wedge C_i, \\ \Delta_i &= 1\{X_i \leq C_i\}. \end{aligned}$$

このモデルの場合,

$$\widehat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n 1_{[0,t]}(X_i)$$

は使えない. X_i が観測されておらず, $T_i = X_i \wedge C_i$ しか観測されていないからである. しかしマルチンゲール理論を用いると, 累積危険関数

$$\int_0^t \frac{f(s)}{1 - F(s-)} ds = \int_0^t \alpha(s) ds$$

の推定ができる.

概略は次のとおりである. 計数過程を

$$N_t = \sum_{i=1}^n 1\{T_i \leq t, \Delta_i = 1\}$$

によって定義すると, 強度は

$$(1.3) \quad \lambda_t = \alpha(t) Y_t$$

によって与えられる. ただし

$$Y_t = n - \sum_{i=1}^n 1_{[0,t)}(T_i)$$

より一般に, Y が整数値であることを忘れて, 一般の非負予測可能過程であるとしたときの (1.3) を積強度モデル (multiplicative intensity model) と呼ぶ. このとき,

$$N_t = \int_0^t dN_s \approx \int_0^t \alpha(s) Y_s ds$$

であるから, 両辺の被積分関数に $\frac{1}{Y}$ を掛けることにより,

$$\int_0^t \frac{1}{Y_s} dN_s \approx \int_0^t \alpha(s) ds$$

という近似が成り立つことが直感的にわかる. 左辺で与えられたものは Nelson-Aalen 推定量と呼ばれる. 実際, 残差

$$\int_0^t \frac{1}{Y_s} (dN_s - \alpha(s) Y_s ds)$$

は確率積分であるから, マルチンゲール理論を用いた解析が可能である.

例 1.5.6 (自己励起過程, ETAS モデル) 強度が

$$\lambda_t = \alpha + \sum_{k:\tau_k < t} \beta \phi(t - \tau_k), \quad \alpha, \beta > 0$$

によって与えられるものは自己励起過程 (*self-exciting process*) と呼ばれる. 例えば $\phi(x) = e^{-x}$ である. これを発展させた ETAS モデル (*epidemic type after shock model*)

$$\lambda_t = \alpha + \sum_{k:\tau_k < t} \frac{\kappa e^{\beta M_k}}{|t - \tau_k + c|^p}$$

は大地震が起こった後の小さな群発地震の発生機構を表すために広く応用されている. ただし M_k は時刻 τ_k に起こった地震のマグニチュードを表す.

例 1.5.7 (自己修正過程) 強度が

$$\lambda_t = \alpha \phi(\beta t - \gamma N_{t-}), \quad \alpha, \beta, \gamma > 0$$

によって与えられるものは自己修正過程 (*self-correcting process*) と呼ばれる. 非負値関数 $\phi(x)$ としては $x \rightarrow \infty$ のときに大きく, $x \rightarrow -\infty$ のときに小さくなるものが選ばれる. このモデルは大地震が起こった後は原因となるストレスが解放されたためしばらくは大地震が起こらなくなり, 長期間にわたって大地震が起こっていないとストレスがたまって危険度が増すことを記述するモデルである.

1.5.2 拡散過程モデルの例

以下, $t \rightsquigarrow W_t$ は常に標準ウィナー過程を表すとする.

例 1.5.8 (Ornstein-Uhlenbeck 過程) 確率微分方程式

$$X_t = X_0 - \int_0^t \alpha X_s ds + \beta W_t$$

の解は Ornstein-Uhlenbeck 過程と呼ばれる. これは $\alpha > 0$ ならばエルゴード的になる.

例 1.5.9 (Vasicek 過程) 確率微分方程式

$$X_t = X_0 - \int_0^t \alpha_1 (X_s - \alpha_2) X_s ds + \beta W_t$$

の解は Vasicek 過程と呼ばれる. これは $\alpha_1 > 0$ ならばエルゴード的になる.

例 1.5.10 (幾何ブラウン運動) μ, σ を定数とし,

$$X_t = \exp(\mu t + \sigma W_t)$$

という正の値をとる確率過程に注目しよう。後に説明する伊藤の公式により、これは

$$X_t = X_0 + \int_0^t \left(\mu + \frac{1}{2} \sigma^2 \right) X_s ds + \int_0^t \sigma X_s dW_s$$

という確率微分方程式の表現をもつ拡散過程であることがわかり、幾何ブラウン運動と呼ばれる。このモデルは数理ファイナンスなどで広く用いられている。

第2章 Cox 部分尤度への最短コース

2.1 準備1：Lenglart の不等式

定理 2.1.1 (Lenglart の不等式) X は原点を出発する càdlàg 適合過程であるとし, A は予測可能 *increasing* 過程であるとする. 任意の有界停止時刻 S に対し $EX_S \leq EA_S$ が成り立つことを仮定する. T は有限停止時刻であるとする. このとき, 任意の $\eta, \delta > 0$ に対し

$$P\left(\sup_{t \in [0, T]} X_t > \eta\right) \leq \frac{\delta}{\eta} + P(A_T \geq \delta)$$

が成り立つ.

特に M が原点を出発する局所二乗可積分マルチンゲールであるとき, 局所化停止時刻列 (T_n) を取ってきて $X = |M^{T_n}|^2$, $A = \langle M^{T_n}, M^{T_n} \rangle$ に対して上記の定理を適用して

$$\begin{aligned} P\left(\sup_{t \in [0, T]} |M_t^{T_n}| > \eta\right) &= P\left(\sup_{t \in [0, T]} |M_t^{T_n}|^2 > \eta^2\right) \\ &\leq \frac{\delta}{\eta^2} + P(\langle M^{T_n}, M^{T_n} \rangle_T \geq \delta) \end{aligned}$$

が成り立つ. $n \rightarrow \infty$ として

$$P\left(\sup_{t \in [0, T]} |M_t| > \eta\right) \leq \frac{\delta}{\eta^2} + P(\langle M, M \rangle_T \geq \delta)$$

を得る.

これらの不等式が著しいのは, 左辺において t に関する \sup が確率の中に入っている点である. 使い方としては「局所二乗可積分マルチンゲールが 0 に確率収束することを示したいときには, その二次変分が 0 に確率収束することを示せばよい」と覚えておけばよい. これを系の形でまとめておく.

系 2.1.2 (Lenglart の不等式の系) 各 $n \in \mathbb{N}$ に対し, M^n は原点を出発する局所二乗可積分マルチンゲールであるとし, T_n は有限停止時刻であるとする. もし $\langle M^n \rangle_{T_n} \rightarrow^p 0$ ならば $\sup_{t \in [0, T_n]} |M_t^n| \rightarrow^p 0$ である.

以下の証明では「予測可能時刻」の概念を用いる. それは停止時刻の特別な場合 (アナウンス列の存在と呼ばれる良い性質をもつ停止時刻) であるが, 詳しくは, 例えば Jacod and Shiryaev (1987) を参照されたい.

定理 2.1.1 の証明. $T_n = T \wedge n$ とするとき

$$\lim_{n \rightarrow \infty} P \left(\sup_{t \in [0, T_n]} X_t > \eta \right) = P \left(\sup_{t \in [0, T]} X_t > \eta \right)$$

かつ

$$\lim_{n \rightarrow \infty} P(A_{T_n} \geq \delta) = P(A_T \geq \delta)$$

であるから, T_n に対して不等式を示せばよい. いいかえれば, T が有界停止時刻であると仮定してよい.

$R = \inf\{s : X_s \geq \eta\}$ および $S = \inf\{s : A_s \geq \delta\}$ とおく. このとき, R は停止時刻であり, S は予測可能時刻である. $\{\sup_{s \leq T} X_s \geq \eta\} \subset \{A_T \geq \delta\} \cup \{R \leq T < S\}$ であるから,

$$P \left(\sup_{t \in [0, T]} X_t > \eta \right) \leq P(R \leq T < S) + P(A_T \geq \delta)$$

である. ここで S は予測可能時刻であるから, ある停止時刻の増大列 (S_n) が存在して $S_n < S$ a.s. on $\{S > 0\}$ かつ $\lim_n S_n = S$ a.s. が成り立つ. (これをアナウンス列とよぶ. Jacod and Shiryaev (1987) の I.2.16 参照.) よって

$$\begin{aligned} P(R \leq T < S) &\leq \lim_n P(R \leq T < S_n) \\ &\leq \lim_n P(X_{R \wedge T \wedge S_n} \geq \eta) \\ &\leq \frac{1}{\eta} \lim_n E(X_{R \wedge T \wedge S_n}) \\ &\leq \frac{1}{\eta} \lim_n E(A_{R \wedge T \wedge S_n}) \end{aligned}$$

である. 最後の不等式で T が有界停止時刻であることを使った. さらに $A_{R \wedge T \wedge S_n} \leq A_{S_n} \leq \delta$ である. 不等式が証明された. \square

2.2 準備 2 : 計数過程のマルチンゲール中心極限定理

定理 2.2.1 各 $n \in \mathbb{N}$ に対し, N^n は強度 λ^n をもつ計数過程であるとする. $H_s^{n,(i)}$, $i = 1, \dots, p$ は予測可能過程であって $t \rightsquigarrow \int_0^t |H_s^{n,(i)}|^2 \lambda_s^n ds$ が局所可積分であるものとする. p 次元の局所二乗可積分マルチンゲール $M^n = (M^{n,(1)}, \dots, M^{n,(p)})^\top$,

$$M_t^{n,(i)} = \int_0^t H_s^{n,(i)} (dN_s^n - \lambda_s^n ds)$$

を考える. T_n は有限停止時刻であるとし, $n \rightarrow \infty$ とするとき

$$\int_0^{T_n} \sum_{i=1}^p |H_s^{n,(i)}|^2 1_{\left\{ \sum_{i=1}^p |H_s^{n,(i)}| > \varepsilon \right\}} \lambda_s^n ds \xrightarrow{p} 0, \quad \forall \varepsilon > 0,$$

$$(2.1) \quad \int_0^{T_n} H_s^{n,(i)} H_s^{n,(j)} \lambda_s^n ds \xrightarrow{p} C^{(i,j)}, \quad i, j = 1, \dots, p$$

であることを仮定する. ただし $C = (C^{(i,j)})_{i,j=1,\dots,p}$ は定数行列である. このとき,

$$M_{T_n}^n \xrightarrow{d} \mathcal{N}_p(0, C)$$

が成り立つ.

証明. 次元の場合を考える. 多次元の場合はクラメール-ウォルドのデヴィアスによる. 以下, 記号

$$M_t^n = \int_0^t H_s^n (dN_s^n - \lambda_s^n ds)$$

を用い, また条件 (2.1) は $C = C^{(1,1)}$ と読まれたい.

停止時刻

$$S_n = \inf\{t \in [0, \infty) : \langle M^n \rangle_t \geq C\}$$

を導入する. このとき, 一般性を失うことなく

$$\langle M^n \rangle_{S_n} = C$$

と仮定してよい. なぜなら N^n と独立な強度が 1 のポアソン過程 $N^{(1)}$ をとり,

$$\tilde{N}_t^n = \int_0^t 1_{[0, T_n]}(s) dN_s^n + \int_0^t 1_{(T_n, \infty)} dN_s^{(1)}$$

とおくと、その強度は

$$\tilde{\lambda}_t^n = 1_{[0, T_n]}(t)\lambda_t^n + 1_{(T_n, \infty)}(t)$$

であり、さらに

$$\tilde{H}_t^n = 1_{[0, T_n]}(t)H_t^n + 1_{(T_n, \infty)}(t)$$

とおくと、対応する $\tilde{M}_t^n = \int_0^t \tilde{H}_s^n (d\tilde{N}_s^n - \tilde{\lambda}_s^n ds)$ に対して、確率1で $\langle \tilde{M}^n \rangle_t \rightarrow \infty$ as $t \rightarrow \infty$ である。 $t \sim \langle \tilde{M}^n \rangle_t$ は連続であるから、 $\langle \tilde{M}^n \rangle$ に対応する \tilde{S}_n は $\langle \tilde{M}^n \rangle_{\tilde{S}_n} = C$ を満たす。定理の仮定も結論もすべて $[0, T_n]$ 上のことからであるから、このような $\tilde{N}^n, \tilde{\lambda}^n, \tilde{H}^n$ を最初からとってきて N^n, λ^n, H^n と書けばよい。

特性関数を計算する。 $z \in \mathbb{R} \setminus \{0\}$ を固定し、

$$G_t^n = \exp\left(\frac{z^2}{2}\langle M^{n, S_n} \rangle_t + izM_t^{n, S_n}\right)$$

とおく。ただし $M_t^{n, S_n} = M_{t \wedge S_n}$ である。

$$(2.2) \quad E[G_{S_n}^n] \rightarrow 1,$$

$$(2.3) \quad M_{T_n}^n - M_{S_n}^n \xrightarrow{p} 0$$

を証明する。これらが示されれば、前者より

$$E[\exp(izM_{S_n}^n)] \rightarrow \exp\left(-\frac{z^2}{2}C\right)$$

であるから $M_{S_n}^n \xrightarrow{d} \mathcal{N}(0, C)$ が従い、後者を用いて $M_{T_n}^n \xrightarrow{d} \mathcal{N}(0, C)$ が得られる。

では (2.2) を示そう。伊藤の公式より

$$\begin{aligned} G_t^n - 1 &= \int_0^{t \wedge S_n} G_{s-}^n (\exp(izH_s^n) - 1) dM_s^n \\ &\quad + \int_0^{t \wedge S_n} G_{s-}^n \left\{ \exp(izH_s^n) - 1 - izH_s^n + \frac{|zH_s^n|^2}{2} \right\} \lambda_s^n ds \end{aligned}$$

である。右辺第1項は二乗可積分マルチンゲールである。実際、

$$\sup_{s \in [0, S_n]} |G_{s-}^n| \leq \exp\left(\frac{z^2}{2}C\right)$$

であることと、よく知られた不等式 $|e^{ix} - 1| \leq |x|$ を使って、二次変分が有界であることが従う。よって $E[G_{S_n}^n] - 1$ は右辺第2項に $t = S_n$ を代入して期待値をとったも

のに等しい. よく知られた不等式 $|e^{ix} - 1 - ix + \frac{x^2}{2}| \leq |x|^3 \wedge |x|^2$ を用いることにより, 任意の $\varepsilon > 0$ に対し, その値は以下によって上から押さえられる:

$$\begin{aligned} & E \left[\int_0^{S_n} |G_{s-}^n| \left| \exp(izH_s^n) - 1 - izH_s^n + \frac{|zH_s^n|^2}{2} \right| \lambda_s^n ds \right] \\ & \leq \exp\left(\frac{z^2}{2}C\right) \left\{ E \left[\int_0^{S_n} |zH_s^n|^3 \mathbf{1}\{|zH_s^n| \leq \varepsilon\} \lambda_s^n ds \right] \right. \\ & \quad \left. + E \left[\int_0^{S_n} |zH_s^n|^2 \mathbf{1}\{|zH_s^n| > \varepsilon\} \lambda_s^n ds \right] \right\} \\ & \leq \exp\left(\frac{z^2}{2}C\right) \left\{ \varepsilon \frac{z^2}{2}C + E \left[\int_0^{S_n} |zH_s^n|^2 \mathbf{1}\{|zH_s^n| > \varepsilon\} \lambda_s^n ds \right] \right\}. \end{aligned}$$

ここで確率変数列 $\int_0^{S_n} |zH_s^n|^2 \mathbf{1}\{|zH_s^n| > \varepsilon\} \lambda_s^n ds$ は有界であるから一様可積分であり, それがゼロに確率収束するという仮定から期待値がゼロに収束することが従う (例えば Billingsley (1986) の Theorem 25.12 を見よ.) つまり右辺の第2項はゼロに収束し, $\varepsilon > 0$ の取り方が任意であったことから $E[G_{S_n}^n] - 1 \rightarrow 0$ が出る.

(2.3) は Lenglart の不等式を用いて示される. \square

応用上の必要から, 上記の定理を少し一般化したものを述べておく. 証明は同様にしてできるので省略する.

定理 2.2.2 各 $n \in \mathbb{N}$ に対し, $N^{n,k}$, $k = 1, \dots, m_n$, は独立な計数過程であってそれぞれの強度は $\lambda^{n,k}$ であるとする. $H^{n,k,(i)}$, $i = 1, \dots, p$ は予測可能過程であって $t \rightsquigarrow \int_0^t |H_s^{n,k,(i)}|^2 \lambda_s^{n,k} ds$ が局所可積分であるものとする. p 次元の局所二乗可積分マルチンゲール $M^n = (M^{n,(1)}, \dots, M^{n,(p)})^\top$,

$$M_t^{n,(i)} = \sum_{k=1}^{m_n} \int_0^t H_s^{n,k,(i)} (dN_s^{n,k} - \lambda_s^{n,k} ds)$$

を考える. T_n は有限停止時刻であるとし, $n \rightarrow \infty$ とするとき

$$\begin{aligned} & \sum_{k=1}^{m_n} \int_0^{T_n} \sum_{i=1}^p |H_s^{n,k,(i)}|^2 \mathbf{1} \left\{ \sum_{i=1}^p |H_s^{n,k,(i)}| > \varepsilon \right\} \lambda_s^{n,k} ds \xrightarrow{p} 0, \quad \forall \varepsilon > 0, \\ & \sum_{k=1}^{m_n} \int_0^{T_n} H_s^{n,k,(i)} H_s^{n,k,(j)} \lambda_s^{n,k} ds \xrightarrow{p} C^{(i,j)}, \quad i, j = 1, \dots, p \end{aligned}$$

であることを仮定する. ただし $C = (C^{(i,j)})_{i,j=1,\dots,p}$ は定数行列である. このとき,

$$M_{T_n}^n \rightarrow^d \mathcal{N}_p(0, C)$$

が成り立つ.

2.3 準備3：Z-推定量の一般論

この節では、一般的な推定方程式の解として定義される推定量（Z-推定量）の一致性を保証する定理、および漸近分布を導出する定理を与える。記号が抽象的なのでこの節だけ読んだのではわかりにくいかもしれないので、前章における例題を

$$\Psi_n(\theta) = \frac{1}{T_n} l'_n(\theta)$$

と読み替えながら読んで頂きたい。ただし、この節で提示する抽象的な議論の方が証明の流れはクリアになるし、応用も利く。

2.3.1 一致性

$\theta \in \Theta \subset \mathbb{R}^p$ とする。 p -次元ランダムベクトル $\Psi_n(\theta) = (\Psi_n^1(\theta), \dots, \Psi_n^p(\theta))^T$ が与えられたとし、対応する Z-推定量 $\hat{\theta}_n$ を、推定方程式

$$\Psi_n^i(\theta) = 0, \quad i = 1, \dots, p$$

の近似的な解として定義する。

定理 2.3.1 あるランダムでない関数 $\theta \mapsto \Psi(\theta)$ であって

$$\sup_{\theta \in \Theta} \|\Psi_n(\theta) - \Psi(\theta)\| \xrightarrow{P} 0$$

および

$$\inf_{\theta: \|\theta - \theta_0\| > \varepsilon} \|\Psi(\theta)\| > 0 = \|\Psi(\theta_0)\|, \quad \forall \varepsilon > 0$$

を満たすものが存在するとする。このとき、 $\Psi_n(\hat{\theta}_n) = o_P(1)$ ならば $\hat{\theta}_n \xrightarrow{P} \theta_0$ が成り立つ。

証明. $\hat{\theta}_n$ の定義から

$$\begin{aligned} \|\Psi(\hat{\theta}_n)\| &\leq \|\Psi(\hat{\theta}_n) - \Psi_n(\hat{\theta}_n)\| + \|\Psi_n(\hat{\theta}_n)\| \\ &\leq \sup_{\theta \in \Theta} \|\Psi(\theta) - \Psi_n(\theta)\| + \|\Psi_n(\hat{\theta}_n)\| = o_P(1). \end{aligned}$$

いま, 任意の $\varepsilon > 0$ に対しある $\delta > 0$ が存在して $\|\theta - \theta_0\| > \varepsilon$ となるような全ての θ に対して $\|\Psi(\theta)\| > \delta$ が成り立つ. よって, イベント $\{\|\hat{\theta}_n - \theta_0\| > \varepsilon\}$ はイベント $\{\|\Psi(\hat{\theta}_n)\| > \delta\}$ に含まれているが, 上で見たようにその確率はゼロに収束する. \square

2.3.2 漸近分布

$\theta \in \Theta \subset \mathbb{R}^p$ とする. p -次元ランダムベクトル $\Psi_n(\theta) = (\Psi_n^1(\theta), \dots, \Psi_n^p(\theta))^T$ および $p \times p$ -ランダム行列 $V_n(\theta) = (V_n^{ij}(\theta))_{i,j=1,\dots,p}$ があって, 次のような関係があると仮定する:

$$(2.4) \quad \Psi_n(\theta) = \Psi_n(\theta_0) + V_n(\tilde{\theta})(\theta - \theta_0).$$

ただし $\tilde{\theta}$ は θ_0 と θ を結ぶ線分上のランダムな点である. すなわち, 関数 $\theta \mapsto \Psi_n(\theta)$ が真値 θ_0 のまわりでテイラー展開できることを仮定するのである. Z -推定量 $\hat{\theta}_n$ を, 推定方程式

$$\Psi_n^i(\theta) = 0, \quad i = 1, \dots, p$$

の近似的な解として定義する.

定理 2.3.2 r_n を正の定数列とする. (2.4) を考える. ある p -次元確率ベクトル $G(\theta_0)$ および $p \times p$ -正則行列 $V(\theta_0)$ が存在して,

$$(2.5) \quad r_n \Psi_n(\theta_0) \rightarrow^d G(\theta_0)$$

が成り立つこと, および, 任意の確率変数列 $\tilde{\theta}_n \rightarrow^p \theta_0$ に対して

$$(2.6) \quad V_n(\tilde{\theta}_n) \rightarrow^p V(\theta_0)$$

が成り立つことを仮定する. このとき, $\Psi_n(\hat{\theta}_n) = o_P(r_n^{-1})$ および $\hat{\theta}_n \rightarrow^p \theta_0$ を満たす任意の $\hat{\theta}_n$ に対して

$$r_n(\hat{\theta}_n - \theta_0) = -r_n V(\theta_0)^{-1} \Psi_n(\theta_0) + o_P(1)$$

が成り立つ. 特に

$$r_n(\hat{\theta}_n - \theta_0) \rightarrow^d -V(\theta_0)^{-1} G(\theta_0)$$

が成り立つ.

証明. (2.4) に $\theta = \hat{\theta}_n$ を代入したときの $\tilde{\theta}$ を $\tilde{\theta}_n$ とおく. すなわち, $\tilde{\theta}_n$ は θ_0 と $\hat{\theta}_n$ を結ぶ線分上の点であって

$$\Psi_n(\hat{\theta}_n) = \Psi_n(\theta_0) + V_n(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0)$$

を満たすものとする. 仮定 (2.6) により, この式はさらに

$$\Psi_n(\hat{\theta}_n) = \Psi_n(\theta_0) + V(\theta_0)(\hat{\theta}_n - \theta_0) + o_P(\|\hat{\theta}_n - \theta_0\|)$$

と書ける. $\hat{\theta}_n$ の定義により左辺は $o_P(r_n^{-1})$ であるから, 右辺の第 2 項を左辺に移項してから両辺に $-r_n V(\theta_0)^{-1}$ をかけることにより

$$r_n(\hat{\theta}_n - \theta_0) = -r_n V(\theta_0)^{-1} \Psi_n(\theta_0) + o_P(1 + r_n \|\hat{\theta}_n - \theta_0\|)$$

を得る. 仮定 (2.5) より右辺の第 1 項は $O_P(1)$ であるから, $r_n(\hat{\theta}_n - \theta_0) = O_P(1)$ であり, 結局右辺の第 2 項は $o_P(1)$ である. これで主張の前半は示された. 後半は (2.5) から直ちに従う. \square

2.4 本論：Cox モデル

この節では確率過程の観測期間 $[0, T]$ は固定されているとする. $t \rightsquigarrow N_t^k$, $k = 1, \dots, n$, は独立な計数過程であるとし, それぞれの強度が

$$\lambda_t^k = Y_t^k e^{\theta^\top Z_t^k} \alpha(t)$$

で与えられることを仮定する. 使う記号を先に書いておこう:

$$\begin{aligned} S_t^{n,0}(\theta) &= \sum_{k=1}^n Y_t^k e^{\theta^\top Z_t^k}, \\ S_t^{n,1}(\theta) &= (\partial_1 S_t^{n,0}(\theta), \dots, \partial_p S_t^{n,0}(\theta))^\top \\ &= \sum_{k=1}^n Z_t^k Y_t^k e^{\theta^\top Z_t^k}, \\ S_t^{n,2}(\theta) &= (\partial_{ij} S_t^{n,0}(\theta))_{i,j=1,\dots,p}, \\ &= \sum_{k=1}^n Z_t^k (Z_t^k)^\top Y_t^k e^{\theta^\top Z_t^k}. \end{aligned}$$

さてこのモデルにおいて、「Cox の部分尤度」と呼ばれるものの対数をとったものをとりあえず天下りの

$$C_n(\theta) = \sum_{k=1}^n \int_0^T (\theta^\top Z_t^k - \log S_t^{n,0}(\theta)) dN_t^k$$

と与えよう。推定関数として

$$\begin{aligned} \Psi_n(\theta) &= \frac{1}{n} (\partial_1 C_n(\theta), \dots, \partial_p C_n(\theta))^\top \\ &= \frac{1}{n} \sum_{k=1}^n \int_0^T \left(Z_t^k - \frac{S_t^{n,1}(\theta)}{S_t^{n,0}(\theta)} \right) dN_t^k \end{aligned}$$

を使うことを提案する。この推定関数の妥当性を議論しよう。アイデアを得るための出発点として、まずベースラインハザード α が仮に既知だったときには対数尤度は

$$\sum_{k=1}^n \int_0^T (\theta^\top Z_t^k + \log(Y_t^k \alpha(t))) dN_t^k - \int_0^T Y_t^k e^{\theta^\top Z_t^k} \alpha(t) dt$$

となるから、推定関数としてはそれを θ で微分した

$$(2.7) \quad \sum_{k=1}^n \int_0^T Z_t^k dN_t^k - \int_0^T Z_t^k Y_t^k e^{\theta^\top Z_t^k} \alpha(t) dt$$

が妥当と考えられる。しかしながら、実際にはベースラインハザードは未知としたいので、右辺の第2項に現れている α を推定関数の定義から取り除きたい。ここで、和をとった計数過程 $\bar{N}_t = \sum_{k=1}^n N_t^k$ の対数尤度が

$$\int_0^T \log S_t^{n,0}(\theta) d\bar{N}_t - \int_0^T S_t^{n,0}(\theta) \alpha(t) dt$$

で与えられることに注意し、これを θ に関して微分する：

$$\int_0^T \frac{S_t^{n,1}(\theta)}{S_t^{n,0}(\theta)} d\bar{N}_t - \int_0^T S_t^{n,1}(\theta) \alpha(t) dt.$$

これに真値 $\theta = \theta_0$ を代入するとマルチンゲールの終点値になるので、ゼロに近い。つまり

$$\int_0^T \frac{S_t^{n,1}(\theta_0)}{S_t^{n,0}(\theta_0)} d\bar{N}_t \approx \int_0^T S_t^{n,1}(\theta_0) \alpha(t) dt = \int_0^T Z_t^k Y_t^k e^{\theta_0^\top Z_t^k} \alpha(t) dt$$

という近似が成り立つ. よって (2.7) の第2項に $\int_0^T \frac{S_t^{n,1}(\theta)}{S_t^{n,0}(\theta)} d\bar{N}_t$ を代入することにより

$$\sum_{k=1}^n \int_0^T Z_t^k dN_t^k - \int_0^T \frac{S_t^{n,1}(\theta)}{S_t^{n,0}(\theta)} d\bar{N}_t$$

が得られるが, これを n で割ったものが Cox の推定関数 $\Psi_n(\theta)$ に他ならない.

解析的に重要なのは, 真値 $\theta = \theta_0$ のもとで $\Psi_n(\theta_0)$ はマルチンゲールであることである. 実際

$$\begin{aligned} \Psi_n(\theta_0) &= \frac{1}{n} \sum_{k=1}^n \int_0^T \left(Z_t^k - \frac{S_t^{n,1}(\theta_0)}{S_t^{n,0}(\theta_0)} \right) dN_t^k \\ &= \frac{1}{n} \sum_{k=1}^n \int_0^T \left(Z_t^k - \frac{S_t^{n,1}(\theta_0)}{S_t^{n,0}(\theta_0)} \right) (dN_t^k - Y_t^k e^{\theta_0^\top Z_t^k} \alpha(t) dt) \end{aligned}$$

である. 定理 2.3 を適用するための準備として $\Psi_n(\theta)$ の θ_0 の周りでのテイラー展開 $\Psi_n(\theta) = \Psi_n(\theta_0) + V_n(\tilde{\theta})(\theta - \theta_0)$ するときの $V_n(\theta)$ が

$$V_n(\theta) = -\frac{1}{n} \sum_{k=1}^n \frac{S_t^{n,2}(\theta) S_t^{n,0}(\theta) - S_t^{n,1}(\theta) S_t^{n,1}(\theta)^\top}{S_t^{n,0}(\theta)^2} dN_t^k$$

によって与えられることに注意しよう.

では漸近理論を展開するための正則条件を用意しよう.

条件 2.4.1 (i) α は非負可測関数で $\int_0^T \alpha(t) dt < \infty$.

(ii) ある実数値関数 $s^0(t; \theta)$, p -次元ベクトル値関数 $s^1(t; \theta)$ および $p \times p$ -行列値関数 $s^2(t; \theta)$ が存在して

$$\sup_{\theta \in \Theta} \sup_{t \in [0, T]} \left\| \frac{1}{n} S_t^{n,l}(\theta) - s^l(t; \theta) \right\| \rightarrow^p 0, \quad l = 0, 1, 2.$$

さらに, $l = 0, 1, 2$ のいずれも $\theta \mapsto s^l(t; \theta)$ は $t \in [0, T]$ について一様に連続であり, $\sup_{t \in [0, T], \theta \in \Theta} \|s^l(t; \theta)\| < \infty$. 特に $l = 0$ は $\inf_{t \in [0, T]} s^0(t; \theta_0) > 0$.

(iii) 任意の $\varepsilon > 0$ に対して $\sum_{k=1}^n \int_0^T \|\xi_t^{n,k}\|_{\{\|\xi_t^{n,k}\| > \varepsilon\}}^2 Y_t^k e^{\theta_0^\top Z_t^k} \alpha(t) dt \rightarrow^p 0$. ただし

$$\xi_t^{n,k} = \frac{1}{\sqrt{n}} \left(Z_t^k - \frac{S_t^{n,1}(\theta_0)}{S_t^{n,0}(\theta_0)} \right).$$

(iv) 行列

$$\Sigma(\theta_0) = \int_0^T \left(s^2(t; \theta_0) - \frac{s^1(t; \theta_0)s^1(t; \theta_0)^\top}{s^0(t; \theta_0)} \right) \alpha(t) dt$$

は正則である.

では一致性を議論しよう.

定理 2.4.2 条件 2.4.1 (i) - (ii) を仮定する. さらに

$$\Psi(\theta) = \int_0^T \left(\frac{s^1(t; \theta_0)}{s^0(t; \theta_0)} - \frac{s^1(t; \theta)}{s^0(t; \theta)} \right) s^0(t; \theta_0) \alpha(t) dt$$

とおいたときに, 任意の $\varepsilon > 0$ に対して $\inf_{\theta: \|\theta - \theta_0\| > \varepsilon} \|\Psi(\theta)\| > 0$ が成り立つことを仮定する. このとき, $\Psi_n(\hat{\theta}_n) = o_P(1)$ ならば $\hat{\theta}_n \rightarrow^P \theta_0$ が成り立つ.

証明. Lengart の不等式より

$$\Psi_n(\theta) = \frac{1}{n} \sum_{k=1}^n \int_0^T S_t^{n,1}(\theta_0) \alpha(t) dt - \frac{1}{n} \sum_{k=1}^n \int_0^T \frac{S_t^{n,1}(\theta)}{S_t^{n,0}(\theta)} dN_t^k + o_P(1).$$

右辺の第 1 項は $\int_0^T s^1(t; \theta_0) \alpha(t) dt$ に確率収束する. 一方, Lengart の不等式より $n^{-1} \sum_{k=1}^n N_T^k = n^{-1} \int_0^T S_t^{n,0}(\theta_0) \alpha(t) dt + o_P(1)$ であるから, 条件 2.4.1 (ii) よりこの値は $O_P(1)$ である. よって再び条件 2.4.1 (ii) より右辺の第 2 項は θ に関して一様に

$$-\frac{1}{n} \sum_{k=1}^n \int_0^T \frac{s^1(t; \theta)}{s^0(t; \theta)} dN_t^k$$

によって近似される. 非積分関数は t, θ に関して一様に有界であるから, これは

$$-\int_0^T \frac{s^1(t; \theta)}{s^0(t; \theta)} s^0(t; \theta_0) \alpha(t) dt$$

に θ に関して一様に確率収束する. すなわち $\sup_{\theta \in \Theta} \|\Psi_n(\theta) - \Psi(\theta)\| \rightarrow^P 0$ である. よって定理 2.3.1 により結論が得られる. \square

定理 2.3 により, 漸近正規性を証明するためには次の補題を示せば十分である.

補題 2.4.3 条件 2.4.1 の全てを仮定する.

(i) $\sqrt{n} \Psi_n(\theta_0) \rightarrow^d \mathcal{N}_p(0, \Sigma(\theta_0))$.

(ii) 任意の確率変数列 $\tilde{\theta}_n \rightarrow^P \theta_0$ に対して $V_n(\tilde{\theta}_n) \rightarrow^P -\Sigma(\theta_0)$.

定理 2.4.4 条件 2.4.1 の全てを仮定する. $\Psi_n(\hat{\theta}_n) = o_P(n^{-1/2})$ かつ $\hat{\theta}_n \rightarrow^p \theta_0$ ならば $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow^d \mathcal{N}_p(0, \Sigma(\theta_0)^{-1})$ が成り立つ.

補題 2.4.3 の証明. 上で見たように, $\sqrt{n}\Psi_n(\theta_0)$ は

$$M_t^n = \frac{1}{\sqrt{n}} \sum_{k=1}^n \int_0^t \left(Z_u^k - \frac{S_u^{n,1}(\theta_0)}{S_u^{n,0}(\theta_0)} \right) (dN_u^k - Y_u^k e^{\theta_0^T Z_u^k} \alpha(u) du)$$

によって与えられるマルチンゲール $t \rightsquigarrow M_t^n$ の終点値である. これにマルチンゲール中心極限定理を適用しよう. 二次変分の収束は

$$\begin{aligned} \langle M^n, M^n \rangle_T &= \frac{1}{n} \sum_{k=1}^n \int_0^T \left(Z_t^k - \frac{S_t^{n,1}(\theta_0)}{S_t^{n,0}(\theta_0)} \right) \left(Z_t^k - \frac{S_t^{n,1}(\theta_0)}{S_t^{n,0}(\theta_0)} \right)^T Y_t^k e^{\theta_0^T Z_t^k} \alpha(t) dt \\ &= \frac{1}{n} \sum_{k=1}^n \int_0^T \left(S_t^{n,2}(\theta_0) - \frac{S_t^{n,1}(\theta_0) S_t^{n,1}(\theta_0)^T}{S_t^{n,0}(\theta_0)} \right) \alpha(t) dt \\ &\rightarrow^p \int_0^T \left(s^2(t; \theta_0) - \frac{s^1(t; \theta_0) s^1(t; \theta_0)^T}{s^0(t; \theta_0)} \right) \alpha(t) dt \end{aligned}$$

と計算される. リンデベルグ条件は (iii) によって仮定されている. よって $\sqrt{n}\Psi_n(\theta_0) = M_T^n \rightarrow^d \mathcal{N}_p(0, \Sigma(\theta_0))$.

一方, Lenglart の不等式より $n^{-1} \sum_{k=1}^n N_T^k = n^{-1} \int_0^T S_t^{n,0}(\theta_0) \alpha(t) dt + o_P(1)$ であるから, 条件 2.4.1 (ii) よりこの値は $O_P(1)$ である. よって再び条件 2.4.1 (ii) より $V_n(\theta)$ は θ に関して一様に

$$-\frac{1}{n} \sum_{k=1}^n \frac{s^2(t; \theta) s^0(t; \theta) - s^1(t; \theta) s^1(t; \theta)^T}{s^0(t; \theta)^2} dN_t^k$$

によって近似できる. 非積分関数は t, θ に関して一様に有界であるから, これは

$$V(\theta) = - \int_0^T \frac{s^2(t; \theta) s^0(t; \theta) - s^1(t; \theta) s^1(t; \theta)^T}{s^0(t; \theta)^2} s^0(t; \theta) \alpha(t) dt$$

によって近似できる. よって $\sup_{\theta} \|V_n(\theta) - V(\theta)\| \rightarrow^p 0$ である. 仮定から $\theta \mapsto V(\theta)$ が連続であることがすぐに出るので, 任意の $\tilde{\theta}_n \rightarrow^p \theta_0$ に対して $V_n(\tilde{\theta}_n) \rightarrow^p V(\theta_0) = -\Sigma(\theta_0)$ であることが従う. \square

関連図書

- [1] Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.
- [2] Billingsley, P. (1986). *Probability and Measure*. (2nd ed.) Wiley
- [3] Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis*. Wiley
- [4] Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes*. Springer
- [5] Karr, A.F. (1991). *Point Processes and Their Statistical Inference*. (2nd ed.) Dekker.
- [6] 西山陽一 (201?). マルチンゲール理論による統計解析の基礎. (執筆中. 約 150 ページの予定.) 近代科学社.

第II部：研究報告

<統計的リスク管理>

1. 「歪みリスク尺度の応用」 塚原英敦
2. 「Weighted Nadaraya-Watson estimation of conditional expected shortfall」 加藤健吾

Applications of Distortion Risk Measures

Hideatsu Tsukahara

(tsukahar@seijo.ac.jp)

Dept of Economics, Seijo University

March 30, 2010

Contents

1. The Class of Distortion Risk Measures (DRMs)
2. Statistical Estimation & Simulation
3. Backtesting of DRMs
4. Capital Allocation with DRMs

1. Distortion Risk Measures (DRM)

For a rv X representing loss, put

- df of X : $F_X(x) := P(X \leq x)$
- quantile of X : $F_X^{-1}(u) := \inf\{x \in \mathbb{R} : F_X(x) \geq u\}$, $0 < u < 1$

Def. A functional $\rho: L^\infty \rightarrow \mathbb{R}$ is called *coherent* if it satisfies

- [PO] (positivity): $X \leq 0$ a.s. $\implies \rho(X) \leq 0$
- [PH] (positive homogeneity): $\forall \lambda > 0$, $\rho(\lambda X) = \lambda \rho(X)$
- [TE] (translation equivariance): $\forall c > 0$, $\rho(X + c) = \rho(X) + c$
- [SA] (subadditivity): $\rho(X + Y) \leq \rho(X) + \rho(Y)$

Add two more axioms:

[LJ] (law invariance): $X \stackrel{L}{=} Y \implies \rho(X) = \rho(Y)$

[CA] (comonotonic additivity):

X and Y are comonotone $\implies \rho(X + Y) = \rho(X) + \rho(Y)$

X_1, \dots, X_d are *comonotone* \Leftrightarrow There exist a rv Z and increasing func's f_1, \dots, f_d s.t. $(X_1, \dots, X_d) \stackrel{L}{=} (f_1(Z), \dots, f_d(Z))$

►► Kusuoka: The class of DRMs coincides with the set of coherent risk measures satisfying law invariance and comonotonic additivity

Distortion function

Any distribution function (df) D on $[0, 1]$;

i.e., right-continuous, increasing on $[0, 1]$. $D(0) = 0$, $D(1) = 1$

For a distortion D , a *distortion risk measure (DRM)* is defined by

$$\rho_D(X) := \int_{[0,1]} F_X^{-1}(u) dD(u) = \int_{\mathbb{R}} x dD \circ F_X(x).$$

[a.k.a. spectral risk measure (Acerbi), weighted VOR (Cherny)]

★ $D_\alpha^{\text{VaR}}(u) = \mathbf{1}_{\{u \geq 1-\alpha\}}$ yields $\text{VaR}_\alpha(X) = F_X^{-1}(1-\alpha)$, $0 < \alpha < 1$, but this D_α^{VaR} is not convex.

Example: Expected Shortfall (ES)

The expected loss that is incurred when VaR is exceeded

$$\text{ES}_\alpha(X) := \frac{1}{\alpha} \int_{1-\alpha}^1 F_X^{-1}(u) du$$

$$= \mathbb{E}(X | X \geq \text{VaR}_\alpha(X))$$

Taking distortion of the form

$$D_\alpha^{\text{ES}}(u) = \frac{1}{\alpha} [u - (1 - \alpha)]_+, \quad 0 < \alpha < 1$$

yields ES as a DRM

Other Examples:

- *Proportional Hazards*: $D_\theta^{\text{PH}}(u) = 1 - (1 - u)^\theta$
- *Proportional Odds*: $D_\theta^{\text{PO}}(u) = \frac{\theta u}{1 - (1 - \theta)u}$
- *Gaussian (Wang transform)*: $D_\theta^{\text{GA}}(u) = \Phi(\Phi^{-1}(u) + \log \theta)$
- *Proportional γ -Odds*: $D_\theta^{\text{PGO}}(u) = 1 - \left[\frac{(1 - u)^\gamma}{\theta - \theta(1 - u)^\gamma + (1 - u)^\gamma} \right]^{1/\gamma}$
- *Positive Poisson Mixture*: $D_\lambda^{\text{PPM}}(u) = \frac{e^{\lambda u} - 1}{e^\lambda - 1}$

2. Statistical Estimation

$(X_n)_{n \in \mathbb{N}}$: strictly stationary process with $X_n \sim F$

\mathbb{F}_n : empirical df based on the sample X_1, \dots, X_n

A natural estimator of $\rho(X)$ is

$$\begin{aligned} \hat{\rho}_n(X) &= \int_0^1 \mathbb{F}_n^{-1}(u) dD(u) \\ &= \sum_{i=1}^n c_{ni} X_{n:i}, \quad c_{ni} := D\left(\frac{i-1}{n}, \frac{i}{n}\right) \end{aligned}$$

Strong consistency

Let $d(u) = \frac{d}{du}D(u)$ for a convex distortion D , and $1 \leq p \leq \infty$, $1/p + 1/q = 1$. Suppose that $(X_n)_{n \in \mathbb{N}}$ is an ergodic stationary sequence, and that $d \in L^p(0, 1)$ and $F^{-1} \in L^q(0, 1)$. Then

$$\widehat{\rho}_n(X) \longrightarrow \rho(X), \text{ a.s.}$$

For a proof, see van Zwet (1980, AP)

[All we need is SLLN and Glivenko-Cantelli Theorem].

Set

$$\sigma(u, v) := [u \wedge v - uv] + \sum_{j=1}^{\infty} [C_j(u, v) - uv] + \sum_{j=1}^{\infty} [C_j(v, u) - uv],$$

$$C_j(u, v) := P(X_1 \leq F^{-1}(u), X_{j+1} \leq F^{-1}(v))$$

Theorem (Asymptotic Normality)

Under the above assumptions, we have

$$\sqrt{n}(\widehat{\rho}_n(X) - \rho(X)) \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

where

$$\sigma^2 := \int_0^1 \int_0^1 \sigma(u, v) d(u) d(v) dF^{-1}(u) dF^{-1}(v) < \infty$$

Assumptions:

- $(X_n)_{n \in \mathbb{N}}$ is strongly mixing with rate $\alpha(n) = O(n^{-\theta-\eta})$ for some $\theta \geq 1 + \sqrt{2}$, $\eta > 0$
- For F^{-1} -almost all u , d is continuous at u
- $|d| \leq B$, $B(u) := Mu^{-b_1}(1-u)^{-b_2}$,
- $|F^{-1}| \leq H$, $H(u) := Mu^{-d_1}(1-u)^{-d_2}$

Assume b_i, d_i & θ satisfy $b_i + d_i + \frac{2b_i+1}{2\theta} < \frac{1}{2}$, $i = 1, 2$

Simulation results for estimating VaR, ES & PO risk measures with GARCH observations ($n = 500$, # of replication = 1000)

$$X_t = 0.0009 + \varepsilon_t, \quad \sigma_t^2 = 0.5 + 0.85\sigma_{t-1}^2 + 0.1\varepsilon_{t-1}^2$$

$\theta = \alpha$	VaR		ES		PO		PH	
	mean	std	mean	std	mean	std	mean	std
0.5	0.0077	0.1679	2.4590	0.2687	1.2134	0.1854	2.2206	0.3119
0.05	5.1429	0.5488	6.6250	0.8048	5.0339	0.5959	8.9421	1.8604
0.01	7.7766	1.1182	8.8885	1.4658	7.3829	1.0806	10.2292	2.2618

- GARCH model:

$$X_n = \sigma_n Z_n, \quad (Z_n) : \text{i.i.d.}$$

$$\sigma_n^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2$$

- ▶▶ If the stationary distribution has a positive density around 0, then GARCH is strongly mixing with exponentially decaying $\alpha(n)$

- Stochastic Volatility model:

$$X_n = \sigma_n Z_n, \quad (Z_n) : \text{i.i.d.}, \quad (\sigma_n) : \text{strictly stationary positive } (Z_n) \text{ and } (\sigma_n) \text{ are assumed to be independent}$$

- ▶▶ The mixing rate of (X_n) is the same as that of $(\log \sigma_n)$

3. Backtesting with general distortion risk measures

Note that, with $d(u) = \frac{d}{du} D(u)$,

$$\begin{aligned} \rho_D(X) &= \int_{-\infty}^{\infty} x \, dD \circ F_X(x) = \int_{-\infty}^{\infty} x d(F_X(x)) \, dF_X(x) \\ &= \mathbb{E}[X d(F_X(X))] \end{aligned}$$

Thus $X d(F_X(X)) - \rho_D(X)$ has mean 0.

- Estimation of Asymptotic Variance

$$\sigma^2 = \iint \sigma(F(x), F(y)) d(F(x)) d(F(y)) \, dx dy$$

where

$$\begin{aligned} \sigma(F(x), F(y)) &= [F(x) \wedge F(y) - F(x)F(y)] \\ &\quad + \sum_{j=1}^{\infty} [F_j(u, v) - F(x)F(y)] + \sum_{j=1}^{\infty} [F_j(y, x) - F(x)F(y)], \end{aligned}$$

and

$$F_j(x, y) = P(X_1 \leq x, X_{j+1} \leq y)$$

- ▶▶ How to estimate this? (to construct confidence intervals)

If we have X_1, \dots, X_N iid with F_X , then we can base the backtesting of our method/model on

$$\begin{aligned} &X_{n+1} d(\widehat{F}_{1:n}(X_{n+1})) - \widehat{\rho}_{D_{1:n}}(X), \\ &\quad \vdots \\ &X_N d(\widehat{F}_{N-n+1:N}(X_N)) - \widehat{\rho}_{D_{N-n+1:N}}(X) \end{aligned}$$

where $\widehat{F}_{k,m}$ and $\widehat{\rho}_{D_{k,m}}$ are estimates based on the sample X_k, \dots, X_m

- ▶▶ If we have dependent data, more explicit model may be necessary

Conditional Approach (McNeil & Frey (2000))

Write $\rho_D^t(X_{t+1})$ for a distortion risk measure with a distortion D for the conditional df $F_t(x) := P(X_{t+1} \leq x \mid \mathcal{F}_t)$, $\mathcal{F}_t := \sigma(X_s : s \leq t)$:

$$\rho_D^t(X_{t+1}) := \int_{[0,1]} F_t^{-1}(u) dD(u)$$

Suppose that for \mathcal{F}_{t-1} -measurable μ_t and σ_t ,

$$X_t = \mu_t + \sigma_t Z_t,$$

where Z_t is iid with finite 2nd moment.

Note that (by translation invariance and positive homogeneity)

$$R_t := \frac{X_{t+1} - \rho_D^t(X_{t+1})}{\sigma_{t+1}} = Z_{t+1} - \rho_D(Z),$$

where Z is a generic rv with the same df G as Z_t 's. Since

$$E\{[Z_{t+1} - \rho_D(Z)]d(G(Z_{t+1}))\} = 0,$$

we can perform backtesting based on $\widehat{R}_t \cdot d(\widehat{G}(Z_{t+1}))$, where

$$\widehat{R}_t := \frac{X_{t+1} - \widehat{\rho}_D^t(X_{t+1})}{\widehat{\sigma}_{t+1}}$$

\widehat{G} may be constructed from

$$z_i = \frac{x_i - \widehat{\mu}_t}{\widehat{\sigma}_t}, \quad i = 1, \dots, t$$

Note that

$$\widehat{G}(Z_{t+1}) = \#\{i: z_i \leq Z_{t+1}\},$$

where

$$Z_{t+1} = \frac{X_{t+1} - \widehat{\mu}_{t+1}}{\widehat{\sigma}_{t+1}}$$

4. Capital Allocation

d investment opportunities (e.g., business units, subportfolios, assets) X_i : loss associated with the i th investments

1. Compute the overall risk capital $\rho(X)$, where $X = \sum_{i=1}^d X_i$ and ρ is a particular risk measure.
2. Allocate the capital $\rho(X)$ to the individual investment possibilities according to some mathematical *capital allocation principle* such that, if κ_i denotes the capital allocated to the investment opportunity with potential loss X_i , we have $\sum_{i=1}^d \kappa_i = \rho(X)$.

►► Find $\kappa = (\kappa_1, \dots, \kappa_d) \in \mathbb{R}^d$ s.t. $\sum_{i=1}^d \kappa_i = \rho(X)$ according to some criterion

Setup

It is convenient to introduce 'weights' $\lambda = (\lambda_1, \dots, \lambda_d)$ (to be interpreted as amount of money invested in each opportunity)

Put $X(\lambda) := \sum_{i=1}^d \lambda_i X_i$ and

$r_\rho(\lambda) := \rho(X(\lambda))$ risk measure function

If ρ is positive homogeneous, then, for $h > 0$

$$r_\rho(h\lambda) = hr_\rho(\lambda)$$

i.e., r_ρ is positive homogeneous of degree 1

Euler's rule: If r_ρ is positive homogeneous and differentiable,

$$r_\rho(\lambda) = \sum_{i=1}^d \lambda_i \frac{\partial r_\rho}{\partial \lambda_i}(\lambda)$$

Euler allocation principle

If r_ρ is a positive homogeneous risk measure function, which is differentiable on the set Λ , then the (per-unit) Euler capital allocation principle associated with r_ρ is

$$\kappa_i(\lambda) = \frac{\partial r_\rho}{\partial \lambda_i}(\lambda)$$

Justification

- Tasche: RORAC compatibility

r_ρ : differentiable risk measure function

κ : capital allocation principle

κ is called *suitable for performance measurement* if for all λ we have

$$\frac{\partial}{\partial \lambda_i} \left(\frac{-E(X(\lambda))}{r_\rho(\lambda)} \right) \begin{cases} > 0 & \text{if } \frac{-E(X_i)}{\kappa_i(\lambda)} > \frac{-E(X(\lambda))}{r_\rho(\lambda)}, \\ < 0 & \text{if } \frac{-E(X_i)}{\kappa_i(\lambda)} < \frac{-E(X(\lambda))}{r_\rho(\lambda)}. \end{cases}$$

- ▶▶ The only per-unit capital allocation principle suitable for performance measurement is the Euler principle.

- Denault: Cooperative game theory

d investment opportunities = d players

If ρ is subadditive, then $\rho(X(\lambda)) \leq \sum_{i=1}^d \rho(\lambda_i X_i)$.

A fuzzy core (Aubin, 1981) is given by

$$\mathcal{C} = \left\{ \kappa \in \mathbb{R}^d : r_\rho(\mathbf{1}) = \sum_{i=1}^d \kappa_i \ \& \ r_\rho(\lambda) \geq \sum_{i=1}^d \lambda_i \kappa_i \ \forall \lambda \in [0, 1]^d \right\}$$

- ▶▶ If r_ρ is differentiable at $\lambda = \mathbf{1}$, then \mathcal{C} consists only of the gradient vector of r_ρ at $\lambda = \mathbf{1}$:

$$\kappa_i = \left. \frac{\partial r_\rho(\lambda)}{\partial \lambda_i} \right|_{\lambda=\mathbf{1}}$$

Examples

- Covariance principle:

$$r_\rho(\lambda) = \sqrt{\text{var}(X(\lambda))} = \sqrt{\lambda' \Sigma \lambda}$$

where Σ is the covariance matrix of (X_1, \dots, X_d) . Then

$$\kappa_i(\lambda) = \frac{\partial r_\rho(\lambda)}{\partial \lambda_i} = \frac{\text{cov}(X_i, X(\lambda))}{\sqrt{\text{var}(X(\lambda))}}$$

In particular, the capital allocated to the i th investment opportunity is

$$\kappa_i = \frac{\text{cov}(X_i, X)}{\sqrt{\text{var}(X)}}$$

- ES contributions:

$$r_\rho(\lambda) = \text{ES}_\alpha(X(\lambda)) = \frac{1}{\alpha} \int_{1-\alpha}^1 F_{X(\lambda)}^{-1}(u) du$$

Then

$$\kappa_i(\lambda) = \frac{\partial r_\rho(\lambda)}{\partial \lambda_i} = \text{E}(X_i | X(\lambda) \geq \text{VaR}_\alpha(X(\lambda)))$$

In particular, the capital allocated to the i th investment opportunity is given by

$$\kappa_i = \text{E}(X_i | X \geq \text{VaR}_\alpha(X))$$

- VaR contributions:

$$r_\rho(\lambda) = \text{VaR}_\alpha(X(\lambda))$$

Then (Tasche, 1999)

$$\kappa_i(\lambda) = \frac{\partial r_\rho(\lambda)}{\partial \lambda_i} = \text{E}(X_i | X(\lambda) = \text{VaR}_\alpha(X(\lambda)))$$

In particular, the capital allocated to the i th investment opportunity is given by

$$\kappa_i = \text{E}(X_i | X = \text{VaR}_\alpha(X))$$

Capital Allocation with DRM

$$r_\rho(\lambda) = \rho_D(X(\lambda)) = \int_{[0,1]} F_{X(\lambda)}^{-1}(u) dD(u)$$

Then, under some regularity conditions,

$$\begin{aligned} \kappa_i(\lambda) &= \frac{\partial r_\rho(\lambda)}{\partial \lambda_i} = \int_{[0,1]} \frac{\partial}{\partial \lambda_i} F_{X(\lambda)}^{-1}(u) dD(u) \\ &= \int_{[0,1]} \text{E}[X_i | X(\lambda) = F_{X(\lambda)}^{-1}(u)] dD(u) \\ &= \int_{\mathbb{R}} \text{E}[X_i | X(\lambda) = x] d(F_{X(\lambda)}(x)) dF_{X(\lambda)}(x) \\ &= \text{E}[X_i d(F_{X(\lambda)}(X(\lambda)))] \end{aligned}$$

Thus, the capital allocated to the i th investment opportunity is

$$\kappa_i = \mathbb{E}[X_i d(F_X(X))]$$

►► We can think of $d(F_X(X))$ as a Radon-Nikodym density: $\mathbb{E}(d(F_X(X))) = 1$ trivially

$$\frac{dQ}{dP} = d(F_X(X)) \implies \kappa_i = \mathbb{E}^Q(X_i)$$

Putting $P(X_i \leq x_i, X \leq x) = C_i(F_{X_i}(x_i), F_X(x))$,

$$\begin{aligned} \kappa_i &= \iint x_i d(F_X(x)) dC_i(F_{X_i}(x_i), F_X(x)) \\ &= \iint F_{X_i}^{-1}(u) d(v) dC_i(u, v) \end{aligned}$$

Example: distortion densities

- Proportional Odds: $d_\theta(u) = \frac{\theta}{(1-u+\theta u)^2}$

- Proportional Hazards: $d_\theta(u) = \theta(1-u)^{\theta-1}$

- Gaussian: $d_\theta(u) = \frac{\phi(\Phi^{-1}(u) + \log \theta)}{\phi(\Phi^{-1}(u))}$

- Proportional γ -Odds: $d_\theta(u) = \frac{\theta}{[\theta - \theta(1-u)^\gamma + (1-u)^\gamma]^{1/\gamma+1}}$

- Positive Poisson Mixture: $d_\lambda(u) = \frac{\lambda e^{\lambda u}}{e^\lambda - 1} \quad (\lambda > 0)$

If an i.i.d. sample $(X_{k,1}, \dots, X_{k,d})$, $k = 1, \dots, n$ is available, then we can estimate κ_i by

$$\hat{\kappa}_i = \iint \mathbb{F}_{n,i}^{-1}(u) d(v) dC_{n,i}(u, v)$$

where

$$\mathbb{F}_{n,i}(x_i) = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{X_{k,i} \leq x_i\}},$$

and putting $X_k := X_{k,1} + \dots + X_{k,d}$, $\mathbb{F}_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{X_k \leq x\}}$,

$$C_{n,i}(u, v) = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{X_{k,i} \leq \mathbb{F}_{n,i}^{-1}(u), X_k \leq \mathbb{F}_n^{-1}(v)\}}$$

Elliptical loss distribution: $E_d(\mu, \Sigma, \psi)$

μ : location vector, Σ : dispersion matrix, ψ : characteristic generator

Assume τ_ρ is the risk measure function of a positive homogeneous, law invariant risk measure ρ . Let $(X_1, \dots, X_d) \sim E_d(\mathbf{0}, \Sigma, \psi)$. Then under an Euler allocation, the relative capital allocation is given by

$$\frac{\kappa_i}{\kappa_j} = \frac{\kappa_i(\mathbf{1})}{\kappa_j(\mathbf{1})} = \frac{\sum_{k=1}^d \sum_{ik} \kappa_{ik}}{\sum_{k=1}^d \sum_{jk} \kappa_{jk}}, \quad 1 \leq i, j \leq d.$$

►► The relative amounts of capital allocated to each investment opportunity are the same as long as we use a positive homogeneous, law invariant risk measure.

6. Concluding Remarks

- Estimation of DRMs is possible, but for some DRMs, we don't get nice asymptotic properties; proportional odds risk measure has some nice features.
- Backtesting procedure can be performed with DRMs. May need more rigorous procedures.
- Euler capital allocation based on DRMs are easy to compute and widely applicable.
- Careful study of portfolio optimization, esp., implementing numerical procedure.
- Extension to dynamic setting

Weighted Nadaraya-Watson estimation of conditional expected shortfall

Kengo Kato *

March 23, 2010

Abstract

This paper addresses the problem of nonparametric estimation of the conditional expected shortfall (CES) which has gained popularity in financial risk management. We propose a new nonparametric estimator of the CES. The proposed estimator is defined as a conditional counterpart of the sample average estimator of the unconditional expected shortfall, where the empirical distribution function is replaced by the weighted Nadaraya-Watson estimator of the conditional distribution function. We establish asymptotic normality of the proposed estimator under an α -mixing condition. The asymptotic results reveal that the proposed estimator has a good bias property. Simulation results illustrate the usefulness of the proposed estimator.

Keywords: α -mixing; conditional expected shortfall; nonparametric estimation; weighted Nadaraya-Watson estimation.

JEL classification numbers: C13, C14, G11.

1 Introduction

This paper addresses the problem of nonparametric estimation of the conditional expected shortfall (CES). The expected shortfall (ES) was first proposed by Artzner et al. (1997) as a risk measure and has gained popularity in financial risk management. Acerbi and Tasche (2002) showed that the ES of their definition is coherent in the sense of Artzner et al. (1999) for a general distribution; whereas the conventional value-at-risk (VaR) measure is not necessarily coherent because of the lack of subadditivity. In this sense, the ES has a theoretical advantage over the VaR.

The purpose of this paper is to propose a new nonparametric estimator of the CES. The proposed estimator is defined as a conditional counterpart of the sample average estimator of the unconditional ES, where the empirical distribution function is replaced by the weighted Nadaraya-Watson (WNW) estimator of the conditional distribution function. We refer to the proposed estimator as the WNW estimator of the CES. Unlike the double kernel estimators of Scaillet (2005) and Cai and Wang (2008), the WNW estimator contains

*Department of Mathematics, Graduate School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan. Email: kkato@hiroshima-u.ac.jp

only a single smoothing. We establish asymptotic normality of the WNW estimator under an α -mixing condition and show that it is design adaptive at interior points of the support of the design distribution.

We compare the WNW estimator with the double kernel estimator of Cai and Wang (2008). As expected, the double kernel estimator contains an additional bias term carried over from the extra smoothing. In addition, by evaluating the variances of the terms which eventually contribute to the asymptotic variances of both estimators, we find that unlike the estimation of the conditional distribution function and the conditional quantile, the double smoothing does not necessarily lead to a higher order variance reduction in the estimation of the CES. Overall, the theoretical performance of the WNW estimator is comparable to that of the double kernel estimator despite the fact that the former is simpler than the latter. The conclusion is parallel to the aforementioned result of Chen (2008). We also conduct simulation experiments to study the finite sample performance of both estimators. In our limited simulation examples, the WNW estimator outperforms the double kernel estimator in the sense of the MSE.

The rest of the paper is organized as follows. In Section 2, we first introduce the definition of the (C)ES and describe our method to estimate the CES. In Section 3, we investigate sampling properties of the proposed estimator. In Section 4, we compare the proposed estimator with the double kernel estimator of Cai and Wang (2008). In Section 5, we report a simulation study. For the technical details of this paper, we refer to the technical report available from <http://www.math.sci.hiroshima-u.ac.jp/stat/TR/TR09/TR09-03.pdf>.

2 Preliminaries

2.1 Definition of CES

In this section, we first introduce the definition of the ES. Suppose we have a random variable Y . In financial risk management, Y is the negative log return of a portfolio. The ES at level $p \in (0, 1)$ of Y is defined as

$$p^{-1}[\mathbb{E}[Y I\{Y \geq q_{1-p}\}] + q_{1-p}\{p - \mathbb{P}(Y \geq q_{1-p})\}],$$

where q_{1-p} is the lower $(1-p)$ -quantile of Y , namely, $q_{1-p} = \inf\{y : \mathbb{P}(Y \leq y) \geq 1-p\}$. The present definition of the ES is due to Acerbi and Tasche (2002). They showed that the ES of their definition is coherent in the sense of Artzner et al. (1999) for a general distribution. If Y has a continuous distribution, the ES reduces to the conventional form

$$\mathbb{E}[Y | Y \geq q_{1-p}],$$

which coincides with the tail conditional expectation of Artzner et al. (1999).

The CES is defined analogously. Suppose we have a response variable Y and an explanatory variable X . A typical example of X is a lagged value of Y . Given $X = x_0$, the CES at level $p \in (0, 1)$ of Y is defined as

$$m_p(x_0) = p^{-1}[\mathbb{E}[Y I\{Y \geq q_{1-p}(x_0)\} | X = x_0] + q_{1-p}(x_0)\{p - \mathbb{P}(Y \geq q_{1-p}(x_0) | X = x_0)\}],$$

where x_0 is a design point, $F(y|x_0)$ is the conditional distribution function of Y given $X = x_0$ and $q_{1-p}(x_0)$ is the conditional lower $(1-p)$ -quantile of Y given $X = x_0$, namely, $q_{1-p}(x_0) = \inf\{y : F(y|x_0) \geq 1-p\}$. It is immediate to see that if $y \mapsto F(y|x_0)$ is continuous, $m_p(x_0)$ reduces to

$$m_p(x_0) = E[Y|Y \geq q_{1-p}(x_0), X = x_0]. \quad (1)$$

Throughout this paper, we assume that the population conditional distribution is continuous so that $m_p(x_0)$ is of the form (1).

2.2 Proposed estimator

Let (Y_t, X_t) , $t = 1, \dots, n$ be observations of (Y, X) . In this section, we introduce our method to estimate the CES and leave some remarks on it. Our basic idea to estimate $m_p(x_0)$ is described as follows. The first step is to estimate the conditional distribution function. Following Hall et al. (1999), we use the WNW estimator:

$$\hat{F}(y|x_0) = \frac{\sum_{t=1}^n p_t(x_0) K_h(X_t - x_0) I(Y_t \leq y)}{\sum_{t=1}^n p_t(x_0) K_h(X_t - x_0)}, \quad (2)$$

where $K(\cdot)$ is a kernel function, h is a bandwidth and $K_h(u) = h^{-1}K(u/h)$. The sequence of nonnegative weights $\{p_t(x_0)\}$ is chosen such that it maximizes $\sum_{t=1}^n \log\{p_t(x_0)\}$ subject to the constraints

$$\sum_{t=1}^n p_t(x_0) = 1, \quad \sum_{t=1}^n p_t(x_0)(X_t - x_0)K_h(X_t - x_0) = 0. \quad (3)$$

The expression of $\{p_t(x_0)\}$ is simplified by introducing the Lagrange multiplier. Let $\hat{\lambda}$ be the maximum point of

$$L_n(\lambda) = \sum_{t=1}^n \log\{1 + \lambda(X_t - x_0)K_h(X_t - x_0)\}.$$

Then, each $p_t(x_0)$ is expressed as

$$p_t(x_0) = \frac{1}{n} \cdot \frac{1}{1 + \hat{\lambda}(X_t - x_0)K_h(X_t - x_0)}.$$

See Fan and Yao (2005), pp.456-457 for a derivation of the expression. Then, the proposed estimator is defined as

$$\hat{m}_p(x_0) = \frac{\sum_{t=1}^n p_t(x_0) K_h(X_t - x_0) Y_t I\{Y_t \geq \hat{q}_{1-p}(x_0)\}}{\sum_{t=1}^n p_t(x_0) K_h(X_t - x_0) I\{Y_t \geq \hat{q}_{1-p}(x_0)\}}.$$

where $\hat{q}_{1-p}(x_0) = \inf\{y : \hat{F}(y|x_0) \geq 1-p\}$ is the lower $(1-p)$ -quantile of $\hat{F}(y|x_0)$. We refer to $\hat{m}_p(x_0)$ as the WNW estimator of the CES.

The proposed estimator can be thought of as a conditional counterpart of the sample average estimator of the unconditional ES which is defined as

$$\frac{\sum_{t=1}^n Y_t I(Y_t \geq \hat{q}_{1-p})}{\sum_{t=1}^n I(Y_t \geq \hat{q}_{1-p})}, \quad (4)$$

where \hat{q}_{1-p} is the lower $(1-p)$ -quantile of the empirical distribution function. Under the condition that the population distribution of Y is continuous, Chen (2008) showed asymptotic normality of the sample average estimator for geometric α -mixing processes.

Instead of the WNW estimator, we may use the NW estimator of the conditional distribution function. We note that Peracchi and Tanase (2008) mentioned a NW type estimator of the CES which they call the fully nonparametric estimator. However, they did not investigate its asymptotic properties. Mimicking the proof of Theorem 1 below, it is not difficult to show that under the same conditions of Theorem 1, the NW estimator of $m_p(x_0)$ is $(nh)^{1/2}$ -consistent and has the same asymptotic variance as $\hat{m}_p(x_0)$ when x_0 is an interior point. However, the NW estimator has well known disadvantages that it is not design adaptive. In some cases of the design distribution, the bias of the NW estimator may become very large. Therefore, we recommend to use the WNW estimator.

3 Sampling properties

In this section, we investigate asymptotic properties of the WNW estimator of the CES. We note that the results of this section are not included in Cai (2001) since $Y_t I\{Y_t \geq \hat{q}_{1-p}(x_0)\}$ depends on the overall data.

In what follows, we use the following notations: $l_j(u|x) = E[Y^j I\{Y \geq u\} | X = x]$ for $j = 1, 2$; $l_1^{(a)}(u|x) = \partial^a l_1(u|x) / \partial x^a$ for $a = 1, 2$; $F^{(a)}(y|x) = \partial^a F(y|x) / \partial x^a$ for $a = 1, 2$; $\mu_j(K) = \int_{-\infty}^{\infty} u^j K(u) du$ for $j = 0, 2$. Let g denote the density of X and x_0 be a interior point of the support of g . The next theorem is the main result.

Theorem 1. *Under conditions (A1)-(A8) (see the technical report), if $h = O(n^{-1/5})$, the map $x \mapsto l_1(q_{1-p}(x_0)|x)$ is twice continuously differentiable in a neighborhood of x_0 and the map $x \mapsto l_2(q_{1-p}(x_0)|x)$ is continuous at x_0 , then as $n \rightarrow \infty$,*

$$(nh)^{1/2} \left\{ \hat{m}_p(x_0) - m_p(x_0) - \frac{h^2 \mu_2(K)}{2} B_p(x_0) \right\} \xrightarrow{d} N \left\{ 0, \mu_0(K^2) \sigma_p^2(x_0) / g(x_0) \right\},$$

where

$$\begin{aligned} B_p(x_0) &= p^{-1} l_1^{(2)}(q_{1-p}(x_0)|x_0) + p^{-1} q_{1-p}(x_0) F^{(2)}(q_{1-p}(x_0)|x_0), \\ \sigma_p^2(x_0) &= p^{-2} l_2(q_{1-p}(x_0)|x_0) - m_p^2(x_0) + (p^{-1} - 1) q_{1-p}(x_0) \{q_{1-p}(x_0) - 2m_p(x_0)\}. \end{aligned}$$

From Theorem 1, we can see that the asymptotic mean squared error (AMSE) of $\hat{m}_p(x_0)$ is

$$\text{AMSE}(x_0; p) = \frac{h^4}{4} \{ \mu_2(K) B_p(x_0) \}^2 + \frac{\mu_0(K^2)}{nhg(x_0)} \sigma_p^2(x_0).$$

Minimizing the AMSE yields the optimal bandwidth

$$h_{\text{opt}}(x_0; p) = n^{-1/5} \left[\frac{\mu_0(K^2) \sigma_p^2(x_0)}{g(x_0) \{ \mu_2(K) B_p(x_0) \}^2} \right]^{1/5}$$

Unfortunately, $h_{\text{opt}}(x_0; p)$ is of a complicated form. The estimation of the unknown quantities in $h_{\text{opt}}(x_0; p)$ is feasible but complex. In practice, we suggest to use the bootstrap approach described in Section 2.3 of Hall et al. (1999), which is simple to implement.

Theorem 1 reveals that the asymptotic bias of $\hat{m}_p(x_0)$ is independent of the design density $g(\cdot)$, that is, $\hat{m}_p(x_0)$ is design adaptive. The design adaptation of $\hat{m}_p(x_0)$ is due to the second part of the property (3) of the weights $\{p_t(x_0)\}$. On the other hand, it is not difficult to see that the asymptotic bias of the NW estimator of $m_p(x_0)$ is

$$\frac{h^2 \mu_2(K)}{2} \left[B_p(x_0) + 2 \frac{g'(x_0)}{g(x_0)} p^{-1} \{l_1^{(1)}(q_{1-p}(x_0)|x_0) + q_{1-p}(x_0) F^{(1)}(q_{1-p}(x_0)|x_0)\} \right].$$

Whereas, the asymptotic variance of the NW estimator is same as that of $\hat{m}_p(x_0)$. It is well known that the dependence on the ratio $g'(x_0)/g(x_0)$ produces a large bias to the estimator in some cases. See Fan (1992) and Section 3.2.4 of Fan and Gijbels (1996). Thus, in terms of the bias property, the WNW estimator is advantageous over the NW estimator.

4 Double kernel estimator

In this section, we compare the WNW estimator with the double kernel estimator of Cai and Wang (2008). The basic idea of Cai and Wang (2008) is plugging some nonparametric estimators of $f(y|x_0)$ and $q_{1-p}(x_0)$ into the expression

$$m_p(x_0) = p^{-1} \int_{q_{1-p}(x_0)}^{\infty} y f(y|x_0) dy.$$

Specifically, combining the WNW method and the double kernel local linear method of Yu and Jones (1998), Cai and Wang (2008) proposed the weighted double kernel local linear estimator of $f(y|x)$:

$$\tilde{f}(y|x_0) = \frac{\sum_{t=1}^n p_t(x_0) K_h(X_t - x_0) W_{h_0}(y - Y_t)}{\sum_{t=1}^n p_t(x_0) K_h(X_t - x_0)},$$

where $W(\cdot)$ is a symmetric kernel function and h_0 is a bandwidth associated with $W(\cdot)$. Let $\tilde{F}(y|x_0)$ be the conditional distribution function with conditional density $\tilde{f}(y|x_0)$:

$$\tilde{F}(y|x_0) = \int_{-\infty}^y \tilde{f}(u|x_0) du.$$

As an estimator of $q_{1-p}(x_0)$, Cai and Wang (2008) proposed to use the solution $\tilde{q}_{1-p}(x_0)$ of the equation $\tilde{F}(\tilde{q}_{1-p}(x_0)|x_0) = 1 - p$, where we note that such $\tilde{q}_{1-p}(x_0)$ always exists since $y \mapsto \tilde{F}(y|x_0)$ is continuous. Let $\tilde{m}_p(x_0)$ denote the corresponding plug-in estimator:

$$\tilde{m}_p(x_0) = p^{-1} \int_{\tilde{q}_{1-p}(x_0)}^{\infty} y \tilde{f}(y|x_0) dy.$$

Cai and Wang (2008) established asymptotic normality of $\tilde{m}_p(x_0)$ for α -mixing processes. We now compare the asymptotic biases and variances of $\hat{m}_p(x_0)$ and $\tilde{m}_p(x_0)$.

[1] Bias: From Theorem 4 of Cai and Wang (2008), the asymptotic bias of $\tilde{m}_p(x_0)$ is

$$\text{bias}\{\tilde{m}_p(x_0)\} : \frac{h^2 \mu_2(K)}{2} B_p(x_0) + \frac{h_0^2 \mu_2(W)}{2} p^{-1} f(q_{1-p}(x_0)|x_0).$$

It is clear that $\tilde{m}_p(x_0)$ contains the additional bias term carried over from the y direction smoothing. Since the first order asymptotic variance of $\tilde{m}_p(x_0)$ does not depend on h_0 , it is reasonable to make h_0 small relative to h as suggested in Cai and Wang (2008). Although the additional term is theoretically negligible if we set $h_0 = o(h)$, it might affect the finite sample performance of $\tilde{m}_p(x_0)$. When h_0 is small relative to h , the additional term improves the bias if $B_p(x_0)$ is negative; otherwise, it has an adverse effect on the estimation accuracy. In a practical situation, we do not know the sign of $B_p(x_0)$ in advance and there is an uncertainty about the effect of the y direction smoothing on the bias. The WNW estimator does not have such an uncertainty and may be preferable in this respect.

[2] Variance: From Theorem 4 of Cai and Wang (2008), the first order asymptotic variances of $\hat{m}_p(x_0)$ and $\tilde{m}_p(x_0)$ are same. However, for the estimation of the conditional distribution function and the conditional quantile, the y direction smoothing improves the variance in a higher order sense; see Azzalini (1981), Yu and Jones (1998), Cai and Roussas (1998), Chen and Tang (2005) and Remark 5 of Cai and Wang (2008), among others. Thus, we need to look at the variances of $\hat{m}_p(x_0)$ and $\tilde{m}_p(x_0)$ more carefully. Since it is difficult to evaluate the exact variances, we evaluate the variances of the terms which eventually contribute to the asymptotic variances of $\hat{m}_p(x_0)$ and $\tilde{m}_p(x_0)$. Although the approximation is formal, we believe that it reflects the true structure of the variance relationship between $\hat{m}_p(x_0)$ and $\tilde{m}_p(x_0)$. From the proof of Theorem 1, the asymptotic variance of $\hat{m}_p(x_0)$ comes from the variance of

$$I_1 = \frac{1}{npg(x_0)} \sum_{t=1}^n b_t(x_0) K_h(X_t - x_0) \{Y_t - q_{1-p}(x_0)\} I\{Y_t \geq q_{1-p}(x_0)\},$$

where $b_t(x_0) = 1/\{1 + \lambda^*(X_t - x_0)K_h(X_t - x_0)\}$ and $\lambda^* = h\mu_2(K)g'(x_0)/\{\mu_2(K^2)g(x_0)\}$. On the other hand, from the proof of Theorem 4 in Cai and Wang (2008), the asymptotic variance of $\tilde{m}_p(x_0)$ comes from the variance of

$$I_2 = \frac{1}{npg(x_0)} \sum_{t=1}^n b_t(x_0) K_h(X_t - x_0) \{Y_t - q_{1-p}(x_0)\} G_{h_0}(Y_t - q_{1-p}(x_0)),$$

where $G(\cdot)$ is the distribution function of $W(\cdot)$ and $G_{h_0}(u) = G(u/h_0)$. Under certain regularity conditions, it is shown that

$$\text{Var}(I_1) - \text{Var}(I_2) = o\{(nh)^{-1}h_0\}. \quad (5)$$

From (5), unlike the estimation of the conditional distribution function and the conditional quantile, the y direction smoothing does not improve the variance at the order of $(nh)^{-1}h_0$ for the estimation of the CES.

Overall, we may conclude that theoretical performance of the WNW estimator is comparable to that of the double kernel estimator despite the fact that the former is simpler than the latter. The conclusion is parallel to the result of Chen (2008) who studied the estimation on the unconditional ES. He compared the sample average estimator with the kernel estimator of Scaillet (2004) which contains the y direction smoothing. He evaluated the variance of the kernel estimator up to the second order and found that the y direction

smoothing does not improve the variance at the second order while it produces an additional bias. He remarked that the simpler sample average estimator is good enough for estimating the ES. In the next section, we study the finite sample performance of both estimators.

5 Simulation study

In this section, we report simulation results which evaluate the finite sample performance of the WNW estimator and the double kernel (DK) estimator of Cai and Wang (2008). Specifically, we consider the following two models.

- ▶ Model 1 [ARCH(1)]: $Y_t = \sigma_t \epsilon_t$, $\sigma_t^2 = 0.30 + 0.65Y_{t-1}^2$, $\epsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$, $X_t = Y_{t-1}$.
- ▶ Model 2: $Y_t = \sin(0.75X_t) + 0.5\epsilon_t$, $X_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$, $\epsilon_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$, $X_t \perp \epsilon_t$.

The ARCH process of Model 1 is shown to be α -mixing with exponential decaying coefficients; see remark (viii) in Fan and Yao (2005), pp. 70. Thus, the process of Model 1 trivially satisfies the mixing conditions of Theorem 1. We note that Model 2 is due to Fan (1992). Throughout this section, we use the Gaussian kernel. The sample size n is 400 for each model. The number of repetitions is 10,000 for each simulation. The computational results of this section were obtained by using the matrix language Ox (Doornik, 2002). To generate ARCH processes, we used the G@RCH 4.2 package (Laurent and Peters, 2006). Figures 1-2 display biases, standard deviations and root mean squared errors of the WNW estimator and the DK estimator at $p = 0.05$ over a set of design points and a set of bandwidth values. Specifically, we choose 0 and the 0.75-quantile of X_t as design points for each model. In addition, we set $h = an^{-1/5}$, where a moves from 0.5 to 1.2 by 0.05 for Model 1 and from 0.6 to 1.3 by 0.05 for Model 2. As Cai and Wang (2008) noted, h_0 is not sensitive to $\tilde{m}_p(x_0)$ as far as h_0 is sufficiently small relative to h . For simplicity, we set $h_0 = 0.1h$. Figures 1-2 show that $\tilde{m}_p(x_0)$ contains an additional bias in comparison with $\hat{m}_p(x_0)$, which corresponds to the observation in the previous section. More importantly, the MSE of $\hat{m}_p(x_0)$ is smaller than that of $\tilde{m}_p(x_0)$ over the set of bandwidth values for both models. Overall, the simulation results support the use of $\hat{m}_p(x_0)$.

References

- Acerbi, C. and Tasche, D. 2002. On coherence of expected shortfall. *Journal of Banking and Finance* 26: 1487-1503
- Artzner, C., Delbaen, F., Eber, J.M. and Heath, D. 1997. Thinking coherently. *RISK* 10: 68-71.
- Artzner, C., Delbaen, F., Eber, J.M. and Heath, D. 1999. Coherent measures of risk. *Mathematical Finance* 9: 203-228.
- Azzalini, A. 1981. A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika* 68: 326-328.

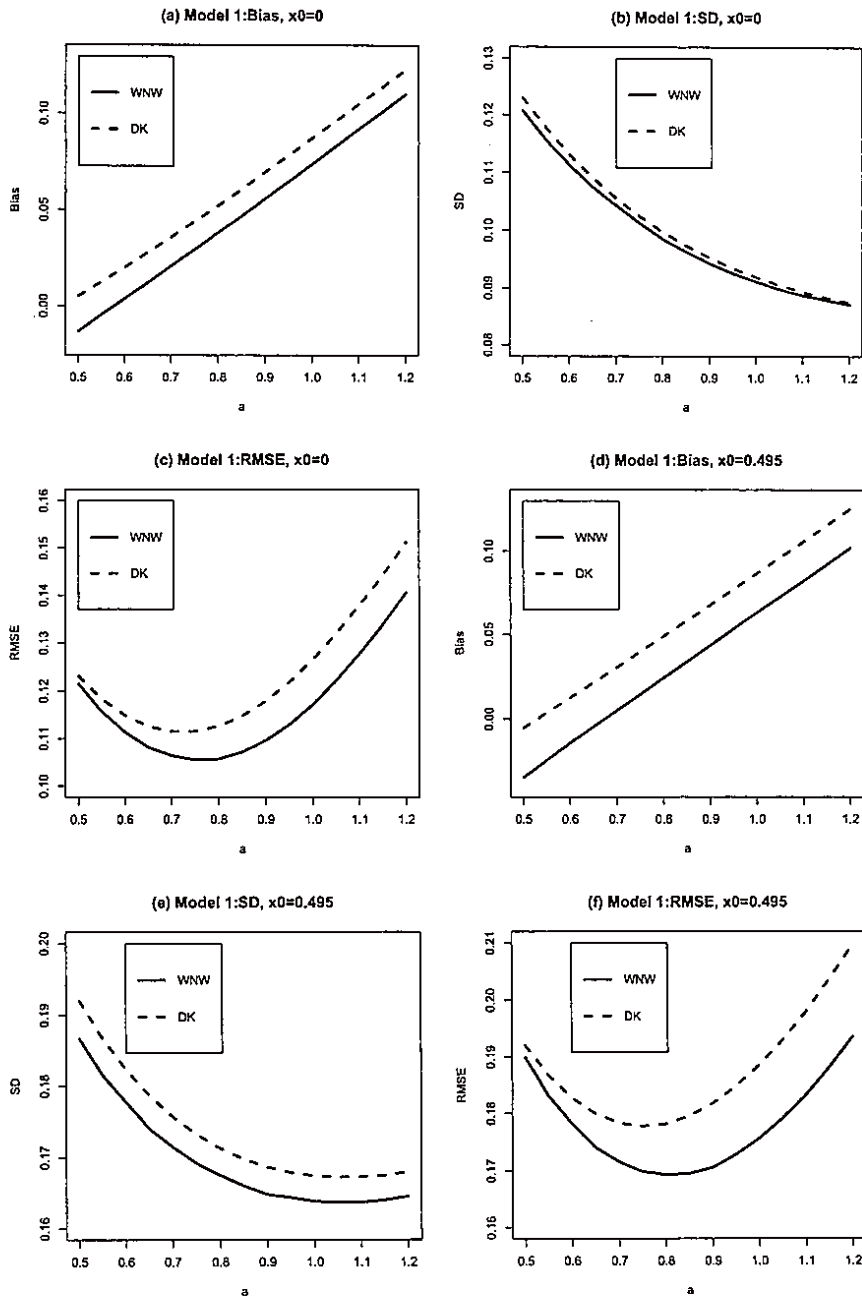


Figure 1: Biases, standard deviations (SD) and root mean squared errors (RMSE) of the WNW estimator and the DK estimator at $p = 0.05$ for Model 1, where $x_0 = \{0, 0.495\}$, $h = an^{-1/5}$ and $h_0 = 0.1h$

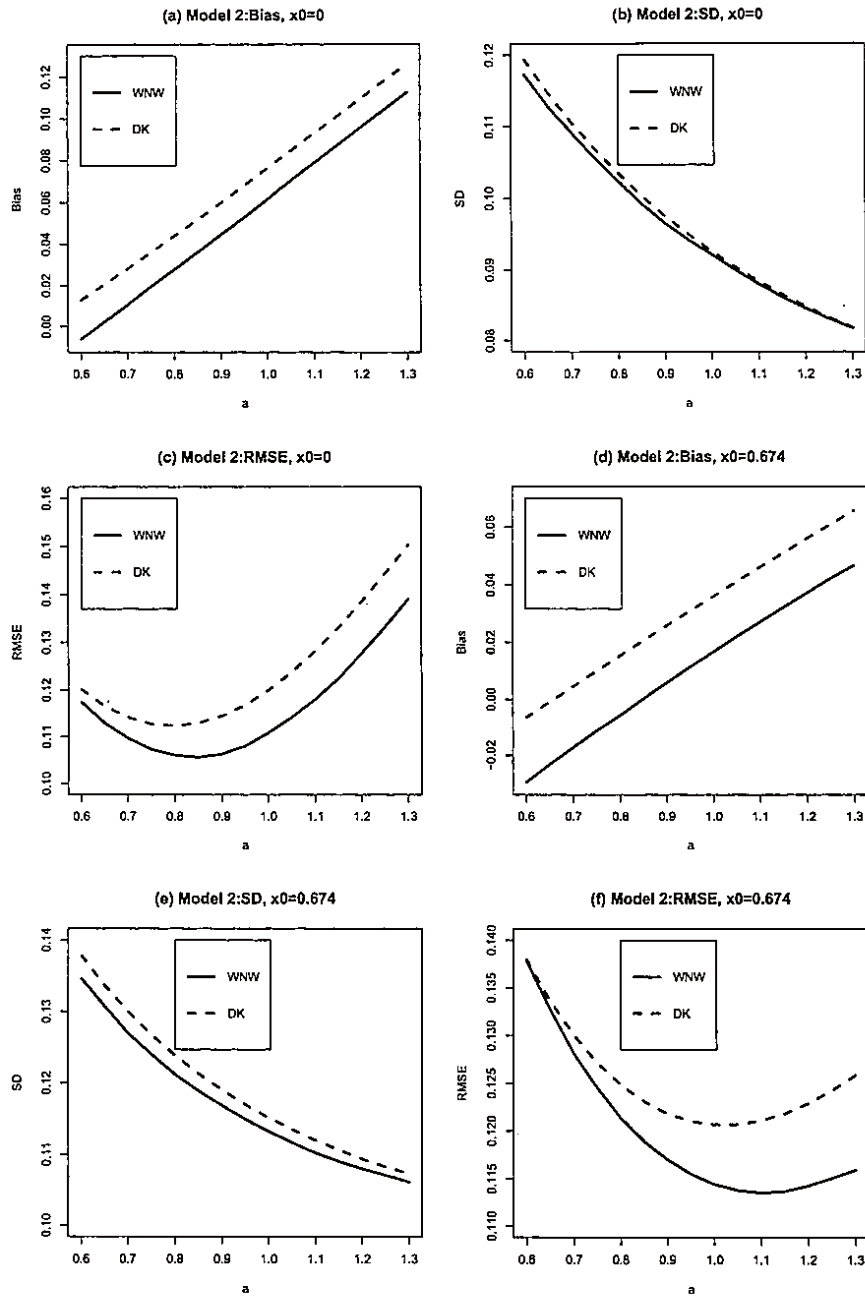


Figure 2: Biases, standard deviations (SD) and root mean squared errors (RMSE) of the WNW estimator and the DK estimator at $p = 0.05$ for Model 2, where $x_0 \in \{0, 0.679\}$, $h = an^{-1/5}$ and $h_0 = 0.1h$

- Cai, Z. 2001. Weighted Nadaraya-Watson regression estimation. *Statistics and Probability Letters* 51: 307-318.
- Cai, Z. and Roussas, G.G. 1998. Efficient estimation of a distribution function under quadrant dependence. *Scandinavian Journal of Statistics* 25: 211-224.
- Cai, Z. and Wang, X. 2008. Nonparametric estimation of conditional VaR and expected shortfall. *Journal of Econometrics* 147: 120-130.
- Chen, S.X. 2008. Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics* 6: 87-107.
- Chen, S.X. and Tang, C.Y. 2005. Nonparametric inference of value at risk for dependent financial returns. *Journal of Financial Econometrics* 3: 227-255.
- Doornik, J.A. 2002. *Object-Oriented Matrix Programming Using Ox*, 3rd edition. Timberlake Consultants Press, London.
- Fan, J. 1992. Design-adaptive nonparametric regression. *Journal of the American Statistical Association* 87: 998-1004.
- Fan, J. and Gijbels, I. 1996. *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.
- Fan, J. and Yao, Q. 2005. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag, New York.
- Hall, P., Wolff, R.C. and Yao, Q. 1999. Methods of estimating a conditional distribution function. *Journal of the American Statistical Association* 94: 154-163.
- Laurent S. and Peters, J.-P. 2006. *G@RCH 4.2, Estimating and Forecasting ARCH Models*. Timberlake Consultants Press, London.
- Peracchi, F. and Tanase, A.V. 2008. On estimating the conditional expected shortfall. *Applied Stochastic Models in Business and Industry* 24: 471-493.
- Scaillet, O. 2004. Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance* 14: 115-129.
- Scaillet, O. 2005. Nonparametric estimation of conditional expected shortfall. *Revue Assurances et Gestion des Risques/Insurance and Risk Management Journal* 74: 639-660.
- Yu, K. and Jones, M.C. 1998. Local linear quantile estimation. *Journal of the American Statistical Association* 93: 228-237.

<信用リスクの計量>

3. 「t 分布 2 ファクターモデルを用いた中小企業 CLO のデフォルト依存関係の分析」吉規寿郎・中川秀敏
4. 「粒子フィルタリングと社債・信用リスク分析」三崎広海
5. 「Self-exciting 性をもつイベント発生強度モデルによる社債ポートフォリオのリスク解析」山中卓・杉原正顯・中川秀敏

t 分布2ファクターモデルを用いた 中小企業CLOのデフォルト依存関係の分析*

吉規 寿郎 (株式会社日本政策金融公庫中小企業事業本部)
中川 秀敏 (一橋大学大学院国際企業戦略研究科)

概要

本研究では、中小企業CLOの組成ポートフォリオのデフォルトリスクを、北野 [?] で導入されている正規分布2ファクターモデルおよび新たに提案する t 分布2ファクターモデルに基づいて分析する。2ファクターモデルを採用することにより、景気変動のような全債務者共通の変動要因(グローバルファクター)に加えて、業種や地域など特定のカテゴリ内のみで共通する変動要因(カテゴリ内共通ファクター)を考慮することが可能になる。また、各ファクターが t 分布に従うと仮定することで、複数の潜在変数の分布について裾部分での強い依存関係が表現可能になることが期待される。具体的には、実際のCLOの組成時点の個社データおよび期中のデフォルトデータに基づいて、両モデルのパラメータの最尤推定を行う。また、推定されたパラメータ値を用いたモンテカルロ・シミュレーションによって裏付け資産プールの損失分布を求めて、モデル間のリスク評価の差異を考察する。分析結果として、カテゴリとファクターの分布に関する他の組合せパターンと比べて、「企業規模」カテゴリ区分に基づく t 分布2ファクターモデルを適用した場合に、デフォルトリスクの高まりが懸念されている案件に対して適度に保守的なリスク評価が与えられたことが確認された。

1. はじめに

1.1. 研究方法および結果の概要

本研究の分析対象は、2004年以降に日本公庫のコーディネートによって組成された日本の中小企業CLOである。2008年3月までに組成された13件の日本公庫CLOは、累計6,000社強が対象となっている。CLO融資を利用した中小企業の実績デフォルト率は、平均的にはスコアリングモデルにより算出されたPD(以下「モデルPD」と言う)の3倍前後で推移しており、格付け会社がモデルPDを保守的に数倍して格付けしていることを勘案すると、全体としては格付け会社が当初に設定していた状況に整合的であると判断できる。その一方で、個々の日本公庫CLOの実績をみると、デフォルトがほとんど発生していない案件と、当初に想定されていたPDを大きく上回るデフォルトが集中して発生して信用補完率が非常に低下している案件に分かれる。後者に該当する2案件の優先受益権及びメザニン受益権が、2009年に入り、

ムーディーズジャパン(以下、「ムーディーズ」と言う)および格付投資情報センター(以下、「R&I」と言う)によって相次いで格下げされており、当初トリプルAを取得していた優先受益権がシングルAクラスまで格下げされている。

中小企業CLOは、以下の点から実証研究の対象として適していると考えられる。まず、通常CLO融資はデフォルト時に回収を見込まない(つまり、LGD=100%)無担保融資として実行されるため、物的担保からの回収も見込んで与信される有担保融資と比べて債務者自身の信用力と純粹に対応していると考えられる。また、いずれも数百社~千社強のポートフォリオであるため、小口分散ポートフォリオの本質的な部分が集約されていて研究対象としては適切なサイズである。加えて、投資家に円滑に販売するための全債務者共通の尺度としてスコアリング審査がベースとなっているため組成当初のモデルPDが明確であり、格付けレポートでも公表されている。さらに、満期までのパフォーマンス

*本稿は吉規・中川 [3] の内容を配布資料用に編集したものである。本稿に関する責任はあくまでも著者個人にあり、本稿の内容は著者の所属する機関とは関係ないことをお断りしておく。また本研究は、日本学術振興会科学研究費補助金基盤研究(A)21243019の補助を一部受けて実施されている。

マンスが格付け会社のモニタリングレポートなどにより開示されているためデータが比較的充実していて、他のリスク分析モデルや事後のパフォーマンスとの比較検証も可能である。

本研究では、中小企業 CLO の裏付け資産プールに関する損失分布を、ファクターモデルによる潜在変数法を適用してモデル化する。具体的には、北野 [?] で導入されている正規分布 2 ファクターモデルに加えて、各ファクターが正規分布ではなく t 分布に従うと仮定した t 分布 2 ファクターモデルを新たに導入する。実際に、日本公庫 CLO の組成時点の個社データおよび期中のデフォルトデータに基づいて、正規分布 2 ファクターモデルと t 分布 2 ファクターモデルのパラメータの最尤推定を行う。そして、推定されたパラメータ値を用いてモンテカルロ・シミュレーションを行い、裏付け資産プールの損失分布を求めリスク評価に関する分析を行う。特に、シミュレーションによるリスク評価の結果を端的に見るためにモデルから推定される格付けに注目し考察を行う。

2 ファクターモデルを採用することにより、景気変動のような全債務者共通の変動要因（グローバルファクター）に加えて、業種や地域など特定のカテゴリ内のみ共通する変動要因（カテゴリ内共通ファクター）を考慮することが可能になる。これによって、組成時に取得される債務者の属性データ（業種、地域、企業規模）に注目して、異なる属性によるカテゴリ区分を用いて分析し結果を比較することで、先行研究よりも適切なカテゴリ区分の選択が行えることが期待できる。また、いくつかの先行研究（北野 [1] など）が示唆するように、潜在変数を特徴付ける各ファクターが t 分布に従うと仮定することで、複数の潜在変数の分布について裾部分での強い依存関係を表現できると考えられる。このことにより、世界的金融危機が叫ばれるような大きなストレスが生じた状況で、多くの企業の信用力が同一方向に変動しやすくなる場合の影響を検証するといった応用が期待できる。

2 ファクターモデルを用いて日本公庫 CLO を分析した結果の概要は、以下のようにまとめられる。(1) パラメータ最尤推定の結果、ファクターの数および分布によらず「地域」「業種」よりも「企業

規模」によるカテゴリ区分を用いた場合に尤度が大きかった。(2) カテゴリ区分を「企業規模」とした t 分布 2 ファクターモデルは統計的に最適であるとは言えないものの、有意にファットテール性があることを示唆する推定結果が得られた。(3) また、同モデルのシミュレーション分析においては、分析期間中に優先受益権がトリプル A からの大幅な格下げを余儀なくされている 2 案件に対して、他のカテゴリ区分とファクターの分布の組合せでは得られなかったほどの相当に低い格付けが推定され、データ分析期間におけるポートフォリオの毀損状況に整合している結果が得られた。

2. t 分布 2 ファクターモデル

t 分布 2 ファクターモデルを導入する主な理由は次の 2 点である。

(1) 北野 [?] の議論および結果をふまえると、カテゴリ区分という考え方を導入することでカテゴリ内相関およびカテゴリ間相関を表現でき、その変動要因をグローバルファクターおよびカテゴリ内共通ファクターに分解できる 2 ファクターモデルは、信用プールのデフォルトリスク評価の精緻化に結びつく可能性があること。

(2) 北野 [1] の数値実験による裾依存性の示唆や吉規・中川 [3] の付録における議論から、グローバルファクターとカテゴリ内共通ファクターが t 分布に従うと仮定することで、潜在変数の周辺分布がファットテール性を持つだけでなく潜在変数の分布間に強い裾依存性が表現できる蓋然性が高いと考えられる。したがって、分布の下裾に位置するファクター実現値がデフォルトと判断されるファクターモデルにおいて、ファクターに t 分布を仮定することでより強いデフォルト依存性が表現可能であると期待されること。

2.1. 定式化

カテゴリ $g \in \{1, \dots, G\}$ に属する債務者 i の t 期における信用力を表す潜在変数 $V_{g,i,t}$ が、カテゴリごとに所与となるデフォルト閾値 θ_g 以下の状態 ($V_{g,i,t} \leq \theta_g$) のとき、債務者 i は t 期においてデフォルト状態であると定義する。

ここで、潜在変数 $V_{g,i,t}$ を以下のファクターで表すことを考える：(1) 景気など全体の変動を表す

全債務者共通のグローバルファクター Y_t (2) カテゴリ内のみで共通する変動を表すカテゴリ内共通ファクター $Z_{g,t}$ (3) 債務者 i 独自の要因を表す固有ファクター $\varepsilon_{i,t}$. 全てのファクター $Y_t, Z_{g,t}, \varepsilon_{i,t}$ は独立であり, 全て自由度 ν ($\nu > 2$) の t 分布に従うと仮定する.

このとき, カテゴリ g に属する債務者 i の t 期における潜在変数 $V_{g,i,t}$ を次のように表すモデルを t 分布 2 ファクターモデルと呼ぶ.

$$V_{g,i,t} = \rho_g(\rho_0 Y_t + \sqrt{1 - \rho_0^2} Z_{g,t}) + \sqrt{1 - \rho_g^2} \varepsilon_{i,t} \quad (1)$$

ただし, $\rho_0, \rho_g \in [0, 1]$ とする.

正規分布の場合と異なり, t 分布に従う確率変数の一次結合がまた t 分布に従うという性質は無いため, (1) の形で与えられる潜在変数 $V_{g,i,t}$ 自体の分布は t 分布にはならない. 異なるカテゴリ g と h の間のカテゴリ内共通ファクター $X_{g,t}$ と $X_{h,t}$ 間の相関は ρ_0^2 , カテゴリ内相関 (同一カテゴリ g に属する債務者の潜在変数 $V_{g,i,t}$ と $V_{g,j,t}$ の相関) は ρ_g^2 , カテゴリ間相関 (異なるカテゴリ g と h に属する債務者間の潜在変数 $V_{g,i,t}$ と $V_{h,j,t}$ の相関) は $\rho_0^2 \rho_g \rho_h$ となる.

$Y_t = y_t, Z_{g,t} = z_{g,t}$ の下でのカテゴリ g に属する債務者の t 期における条件付きデフォルト確率を $p_g(y_t, z_{g,t})$ と表記すると, $\varepsilon_{i,t}$ が自由度 ν の t 分布に従うので次式のとおりとなる.

$$p_g(y_t, z_{g,t}) = T_\nu \left(\frac{\theta_g - \rho_g(\rho_0 y_t + \sqrt{1 - \rho_0^2} z_{g,t})}{\sqrt{1 - \rho_g^2}} \right) \quad (2)$$

ただし, $T_\nu(\cdot)$ は自由度 ν の t 分布関数を表す.

2.2. 最尤推定法

(2) 式に含まれるパラメータ

$$\rho_0, \{\rho_g\}_{g=1, \dots, G}, \{\theta_g\}_{g=1, \dots, G}, \nu$$

の推定には, 北野 [?] で用いられている正規分布 2 ファクターモデルに関する最尤法を参考とする.

G 個のカテゴリのうちカテゴリ g に属する t 期初 ($t = 1, 2, \dots, T$) の生存債務者数を $N_{g,t-1}$, t 期中に発生したデフォルト債務者数を $D_{g,t}$ とする. $Y_t, Z_{g,t}$ はそれぞれ t に関して独立かつ自由度 ν の t 分布に従うとする. t 期において

$Y_t = y_t, Z_{g,t} = z_{g,t}$ の条件の下ではデフォルト事象は条件付き独立であるので, デフォルト債務者数 $D_{g,t}$ を対象事象の発生回数は, デフォルト発生確率 $p_g(y_t, z_{g,t})$ を対象事象の発生確率, t 期初の生存債務者数 $N_{g,t-1}$ を試行回数と見なす 2 項分布に従う.

$f(D_{g,t}; N_{g,t-1}, p_g(y_t, z_{g,t}))$ をこの 2 項分布の密度関数とする. これを y_t および $z_{g,t}$ に関して, 自由度 ν の t 分布の密度関数によって積分して無条件化することで, t 分布 2 ファクターモデルに関する対数尤度関数

$$\begin{aligned} \ell(\rho, \theta, \nu) &= \sum_{t=1}^T \log \left[\int_{-\infty}^{\infty} F(y_t) t_\nu(y_t) dy_t \right], \\ F(y_t) &:= \prod_{g=1}^G \int_{-\infty}^{\infty} f(D_{g,t}; N_{g,t-1}, p_g(y_t, z_{g,t})) t_\nu(z_{g,t}) dz_{g,t} \end{aligned}$$

を得る. ただし, $t_\nu(\cdot)$ は自由度 ν の t 分布の密度関数とし, 全カテゴリに関するファクターウエイトおよびデフォルト閾値を要素とするベクトルをそれぞれ $\rho = (\rho_0, \rho_1, \dots, \rho_G)$ および $\theta = (\theta_1, \dots, \theta_G)$ とする.

2.3. CLO 格付けの検証方法

CLO は優先・メザニン・劣後などのトランシェに区分され, 裏付け資産が生み出すキャッシュフローは優先順位に従って配分される. 各トランシェにとっては, 下位にあるトランシェが「信用補完」となるため, 上位トランシェの信用力が相対的に高く, 下位トランシェの信用力が相対的に低くなる. CLO の格付けに際しては, 裏付け資産である中小企業ポートフォリオについて損失分布を得たうえで, 各トランシェの上下限に切り分けてトランシェごと期待損失率を計算し, 表 1 で与えられるような格付けと期待損失率の対応テーブルに照らして格付けが付与される.

本研究では, 前節の手法により推定されたパラメータに基づき, 対象 CLO の裏付け資産について以下のような方法でモンテカルロ・シミュレーションを行い, 損失分布を生成する.

1. 各試行ごとに, 正規分布もしくは自由度 ν の t 分布に従う乱数を, グローバルファクター

Y_t 用に1つ、カテゴリ内共通ファクター $Z_{g,t}$ 用にカテゴリ数（本研究ではどのモデルでも3つとしている）、固有ファクター $\varepsilon_{i,t}$ 用に各CLOの債務者数だけ発生させる

2. 1. で発生した乱数と最尤推定により得られたパラメータを(2)式に代入して債務者ごとのデフォルト確率を求める
3. 2. で求めたデフォルト確率と、モデルPDを保守的に修正した債務者ごとのPDとを比較して各債務者のデフォルト時刻を求める
4. 3. で求めたデフォルト時刻が満期時点前の債務者について、その時点での約定残高（これを当該債務者のデフォルト金額と見なす）の合計をCLO組成金額で除したものを、当該試行での裏付け資産プールの損失率とする
5. 1~4を100万回繰り返すことで、CLOポートフォリオの損失分布を生成する

あるトランシェの境界を、総資産を1とした比率で下限 α 、上限 β と表す ($0 \leq \alpha < \beta \leq 1$) と表す。また、試行 j 回目の裏付け資産プールの損失率を $p^{(j)}$ とする。このとき、試行 j 回目における対象トランシェの損失率は、下限 α と上限 β およびプール損失率 $p^{(j)}$ により

$$\min\left\{1, \max\left(0, \frac{p^{(j)} - \alpha}{\beta - \alpha}\right)\right\}$$

と計算される。これを100万回繰り返しその平均を対象トランシェの期待損失とし、表1に照らし合わせてモデルが妥当とする格付けを推定する。

3. 研究結果

3.1. 実データによるパラメータの最尤推定

まず、日本公庫CLO13件のうち、デフォルト先の詳細な情報が開示されているシンセティックCLOのデフォルト実績データを、2007年3月から2009年6月まで28カ月分用いて、2.2節で述べた最尤推定の方法に基づいて、正規分布2ファクターモデルおよび t 分布2ファクターモデルのパラメータを最尤推定する。

¹当初データには売上規模による債務者の分布が開示されているが、期中のデフォルト先については企業の特定を避けるため、当初借入金額のみの公表となっている。このため、パラメータの推定においては、借入金額がある程度企業規模を表していると仮定して借入金額の規模によるカテゴリ区分に基づいて推定を行う。一方、CLOの評価においては、借入金額の規模によるカテゴリ区分に基づいて分析すると金額ベースでのCLO評価にバイアスがかかってしまうため、売上規模ベースで分析を行う。

本研究では、カテゴリ区分として、(1) 地域（大都市<東京都、大阪府および愛知県>、都市圏<先の3都府県を除く関東、関西および東海>、その他の地方）、(2) 業種（製造業、卸売・小売業、その他）、(3) 企業規模¹（当初借入金額20百万円未満、同20百万円以上40百万円未満、同40百万円以上）を用いる。なお、日本公庫CLOにおいては、地域、業種、企業規模について、より細かい区分での開示が行われているが、サンプル数の制約から、いずれのカテゴリも3区分としている。

各カテゴリのデフォルト閾値 $\theta_1, \theta_2, \theta_3$ 、グローバルファクターおよびカテゴリ内共通ファクターに対するファクターウエイトを表すパラメータ $\rho_0, \rho_1, \rho_2, \rho_3$ ならびに自由度 ν (t 分布2ファクターモデルのみ)の最尤推定結果は表3および表4のとおりである。各パラメータについて最尤推定値と推定量の標準偏差である標準誤差（カッコ内の数値）を示している。

1ファクターモデルに対する推定結果は吉規・中川[3]の付録を参照されたい。全体的に言えることは、正規分布か t 分布か、あるいは1ファクターか2ファクターかという違いによらず、相関の大きさはほぼ同じ傾向が示されていることである。例えば、カテゴリ区分を地域とした場合は、地方でのカテゴリ内相関が最も高く、大都市、都市圏では同じ水準となり、カテゴリ区分を業種とした場合、その他の業種でのカテゴリ内相関が最も高く、卸・小売業がそれに次ぎ、製造業が最も低い結果となった。

また、どのカテゴリ区分が有用であるかという点について対数尤度を比較すると、対数尤度はカテゴリ区分を企業規模とした場合が最も高く、地域による区分がそれに次ぎ、業種による区分で最も低くなっている。特に、企業規模が大きいカテゴリでのカテゴリ内相関 ρ_3^2 が相対的に高くなっていることが分かる。2009年上半期に比較的規模の大きい（資本金40百万円以上）中小企業にデフォルトが集中して発生していたことがあり、そのこ

とが企業規模の区分の大きいカテゴリでの相関が大きくなったことにつながっている可能性もある。

さらに、どのモデルが最良かということについて AIC に注目して比較すると、 t 分布 1 ファクターモデルおよび正規分布 2 ファクターモデルはほぼ同様の水準で高いが、 t 分布 2 ファクターモデルは AIC がやや大きいことが確認される。このことは、AIC 最小をモデル選択基準とする場合には、パラメータ数が最も多い t 分布 2 ファクターモデルが必ずしも最良とは主張できないことを示唆しているが、モデル選択基準については様々な議論があり、今後の課題としたい。

また、 t 分布 2 ファクターモデルの自由度パラメータ ν の推定値を見ると、いずれのカテゴリ区分においても 6~8 前後となっている。正規分布を示す $\nu = \infty$ からはかなり離れた結果であり、実際の中小企業 CLO ポートフォリオの損失分布は正規分布 2 ファクターモデルが想定する損失分布よりも裾の厚い（ファットテール性のある）ものになる可能性が示された。

しかし、ファットテール性に寄与するもう 1 つのパラメータ ρ の推定値は各カテゴリ区分で見ると、その大小関係の傾向は両モデルで同様であるものの、 t 分布 2 ファクターモデルよりも正規分布 2 ファクターモデルの方が ρ_1, ρ_2 では高くなっており、先行研究のように ρ が一定の下で自由度 ν だけ低下するという結果にはなっていない。そこで、本節で得られた自由度パラメータと相関関係のバランスについて、次小節においてモンテカルロ・シミュレーションによって得られる損失分布を通じて確認することとする。

3.2. 既発行 CLO に対する損失分布シミュレーション分析の結果

本研究では、日本公庫 CLO13 案件のうち、地域金融機関平成 18 年 9 月 CLO および 19 年 6 月 CLO 並びに地域金融機関 CLO シンセティック型（株式会社たんぼぼ 2007）について、前小節で得られたパラメータ推定値を代入した正規分布 2 ファクターモデルおよび本研究で提案する t 分布 2 ファクターモデルを用いて、2.3 節で述べた方法でモンテカルロ・シミュレーションにより損失分布を算出する。

これらの案件の概要は表 2 のとおりである。この 3 案件は、他の案件に比べて実績累積デフォルト率が非常に高くなっており、うち 2 案件（18 年 9 月 CLO および 19 年 6 月 CLO）では大幅な格下げも起きていて、投資家からも多くの質問が寄せられている特徴的な案件である。

また、シミュレーションの設定としては以下の 4 通りの組合せをそれぞれ試行し、裏付け資産プールの損失分布に関する VaR 値や、得られた損失状況について各トランシェの上下限で切り分けて求めたトランシェごとの期待損失率を表 3 に基づいて変換して得られる推定格付けについて考察する。

カテゴリ区分を「地域」とする正規分布 2 ファクターモデル（「地域×正規」）

カテゴリ区分を「地域」とする t 分布 2 ファクターモデル（「地域× t 」）

カテゴリ区分を「企業規模」とする正規分布 2 ファクターモデル（「規模×正規」）

カテゴリ区分を「企業規模」とする t 分布 2 ファクターモデル（「規模× t 」）

「地域×正規」に対するシミュレーションを行うのは、格付け会社が債務者の地域もしくは業種の集中に配慮しながら正規コピュラモデルを使用して中小企業 CLO の格付けを行っていることに着目して、ある程度は格付け会社の評価に近い結果が出ることを期待したことが理由である。同じカテゴリに対してモデルのアウトプットの違いを見るために、「地域× t 」についてもシミュレーションを行う。

また、前小節のパラメータ推定においていずれのモデルでも「企業規模」によるカテゴリ区分の場合に、対数尤度が最大（AIC が最小）であったことから、「規模×正規」および「規模× t 」についてシミュレーションを行う。

ただし、このシミュレーションによる損失分布の分析は、上記 3 案件を含むデフォルト実績データを用いてパラメータ推定をしておいて、各 3 案件の組成時点まで遡ってリスク評価を企図するものである。その意味で事後的な分析であり、フォワード・ルッキングなリスク評価や実際の格付けの妥当性についての議論に結びつけるには注意が

不可欠である。

しかし、事後的であって、なおかつ特徴的な3案件に対してだけであったとしても、 t 分布2ファクターモデルが正規分布2ファクターモデルと異なるリスク評価を与えること、または同じモデルであってもカテゴリ区分の違いによるリスク評価の差があることが確認できれば、 t 分布2ファクターモデルについて考える意義が十分にあるものとする。

3.2.1. 平成18年9月CLOおよび平成19年6月CLOの分析結果

地域金融機関平成18年9月CLOおよび平成19年6月CLOは両案件ともに、当初の想定を大きく上回るデフォルトが発生している（年率換算で3%前後）。このペースでデフォルトが発生し続けると信用補完となっている劣後受益権部分を超過して優先およびメザニン受益権に棄損が及ぶ可能性がある。このため、両受益権は、2009年1月に1~2ノッチ、さらに2009年5月末には再度格下げされて、2009年11月現在では、優先受益権がシングルAクラスに、メザニン受益権はトリプルBクラスになっている。

損失分布のシミュレーション結果に基づく期待損失率および分布の上裾部分の厚みを示すVaR値（いずれも、裏付け資産プールに関して当初原債権金額に対する満期までの累積デフォルト損失率に対する値）に関して、平成18年9月CLOについては表5の上段、平成19年6月CLOについては表5の中段にまとめている。

また、各シミュレーションによって得られた期待損失率に基づく推定格付けは、平成18年9月CLOについては表6の上段、平成19年6月CLOについては表6の中段にまとめている。

いずれのカテゴリ区分でも t 分布2ファクターモデルの方が正規分布2ファクターモデルよりも明らかに保守的なリスク評価を与えているのは、ファットテール性または裾依存性が寄与している可能性が高いと考えられる。また、「企業規模」カテゴリ区分の方が保守的な評価を与えているのは、現下のような強いストレス下で、借入金額が比較的多い企業規模の大きい層にデフォルトが集中して、金額ベースでの損失率が件数ベースの損失率

よりも大幅に高くなった推定結果がそのまま反映していると推察される。しかし、格付け会社が主に注目している地域や業種以外のカテゴリ区分が、リスク評価に有効となりうることを示唆していると思えることはできると考える。

以上をまとめると、「規模× t 」の組合せについては、18年9月CLOおよび19年6月CLOのデータ分析期間におけるポートフォリオの毀損状況に整合している結果が得られた一方で、それ以外の組合せに対する結果は同期内のポートフォリオの毀損状況とは整合していないことがうかがえる。事後的かつ限られたサンプルに対する分析に基づく考察であるので一般的な断定はできないが、ファクターの分布だけでなくカテゴリ区分の与え方も、ポートフォリオのデフォルトリスク計測において重要となると考えられる。

3.2.2. 株式会社たんぼぼ2007の分析結果

地域金融機関CLOシンセティック型（株式会社たんぼぼ2007）についても、デフォルトが年率3%を超える水準で発生している。しかし、原債権の平均年数が1.56年と短いことから信用補完水準には相当の余裕がある。

同CLOについての損失分布シミュレーション結果に基づく裏付け資産プールの期待損失率およびVaR値は、表5の下段にまとめている。

また、各シミュレーションによって得られた期待損失率に基づく推定格付けは、表6の下段にまとめている。

たんぼぼ2007に関しては組成から2年半たった現時点でも格付けの見直しは行われておらず、2ファクターモデルで分析した結果はいずれのパターンにおいても現状のデフォルト状況と整合的であることが示唆されている。この結果は、18年9月CLOおよび19年6月CLOに対する分析において「規模× t 」が保守的なリスク評価を与えていたこととは対照的であり、「規模× t 」パターンが必ずしも保守的なリスク評価を与えるわけではないことを表している。

また、一般的な言及はできないものの、今回取り上げなかった残りの案件についてはメザニン毀損可能性がたんぼぼ2007と同程度かそれより低いより状況であることを考えると、いずれのパターン

のシミュレーション分析においてもたんぼぼ 2007 と同様の結果が得られると推察される。

参考文献

- [1] 北野利幸, 「デフォルト相関に関する t 分布ファクターモデル～CDO 評価への応用」, ジャフイー・ジャーナル『金融工学と証券市場の計量分析 2006』, 83-117 (2006)
- [2] 北野利幸, 「デフォルト実績データによるデフォルト依存関係の推定～2 ファクターモデルによる資産相関の最尤推定～」, 日本オペレーションズ・リサーチ学会和文論文誌, 50, 42-67 (2007)
- [3] 吉規寿郎, 中川秀敏, 「 t 分布 2 ファクターモデルを用いた中小企業 CLO のデフォルト依存関係の分析」, ジャフイー・ジャーナル『定量的信用リスク評価とその応用』, 掲載予定 (2010)

表 1: 格付けと期待損失率の対応テーブル (Moody's 公表). 各年限ごとに左列の格付け付与に必要な期待損失率の上限。

	1年	2年	3年
Aaa	0.0000%	0.0001%	0.0004%
Aa1	0.0003%	0.0017%	0.0055%
Aa2	0.0007%	0.0044%	0.0143%
Aa3	0.0017%	0.0105%	0.0325%
A1	0.0032%	0.0204%	0.0644%
A2	0.0060%	0.0385%	0.1221%
A3	0.0214%	0.0825%	0.1980%
Baa1	0.0495%	0.1540%	0.3080%
Baa2	0.0935%	0.2585%	0.4565%
Baa3	0.2310%	0.5775%	0.9405%

表 2: 分析対象とした日本公庫 CLO の概要: 地域金融機関平成 18 年 9 月 CLO および平成 19 年 6 月 CLO はキャッシュ (債権譲渡) 方式で期間は 5 年, 地域金融機関 CLO シンセティック型 (株式会社たんぼぼ 2007) は期間が 3 年である。

	18 年 9 月 CLO	19 年 6 月 CLO	たんぼぼ 2007
債権元本総額	12,053 百万円	13,488 百万円	30,484 百万円
債務者数	338 社	355 社	1,190 社
原債権の平均年限	約 2.6 年	約 2.6 年	約 1.56 年
優先トランシェ	10,500 百万円	11,900 百万円	26,900 百万円
格付け	Aaa	Aaa	Aaa
(信用補完率)	(12.9%)	(11.8%)	(11.8%)
メザニントランシェ	250 百万円	180 百万円	480 百万円
格付け	A3	A3	A2
(信用補完率)	(10.8%)	(10.4%)	(10.2%)
信託設定日	2006 年 9 月 27 日	2007 年 6 月 28 日	2007 年 2 月 28 日
予定償還日	2011 年 10 月 17 日	2012 年 7 月 17 日	2010 年 4 月 30 日

表 3: 日本公庫 CLO データに対する正規分布 2 ファクターモデルの最尤推定結果. カッコ内の数値は各推定値の標準誤差. また, [] 内の数値は AIC (赤池情報量規準)

	地域	業種	企業規模
カテゴリ 1	大都市	製造業	20 百万円未満
カテゴリ 2	都市圏	卸・小売	20 百万円以上 40 百万円未満
カテゴリ 3	地方	その他	40 百万円以上
θ_1	-2.715 (0.78)	-3.040 (1.03)	-3.302 (1.05)
θ_2	-2.750 (0.32)	-2.591 (0.58)	-2.668 (0.14)
θ_3	-3.161 (0.83)	-3.108 (1.01)	-3.195 (1.01)
ρ_0	0.707 (1.00)	0.707 (1.00)	0.707 (1.00)
ρ_1	0.041 (1.00)	0.045 (1.00)	0.114 (1.00)
ρ_2	0.041 (1.00)	0.051 (1.00)	0.114 (1.00)
ρ_3	0.060 (1.00)	0.066 (1.00)	0.284 (1.00)
対数尤度 [AIC]	-7.67 [29.3]	-8.44 [30.9]	-5.04 [24.1]
カテゴリ内相関平均	0.0023	0.0030	0.0354
ρ_1^2	0.0017	0.0021	0.0129
ρ_2^2	0.0017	0.0026	0.0129
ρ_3^2	0.0036	0.0044	0.0804
カテゴリ間相関平均	0.0011	0.0014	0.0129
$\rho_0^2 \rho_1 \rho_2$	0.0009	0.0012	0.0064
$\rho_0^2 \rho_2 \rho_3$	0.0012	0.0017	0.0161
$\rho_0^2 \rho_3 \rho_1$	0.0012	0.0015	0.0161

表 4: 日本公庫 CLO データに対する t 分布 2 ファクターモデルの最尤推定結果. カッコ内の数値は各推定値の標準誤差. また, [] 内の数値は AIC (赤池情報量規準)

	地域	業種	企業規模
カテゴリ 1	大都市	製造業	20 百万円未満
カテゴリ 2	都市圏	卸・小売	20 百万円以上 40 百万円未満
カテゴリ 3	地方	その他	40 百万円以上
θ_1	-3.798 (0.75)	-4.694 (1.00)	-5.211 (1.00)
θ_2	-3.861 (0.62)	-3.607 (1.00)	-3.552 (1.00)
θ_3	-5.236 (0.84)	-4.818 (1.00)	-4.075 (1.00)
ρ_0	0.707 (1.00)	0.707 (1.00)	0.707 (1.00)
ρ_1	0.036 (1.00)	0.032 (1.00)	0.069 (1.00)
ρ_2	0.036 (0.88)	0.056 (1.00)	0.054 (1.00)
ρ_3	0.208 (0.96)	0.109 (1.00)	0.376 (1.00)
ν	6.91 (1.03)	6.94 (1.00)	7.88 (1.00)
対数尤度 [AIC]	-7.58 [31.2]	-8.41 [32.8]	-4.96 [25.9]
カテゴリ内相関平均	0.0153	0.0053	0.0497
ρ_1^2	0.0013	0.0010	0.0048
ρ_2^2	0.0013	0.0032	0.0029
ρ_3^2	0.0432	0.0118	0.1414
カテゴリ間相関平均	0.0027	0.0019	0.0083
$\rho_0^2 \rho_1 \rho_2$	0.0006	0.0009	0.0019
$\rho_0^2 \rho_2 \rho_3$	0.0037	0.0031	0.0101
$\rho_0^2 \rho_3 \rho_1$	0.0037	0.0017	0.0130

表 5: 各 CLO に対して、表 3, 表 4 の「地域」および「企業規模」区分で推定した両モデルのパラメータ値を用いて、100 万回のモンテカルロ・シミュレーションによって求めた原債権の損失分布の期待損失および VaR 値の比較。表中の値は累積デフォルト損失額（上段）当初原債権金額に対する満期までの累積デフォルト損失率（中段）およびその年率換算値（下段）。

◆ 18 年 9 月 CLO

	モデル	期待損失	80%VaR	90%VaR	95%VaR	99%VaR	99.9%VaR
地	正規分布	751 百万円	885 百万円	963 百万円	1,029 百万円	1,159 百万円	1,309 百万円
	2ファクター モデル	6.23% (2.27%)	7.34% (2.67%)	7.99% (2.90%)	8.54% (3.10%)	9.61% (3.50%)	10.86% (3.95%)
域	t 分布	751 百万円	910 百万円	1,014 百万円	1,109 百万円	1,320 百万円	1,666 百万円
	2ファクター モデル	6.23% (2.27%)	7.55% (2.74%)	8.41% (3.06%)	9.20% (3.35%)	10.95% (3.98%)	13.82% (5.03%)
借 入 規 模	正規分布	593 百万円	799 百万円	965 百万円	1,120 百万円	1,463 百万円	1,918 百万円
	2ファクター モデル	4.92% (1.79%)	6.63% (2.41%)	8.00% (2.91%)	9.29% (3.38%)	12.14% (4.41%)	15.91% (5.79%)
	t 分布	594 百万円	789 百万円	985 百万円	1,195 百万円	1,765 百万円	2,841 百万円
	2ファクター モデル	4.92% (1.79%)	6.55% (2.38%)	8.17% (2.97%)	9.91% (3.60%)	14.64% (5.32%)	23.57% (8.57%)

◆ 19 年 6 月 CLO

	モデル	期待損失	80%VaR	90%VaR	95%VaR	99%VaR	99.9%VaR
地	正規分布	779 百万円	924 百万円	1,009 百万円	1,081 百万円	1,223 百万円	1,390 百万円
	2ファクター モデル	5.78% (2.10%)	6.85% (2.49%)	7.48% (2.72%)	8.02% (2.92%)	9.07% (3.30%)	10.31% (3.75%)
域	t 分布	780 百万円	957 百万円	1,074 百万円	1,181 百万円	1,420 百万円	1,825 百万円
	2ファクター モデル	5.79% (2.10%)	7.09% (2.58%)	7.96% (2.90%)	8.76% (3.18%)	10.53% (3.83%)	13.53% (4.92%)
借 入 規 模	正規分布	615 百万円	828 百万円	999 百万円	1,161 百万円	1,516 百万円	1,996 百万円
	2ファクター モデル	4.56% (1.66%)	6.14% (2.23%)	7.41% (2.69%)	8.61% (3.13%)	11.24% (4.09%)	14.80% (5.38%)
	t 分布	614 百万円	815 百万円	1,016 百万円	1,229 百万円	1,798 百万円	2,892 百万円
	2ファクター モデル	4.56% (1.66%)	6.04% (2.20%)	7.53% (2.74%)	9.11% (3.31%)	13.33% (4.85%)	21.44% (7.80%)

◆ たんぽぽ 2007

	モデル	期待損失	80%VaR	90%VaR	95%VaR	99%VaR	99.9%VaR
地	正規分布	1,199 百万円	1,427 百万円	1,581 百万円	1,720 百万円	2,014 百万円	2,392 百万円
	2ファクター モデル	3.93% (2.25%)	4.68% (2.67%)	5.18% (2.96%)	5.64% (3.22%)	6.61% (3.78%)	7.85% (4.48%)
域	t 分布	1,194 百万円	1,393 百万円	1,527 百万円	1,651 百万円	1,938 百万円	2,441 百万円
	2ファクター モデル	3.92% (2.24%)	4.57% (2.61%)	5.01% (2.86%)	5.42% (3.09%)	6.36% (3.63%)	8.01% (4.58%)
借 入 規 模	正規分布	939 百万円	1,183 百万円	1,361 百万円	1,526 百万円	1,879 百万円	2,350 百万円
	2ファクター モデル	3.08% (1.76%)	3.88% (2.22%)	4.47% (2.55%)	5.00% (2.86%)	6.16% (3.52%)	7.71% (4.41%)
	t 分布	941 百万円	1,152 百万円	1,308 百万円	1,458 百万円	1,815 百万円	2,467 百万円
	2ファクター モデル	3.09% (1.76%)	3.78% (2.16%)	4.29% (2.45%)	4.78% (2.73%)	5.96% (3.40%)	8.09% (4.62%)

表 6: 各 CLO に対する当初格付けと本研究両モデルによる CLO の評価：優先およびメザニントランシェの想定格付けは、ムーディーズの「理想化された期待損失テーブル」(表 1) とシミュレーションで算出した累積期待損失率を比較して試算している。毀損回数は 100 万回の試行のうち累積損失がトランシェまで及んだ回数を表す。また、当初格付け欄に記載した累積期待損失率は、同テーブルを基に各当初格付けを得るために要求される期待損失率の上限。

◆ 18 年 9 月 CLO		優先受益権		メザニン受益権	
		累積期待損失率 (毀損回数)	想定格付け	累積期待損失率 (毀損回数)	想定格付け
	当初格付け	[0.00033%以下]	[Aaa]	[0.1691%以下]	[A3]
地域	正規分布 2ファクター	0.00001% (16)	Aaa	0.06224% (3,203)	A2
	t分布 2ファクター	0.00275% (1,444)	Aa1	0.63967% (18,546)	Baa3
	借入規模	正規分布 2ファクター	0.01193% (6,320)	Aa3	1.51030% (29,062)
	t分布 2ファクター	0.06328% (15,136)	A2	2.65857% (43,562)	投資不適格
◆ 19 年 6 月 CLO		優先受益権		メザニン受益権	
		累積期待損失率 (毀損回数)	想定格付け	累積期待損失率 (毀損回数)	想定格付け
	当初格付け	[0.00033%以下]	[Aaa]	[0.1691%以下]	[A3]
地域	正規分布 2ファクター	0.00004% (88)	Aaa	0.04730% (1,382)	A1
	t分布 2ファクター	0.00731% (4,503)	Aa2	0.83184% (14,192)	Baa3
	借入規模	正規分布 2ファクター	0.01500% (8,657)	Aa3	1.34379% (19,538)
	t分布 2ファクター	0.06832% (19,486)	A2	2.53904% (32,629)	投資不適格
◆ たんばぼ 2007		優先受益権		メザニン受益権	
		累積期待損失率 (毀損回数)	想定格付け	累積期待損失率 (毀損回数)	想定格付け
	当初格付け	[0.00008%以下]	[Aaa]	[0.03038%以下]	[A2]
地域	正規分布 2ファクター	0.00000% (0)	Aaa	0.01099% (1,488)	A1
	t分布 2ファクター	0.00003% (12)	Aaa	0.02158% (1,600)	A2
	借入規模	正規分布 2ファクター	0.00000% (0)	Aaa	0.02436% (2,830)
	t分布 2ファクター	0.00004% (14)	Aaa	0.02673% (1,535)	A2

粒子フィルタリングと社債・信用リスク分析

三崎 広海*

概要

信用リスク・モデルの一種である誘導型モデルにおいて、粒子フィルタリングの手法を用いてインプライド・デフォルト強度を推定する。デフォルト強度を状態変数に含めることで、デフォルト強度と社債の市場価格の関係が一般化状態空間モデルとして記述され、粒子フィルタによる推定が可能となる。実証分析は Takahashi and Sato (2001) のモデルを拡張し、日本国内市場で観測される固定利付普通社債に対してモンテカルロ・フィルタを適用した。結果はデフォルト強度の推定値については概ね合理的であり、この種の推定問題に対するモンテカルロ・フィルタの実用性が示された。

1 はじめに

誘導型の信用リスク・モデルは intensity model などとも呼ばれ、デフォルトを引き起こす内在的なメカニズムは問題にせず、デフォルトの“起こりやすさ”を表す過程をモデル化するものであり、デフォルトはある点過程のジャンプの発生として定義される。このタイプのモデルは無リスク債券の期間構造に関する議論と数学的に類似しており、状態変数を基礎としたモデルと整合的である。誘導型モデルにおいてジャンプの発生確率を与える変数をデフォルト強度 (default intensity) といい、これが大きいほど信用リスクが高いみなすことができる。

デフォルト強度の推定は誘導型モデルにおける研究課題の一つであり、財務情報などに基づきデフォルト強度を非説明変数として推定する方法もあるが、本報告では、社債の市場価格を所与として、そこに織り込まれているデフォルト強度、すなわちインプライド・デフォルト強度 (implied default intensity) を推定する。このようにして得られたデフォルト強度は、当該企業を対象とするクレジット・デフォルト・スワップ (CDS) のレート設定等に利用できると考えられる。

さて、短期金利とデフォルト強度の過程が与えられれば、Duffie and Singleton (1999) の公式によりゼロクーポン社債価格が求められる。このようなデフォルト強度と社債価格の関係は、後述のように状態空間モデルの枠組みで記述することができる。そのため、観測価格からデフォルト強度を推定する問題は、フィルタリングの問題に帰着される。

線形ガウス状態空間モデルに対する状態推定の方法としては従来よりカルマン・フィルタやその拡張が利用されてきた。Duffee (1999) は3ファクターの CIR 過程を状態変数として、短期金利 r と強度 λ をその線形結合とするモデルに対し、拡張カルマン・フィルタを用いてデフォルト率の推定を行った。また、Chen and Scott (2003) も CIR 過程のマルチファクターモデルに対してカルマン・フィルタを用いて推定を行った。

しかし、カルマン・フィルタを用いるためには、状態方程式・観測方程式ともに線形ガウスであり、観測対象の理論価格の解析解が求められる必要がある。また、カルマン・フィルタで推定可能なモデルのクラスでは、 r と λ の非負性と負の相関の両立が保証されないなどの問題がある。そこで、数値的な方法である粒子フィルタを用いることで、非負性と負の相関の両立を含むような、よ

*東京大学大学院経済学研究科博士課程

以上のようなデフォルト強度の過程と観測される債券の価格系列の関係は、状態空間モデルで表現することができる。状態空間モデルとは、次のような状態モデルと観測モデルの組として表される時系列モデルである。

$$x_t = f(x_{t-1}, v_t) \quad \text{状態モデル} \quad (4)$$

$$y_t = h(x_t, w_t) \quad \text{観測モデル} \quad (5)$$

ここで、 x_t は n_x 次元の未知の状態ベクトル、 y_t は n_y 次元の観測ベクトル、 v_t は $q(v)$ に従う n_v 次元のシステムノイズベクトル、 w_t は $r(w)$ に従う n_w 次元の観測ノイズベクトルを表す。観測値から x_t を推定することを状態推定という。

具体的には、状態モデルにおける x_t を状態変数 X_t と考え、短期金利や強度を X_t の関数とする。また観測データは国債や社債価格等の金融時系列データであり、観測モデルは短期金利や強度と観測データの論理的な関係を表す関数であると考えることができる。このようにすると、状態推定によって、市場価格に基づくインプライド・デフォルト強度を推定することができる。

3 モンテカルロ・フィルタ

本研究では Kitagawa (1996) に提案されたモンテカルロ・フィルタにより、非線形非ガウスの状態空間モデルの状態推定を行う。モンテカルロ・フィルタでは以下の各分布を多数の実現値 (= “粒子”) で近似する。

$$\begin{aligned} p_t^{(1)}, \dots, p_t^{(M)} &\sim p(x_t | Y_{t-1}) && \text{予測分布} \\ f_t^{(1)}, \dots, f_t^{(M)} &\sim p(x_t | Y_t) && \text{フィルタ分布} \\ s_{i|T}^{(1)}, \dots, s_{i|T}^{(M)} &\sim p(x_t | Y_T) && \text{平滑化分布} \\ v_t^{(1)}, \dots, v_t^{(M)} &\sim p(v_t) && \text{システムノイズの分布} \end{aligned}$$

ここで $Y_t := \{y_1, \dots, y_t\}$ 、 M は粒子数である。

モンテカルロ・フィルタのアルゴリズムは、一期先予測とフィルタの繰り返しで構成されるが、理論の詳細については Kitagawa (1996) を参照されたい。アルゴリズムをまとめると次のようになる。

1. 適当な初期分布 $p_0(x)$ から $j = 1, \dots, M$ について乱数 $f_0^{(j)}$ を生成する。
2. 以下を $t = 1, \dots, T$ について繰り返す。
 - (a) $j = 1, \dots, M$ について $v_t^{(j)} \sim q(v)$ を生成する。
 - (b) $j = 1, \dots, M$ について $p_t^{(j)} = f(f_{t-1}^{(j)}, v_t^{(j)})$ を計算する。
 - (c) $j = 1, \dots, M$ について $\alpha_t^{(j)} = p(y_t | p_t^{(j)})$ を計算する。
 - (d) $j = 1, \dots, M$ について $f_t^{(j)}$ を

$$f_t^{(j)} = \begin{cases} p_t^{(1)} & \text{w.p. } \alpha_t^{(1)} / \sum_{i=1}^M \alpha_t^{(i)} \\ \vdots & \vdots \\ p_t^{(M)} & \text{w.p. } \alpha_t^{(M)} / \sum_{i=1}^M \alpha_t^{(i)} \end{cases}$$

によってリサンプリングする。

とし、 $\epsilon = 5 \text{ b.p.}$ とおいている。デフォルト強度は $\lambda(X_u) = X_{3u}$ である。 $a, b, c, \theta_2, \theta_3$ は正定数のパラメータである。また、リスクの市場価格はゼロと仮定する。

r 及び λ の過程を所与とすると、時刻 t における満期 T のゼロクーポン債価格は X_t の関数となり、

$$P(X_t, t; T) = E_t^Q \left[\exp \left\{ - \int_t^T r(X_u) du \right\} \right] \quad (16)$$

である。ただし E_t^Q は Q の条件付期待値を表す。またゼロクーポン社債価格は前述のように

$$V(X_t, t; T) = E_t^Q \left[\exp \left\{ - \int_t^T (r + \lambda)(X_s) ds \right\} \right] \quad (17)$$

となる。このようなモデルの場合、式 (16) 及び (17) は解析的に求まらないためモンテカルロ・シミュレーションにより数値的に計算する必要がある。

理論価格 (16)(17) は直接市場で観測できるデータではないので、観測モデルにおいては、これらの価格と観測可能な金融時系列データとの論理的な関係を利用する。そのようなデータとして LIBOR と金利スワップがある。具体的には、満期まで τ_n 期間の LIBOR は

$$L_t(X_t, \tau_n) = \left(\frac{1}{P(X_t, t; t + \tau_n)} - 1 \right) \frac{1}{\tau_n} \quad (18)$$

で求められる。また満期まで τ_n 期間で間隔 δ でキャッシュフローの発生するスワップの金利は

$$S_t(X_t, \tau_n) = \frac{1 - P(X_t, t; t + \tau_n)}{\delta \sum_{i=1}^{\tau_n/\delta} P(X_t, t; t + i\delta)} \quad (19)$$

となる (Björk (2004) 等参照)。さらに本稿では、固定金利 C を支払う普通社債を観測データとする。額面 1、満期 T で、将来の利払い日 t_j において固定利率 C に基づく利息を支払う社債の理論価格を $V_t(X_t, T, C, \{t_j\})$ とかくと、無裁定の議論より以下のように表現される。

$$V_t(X_t, T, C, \{t_j\}) = C \sum_j V(X_t, t; t_j) + V(X_t, t; T). \quad (20)$$

本報告では LIBOR は 6ヶ月と 12ヶ月、円スワップは 2, 3, 5, 7 年物を用いる。以上をまとめて、7次元の観測ベクトル Y を

$$Y_{n,t} = \begin{cases} L_t(X_t, \tau_n) + u_{n,t}, & n = 1, 2 \text{ のとき} \\ S_t(X_t, \tau_n) + u_{n,t}, & n = 3, \dots, 6 \text{ のとき} \\ V_t(X_t, T, C, \{t_j\}) + u_{n,t}, & n = 7 \text{ のとき} \end{cases} \quad (21)$$

とする。ここで $(\tau_1, \dots, \tau_6) = (0.5, 1, 2, 3, 5, 7)$ であり、 $u_{n,t}$ は $i.i.d.$ に独立な正規ノイズである。実証分析に用いるデータは、日本航空が 1996 年に発行し 2006 年 12 月 6 日に満期を迎える年利率 2.9% の固定利付普通社債を対象にして行う。期間と観測頻度は 1998 年 11 月から 2006 年 11 月までの 418 時点の週次データとする。図 1 にそれらの時系列を示す。データベースは Datastream と Bloomberg を利用した。

実際の推定においては、モンテカルロ・フィルタの適用のために、状態変数の過程を離散化する必要がある。ここではオイラー近似により

$$X_{1,k+1} = X_{1,k} + a(X_{2,k} - X_{1,k})\Delta t + \sigma_1 w_{1k} \sqrt{\Delta t} \quad (22)$$

$$X_{2,k+1} = X_{2,k} + b(\theta_2 - X_{2,k})\Delta t + (\sigma_{12} w_{1k} + \sigma_2 w_{2k}) \sqrt{\Delta t} \quad (23)$$

$$X_{3,k+1} = X_{3,k} + c(\theta_3 - X_{3,k})\Delta t + \sigma_3 \sqrt{X_{3,k}} \left[\rho w_{1k} + \sqrt{1 - \rho^2} w_{3k} \right] \sqrt{\Delta t} \quad (24)$$

a	b	c	θ_2	θ_3			
0.41	0.12	0.066	0.043	0.0084			
σ_1	σ_{12}	σ_2	σ_3	ρ			
0.006	-0.0027	0.029	0.041	0.03			
u_1	u_2	u_3	u_4	u_5	u_6	u_7	AIC
0.00027	0.00028	0.00015	0.00014	0.00042	0.0007	0.0021	-34926

表 1: 推定値

a	b	c	θ_2	θ_3			
0.17	0.076	0.035	0.016	0.0057			
σ_1	σ_{12}	σ_2	σ_3	ρ			
0.0035	0.0016	0.009	0.019	0.25			
u_1	u_2	u_3	u_4	u_5	u_6	u_7	
0.00017	0.00013	9.6×10^{-5}	0.00012	0.00027	0.00033	0.0011	

表 2: 漸近標準誤差

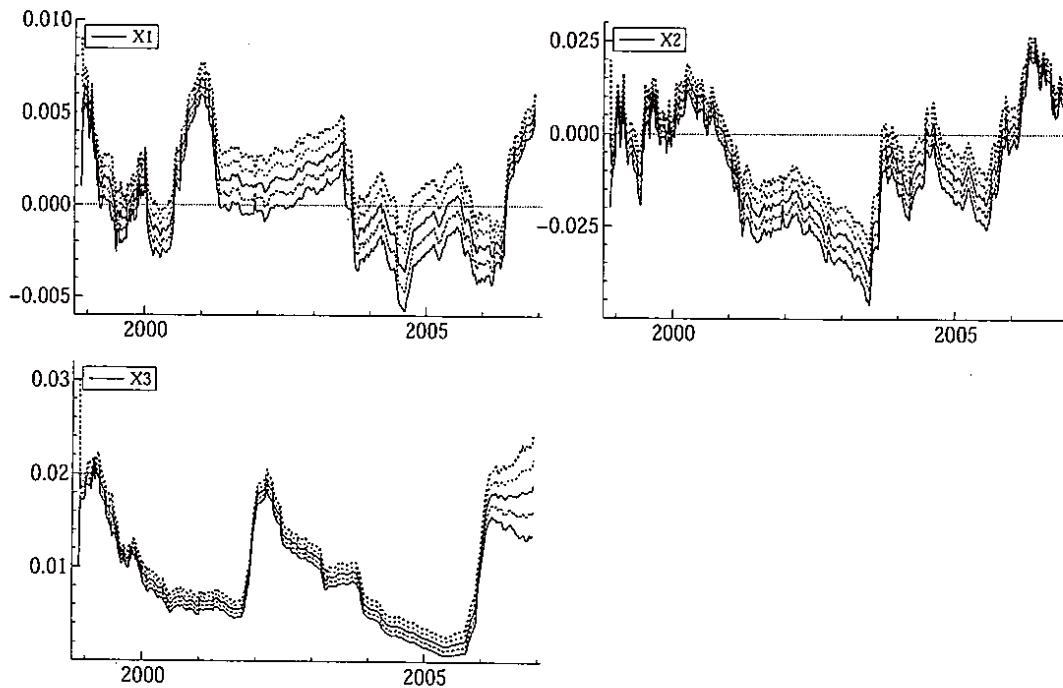


図 2: 各ファクターのフィルタ分布

- [14] 中川秀敏 (2008), 信用リスク・モデルの観望とその新展開 — トップダウン・アプローチによるデフォルトの依存関係のモデル化 —, *現代ファイナンス*, 23, 3-33.
- [15] Nelder, J. A. and R. Mead (1965), A Simplex Method for Function Minimization, *Computer Journal*, 7, 308-313.
- [16] Schönbucher, P. J. (2003), *Credit Derivatives Pricing Models: Models, Pricing and Implementation*, Wiley.
- [17] Takahashi, A. and S. Sato (2001), A Monte Carlo Filtering Approach for Estimating the Term Structure of Interest Rates, *Annals of the Institute of Statistical Mathematics*, 53(1), 50-62.
- [18] 高橋明彦, 佐藤整尚 (2002), モンテカルロフィルタを用いた金利モデルの推定, *統計数理*, 50(2), 133-147.
- [19] Yano, K. (2007), A Self-organizing State Space Model and Simplex Initial Distribution Search, *Computational Statistics*, 22, 197-216.
- [20] Zhou, Y., X. Qin, Q. Shang and Z. Liu (2008), Reduced Credit Risk Measurement Model with Particle Filter Approach, *3rd IEEE Conference on Industrial Electronics and Applications*, 203-207.

Self-exciting 性をもつイベント発生強度モデルによる 信用ポートフォリオのリスク解析¹

1) 山中 卓, 2) 杉原 正顯, 3) 中川 秀敏

1),2) 東京大学大学院情報理工学系研究科, 3) 一橋大学大学院国際企業戦略研究科

*Email: Suguru.Yamanaka@mist.i.u-tokyo.ac.jp

2010年3月30日

1 はじめに

本稿では複数の信用ポートフォリオの信用リスクを同時に評価するための新しいモデルを提案する。提案モデルの枠組みは Giesecke et al. [1] や Nakagawa [4] で研究されている top-down アプローチに基づく。具体的には、経済全体での信用イベント発生強度を self-exciting 性をもつ確率過程でモデル化し、thinning によって部分ポートフォリオのイベント発生強度を特定する。提案モデルによって、信用リスクの伝播をとらえた上で、複数のポートフォリオ・クレジット・デリバティブのリスク解析を同時に行うことができる。

Top-down アプローチは、ポートフォリオのデフォルト・リスクのモデル化の方法として Giesecke et al.[1] によって提案された。Top-down アプローチでは、ポートフォリオを構成する個々の企業のデフォルト発生強度のモデル化を直接は行わず、ポートフォリオ全体からのデフォルト発生強度過程を用いてモデル化し、ポートフォリオ内の個々の企業のデフォルト強度は thinning という方法によって与える。強度過程の具体的なモデルとして Giesecke et al.[1] は self-exciting 性（デフォルト発生時点でジャンプする性質）をもつ確率過程を考えている。Giesecke et al. [1] の枠組みを用いて、ポートフォリオ・クレジット・デリバティブのリスク解析や価格評価を行った研究として、例えば債務担保証券 (collateralized debt obligation, CDO) のリスク解析を行った Giesecke and Kim[2] が挙げられる。

Giesecke et al. [1] や Giesecke and Kim [2] がデフォルト発生だけに注目したのに対し、デフォルト発生だけでなく信用格付の変更のモデル化も行った研究に、中川 [3] や Nakagawa[4] がある。中川 [3] は格付変更（格上げ、格下げ）とデフォルト発生を self-exciting 性強度によってモデル化した。Nakagawa [4] は self-exciting 性だけでなく mutually exciting 性をもつ強度モデルを提案している。さらに、中川 [3] や Nakagawa[4] はマルチ・ダウングレード・プロテクションというクレジット・デリバティブの価格評価を行っている。

本稿では、複数の信用ポートフォリオのリスク解析を行うための、信用イベント発生強度モデルおよび thinning のモデルを提案する。本稿では、経済全体の信用イベント発生を state-dependent 性をもつ self-exciting 性強度過程でモデル化し、ポートフォリオの強度はポートフォリオの信用力、すなわちポートフォリオ内の各格付の企業の割合に依存した thinning によって得られるとする。経済全体のイベント発生を self-exciting 性強度過程でモデル化することによって、複数の信用ポートフォリオ間の信用リスクの伝播をとらえることができる。また、本モデルはポートフォリオ・クレジット・デリバティブの

¹本稿は Yamanaka et al. [7] の簡略版である。本研究は文部科学省グローバルCOEプログラム「数学新展開の研究教育拠点」の支援を受けた。

スク解析に適している。

本稿の構成は以下の通りである。第2節では、モデルの定式化を行う。第3節では、シミュレーションアルゴリズムの概要を述べる。第4節では、融資ポートフォリオのリスク解析を行った数値実験の結果を紹介する。第5節でまとめを述べる。

2 モデル

2.1 経済全体の信用イベント発生強度

経済全体の企業の集合を S^* で表すことにする。各企業は経済全体の部分ポートフォリオ S_i ($i = 1, 2, \dots, I$) のいずれかに属するとする。すなわち、 $S^* = \bigcup_{i=1}^I S_i$, $S_i \cap S_j = \emptyset$ ($i \neq j$) とする。また、各企業には信用格付が付与されているとする。格付は $1, 2, \dots, K$ および $K+1$ で表すことにする。ただし、格付 $k=1$ は最も信用力の高い格付に対応し、順に格付が低くなり、 $k=K$ が最も信用力の低い格付で、 $k=K+1$ がデフォルト状態に対応するものとする。 $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$ をフィルトレーション付完備確率空間とする。 $\{\mathcal{F}_t\}$ は右連続性と完備性を満たすとする。 $l \in \{1, 2, \dots, L\}$ で信用イベントのタイプを表す。本稿では簡潔に $L=3$ とし、イベント1が格上げ、イベント2が格下げ、イベント3がデフォルトに対応するとする。各 l について、 $0 < T_1^l < T_2^l < \dots$ をイベント l の発生時刻列を表す $\{\mathcal{F}_t\}$ -適合な点過程とする。ただし、 $T_n^l < \infty$ a.s. $\forall n \in \mathbb{N}$ とする。また、異なるイベントが同時刻に起こることはないとする。デフォルトした企業は消滅し、企業が新たに誕生することはないとする。イベント l の計数過程を $N_t^l = \sum_{n \geq 1} 1_{\{T_n^l \leq t\}}$ で表し、各 N_t^l は強度 λ_t^l を持つとする²。本稿では λ_t^l として、次の確率過程を考える：

$$d\lambda_t^l = \kappa_t^l (c_t^l - \lambda_t^l) dt + dJ_t^l, \quad (1)$$

$$J_t^l = \sum_{n \geq 1} (\min(\delta^l \lambda_{T_n^l}^l, \gamma^l) 1_{\{T_n^l \leq t\}}), \quad (2)$$

$$\kappa_t^l = \kappa^l \lambda_{N_t^l}^l, \quad c_t^l = c^l \lambda_{N_t^l}^l. \quad (3)$$

ただし $\kappa^l > 0$, $c^l \in (0, 1)$, $\delta^l > 0$, $\gamma^l \geq 0$, $\lambda_0^l > 0$ とする³。

図1は強度モデル(1)–(3)のサンプル・パスである。イベント発生時刻では強度の値がジャンプし (self-exciting 性)、イベント時刻の間では強度自身の値に依存して動くことが分かる (state-dependent 性)⁴。図2は格上げ件数の実データと、そのデータに対して推定した格上げ強度モデルの実現パスを重ねたものである。格上げの発生頻度と格上げ強度が相重なって高くなっており、self-exciting 性をもつ強度モデルによる格付変更のモデル化の有用性が示唆される⁵。

² λ_t^l は非負の $\{\mathcal{F}_t\}$ -発展的測過程で、 $N_t^l = \int_0^t \lambda_s^l ds$ が $\{\mathcal{F}_t\}$ -マルチンゲールになるものである。本稿では $\int_0^t \lambda_s^l ds < \infty$, $\forall t > 0$ a.s. を仮定する。

³ 本強度モデルはジャンプ幅に上限があるという点で Giesecke and Kim [2] のデフォルト強度モデルと異なる。

⁴ イベント発生の合間 $T_n^l \leq t < T_{n+1}^l$ において、強度(1)–(3)は

$$\lambda_t^l = c^l \lambda_{T_n^l}^l + (1 - c^l) \lambda_{T_n^l}^l \exp(-\kappa^l \lambda_{T_n^l}^l (t - T_n^l))$$

である。

⁵ 中川 [5] は self-exciting 性強度モデルの拡張である mutually exciting 性強度モデルを用いて、日本の格付変更に関する self-exciting 性があることを示す結果を得ている。

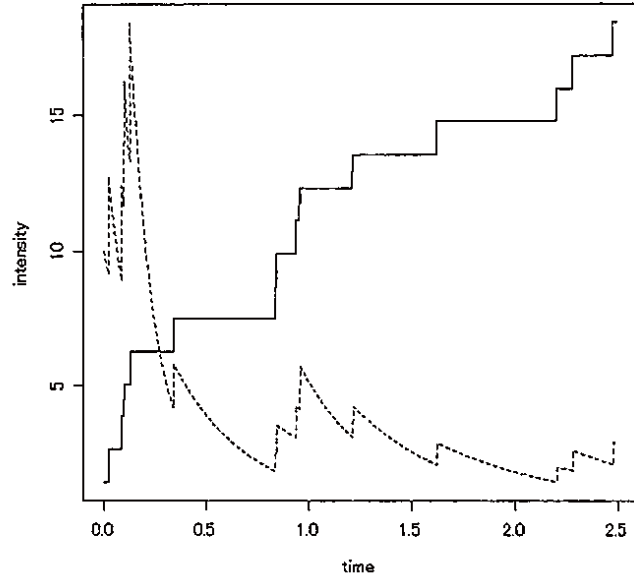


図 1: 強度モデル (1) - (3) のサンプルパスと累積イベント件数. (実線が累積イベント件数, 破線がイベント強度. モデルのパラメタは $\kappa = 0.5$, $c = 0.1$, $\delta = 0.4$, $\gamma = 50.0$, $\lambda_0 = 10.0$.)

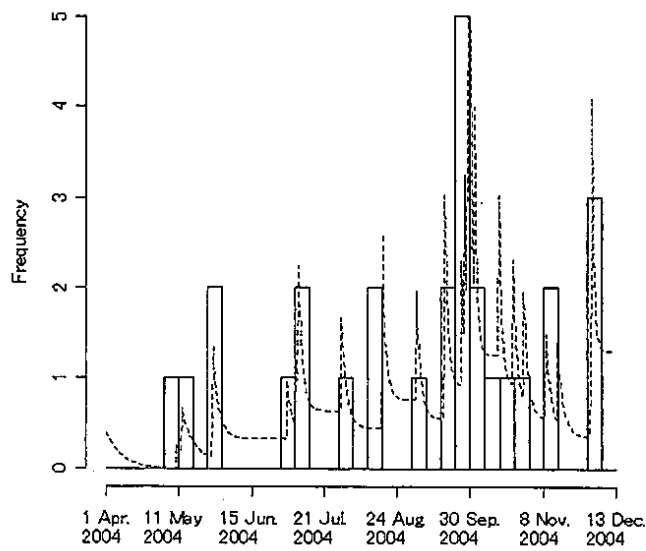


図 2: 格上げ件数の推移と格上げ強度モデルの実現パス (格付は R&I の発行体格付, パラメタは格上げ時刻データに関する最尤推定.)

2.2 Thinning

本小節では部分ポートフォリオの格付分布に基づいた thinning によって、部分ポートフォリオのイベント強度を得る。Thinning を行うために、時刻 t において格上げ（格下げ）が起こったときにそれがポートフォリオ S_i 内の格付 k の企業の格付 k' への格上げ（格下げ）である条件付き確率を表す確率変数 $Z_t^{(i,1)}(k, k')$ と $Z_t^{(i,2)}(k, k')$ 、および時刻 t にデフォルトが発生した場合にそれがポートフォリオ S_i 内の格付 k の企業のデフォルトである条件付き確率を意味する確率変数 $Z_t^{(i,3)}(k)$ を考える⁶。具体的には、 $Z_t^{(i,1)}(k, k')$ 、 $Z_t^{(i,2)}(k, k')$ および $Z_t^{(i,3)}(k)$ として、次のようなポートフォリオ内の格付分布に基づいたものを考えることにする：

$$Z_t^{(i,1)}(k, k') = \frac{X_t^{(i)}(k)}{\sum_{\bar{k}=(k-k')+1}^K X_t^*(\bar{k})} z_{k-k'}^1 \mathbf{1}_{\{\sum_{\bar{k}=(k-k')+1}^K X_t^*(\bar{k}) > 0\}} \quad (4)$$

$$(k = k' + 1, k' + 2, \dots, K),$$

$$Z_t^{(i,2)}(k, k') = \frac{X_t^{(i)}(k)}{\sum_{\bar{k}=1}^{K-(k'-k)} X_t^*(\bar{k})} z_{k'-k}^2 \mathbf{1}_{\{\sum_{\bar{k}=1}^{K-(k'-k)} X_t^*(\bar{k}) > 0\}} \quad (5)$$

$$(k = 1, 2, \dots, k' - 1),$$

$$Z_t^{(i,3)}(k) = \frac{X_t^{(i)}(k)}{X_t^*(k)} z_k^3 \mathbf{1}_{\{X_t^*(k) > 0\}} \quad (k = 1, 2, \dots, K). \quad (6)$$

ただし、 $X_t^*(k)$ は時刻 t における格付 k の企業数であり、 $X_t^{(i)}(k)$ は時刻 t におけるポートフォリオ S_i 内の格付 k の企業数を表す。また、 $z_m^l (l = 1, 2, m = 1, 2, \dots, K-1)$ は格付変更が起こった場合にそれが m 段階の格付変更である条件付き確率を意味する定数で、 $0 \leq z_m^l \leq 1$ 、 $z_1^l > z_2^l > \dots > z_{K-1}^l (l = 1, 2)$ および $\sum_{m=1}^{K-1} z_m^l = 1$ を満たす。定数 z_k は、デフォルトが発生した時にそれが格付 k の企業のデフォルトである条件付き確率を表し、 $0 \leq z_k \leq 1 (k = 1, 2, \dots, K)$ 、 $z_1^3 < z_2^3 < \dots < z_K^3$ および $\sum_{k=1}^K z_k = 1$ を満たすとす。

$N_t^{(i,1)}(k, k')$ ($N_t^{(i,2)}(k, k')$) をポートフォリオ S_i 内の格付 k の企業の格付 k' への格上げ（格下げ）計数過程とし、 $N_t^{(i,3)}(k)$ をポートフォリオ S_i 内の格付 k の企業のデフォルト計数過程とする。このとき、 $Z_t^{(i,1)}(k, k')$ 、 $Z_t^{(i,2)}(k, k')$ 、 $Z_t^{(i,3)}(k)$ を用いて $N_t^{(i,1)}(k, k')$ 、 $N_t^{(i,2)}(k, k')$ 、 $N_t^{(i,3)}(k)$ の強度は次のように与えられる：

$$\lambda_t^{(i,1)}(k, k') = Z_t^{(i,1)}(k, k') \lambda_t^1, \quad (7)$$

$$\lambda_t^{(i,2)}(k, k') = Z_t^{(i,2)}(k, k') \lambda_t^2, \quad (8)$$

$$\lambda_t^{(i,3)}(k) = Z_t^{(i,3)}(k) \lambda_t^3. \quad (9)$$

⁶Giesecke et al. [1] の命題から、たとえば、 $Z_t^{(i,3)}(k)$ は次のように定義される：

$$Z_t^{(i,3)}(k) = \lim_{\varepsilon \rightarrow 0} Z_t^{(i,3)}(k, \varepsilon),$$

$$Z_t^{(i,3)}(k, \varepsilon) = \sum_n \frac{P[\{T_n^3 \in \tau^3(S_i)\} \cap \{T_n^3 \in \tau^3(k)\} \cap \{T_n^3 \leq t + \varepsilon\} \mid \mathcal{F}_t]}{P[T_n^3 \leq t + \varepsilon \mid \mathcal{F}_t]} \mathbf{1}_{\{T_{n-1}^3 < t \leq T_n^3\}}.$$

ただし、 $\tau^3(S_i)$ はポートフォリオ S_i で起こったタイプ 3 のイベント発生時刻の集合を表し、 $\tau^3(k)$ は格付 k の企業のイベント l の発生時刻の集合を表す。 $Z_t^{(i,l)}(k, k')$ ($l = 1, 2$) も同様に定められる。

3 シミュレーション・アルゴリズム

本節では第2節のモデルに基づく信用イベント発生シミュレーション・アルゴリズムの概略を述べる。ステップ2, 3はOgata [6]のアルゴリズムに基づく。

[0, H]上のイベント発生時刻を生成するアルゴリズム

1. [モデルのパラメータおよび初期値を設定する]

- 以下を設定する：モデルのパラメータ $(\kappa^l, c^l, \delta^l, \gamma^l, \lambda_0^l)$ ($l = 1, 2, 3$), z_m^l ($l = 1, 2, m = 1, 2, \dots, K-1$), z_k ($k = 1, 2, \dots, K$), 現時刻 $S = 0$, シミュレーションの満期 $H (> 0)$, ポートフォリオ内の企業数 $X_T^{(i)}(k) = X_0^{(i)}(k)$ ($i = 1, 2, \dots, I, k = 1, 2, \dots, K$), イベント l の発生件数 $N^l = 0$ ($l = 1, 2, 3$), イベント l の最終発生時刻 $T_{N^l}^l = 0$ ($l = 1, 2, 3$).

2. [イベント発生時刻の候補 T を生成する.]

- $\Lambda = \sum_{l=1}^3 \lambda_S^l$ として, 指数分布に従う乱数 $\varepsilon \sim \exp(\Lambda)$ を発生させる.
- $T = S + \varepsilon$ とする. $T > H$ であればアルゴリズムを終了する.

3. [イベント時刻候補 T を採択するか棄却するかを決定する]

- λ_T^l / Λ の確率で時刻 T をイベント l の発生時刻として採択する. いずれのイベント時刻としても採択されなければ, ステップ5に行く.

4. [Thinning]

- 採択されたイベントタイプ l について, (4),(5),(6)の確率で, イベントの詳細(イベントがどのポートフォリオの何格から発生したか, 何格への格付変更か)を決める. 特定されたイベントに従って, $X_T^{(i)}(k), N^l, T_{N^l}^l$ を更新する.

5. $S = T$ として, ステップ2に行く.

4 数値例

本節では, 本モデルに基づく融資ポートフォリオのリスク解析に関する数値結果のいくつかを述べる. また, 日本の格付け変更データに対して試みたパラメータ推定の結果も述べる.

4.1 強度モデルの推定

各イベント発生時刻列のデータから, 各イベント発生強度のパラメータ $(\kappa_l, c_l, \gamma_l, \delta_l, \lambda_0^l)$ を推定するために最尤法を用いることにする. イベント l の時間区間 $[0, H]$ 内のイベント発生時刻 $(0 <) T_1^l < T_2^l < \dots < T_{N^l}^l (< H)$ のデータが得られているとする. このとき, イベント l の強度モデルの尤度関数は以下のようになる:

$$\sum_{n=1}^{N^l} \log \lambda_{T_n^l}^l - \int_0^H \lambda_s^l ds. \quad (10)$$

また、推定されたモデルのデータへの適合度をみるために、中川 [5] を参考にして Kolmogorov-Smirnov 検定を行う。具体的にはまず、各タイプのイベント発生時刻列 $\{T_n^l\}_{n=1}^N$ を事後推定された強度 $\{\lambda_t^l\}$ を用いて

$$A_n^l := \int_0^{T_n^l} \lambda_s^l ds \quad (11)$$

と変換する。こうして得られた $\{A_n^l\}_{n=1}^N$ が標準 Poisson 過程のジャンプ時刻とみなせることを利用して検定を行う。すなわち、 $\{A_{n+1}^l - A_n^l\}_{n=1}^{N-1}$ は独立同分布で、指数 1 の指数分布に従うことを帰無仮説として、これが棄却されない時に、モデルがデータに適合していることにする。

推定に用いたデータは 2004 年 4 月 1 日 2009 年 4 月 1 日までの間に R&I の公表した日本企業の発行体格付け変更データである。ここでは、BB+以下の格付けを疑似デフォルト状態とした。すなわち、格付けは AAA, AA+, AA, AA-, A+, A, A-, BBB+, BBB, BBB- の 10 段階になる ($K = 10$)。また、休日を除いた上で 2004/4/1, 2005/4/1, ... が $t = 0, 1, \dots$ に対応するように時刻の変換を行った。また、同日に起こった複数の格付け変更に対しては、一様乱数を用いて変更時刻をずらした。

表 1 が得られたパラメタ推定値と p 値である。p 値から、推定されたモデルは有意水準 5% で棄却されないことが分かる。

表 1: 強度モデルのパラメタ推定値

	κ	c	δ	γ	λ_0	p-value
格上げ ($l = 1$)	1.745	0.350	1.2(固定)	90.804	26.486	0.062
格下げ ($l = 2$)	1.643	0.281	1.2(固定)	168.839	82.676	0.063
デフォルト ($l = 3$)	3.450	0.503	1.2(固定)	23.384	1.181	0.974

4.2 Thinning モデルのパラメタ推定

細分化モデル (4), (5), (6) のパラメタ z_m^1, z_m^2, z_k^3 を次のように与える:

$$z_m^l = \frac{\exp(a^l(K - m))}{\sum_{\tilde{m}} \exp(a^l(K - \tilde{m}))} \quad (l = 1, 2), \quad (12)$$

$$z_k^3 = \frac{\exp(a^3 k)}{\sum_{\tilde{k}} \exp(a^3 \tilde{k})}. \quad (13)$$

z_m^1, z_m^2 および z_k^3 を推定するために、 m 段階の格上げおよび格下げが起こる割合である \hat{z}_m^1, \hat{z}_m^2 とデフォルト発生元の格付けが k である割合である \hat{z}_k^3 を格付け変更データから算出する。そして、 $\sum_m (z_m^l - \hat{z}_m^l)^2$ ($l = 1, 2$), $\sum_k (z_k^3 - \hat{z}_k^3)^2$ を最小にする a^1, a^2, a^3 を求め、それに対応する z_m^1, z_m^2 および z_k^3 を推定値とする。推定には 1998 年 4 月 1 日から 2009 年 4 月 1 日までの格付け変更データを用いた。表 2, 4 は $\hat{z}_m^1, \hat{z}_m^2, \hat{z}_k^3$ の値である。推定の結果、 $a^1 = 2.327, a^2 = 1.979, a^3 = 1.238$ が得られ、推定値 \hat{z}_m^1, \hat{z}_m^2 および \hat{z}_k^3 は表 3, 5 となった。

表 2: 格付変更幅の割合

m	\hat{z}_m^1	\hat{z}_m^2
1	0.9024	0.8618
2	0.0732	0.1290
3	0.0244	0.0046
4	0	0.0046
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0

表 3: Thinning モデル (4), (5) のパラメタ推定値

m	\hat{z}_m^1	\hat{z}_m^2
1	0.9024	0.8618
2	0.08808	0.1191
3	0.008597	0.01646
4	8.400×10^{-4}	0.002276
5	8.289×10^{-5}	3.154×10^{-4}
6	8.992×10^{-6}	4.444×10^{-5}
7	1.780×10^{-6}	6.999×10^{-6}
8	1.076×10^{-6}	1.824×10^{-6}
9	1.007×10^{-6}	1.109×10^{-6}

表 4: デフォルト発生元の格付けの割合

k	\hat{z}_k^3
1	0
2	0
3	0
4	0
5	0
6	0
7	0
8	0.03
9	0.26
10	0.71

表 5: Thinning モデル (6) のパラメタ推定値

k	\hat{z}_k^3
1	1.009×10^{-5}
2	3.531×10^{-5}
3	1.223×10^{-4}
4	4.222×10^{-4}
5	0.001456
6	0.005022
7	0.01732
8	0.05972
9	0.2059
10	0.7100

4.3 融資ポートフォリオのリスク解析

本小節では、融資ポートフォリオのリスク解析に関する数値結果を紹介する。ここでは、3つのポートフォリオ1,2,3を同時に考える。融資ポートフォリオの格付分布は表6のように設定した。すなわち、ポートフォリオ1から順に信用力の高いポートフォリオになる。モデルのパラメタ値は表1, 3, 5の値を用いた。

表 6: 各ポートフォリオの格付け分布

rating	1	2	3	4	5	6	7	8	9	10
portfolio 1	15	15	15	15	15	5	5	5	5	5
portfolio 2	10	10	10	10	10	10	10	10	10	10
portfolio 3	5	5	5	5	5	15	15	15	15	15
residual	10	10	10	10	10	10	10	10	10	10

図3と表7は各ポートフォリオの1年間の損失分布関数およびリスク指標である。図3と表7から、ポートフォリオの信用力が低いほど損失が大きくなることから分かる。これは、thinningのモデルの仮定である $\lambda_1^3 < \lambda_2^3 < \dots$ から低格付けの企業のデフォルト発生強度が大きくなるためである。

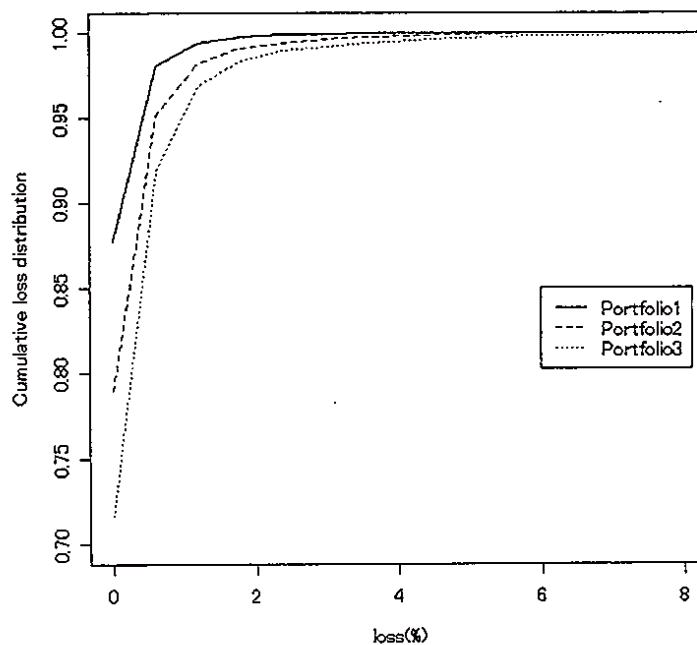


図 3: 各ポートフォリオの累積損失分布関数 (1年間)

本モデルでは複数のポートフォリオを同時に扱え、異なる融資ポートフォリオ間のリスク伝播を分析

表 7: 各ポートフォリオのリスク指標

	One year loss		
	Portfolio1	Portfolio2	Portfolio3
Average	0.09%	0.18%	0.27%
Max	6.00%	9.60%	15.00%
95%VaR	0.60%	0.60%	1.20%
95%ES	0.95%	1.68%	2.23%
99%VaR	1.20%	1.80%	3.00%
99%ES	1.79%	3.25%	4.66%

することができる。表 8 は、ポートフォリオ 2 の 1 年間の損失が 0% - 0.75%, 0.75% - 1.5%, 1.5% - のそれぞれの水準である場合のポートフォリオ 1 のリスク指標である。表 8 より、ポートフォリオ 2 の損失が大きくなるほど、ポートフォリオ 1 の損失も大きくなることが分かる。

表 8: ポートフォリオ 2 の損失に応じたポートフォリオ 1 のリスク指標

One year loss of Port.2	One year loss					
	Portfolio 1					
	Average	Max	95%VaR	95%ES	99%VaR	99%ES
0%-0.75%	0.07%	3.60%	0.60%	0.74%	1.20%	1.28%
0.75%-1.5%	0.24%	4.80%	1.20%	1.52%	1.80%	2.20%
1.5%-	0.93%	6.00%	3.00%	3.30%	3.60%	3.99%

表 9 は、最初の 1 年間のポートフォリオ 2 の損失で条件付けた次の 1 年間の各ポートフォリオの期待損失率である。ポートフォリオ 2 の損失が大きくなると、次の 1 年間の各ポートフォリオの損失が大きくなることが分かる。このように、あるポートフォリオの過去の損失が他のポートフォリオの将来の損失に与える影響を本モデルで分析できる。

表 9: ポートフォリオ 2 の過去の損失に条件付けられた各ポートフォリオの期待損失率。

First one year loss of Port.2	Expected loss of second one year			
	Economy	Port1	Port2	Port3
0%-0.75%	0.26%	0.13%	0.27%	0.39%
0.75%-1.5%	1.56%	0.78%	1.55%	2.37%
1.5%-	3.72%	1.85%	3.56%	5.74%

5 まとめ

Top-down アプローチに基づいて複数の社債ポートフォリオの信用リスクを評価するモデルを提案した。具体的には、経済全体からの信用イベントの発生を self-exciting 性をもつ強度モデルを用いてモデル化し、部分ポートフォリオのイベント発生強度は、ポートフォリオ内の格付分布に依存した thinning によって与えた。モデルに基づく信用イベント発生シミュレーションアルゴリズムの概要を述べ、融資ポートフォリオのリスク解析に関する数値例の結果のいくつかを紹介した。本モデルは複数の信用ポートフォリオを同時に扱うことが容易に可能であり、また信用リスクの伝播をとらえているので、ポートフォリオ・クレジット・デリバティブのリスク解析に有用であると考えられる。

参考文献

- [1] K. Giesecke, L. R. Goldberg, X. Ding: A Top-down Approach to Multi-Name Credit. *Working paper* (2005), Stanford University.
(<http://www.stanford.edu/dept/MSandE/cgi-bin/people/faculty/giesecke/giesecke.php>).
- [2] K. Giesecke and B. Kim: Risk Analysis of Collateralized Debt Obligations. *Working paper* (2009), Stanford University.
(<http://www.stanford.edu/dept/MSandE/cgi-bin/people/faculty/giesecke/giesecke.php>).
- [3] 中川 秀敏: トップダウン・アプローチによるマルチ・ダウングレード・プロテクションの評価. MTEC ジャーナル特別号 (2008), pp. 451–481.
- [4] H. Nakagawa: Modeling of Contagious Downgrades and Its Application to Multi-Downgrade Protection. *Working paper* (2009), Hitotsubashi University.
(<https://sites.google.com/site/icsnakagawah/working-papers-tesuto-ban>).
- [5] 中川 秀敏: 相互作用型の格付変更強度モデルによる格付変更履歴データの分析. *Working paper* (2009), 一橋大学.
(<https://sites.google.com/site/icsnakagawah/working-papers-tesuto-ban>).
- [6] Y. Ogata: On Lewis' Simulation Method for Point Processes. *IEEE Trans. Inform. Theory*, vol. IT-27 (1981), pp. 23–31.
- [7] S. Yamanaka, M. Sugihara and H. Nakagawa: Modeling of Contagious Credit Events and Risk Analysis of Collateralized Debt Obligations. *Working paper* (2010).
(<http://www.sr3.t.u-tokyo.ac.jp/~yamanaka/working-papers>)

< マイクロ市場の計量-I >

6. 「株式市場における終値形成について」 太田亘
7. 「高頻度データ分析：取引時間と価格の相互依存性について」 林高樹

取引開始前および取引開始後の価格発見

太田 亘

大阪大学大学院経済学研究科

1 はじめに

証券取引所の取引時間中は、取引を通じて価格に情報が織り込まれて行く。例えば Kyle (1985) のモデルでは、情報をもっていない投資家に混ざって情報を持っている投資家が注文を出し、取引参加者が価格や注文情報から自分の知らない情報を推測した上で売買をする過程を通じて、価格に情報が反映して行く。では、取引をせず単に気配を提示しているだけの場合、気配に情報が反映されるだろうか。この問題について Biais *et al.* (1999) は、パリ証券取引所の取引開始前に配信される気配には新情報が反映しており、それを通じて取引参加者間で学習が行われている、という実証結果を示している。その他の株式市場について、Cao *et al.* 2000, Barclay *et al.* (2003, 2008), Madhavan and Panchapagesan (2000), Chen *et al.* (2009) などの研究が、取引開始前の気配の配信により学習・情報生産が行われていることを肯定している。本稿では、東京証券取引所 (東証) においても寄付き前に学習・情報生産が行われているかを、Biais *et al.* と同様の方法を用いて分析する。さらに、同様の方法を用いて、取引時間中の価格形成の効率性についても分析する。

東証では、一日の取引開始前に寄前気配が配信されるばかりでなく、昼休みがあり、昼休み終了前にも寄前気配が配信される。昼休みの中断は夜間の中断よりも短いため、昼休み終了前の価格発見は1日の取引開始前の価格発見よりも容易であると推測されるが、両者を比較することにより、1日の取引開始前の価格発見のスピードに関して分析することが可能となる。また Biais *et al.* では、当日終値を基準にした分析をしているため、終値の価格形成に対する分析ができない。それに対し東証では、一日の終値以外にも、前場終値の形成などがコールオークションにより行われ、しかも Amihud and Mendelson (1991) により日中のコールオークションによる価格形成は効率的であることが知られている。そのため、終値以外の価格を基準とすることで取引終了前の価格形成に対する分析を行うことが可能となる。

以下ではまず、分析方法を説明し、次にサンプルについて説明する。その後、取引開始前の気配形成および取引時間中の価格形成についての分析結果をまとめる。

2 分析方法

本稿では、unbiasedness regression により、寄付き前の学習および寄付き後の価格発見を分析する。unbiasedness regression は、証券の本源的価値を v 、時点 0 での本源的価値の推定値を $E(v|I_0)$ 、時点 t での価格を P_t としたとき、

$$v - E(v|I_0) = \alpha_t + \beta_t(P_t - E(v|I_0)) + \epsilon_t$$

とする regression である。Biais et al. (1999) は、取引日 T の終値を $P_{c,T}$ おいたとき、 $v = P_{c,T}$ 、 $E(v|I_0) = P_{c,T-1}$ として、

$$\frac{P_{c,T} - P_{c,T-1}}{P_{c,T-1}} = \alpha_t + \beta_t \frac{P_t - P_{c,T-1}}{P_{c,T-1}} + \epsilon_t$$

とする回帰式を 1 分ごとと推計している。このとき

$$\beta_t = 1 + \frac{\text{Cov}(P_{c,T} - P_t, P_t - P_{c,T-1})}{\text{Var}(P_t - P_{c,T-1})}$$

である。価格発見が全く行われておらず、 $\text{Cov}(\epsilon_t, P_{c,T}) = 0$ となる ϵ_t に対して $P_t = P_{c,T-1} + \epsilon_t$ であるとき、 $\beta = 0$ となる。それに対して価格発見が行われていれば $\beta > 0$ となる。さらに、 t 時点の価格が効率的であれば、 $\text{Cov}(P_{c,T} - P_t, P_t - P_{c,T-1}) = 0$ であり、 $\beta = 1$ となる。 $\text{Cov}(P_{c,T} - P_t, P_t - P_{c,T-1}) < 0$ のとき $\beta < 1$ であり、この場合には情報効率的な価格に比べて過剰反応が起こっている可能性があり、逆に $\text{Cov}(P_{c,T} - P_t, P_t - P_{c,T-1}) > 0$ のとき $\beta > 1$ であり、この場合には情報効率的な価格に比べて過小反応が起こっている可能性がある。また Barclay et al. (2003) が指摘しているように、「真の価格」が観察不可能である場合には、係数が過小に推計されるため、注意が必要である。

Biais et al. (1999) はパリ証券取引所に上場されている流動性の高い銘柄について unbiasedness regression を行い、指数については取引開始前から価格発見が行われ、取引開始時に $\beta = 1$ となる、個別銘柄についても取引開始前から価格発見が行われ、取引開始後 30 分程度経過したときに $\beta = 1$ となり、その後は $\beta = 1$ が継続する、という結果を得ている。

3 サンプル

本稿での分析には、日経メディアマーケティング提供のティックデータおよびポートフォリオ・マスターを用いる。期間は 2004 年および 2005 年の 2 年間であり、この期間に東証に上場されていた普通株を分析対象とする。但し、午前の取引しかない年末年始、システムトラブルのあった 2005 年 11 月 1 日およびその前日、各銘柄の権利落ち日の前日を分析対象外とする。期間中の最低価格が 200 円を超えている銘柄について、上記分析除外日以外のうち売買のあった日が 400 日を超える銘柄のうち、一日の平均売買金額の上位 400 銘柄を分析対象とする。以下、必要に応じて売買高グループごとの結果を報告している場合があるが、売買高グループは一日の平均売買金額により 100 銘柄ごと 4 つに分割したグループである。

unbiasedness regression を、各銘柄・各時刻(分単位)ごとに行う。unbiasedness regression では、本源的価値 v および基準となる時点 0 における本源的価値の推定値 $E(v|I_0)$ を選択する必要がある。Biais et al. (1999) は、本源的価値に当日終値、0 時点における本源的価値の推定値に前日終値を用いている。本稿でも同様の回帰分析を行う。それと同時に、一日の終値(後場終値)以外に、後場始値、前場終値、前場始値を本源的価値の推定値とし、それぞれの 1 日前の価格を 0 時点における本源的価値の推定値とした推計も行う。また、Biais et al. では個別銘柄の回帰において、マーケットモデルの残差を被説明変数に用いているが、ここでは残差ではなく収益率を用いている。

気配を 1 分ごとに計測するが、 P_t には、寄り付き前であれば各分最後に提示されていた寄り付き前気配の中値、寄り付き後には各分最後に提示されたいいた気配の中値を用いる。売り気配と買い気配の両方

が一般気配として提示されていなかった時刻は、サンプルから除外する。例えば特別気配が提示されていた場合には回帰分析から除かれる。さらに寄前気配については、寄付き時刻が9時0分または12時30分よりも遅れた場合に、サンプルから除外する。特別気配が提示されているときには、価格調整が取引ルールによって遅れ、 β を高くするためである。

4 寄付き前の学習

表1は、寄付き前の気配について、各銘柄ごと4つの基準とする価格ごとに unbiasedness regression を行い、係数の平均をまとめた表である。但し、前場・後場それぞれ寄付き前5分より前は5分ごと、5分前以降は1分ごとの結果を掲載している。

まず9時前の前場寄付き前は、どの価格を基準としても、8時59分以前は係数が1よりも小さく、しかも1%水準で有意に1と異なる。また、9時に近づくほど係数の平均は1に近づいており、9時0分の気配を用いた場合の係数の平均は1よりも大きい。これは、先行研究と同様、取引を伴わない気配の配信のみで情報伝達または価格発見が行われていることを示唆している。また、係数が1よりも小さいことは寄付き前の気配の変動が大きいことを意味しているが、取引が行われないため価格を大きく動かしながら均衡価格を探索しているという解釈も可能である。

後場寄付き前の推計結果にも同様の傾向があるが、前場寄付き前に比べて係数の平均は1により近い。例えば、後場終値を基準とした場合、8時59分における係数の平均は0.706であるのに対し、12時5分から12時29分の係数の平均は0.9を上回っている。後場寄付き前の気配の配信は、前場における取引が終了してから1時間30分も経過していない状態で行われるのに対し、前場寄付き前の気配の配信は、前取引日の取引終了時刻15時から17時間以上経過した状態で行われる。中断時間が長いほどその間に新情報が発生する可能性が高いのであれば、それを反映して前場寄付き前の価格発見が困難となり、unbiasedness regression の係数が低くなるのに対し、後場寄付き前は中断時間が短いために係数が高くなっている可能性がある。また、前場寄付き前8時59分の係数は、後場寄付き前の係数の平均の最小値よりも小さい。これは、前場寄付き前の価格発見は、寄付き直前に集中して行われることを示唆している。

係数の大きさが売買高の影響を受けるかを検証するために、400銘柄を平均売買高により4グループに分け、グループ間で平均に差があるかを検定してみると、前場寄付き前の気配において、後場始値・前場終値・前場始値を基準とした場合に、平均に差があり、後場寄付き前の気配については、前場終値を基準とした場合にのみ売買高により差がある、という結果であった。売買高による差をより詳しく見るために、表2は、前場寄付き前については前場終値を基準とした推計値について、後場寄付き前については後場終値を基準とした推計値について、売買高グループごと平均をまとめている。また括弧内にはグループごとのRMSEの平均を示している。これより、前場寄付き前には、売買が活発な銘柄において、係数が低くなる傾向があることがわかる。寄付き前の気配の変動が売買高により異なる原因として、情報生産活動や価格操作を狙った注文の流入などが考えられるが、より詳しい分析は今後の課題である。

5 取引時間中における価格形成

表3は、取引時間中において各銘柄ごと unbiasedness regression を行い、その係数の平均をまとめたものである。表1と同様、基準となる価格として後場終値、後場始値、前場終値、前場始値を用い、それぞれに係数の平均を報告している。いずれの価格を基準にした場合も、9時の係数の平均は1よりも大きく、しかも1と有意に異なる。その後、前場始値を基準とした場合を除き、係数の平均は時間の経過とともに若干低下する傾向を示し、特に後場終値を基準とした場合には、9時1分から9時15分まで係数の平均は1と有意に異ならない。このように取引開始後15分程度は係数が1近辺にとどまっているが、9時20分、25分、30分については、係数の平均はどの価格を基準とした場合も1よりも大きく、10時55分以降は係数の値が低下し再び1に近づいている。これは、取引開始時には価格形成に遅れがみられるが、その遅れは15分程度で修正されるものの、その後再び価格変化が過小となり、前場終了間際にはまた適切な価格形成が行われることを示唆している。

前場において、9時15分以降、価格調整が遅くなるが、この期間は売買高グループ間で係数の大きさに差がある。表4は売買高グループごとの係数の平均およびRMSEの平均を示しているが、表2と同様、前場については前場終値を基準とした推計結果、後場については後場終値を基準とした推計結果についてまとめている。表4より、9時15分以降の売買高グループごとの差は、売買高が大きなグループで係数の平均が大きいことにより発生していることがわかる。この原因の一つの可能性は次のようなものである。取引開始後の価格発見が数分で終了すると、ビッド・アスク・スプレッドが狭くなり、最良気配における数量が大きくなり板の厚みが増す、それにより価格変化が遅くなり、unbiasedness regression の係数が大きくなる。売買が不活発な銘柄では、板が十分に厚くならないため、価格変化の遅れはみられない。但し、これに関する詳細な分析は今後の課題である。

表3より、後場の価格形成についても前場と同様、取引開始後に係数が1より大きく、売買高グループごとに係数の平均が異なる傾向がある。但し、前場と異なり、基準とする価格により結果が違ふという点で頑健な結果ではない。一方、後場終了間際において、後場終値を基準とした場合の係数は、平均が1よりも小さく、さらに売買高グループにより平均が異なる。表4が売買高グループごとの係数の平均を示しているが、売買高が高い銘柄で係数がより小さい傾向があることがわかる。これは、売買が活発な銘柄において、1日の取引終了前に価格変動が大きくなっている可能性を示唆している。考えられる理由の一つは、売買が活発な銘柄の取引には機関投資家が多数参加しており、特に終値で取引したいという需要のある機関投資家が取引終了間際に流動性を需要するために価格変動が大きくなっている、というものである。一方、後場終値が本源的価値の代理変数として望ましくないことから発生している可能性もある。後者の場合には、後場終値を用いたすべての推計結果に対する解釈において注意が必要になる。

6 References

- Amihud, Y., and H. Mendelson, 1991, Volatility, efficiency, and trading - evidence from the Japanese stock-market, *Journal of Finance* 46, 1765-1789.
- Barclay, Michael J., and Terrence Hendershott, 2003, Price discovery and trading after hours, *Review of Financial Studies* 16, 1041-73.
- Barclay, M. J., and T. Hendershott, 2008, A comparison of trading and non-trading mechanisms for price discovery, *Journal of Empirical Finance* 15, 839-849.
- Biais, Bruno, Pierre Hillion, and Chester Spatt, 1999, Price discovery and learning during the preopening period in the Paris Bourse, *Journal of Political Economy* 107, 1218-48.
- Cao, Charles, Eric Ghysels, and Frank M. Hatheway, 2000, Price discovery without trading: Evidence from the nasdaq preopening, *Journal of Finance* 55, 1339-65.
- Chen, Tao, Jun Cai, and Richard Y. K. Ho, 2009, Intraday information efficiency on the Chinese equity market, *China Economic Review* 20, 527-41.
- Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315-35.
- Madhavan, Ananth, and Venkatesh Panchapagesan, 2000, Price discovery in auction markets: A look inside the black box, *Review of Financial Studies* 13, 627-58.

表 1: 寄付き前

Time	PM Close	PM Open	AM Close	AM Open
835	0.128 *	0.311 *	0.321 *	0.458 **
840	0.167 *	0.366 *	0.375 *	0.516 *
845	0.218 *	0.422 *	0.429 *	0.565 **
850	0.28 *	0.487 *	0.494 *	0.625 **
855	0.414 *	0.609 **	0.611 **	0.715 **
856	0.447 *	0.635 *	0.636 *	0.732 **
857	0.497 *	0.67 **	0.67 **	0.756 **
858	0.566 *	0.719 **	0.717 **	0.789 **
859	0.706 *	0.824 **	0.82 **	0.878 **
900	1.05 *	1.05 *	1.038 *	0.789 **
1205	0.947 *	0.97 *	0.407 **	0.984
1210	0.904 *	0.931 *	0.411 *	0.935 *
1215	0.918 *	0.945 *	0.447 *	0.957 *
1220	0.927 *	0.954 *	0.517 *	0.978
1225	0.933 *	0.959 *	0.662 *	0.992
1226	0.937 *	0.962 *	0.703 *	0.998
1227	0.944 *	0.968 *	0.758 *	1.009
1228	0.951 *	0.973 *	0.85 **	1.019
1229	0.985 *	1.002 *	1.123 **	1.072 *
1230	1.000	0.951	1.31 **	1.093 *

各銘柄各時刻(分)において unbiasedness regression を行い、係数の平均を示している。*は、係数が1%水準で1より有意に異なることを表す。+は、売買高グループごとの係数が1%水準で有意に異なることを表す。

表 2: 寄付き前

Time	Total		売買高グループ							
	β	(RMSE)	1(high)		2		3		4(low)	
	β	(RMSE)	β	(RMSE)	β	(RMSE)	β	(RMSE)	β	(RMSE)
前場終値基準										
835	0.321 *	(.0026)	0.328 *	(.0025)	0.27 *	(.0025)	0.314 *	(.0027)	0.37 *	(.0028)
840	0.375 *	(.0026)	0.387 *	(.0025)	0.329 *	(.0025)	0.372 *	(.0027)	0.412 *	(.0028)
845	0.429 *	(.0026)	0.433 *	(.0025)	0.374 *	(.0025)	0.436 *	(.0027)	0.474 *	(.0028)
850	0.494 *	(.0026)	0.474 *	(.0025)	0.436 *	(.0025)	0.511 *	(.0027)	0.553 *	(.0028)
855	0.611 **	(.0026)	0.586 *	(.0025)	0.542 *	(.0025)	0.652 *	(.0027)	0.665 *	(.0028)
856	0.636 *	(.0026)	0.615 *	(.0025)	0.569 *	(.0025)	0.675 *	(.0027)	0.687 *	(.0028)
857	0.67 **	(.0026)	0.644 *	(.0025)	0.598 *	(.0025)	0.715 *	(.0027)	0.722 *	(.0028)
858	0.717 **	(.0026)	0.69 *	(.0025)	0.643 *	(.0025)	0.768 *	(.0027)	0.767 *	(.0028)
859	0.82 **	(.0026)	0.792 *	(.0025)	0.761 *	(.0025)	0.86 *	(.0027)	0.868 *	(.0028)
900	1.038 *	(.0026)	1.04 *	(.0025)	1.05 *	(.0025)	1.037 *	(.0027)	1.027 *	(.0028)
後場終値基準										
1205	0.947 *	(.0113)	0.959 *	(.0103)	0.928 *	(.0112)	0.949 *	(.0113)	0.953 *	(.0121)
1210	0.904 *	(.0116)	0.913 *	(.0107)	0.863 *	(.0117)	0.92 *	(.0115)	0.921 *	(.0123)
1215	0.918 *	(.0113)	0.935 *	(.0104)	0.896 *	(.0113)	0.921 *	(.0114)	0.92 *	(.0123)
1220	0.927 *	(.0111)	0.951 *	(.0101)	0.92 *	(.011)	0.92 *	(.0113)	0.917 *	(.0122)
1225	0.933 *	(.0109)	0.954 *	(.01)	0.927 *	(.0107)	0.932 *	(.011)	0.92 *	(.012)
1226	0.937 *	(.0109)	0.956 *	(.0099)	0.932 *	(.0106)	0.936 *	(.0109)	0.923 *	(.0119)
1227	0.944 *	(.0108)	0.962 *	(.0098)	0.94 *	(.0105)	0.942 *	(.0108)	0.933 *	(.0118)
1228	0.951 *	(.0107)	0.966 *	(.0098)	0.946 *	(.0105)	0.949 *	(.0107)	0.941 *	(.0117)
1229	0.985 *	(.0104)	0.992	(.0095)	0.984 *	(.0101)	0.98 *	(.0105)	0.983	(.0114)
1230	1.000	(.0102)	0.999	(.0093)	0.999	(.0099)	0.997	(.0103)	1.007	(.0112)

各銘柄各時刻(分)において unbiasedness regression を行い、係数の平均を示している。()内は、RMSE の平均を示している。*は、係数が1%水準で1より有意に異なることを表す。+は、売買高グループごとの係数が1%水準で有意に異なることを表す。

表 3: 寄付き後

Time	PM Close	PM Open	AM Close	AM Open
900	1.05 *	1.05 *	1.038 *	.789 **+
901	1.018	1.041 *	1.031 *	.899 **+
902	1.001	1.037 *	1.028 *	0.929 **+
905	1.008	1.044 *	1.036 *	1.007 +
910	1.000 +	1.039 *	1.034 *	1.041 **+
915	1.008 +	1.039 **+	1.034 **+	1.075 **+
920	1.017 **+	1.044 **+	1.04 **+	1.101 **+
925	1.022 **+	1.046 **+	1.043 **+	1.125 **+
930	1.031 *	1.051 *	1.048 *	1.155 **+
1055	0.994	1.01 *	1.007 *	1.075 *
1056	0.994	1.009 *	1.006 *	1.074 *
1057	0.993	1.008 *	1.006 *	1.073 *
1058	0.992	1.008 *	1.006 *	1.071 *
1059	0.986 *	1.006 *	1.004 *	1.064 *
1230	1.000	0.951	1.31 **+	1.093 *
1231	1.001	1.116 **+	1.281 **+	1.091 *
1232	1.002	1.127 **+	1.251 **+	1.089 *
1235	1.006 *	1.113 *	1.195 **+	1.089 *
1240	0.998	1.061 *	1.131 **+	1.075 *
1245	0.997	1.072 **+	1.128 **+	1.073 *
1430	0.994 **+	1.052 *	1.056 *	1.052 *
1435	0.994 **+	1.054 *	1.058 *	1.05 *
1440	0.993 **+	1.04 *	1.046 *	1.047 *
1445	0.995 **+	1.031 *	1.038 *	1.047 *
1450	0.995 **+	1.029 *	1.034 *	1.045 *
1455	0.994 **+	1.021 *	1.027 *	1.042 *
1456	0.995 **+	1.02 *	1.026 *	1.043 *
1457	0.996 **+	1.02 *	1.026 *	1.043 *
1458	0.997 **+	1.02 *	1.027 *	1.043 *
1459	1.000 +	1.017	1.024 *	1.044 *

各銘柄各時刻(分)において unbiasedness regression を行い、係数の平均を示している。*は、係数が1%水準で1より有意に異なることを表す。+は、売買高グループごとの係数が1%水準で有意に異なることを表す。

表 4: 寄付き後

Time	売買高グループ									
	Total		1(high)		2		3		4(low)	
	β	(RMSE)	β	(RMSE)	β	(RMSE)	β	(RMSE)	β	(RMSE)
前場終値基準										
900	1.038 *	(.0026)	1.04 *	(.0025)	1.05 *	(.0025)	1.037 *	(.0027)	1.027 *	(.0028)
901	1.031 *	(.0026)	1.027 *	(.0025)	1.046 *	(.0025)	1.029 *	(.0027)	1.024 *	(.0028)
902	1.028 *	(.0026)	1.024 *	(.0025)	1.042 *	(.0025)	1.026 *	(.0027)	1.022 *	(.0028)
905	1.036 *	(.0026)	1.032 *	(.0025)	1.044 *	(.0025)	1.041 *	(.0027)	1.026 *	(.0028)
910	1.034 *	(.0026)	1.041 *	(.0025)	1.041 *	(.0025)	1.034 *	(.0027)	1.018 *	(.0028)
915	1.034 **	(.0026)	1.043 *	(.0025)	1.042 *	(.0025)	1.032 *	(.0027)	1.019 *	(.0028)
920	1.04 **	(.0026)	1.049 *	(.0025)	1.043 *	(.0025)	1.04 *	(.0027)	1.025 *	(.0028)
925	1.043 **	(.0026)	1.053 *	(.0025)	1.046 *	(.0025)	1.047 *	(.0027)	1.027 *	(.0028)
930	1.048 *	(.0026)	1.055 *	(.0025)	1.049 *	(.0025)	1.049 *	(.0027)	1.038 *	(.0028)
後場終値基準										
1430	0.994 **	(.0054)	0.982 *	(.0051)	0.994	(.0052)	0.993 *	(.0056)	1.008 *	(.0059)
1435	0.994 **	(.0051)	0.982 *	(.0049)	0.991 *	(.0049)	0.994 *	(.0053)	1.008 *	(.0055)
1440	0.993 **	(.0048)	0.982 *	(.0046)	0.99 *	(.0046)	0.992 *	(.0049)	1.005	(.0052)
1445	0.995 **	(.0045)	0.987 *	(.0042)	0.992 *	(.0043)	0.995	(.0046)	1.005	(.0048)
1450	0.995 **	(.0041)	0.99 *	(.0038)	0.991 *	(.0039)	0.995 *	(.0042)	1.005	(.0044)
1455	0.994 **	(.0035)	0.99 *	(.0033)	0.99 *	(.0034)	0.994 *	(.0036)	1.003	(.0038)
1456	0.995 **	(.0034)	0.991 *	(.0032)	0.992 *	(.0032)	0.995 *	(.0035)	1.003	(.0036)
1457	0.996 **	(.0032)	0.992 *	(.003)	0.992 *	(.0031)	0.996	(.0033)	1.004	(.0034)
1458	0.997 **	(.003)	0.994 *	(.0029)	0.994 *	(.0029)	0.998	(.0031)	1.004 *	(.0032)
1459	1.000 +	(.0026)	0.998	(.0025)	0.998	(.0025)	1.001	(.0027)	1.005 *	(.0028)

各銘柄各時刻(分)において unbiasedness regression を行い、係数の平均を示している。()内は、RMSE の平均を示している。*は、係数が1%水準で1より有意に異なることを表す。+は、売買高グループごとの係数が1%水準で有意に異なることを表す。

高頻度データ分析：取引時間と価格の相互依存性について

林 高樹*
慶應義塾大学

「桜の季節に計量ファイナンス」研究集会

First draft: March 30, 2010

This version: April 30, 2010

概要

H.-Yoshida(06,08), H.-Jacod-Yoshida(08) では、(過去の取引履歴にも依存しうるような) ランダムな取引時間にある種の可予測性 (predictability) 条件を付与することで、累積共分散やべき乗変動に対する推定量の漸近混合正規性を導いている。一方、最近、取引時間が“内生的”な場合、すなわち価格変動と同時決定するようなケースに関して実現ボラティリティの漸近分布を調べる研究も行われ始めている。本稿では、これら最近の研究動向を簡潔にレビューし、また高頻度データによる相互依存性の評価方法について提案し、実証分析を行う。

1 はじめに

本研究は、高頻度データを活用して証券価格とその取引時間との相互依存性を調べることを目的とする。その最初のステップとして、本稿では、まず著者が最近関わっている研究を中心に、文献において、どのように依存性がモデルによって表現されているかをレビューする。次に、既存研究を踏まえ、そのような依存性の程度を実際のデータにより計量する方法を考える。

本稿では“取引時点 (または取引時刻)”は実際に取引が発生しただちにデータベースに記録された時間と考え、“サンプリング時点,” “観測時点”と区別しないで使用する。また、一つの取引が発生してから次の取引が発生するまでの時間間隔を“取引間隔”あるいは“デュレーション”と呼ぶことにする。

本稿では、特に断らない限り、“実現ボラティリティ”(realized volatility), “実現べき乗変動”(realized power variation) に関する文献の標準的設定に倣い、証券価格はボラティリティ過程 σ_s を持つ連続 Itô semimartingale

$$(1.1) \quad X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s$$

に従って変動し、それが離散時点で観測される (=取引される) と仮定する。また、確率過程 μ_s, σ_s に対して標準的な正則条件も満たされていると仮定する。一方、マイクロストラクチャ・ノイズについては仮定しない。以下では、実現ボラティリティを RV, 実現べき乗変動を RPV, 実現ボラティリティの推定対象である累積分散 $\int_0^t \sigma_s^2 ds$ を IV と書くことにする。

取引時刻列 $(T(n, i))_{i \in \mathbb{Z}_+}$ については、特に断らない限り、 $T(n, 0) = 0$ から出発する停止時刻 (stopping time) の列で、 $T(n, i) \uparrow \infty$ a.s., as $i \rightarrow \infty$ ($n \geq 1$) とする。また、記号として、は次セクションで紹介する HJY(08) を参考に、取引間隔・デュレーション $\Delta(n, i) := T(n, i) - T(n, i-1)$, 価格変化 $\Delta_i^n X := X_{T(n, i)} - X_{T(n, i-1)}$, 取引区間数 (すなわち、収益率データ数) $N_t^n = \max\{i \geq 0; T(n, i) \leq t\}$, 最大取引間隔 $\pi_t^n := \sup_{i=1, \dots, N_t^n+1} \Delta(n, i)$

〒223-8526 横浜市港北区日吉4-1-1 協生館. Email: takaki@kbs.keio.ac.jp

を導入する.*1. RV/RPV に関する既存研究がそうであるように、高頻度データ分析の文脈から本稿においても $\pi_t^n \xrightarrow{P} 0$ (as $n \rightarrow \infty$) なる設定を考える。

2 取引時間の依存性の表現

本セクションでは、価格過程に対する連続 (Itô) semimartingale モデルの下での、取引時間の確率モデル表現について、著者が関わった研究を中心に、最近の研究について取り上げる。特に、これらの研究の結論にではなく、価格過程と取引時間との依存関係に焦点を当てる。なお、以下では、応用的観点から、数学的厳密性を犠牲にしながらもより直感的な説明を心がける。厳密な定義、議論に興味のある読者は、引用文献を直接参照されたい。

2.1 H.-Jacod-Yoshida(08) の可予測性

H.-Jacod-Yoshida([4]) は、RPV に関する拡張研究である (以下、HJY(08) ないしは HJY と呼ぶ)。既存の RPV 研究 (例, [1]) は取引間隔が等間隔に並んでいるものが大部分を占め、非等間隔の取引間隔、取引過程とは独立なランダムな取引間隔あるいは非ランダムな取引間隔を扱った ([2]) 研究もあったが、HJY は取引時間・取引間隔が価格過程に依存するような設定に一般化した。

まず X_t は 1 次元の連続 Itô semimartingale (1.1) とする。 σ_t としては 1 次元の (連続パスを持つとは限らない) semimartingale とする (詳細は同論文 p.3 を参照せよ)。一方、取引時点 (サンプリング・スキーム) としては先述の停止時刻列を出発点として、以下のような構造のもの考える (原著の直接引用) :

[Assumption(C)] There is a sub-filtration $(\mathcal{F}_t^0)_{t \geq 0}$ of $(\mathcal{F}_t)_{t \geq 0}$, with respect to which W and b and σ are adapted, and such that any (\mathcal{F}_t^0) -martingale is also an (\mathcal{F}_t) -martingale, and such that for all $n, i \geq 1$, $T(n, i)$ is an (\mathcal{F}_t) -stopping time which, conditionally on $\mathcal{F}_{T(n, i-1)}$, is independent of the σ -field $\mathcal{F}^0 = \vee_{t > 0} \mathcal{F}_t^0$.

このようなスキームの例として、非確率的なスキーム、(確率過程 (W, X, b, σ) に) 独立なスキームなどの trivial なケースは勿論のこと、次のような trivial でないスキームが含まれる。

- (i) $T(n, i) \in \mathcal{F}_{T(n, i-1)}$ ($\mathcal{F}_t^0 = \mathcal{F}_t$ のケース)。
- (ii) “混合再生スキーム” (“Mixed renewal schemes”) :

$$T(n, i + 1) = T(n, i) + \frac{1}{r_n} v_{T(n, i)}^n \varepsilon(n, i + 1),$$

ここで、 $(\varepsilon(n, i) : i, n \geq 1)$ は、 (Ω, \mathcal{F}, P) 上で定義される *i.i.d.* の正の確率変数列で、 \mathcal{F}^0 と独立であるとする (ある種のモーメント条件を満たすとする)。 (v^n) は正の (\mathcal{F}_t) -適度な確率過程 (の列) である。 r_n は $N_t^n = O_P(r_n)$ であるような deterministic なレートであり、詳細は後述する (例えば、定間隔サンプリングのケースにおいては $r_n = 1/\Delta_n$)

ここで、 \mathcal{F}^0 は安定収束が成立するのに必要な粗さの σ -field として導入されたものである (同論文の証明した CLT における安定収束は一般に元の大きさの \mathcal{F} では成立しない。) 以下では、仮定 (C) を便宜上 “HJY-predictability” と呼ぶことにする。

*1 文献によっては、例えば、最大区間 π_t^n の定義は—特に最終観測区間の扱いに関して—多少異なったりするが、以下の議論においては本質的ではない。

同論文では、RPV を一般化した次のような統計量を考えた。 $p > 0, q \geq 0$ に対して、

$$(2.1) \quad V(p, q)_t^n := \sum_{i=1}^{N_t^n} \Delta(n, i)^{q+1-\frac{p}{2}} |\Delta_i^n X|^p.$$

そして、適当な正則条件の下での $\tau_n^q V(p, q)_t^n$ に対する大数の法則、中心極限定理を導いた。その変形バージョンについて後述する。

2.2 H.-Yoshida(06,08) の可予測性

次に、H.-Yoshida([6], [7]) において用いられた取引時間の表現を紹介する (以下、HY(06), HY(08) 等と書く)。HY においては、証券 X, Y の取引価格は連続 semimartingale であり、取引時間 $(S^i)_{i \in \mathbb{Z}_+}, (T^j)_{j \in \mathbb{Z}_+}$ はそれぞれが時点ゼロでゼロから出発し、 i, j の増大に伴って発散するような停止時刻の単調増加列であるとする。一連の研究では、インデックス $n \rightarrow \infty$ に従って (S^i) や (T^j) の各々の最大取引間隔がゼロに縮小してゆく状況において、非同期離散観測される X, Y から、両者の共分散を一致的に推定できるような統計量 (“nonsynchronous covariance estimator”, 以下、“HY 推定量”と呼ぶ) を提案し、さらに HY 推定量の漸近分布を導出している。

その中で、HY 推定量の漸近混合正規性を導くための十分条件として、 X, Y の特性量に対する仮定や、 (S^i) and (T^j) に対する仮定が必要となってくる。特に、 (S^i) and (T^j) の、情報集合 (filtration) に対する measurability に関する仮定が重要な役割を果たす (原著の直接引用):

[A2]: For every $n, i \in \mathbb{N}$, S^i and T^i are $(\mathcal{G}_t^{(n)})$ -stopping times, where $\{\mathcal{G}_t^{(n)}\}_{t \in \mathbb{R}_+}$ is the filtration given by $\mathcal{G}_t^{(n)} = \mathcal{F}_{(t-b_n^i)_+ \vee 0}$ for $t \in \mathbb{R}_+$.

ここで、 $\xi < 1$ であり、 (b_n) は HY 推定量の推定誤差過程の 2 次変動を $n \rightarrow \infty$ において (trivial でない極限に) 収束させるような deterministic なスケーリング係数 (の逆数) である。サイズ n の等間隔サンプリングの場合には、 $b_n \propto 1/n$ である。換言すれば、観測時刻 S^i, T^i が、現在時刻 t における完全な情報集合 \mathcal{F}_t よりも少しだけ小さい $\mathcal{G}_t^{(n)}$ (2つの filtration は極限において差は解消される) に関する停止時刻になっているという仮定である。すなわち、 S^i, T^i はランダムで、価格の履歴 X, Y に依存しても良いが、それでも局所的には、deterministic に振舞う (予測可能である) ことが規定されることになる。^{*2}

同論文は [A2] 条件を “strong predictability” と呼んでいる。以下では、“HY-predictability” と呼ぶことにする。

HJY-predictability, HY-predictability とも、既存研究において deterministic ないしは、価格過程とは独立なランダムな観測時刻列によってサンプルされた価格変動を元に計測してきた RV/RCV や RPV に関する極限定理を、観測時刻列が価格過程の過去の履歴に依存するような状況に一般化するために導入された条件である。

さて、RV/RPV 型の統計量が持つ漸近的挙動を考える手がかりとして、 X の (絶対値なし) “3 次変動” を取り上げよう。

等間隔に観測時刻列が並んでいる場合、連続 Itô semimartingale X (式 (1.1)) に対して、ボラティリティがやはり式 (1.1) のようなダイナミクスを持つ確率ボラティリティ・モデルであっても、

$$(2.2) \quad \sqrt{n} \sum_{i=1}^n (\Delta_i^n X)^3 \xrightarrow{P} 0,$$

^{*2} [A2] のほか、例えば、 $\frac{4}{5} \vee \xi < \xi' < 1$ を満たす ξ' に対して、[A4]: $\tau_n(t) = o_P(b_n^{\xi'})$ ($\forall t$)、なる条件—時点 t までの最大区間幅 $\tau_n(t)$ は “平均” 区間幅 b_n に比べて大き過ぎない—を課している (注: 原著の記号をそのまま使用)。ここで、[A4] 条件は sub-filtration $(\mathcal{G}_t^{(n)})$ の (\mathcal{F}_t) に対する (時間差による) 相対的大きさを規定する ξ の値に間接的に依存していることにも注意せよ。

となることが示される (cf. [9], [10]). 非等間隔, あるいは独立な観測時刻列であっても同様である. さらに, 価格と観測時刻間の従属性を許す RPV の収束を扱う HJY も同様である. “predictability”条件は, 取引時間を価格に (少しだけ) 先行させる条件であるためである. 次に紹介する [13], [10] においてはこの“3 次変動”がゼロへ収束しない状況に着目する.

2.3 その他の既存研究

深澤 (09)

[13] では, (\mathcal{F}_t) -停止時刻列 $(T(n, i))$ と X が独立でないケースにおいて, RV の漸近分布を求めた.

同論文の中では, (十分な可積分性を持つ) $(T(n, i))$ において離散時点サンプルされるブラウン運動 $(W_{T(n, i)})$ に対して以下のような局所的な変動分布のモーメントに関する仮定 (“local homogeneity”) を置く. 正数列 (ϵ_n) , $\epsilon_n \rightarrow 0$ と (\mathcal{F}_t) -適合 (左) 連続過程 $\kappa_t^{(3)}, \kappa_t^{(4)}$ が存在し (後者は常に正), $i = 0, 1, \dots, N_t^n$ に対して一様に

$$\begin{aligned} G_{i,n}^3 / G_{i,n}^2 &= \kappa_{T(n,i)}^{(3)} \epsilon_n + o_p(\epsilon_n), & G_{i,n}^4 / G_{i,n}^2 &= \kappa_{T(n,i)}^{(4)} \epsilon_n^2 + o_p(\epsilon_n) \\ G_{i,n}^{2k} / G_{i,n}^2 &= o_p(\epsilon_n^k), & k &= 3, 4, 6. \end{aligned}$$

ただし, $G_{i,n}^k := E \left[(\Delta_{i+1}^n W)^k \middle| \mathcal{F}_{T(n,i)} \right]$ である. $\kappa^{(3)}$ は変動分布の (局所的な) 歪度, $\kappa^{(4)}$ は尖度を表す. ここでスケール ϵ_n はオーダー的には, HJY(08) での “ $\tau_n^{-1/2}$ ” に, HY(06), HY(08) での “ $b_n^{1/2}$ ” — 直感的にはサンプルサイズを n として “ $n^{-1/2}$ ” に一対応している.

同論文では, Itô の公式を適用することで得られる

$$(2.3) \quad E \left[\Delta(n, i) \Delta_i^n W \middle| \mathcal{F}_{T(n, i-1)} \right] = \frac{1}{3} E \left[(\Delta_i^n W)^3 \middle| \mathcal{F}_{T(n, i-1)} \right]$$

なる関係を用いて

$$(2.4) \quad \epsilon_n^{-1} \sum_{i=1}^{N_t^n} \Delta(n, i) \Delta_i^n W \xrightarrow{P} \frac{1}{3} \int_0^t \kappa_s^{(3)} ds$$

であることを導く. この非ゼロの確率過程への収束が, 同論文で得られた RV の誤差の漸近分布のバイアスを発生させることになる. 続編 [3] において, モデルを拡張して議論を行っている.

Li, Mykland, Renault, Zhang, and Zheng (09)

[10] は, RV に関するこれまでの既存研究では無視されていた (絶対値なし) “3 次変動” に trivial でない極限過程を明示的に仮定することによって, 価格と取引時間の従属性を考慮に入れた RV の漸近分布を導出した. 具体的には, 対数価格過程 X (それぞれ (\mathcal{F}_t) -適合で局所有界な, ドリフト過程 μ_t , ボラティリティ過程 $\sigma_t \geq c > 0$, $\exists c > 0$) に対して, 適当な確率過程 u, v が存在して

$$(2.5) \quad \sqrt{n} \sum_{i=1}^{N_t^n} (\Delta_i^n X)^3 \xrightarrow{P} \int_0^t v_s ds, \quad \forall t \geq 0$$

$$(2.6) \quad n \sum_{i=1}^{N_t^n} (\Delta_i^n X)^4 \xrightarrow{P} \int_0^t u_s ds, \quad \forall t \geq 0$$

であると仮定する。また、最大区間長 $\pi_t^n = o_P(n^{-2/3-\varepsilon})$ ($\exists \varepsilon > 0$) とする。この時、追加的な正則条件の下で、

$$(2.7) \quad \sqrt{n}(\text{RV} - \text{IV}) \xrightarrow{\mathcal{F}\text{-stably}} \underbrace{\frac{2}{3} \int_0^t \frac{v_s}{\sigma_s^2} dX_s}_{\text{漸近バイアス}} + \int_0^t \sqrt{\frac{2}{3}u_s - \frac{4}{9}\frac{v_s^2}{\sigma_s^2}} dB_s$$

となる。ここで、 B は元の σ -field とは独立な標準ブラウン運動である。^{*3}

同論文では、漸近バイアスの大きさを評価することにより、取引時間の endogeneity を評価する検定法を提案し、実証分析を行った。

ボラティリティ推定の文脈の中で取引時間と価格の相互依存性を論じた研究としては、この他、[11], [12] などがある。

3 取引価格と取引時間の依存性の定量的評価

3.1 方法論

価格過程 X が取引時点 $T(n, i)$ (あるいは取引間隔 (n, i)) と互いに独立でない場合一体どのようにデータから観測される変量によってそれを把握することが出来るであろうか。最も単純な例として考えられるのは収益率とデュレーションの間に相関が見られるケースであり、高頻度データの利用という設定から、最大観測区間幅 $\pi_t^n \rightarrow 0$ における漸近的挙動を考え

$$(3.1) \quad (\text{適当なスケール}) \times \sum_{i=1}^{N_t^n} \Delta(n, i) \Delta_i^n X \xrightarrow{P} 0,$$

かどうかを見極めることであろう。上で紹介した [13] においても、漸近分布を導出する中で極限がゼロに行かないケースを扱っている。

一方、収益率の歪度がゼロでないケースも考えられる。すなわち、

$$(3.2) \quad (\text{適当なスケール}) \times \sum_{i=1}^{N_t^n} (\Delta_i^n X)^3 \xrightarrow{P} 0.$$

このような極限がゼロに収束しないケースが価格と取引時点の非独立性、同時決定性 (endogeneity) につながる点を指摘しているのが、上述のように [13], [10] である。ここで、現在考えている連続 semimartingale の設定においては、(3.1)(3.2) における“(適当なスケール)”とは“ \sqrt{n} ”である。

なお、(3.1) と (3.2) との関連性については、[13] にもある通り $X = W$ のケースにおける先述の等式 (2.3) が直観的な理解に役立つであろう。

本稿では、HJY(08) の成果をベースに、(3.1)(3.2) を包含するクラスとして、次の形を持つ統計量を取り上げる。べき指数 α, β ($\alpha \geq 1 - \beta/2, \beta > 0$) に対して、

$$(3.3) \quad U(\alpha, \beta)_t^n := \sum_{i=1}^{N_t^n} \Delta(n, i)^\alpha (\Delta_i^n X)^\beta$$

を定義する (勿論、右辺が定義されるような β の選択を前提とする)。これに適当な大きさの (非確率的な) スケール変換を行うことによって、最大区間幅が縮小 ($\pi_t^n \rightarrow 0$) しても発散することなく有限な大きさに留まるようにす

^{*3} 但し、 (\mathcal{F}_t) の大きさに制約を与えている—高々有限個の連続 martingale で生成される filtration であるとする。

る。この統計量は上記 RPV を一般化した (2.1) 内の絶対値をはずしたバージョンで、やはり HJY(08) においても言及されているものである。^{*4}

(3.3) は、取引時点 $T(n, i)$ において離散観察された価格変動列 $(\Delta_i^n X)$ のみから計算される RV/RPV 型統計量からの自然な一般化である。

高頻度データを用いて両者の依存性を定量評価するための具体的方法論として、同論文で示された漸近定理を適用する。同論文に従い、準備として、先述の filtration に関する仮定 (C) に加えて、次のような取引間隔の履歴を記述するような確率過程

$$(3.4) \quad A(q)_t^n := r_n^{q-1} \sum_{i=1}^{N_t^n} \Delta(n, i)^q$$

を導入し (任意の $q \geq 0$ に対して), その漸近的挙動に関する仮定を与える。ここで, r_n は, $A(q)_t^n / r_n^{q-1} = O_P(1)$ であるような deterministic な正の増加列である。

[Assumption(D(q))] ある (必然的に, 非負の) (\mathcal{F}_t^0) -optional process $a(q)$ が存在して,^{*5}

$$(3.5) \quad A(q)_t^n \xrightarrow{P} \int_0^t a(q)_s ds, \quad \forall t \geq 0.$$

定義 3.4 より, 特に, $A(0)_t^n = N_t^n / r_n$, $A(1)_t^n \xrightarrow{P} t$ である。また, 等間隔サンプリングのケース ($\Delta(n, i) \equiv \Delta_n$, $T(n, i) = i\Delta_n$) においては, $A(q)_t^n = \Delta_n [t/\Delta_n]^q \xrightarrow{P} t^q (\forall q)$ となる。 $q = 2$ のケース $A(2)_t^n$ は Mykland らのグループが呼ぶところの“quadratic variation of time”であり, 彼らは累積分散に関する推測問題を扱うためにこれを導入した。また, HY の一連の研究 ([5], HY(06), HY(08)) においても, nonsynchronous covariance estimator の漸近的挙動を調べる際にも同様な $q = 2$ に相当する関数 (彼らの場合には非同期の二時系列を併せて記述する必要から 4 個の関数) を導入した。HJY(08) はより一般のべき乗変動に関する漸近論を論ずる必要性から, 必然的にそれを任意の q 次に一般化した。

以上の設定の下で, 次のような極限定理が導かれる:^{*6}

定理 1 (Law of Large Numbers) ある正則条件の下で,

$$(3.6) \quad r_n^{\alpha-1+\frac{\beta}{2}} U(\alpha, \beta)_t^n \xrightarrow{\text{u.c.p.}} \int_0^t \rho_{\sigma_s}(\beta) a(\alpha + \frac{\beta}{2})_s ds =: U(\alpha, \beta)_t.$$

ここで, $\rho_a(\gamma) := E[Z^\gamma]$, $Z \sim N(0, a^2)$ とする。^{*7}

定理 2 (Central Limit Theorem) ある正則条件の下で,

$$(3.7) \quad \sqrt{r_n} \left(r_n^{\alpha-1+\frac{\beta}{2}} U(\alpha, \beta)_t^n - U(\alpha, \beta)_t \right) \xrightarrow{\mathcal{F}^0\text{-stably}} \int_0^t \sqrt{(\rho_{\sigma_s}(2\beta) - \rho_{\sigma_s}(\beta)^2) a(2\alpha + \beta)_s} dB_s$$

^{*4} ただし, 同論文においては式 (3.3) を含むより一般的なクラスが定義されている。

^{*5} 同論文では, Assumption (C) や $T(n, i)$, π_t^n の漸近的挙動 (セクション 1 で述べた仮定) が成立していることを Assumption(D(q)) のステートメントに含めている。

^{*6} “u.c.p.”は, “convergence in probability, locally uniformly in time”を意味する。

^{*7} 但し, a がランダムな場合には, 条件付正規分布に読み替えるものとする。

ここで、 B は \mathcal{F} に独立な標準ブラウン運動である。極限過程は \mathcal{F}^0 を与えた下で、平均ゼロ、条件付分散

$$(3.8) \quad \int_0^t (\rho_{\sigma_s}(2\beta) - \rho_{\sigma_s}(\beta)^2) a(2\alpha + \beta)_s ds$$

を持つ連続 (conditional) Gaussian martingale である。

定理 1 の極限において、 $\rho_{\sigma}(\beta) = 0 \Leftrightarrow U(\alpha, \beta)_t = 0$ は明らかである (trivial なケースを排するため $a(q) \neq 0$ とする)。上記 2 つの定理を得る前提として、取引時刻列に対しては先に述べた“HJY-predictability”の成立を前提としている。従って、もしこの predictability が成立しない状況—価格と取引時刻とが同時決定される状況 (endogeneity) においては、大数の法則の主張する収束は得られない。直感的には、predictability が成立しない状況においては、高頻度領域 (微小区間) において実現収益率はもはや平均ゼロの正規分布では近似されず、よってその奇数モーメントは一般に非ゼロの値を取るであろうから、上記大数の法則は成立しないと理解される。

以下では、 X が連続 Itô semimartingale であるなど、predictability を除き、定理 1, 2 の十分条件が満たされている状況を想定する。もし本当に predictability が成立するのであれば、任意の奇数 β に対して計測される $U(\alpha, \beta)_t^n$ は、 n が十分大きい時には、 $r_n^{\alpha-1+\frac{\beta}{2}}$ によるスケール変換後も大数の法則に従いその極限であるゼロに十分に近くなければならないはずである。従って、 $U(\alpha, \beta)_t = 0$ を帰無仮説に設定し、これがデータによって支持されるのかを検定すればよい。

しかも、 $\rho_{\sigma_s}(\beta) = 0$ の場合には、大数の法則を再び適用することによって、中心極限定理における極限分布の条件付分散 ($\sigma_s, a(\cdot)$ に依存するランダムな量) に対する一致推定量を観測データにより構成することが出来る。具体的には、 $\rho_{\sigma_s}(\beta) = 0$ の場合、大数の法則によって、

$$r_n^{2\alpha-1+\beta} U(2\alpha, 2\beta)_t^n \xrightarrow{u.c.p} \int_0^t \rho_{\sigma_s}(2\beta) a(2\alpha + \beta)_s ds \equiv U(2\alpha, 2\beta)_t$$

となることから、この prelimit の観測量を分散の推定量に用いれば、スチューデント化された統計量

$$(3.9) \quad Z(\alpha, \beta)_t^n := \frac{U(\alpha, \beta)_t^n}{\sqrt{U(2\alpha, 2\beta)_t^n}} \xrightarrow{L} N(0, 1)$$

は明らかである。従って、帰無仮説 $U(\alpha, \beta)_t = 0$ の下で、この“feasible CLT”を適用することができる。

さらにこれはボラティリティ σ_t のモデルや r_n に依存しないという実用上の特長を持つ。

ところで、高頻度データを分析する際に、異なる日のデータを結合して一本の時系列データとして一度に分析を行うデータ処理も頻繁に行われるが、ここでは、日毎の一日内データセットを分析を個別に行ってからその結果を集計する方法を取る。具体的には、[10] に倣い、次のような手順で p 値の合成を行うことにする。いま、各日のデータが独立に生起すると仮定する。

まず第 i 日目の高頻度データより $Z(\alpha, \beta)_T^n$ ($T = 1$ 日) を計算し、その p 値 p_i を計算する ($i = 1, \dots, n$)。この時、帰無仮説が正しければ、

$$-2 \sum_{i=1}^n \log(p_i) \sim \chi^2(2n)$$

となることから、

$$P_{\text{combined}} = P \left[\chi^2(2n) > -2 \sum_{i=1}^n \log(p_i) \right]$$

によって期間全体の p 値を定義すれば良い (Fisher's combined p -value)

次のセクションでは、 (α, β) の組み合わせ 4 通りについて、実証分析を試みる。

3.2 実証分析例

使用データ

本稿では外国為替市場と株式市場（指数先物）の二種類の高頻度データを使用する。

外国為替市場データは ICAP 社の EBS (Electronic Brokerage System) のヒストリカル・データ『Data Mine』(レベル 5.0), 2007 年 7 月 26 日～同 12 月 5 日まで, 2008 年 9 月の“リーマン・ショック”をはさんだ, 約定データ約 4 か月分である。これは, 全取引記録を 250 ミリ秒毎にスライスしたデータである。各記録時点の ask (bid) サイドの価格データは, その直前の 250 ミリ秒間に発生した全ての ask (bid) サイドの取引の中の最安値 (最高値) を記録したものである。^{*8} 株式市場データは日経 NEEDS『ティックデータ』ファイル (5 本気配版) の中から日経平均先物 2008 年 6 月限, 9 月限, 12 月限の 3 限月, 各々が期近物になってからの約定データ約 3 カ月分である。データの解像度は 1 秒である。同一秒内に発生した複数の取引は, 取引時刻としては同一の値を持つが, 取引発生順に連番が付与されている。

EBS データに対しては, 分析に先立ち次のような前処理を行う。まず, 同一時刻に ask サイド, bid サイドの双方に約定レコードが存在する場合, 約定価格として両サイドの幾何平均を取り, それをその時点での (単一の) 約定価格と見なす。これは, これらの二つの約定価格の内, どちらのサイドが先に発生したか再現できないが故の措置でもある。次に, 為替レートは週末はやそれ以外の銀行休業日により, 取引が実質的に行われない日が不規則に生じることから, 1 ファイルが 30 レコード未満の日は分析対象から機械的に除去する (ファイルサイズにして約 1000 バイト以下)。

一方, NEEDS データに関しては, 同一記録時刻内の連番を川いることができるので, 物理時間ではなく順番のみが大事となる“ティック時間”でのサブサンプリング (subsampling) を行う際に, 1 時点 1 レコードに集約する必要はない。

どちらのデータセットに対しても, ボラティリティの 1 日内変動を事前に除去したりして“非定常な”あるいは deterministic な成分を除くような 1 日内季節調整は (diurnal adjustment) は行わない。また, 指数先物データは, 板寄せ前後の時間帯, 特に市場のオープン直後やクローズの直前の一定時間のデータを除去したりもしない。

上述のように, 本稿で提案する分析は, 価格データが連続 Itô semimartingale であることを前提としている。従って, 今回のような特に高頻度のデータセットを用いる際には, ビッド・アスク・バウンス, 価格の離散性などに代表されるマイクロストラクチャ・ノイズを除去せねばならない。

為替データに関しては, [10] による“preaveraging”アプローチ ([8] の簡易版) を適用する。すなわち, (予め一記録時点 1 レコードに前処理されている) データ内の各記録時点においてその点を含む直前の P 個の価格データの単純平均を計算し (local averaging), 次にその局所平均の時系列を全サンプルに対して, その中から K ($P \ll K$) 個ずつデータを飛ばしながらサブサンプルを生成する。一方, 株価指数先物データに関しては, 全取引の価格と出来高がデータセットに記録されていることから, 少し異なる“preaveraging”を施す。すなわち, 価格の単純平均の代わりに, 直前の K 個の約定価格を出来高加重平均 (VWAP) した値を, (データの使用が冗長にならないよう) K 個飛ばして採取してサブサンプルを生成することにする。

以上で得られた各データセットのサブサンプルに対して, 前セクションの手順に従って分析を行う。

^{*8} あくまでも 250 ミリ秒毎のスナップショットであるため, データに記録されている取引枚数や件数を合計しても, 必ずしも実際の取引量には達しない。

分析結果

具体的には、今回は次の4つのスチューデント化された変量 $Z(\alpha, \beta)_T^2$ に対して combined p -value を計算した:
(1) $\alpha = 1/2, \beta = 1$; (2) $\alpha = 1/2, \beta = 3$; (3) $\alpha = 1, \beta = 1$; (4) $\alpha = 1, \beta = 3$.

外国為替データは、USDJPY を用い、データ期間として、(I) リーマン・ショック前 (2008年7月26日—同9月13日)、(II) リーマン・ショック後 (2008年9月14日—同12月5日)、(III) 全期間の3通りを用意した。使用データの概要を、表1に示す。以下、“preaveraging”のパラメータ $P = 5, K = 20$ のケースの結果を示す。^{*9}

前セクションで述べた手順に従い、まず $Z(\alpha, \beta)_T^2$ ($T = 1$ 日) を各日のデータより計算した。図1にUSDJPYに対して計算されたものを時系列プロットしたものを示す。プロットから、上記4つのケースのいずれの場合にも、データ数102個の中で有意 (絶対値が ± 2 を超える) 日は、1日程度であることが見て取れる。従って、ここまでの結果を見れば、データ1日分を使って有意水準5%程度の (両側) 検定を行ったとしても殆ど全てのケースにおいて帰無仮説は棄却されなかったであろうことが分かる。次に、combined p -value を計測するに先立ち、その前提条件となる p -値系列の独立性をチェックするため、各日の $Z(\alpha, \beta)_T^2$ より計算された p 値を集めた時系列 $\{p_i, i = 1, \dots, 102\}$ に対して、系列相関の有無を確認した。図2はラグ1から20までのUSDJPYデータの自己相関係数のプロット (コログラム) である。同図より、同時系列の無相関性が読み取れることから、少なくとも、独立性を仮定して分析を進めても不自然ではないことが確認される (なお、ラグ総数20にてLjung-Box統計量を計算したところ、期間(I)においてパラメータ・セット(2)の1ケース-12ケース中にて有意水準5%で独立性仮説が棄却された)。

3つの期間(I)–(III)、4つの (α, β) の組み合わせに対して計算された χ^2 の値、自由度、その combined p -value を表3に示す。

一方、日経平均先物の使用データの概要を表2に、 $K = 50$ と設定した場合の combined p -value の計算結果を表4に示す。日経平均先物の場合についての p 値系列の独立性のチェックに関しては、ページ数の関係からグラフは省略する。Ljung-Box統計量は、限月(I)のパラメータセット(1)(3)、限月(II)の同(3)の3ケース (12ケース中) において有意水準5%で独立性が棄却されたものの、全般として独立性を仮定しても問題ないと判断された。^{*10}

表3-4より、次のようなことが観察される。まず、今回分析したUSDJPY為替レートには、慣用的な有意水準の下では帰無仮説が棄却されない大きさの p 値を持ち、価格と取引時間間の相互依存性は認められない。一方、日経平均先物は、1ケースを除く全てのケースにて p 値が極めて小さく、相互依存性の存在を強く示唆している。

パラメータ (α, β) の組合せの違いによる相違は今回の結果からは明確には読み取れず、それよりもデータの種類による相違 (USDJPY か日経平均先物か) の方が圧倒的に大きい。

期間による相違に関しては、為替データにおいて、リーマン・ショックの前後で結果に大きな差異は認められない。同様に日経平均先物データにおいても、異なる限月間においても p 値の表れ方に大きな差異は認められない。

さて、USDJPYデータと日経平均先物データとの間の有意性に関する顕著な相違の原因であるが、もし、データの処理方法、分析方法が適切なものであったとするならば、まず、両市場の市場参加者の相違—例えば、ノイズトレーダーの多寡、指値注文の割合の多寡など—が考えられる。また、データの質—EBS為替データの約定レートが250ミリ秒間の最高値・最安値を表し、直前の値による補間値“previous tick interpolation”ではないなど—がこ

^{*9} これは、表1によれば、今回使用したUSDJPYデータにおいて、平均しておよそ2分間隔での実時間軸上のサブサンプルに相当する。この“preaveraging”処理によって、マイクロストラクチャ・ノイズ (特にビッド・アスク・バウンス) がおおむね除去されたことは、realized volatility の大きさを計測することでも確認される。

^{*10} これは、表2から、およそ1分半~2分間毎の実時間軸上のサブサンプルに相当するが、単純平均ではなく、出来高加重平均 (VWAP) を用いているがゆえ、マイクロストラクチャ・ノイズ (特にビッド・アスク・バウンス) がおおむね除去されることが期待される。また、(対数) 収益率のヒストグラム形状もファットテール性は残るものの正規分布に近い形状になることも確認される。

のような差異をもたらしているのかもしれない。あるいは、データの処理方法、特にマイクロストラクチャ・ノイズの処理方法の違いが影響している可能性も否定できない。リーマン・ショック前後という“100年に一度”の金融動乱の時期を対象としたことのよし悪しも今後に向けて留意すべき点である。

当然ながら、実際の市場には異なる取引戦略、資金量、リスク許容度を持つ参加者が参加しているため、実際の“取引成立時刻”が理論上の単一種類の“停止時刻”列のみから成っていることはあり得ない。いわゆる“指値注文”の約定は、理論上の hitting time に相当するが、実際には、市場には“成行注文”も混在しているはずであり、データによって計測される取引時間と価格の相互依存性の“度合”は、先に考えたような純粋な相関性の有無ではない。そして、指値注文の割合は、ティックサイズの大きさや取引ルールなどの文字通り“マイクロストラクチャ、”市場参加者の質・量、市場の流動性などにも関わっていると考えられる。すなわち、 $Z(\alpha, \beta)_t^*$ や $U(\alpha, \beta)_t^*$ は、指値注文の割合や、戦略を持たないノイズトレーダーの割合など、市場やその参加者の特徴を表す統計量であると考えるのが自然であろう。^{*11}

4 おわりに

本研究は、HJY(08)の理論をベースに、証券価格とその取引時間との相互依存性を調べる方法論を提案し、リーマン・ショック前後におけるUSDJPY、日経平均先物の二種類の高頻度データを用いて実証分析を行った。USDJPYデータは相互依存性が認められないのに対し、日経平均先物は相互依存性が強いという観察結果が得られた。

これは、限られた期間における、特定の時系列データに対して得られたもので、もとより最終結論ではありえない。今後は、今回と同じデータセットに対して、異なる“preaveraging”パラメータ (P, K) や異なるマイクロストラクチャ・ノイズ処理方法を試し、今回の暫定的な結果が同じように成立するのかを確認する必要がある。さらに、別の期間や別のデータセットに対しても同様の結果が得られるのかも調べる必要がある。本稿での実証分析は、取引時間と価格の相互依存性を調べる研究の出発点に過ぎない。

謝辞

本稿作成に先立ち、東京大学吉田朋広教授には技術的側面その他幅広くご教授賜り、また議論させて頂いた。本稿は、科学研究費（基盤研究（A）：No.21243019「ファイナンス計量分析の新展開と金融市場」、研究代表者国友直人東京大学教授）の資金援助の下で作成された。また、分析に用いたティックデータの一部は、石井記念証券研究振興財団研究調査助成金によって購入されたものである。ここに謝意を表す。本稿に含まれる誤りの責任はすべて著者に帰する。

^{*11} この点に関して、東京大学国友直人教授よりコメントを頂いた。

USDJPY 期間	日数			レコード件数		1日当たり		1分当たり
	総数	取引有	処理後	総数	処理後	平均件数	最大件数 (実現日)	平均件数
(I)	50	42	38	595,023	563,559	14,167	25,493 (Sep/9)	9.84
(II)	81	70	64	1,277,793	1,192,902	18,254	43,990 (Oct/8)	12.68
(III)	131	112	102	1,872,816	1,756,461	16,722	43,990 (Oct/8)	11.61

表1 USDJPY データの概要: (I) リーマン・ショック前 (2008/7/26-2008/9/13), (II) リーマン・ショック後 (2008/9/14-2008/12/5), (III) 全期間 (=I)+(II). 前処理として, 同時刻の売り買い約定価格は幾何平均によって一つに統合, またレコード件数が 30 件未満の日は分析用データセットから除去.

日経平均先物 限月	データ期間		レコード件数	1日当たり		1分当たり
	日数	日数	総数	平均件数	最大件数 (実現日)	平均件数
(I)	Mar/14 -Jun/13	61	622,821	10,210	18,418 (Mar/17)	23.74
(II)	Jun/13 -Sep/12	64	726,882	11,358	16,544 (Sep/2)	26.41
(III)	Sep/12 -Dec/12	60	990,821	16,514	29,890 (Oct/27)	38.40

表2 日経平均先物データの概要: (I)2008年6月限, (II)2008年9月限, (III)2008年12月限, 各々が期近物となった日(直前の期近物の特別清算日(SQ))から, それ自身の最終取引日(SQ前日)までの期間.

USDJPY $Z(\alpha, \beta)_T^n$	(I)			(II)			(III)
	d.f.	χ^2	p-value	d.f.	χ^2	p-value	p-value
(1) $\alpha = 1/2, \beta = 1$	76	76.23	0.4709	128	136.07	0.2961	0.3305
(2) $\alpha = 1/2, \beta = 3$	76	86.64	0.1897	128	133.14	0.3600	0.2134
(3) $\alpha = 1, \beta = 1$	76	75.19	0.5046	128	124.40	0.5734	0.5739
(4) $\alpha = 1, \beta = 3$	76	92.49	0.0960	128	132.58	0.3729	0.1487

表3 USDJPY データによる検定結果. 局所平均パラメータ $P = 5$, サブサンプル・パラメータ $K = 20$.

日経平均先物 $Z(\alpha, \beta)_T^n$	(I)			(II)			(III)		
	d.f.	χ^2	p-value	d.f.	χ^2	p-value	d.f.	χ^2	p-value
(1) $\alpha = 1/2, \beta = 1$	122	170.34	0.0025	128	142.87	0.1743	120	194.97	0.0000
(2) $\alpha = 1/2, \beta = 3$	122	181.72	0.0004	128	161.35	0.0246	120	168.75	0.0022
(3) $\alpha = 1, \beta = 1$	122	176.92	0.0009	128	170.57	0.0071	120	173.85	0.0010
(4) $\alpha = 1, \beta = 3$	122	183.61	0.0003	128	173.76	0.0044	120	167.90	0.0026

表4 日経平均先物データによる検定結果. VWAP 計算およびサブサンプルの共通パラメータ $K = 50$.

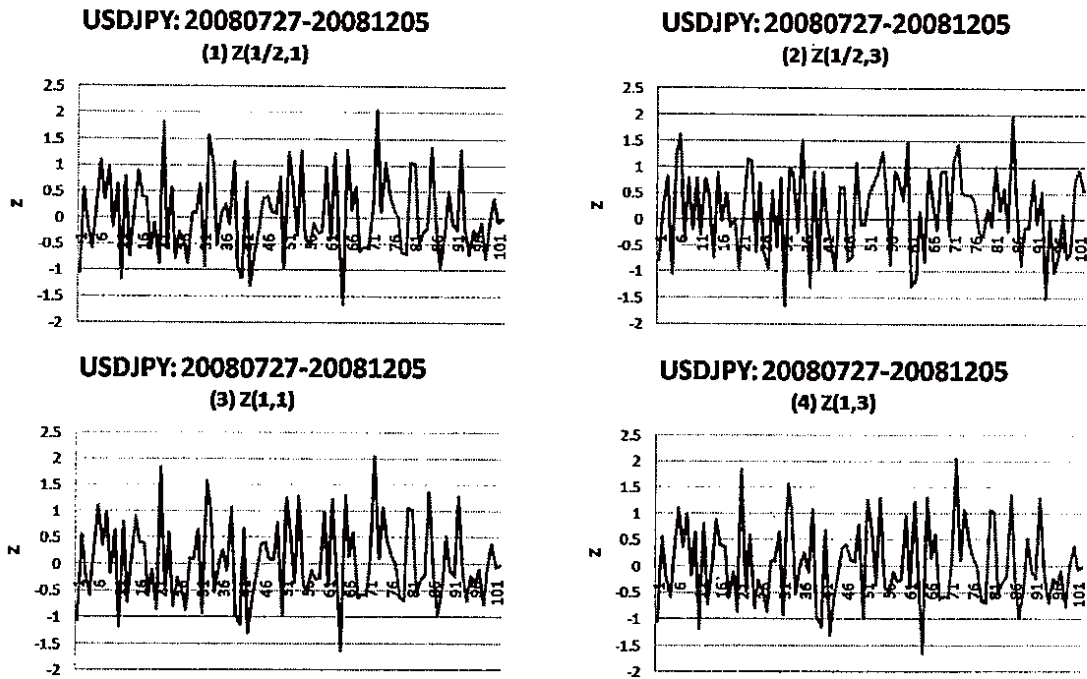


図1 $Z(\alpha, \beta)_T^n$ ($T = 1$ 日) の時系列プロット (横軸は営業日, 1~102) : USDJPY データのケース.

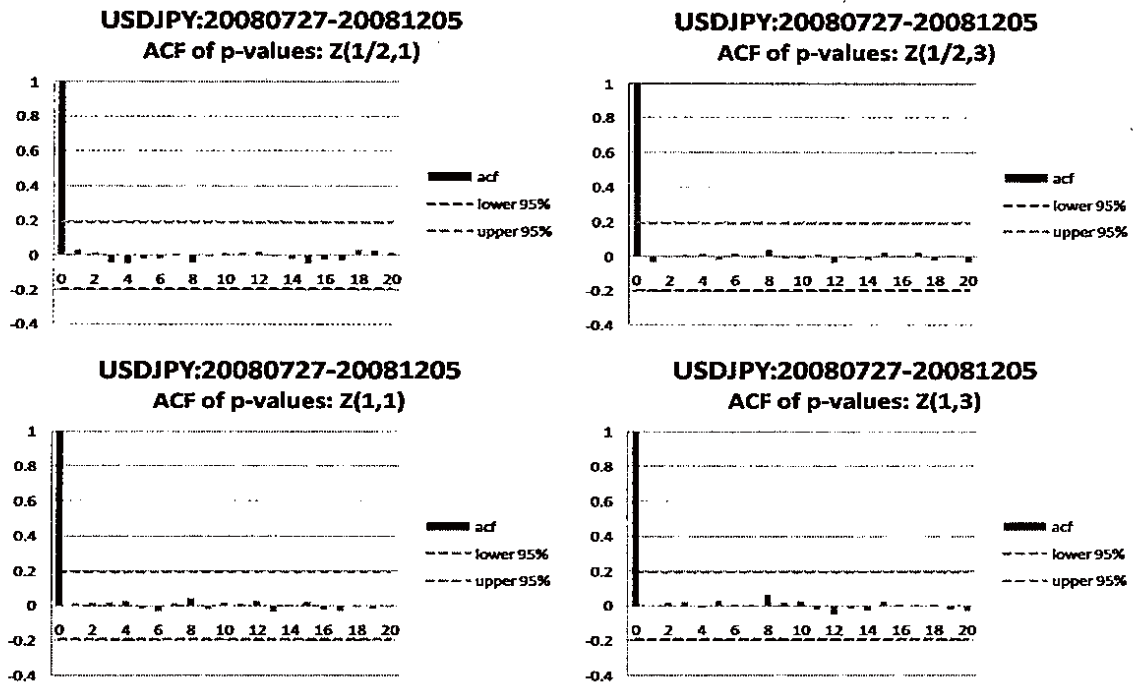


図2 p 値系列 $\{p_i = 1, \dots, 102\}$ の自己相関プロット: USDJPY データのケース.

参考文献

- [1] Ole E. Barndorff-Nielsen, Sven Erik Graversen, Jean Jacod, Mark Podolskij, and Neil Shephard. A central limit theorem for realized power and bipower variations of continuous semimartingales. In *From Stochastic Calculus to Mathematical Finance*. Springer, Berlin.
- [2] Ole E. Barndorff-Nielsen and Neil Shephard. Power variation and time change. *Theory Probab. Appl.*, Vol. 50, pp. 1–15, 2005.
- [3] Masaaki Fukasawa. Realized volatility with stochastic sampling. *Stochastic Process. Appl.*, Vol. to appear, , 2009.
- [4] Takaki Hayashi, Jean Jacod, and Nakahiro Yoshida. Irregular sampling and central limit theorems for power variations: the continuous case. preprint, 2008.
- [5] Takaki Hayashi and Nakahiro Yoshida. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, Vol. 11, No. 2, pp. 359–379, 2005.
- [6] Takaki Hayashi and Nakahiro Yoshida. Nonsynchronous covariance estimator and limit theorem. Research Memorandum No.1020, Institute of Statistical Mathematics, 2006.
- [7] Takaki Hayashi and Nakahiro Yoshida. Nonsynchronous covariance estimator and limit theorem ii. Research Memorandum No.1067, Institute of Statistical Mathematics, 2008.
- [8] Jean Jacod, YingYing Li, Per A. Mykland, Mark Podolskij, and Mathias Vetter. Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Process. Appl.*, Vol. 119, pp. 2249–2276, 2009.
- [9] Silja Kinnebrock, Mark Podolskij, and Mathias Vetter. A note on the central limit theorem for bipower variation of general functions. *Stochastic Process. Appl.*, Vol. 118, pp. 1056–1070, 2008.
- [10] Yingying Li, Per A. Mykland, Renault Eric, Lan Zhang, and Xinghua Zheng. Realized volatility when sampling times can be endogeneous. preprint, 2009.
- [11] Peter C. Phillips and Jun Yu. Information loss in volatility measurement with flat price trading. preprint, 2008.
- [12] C.Y. Robert and M. Rosenbaum. Volatility estimation under endogenous microstructure noise. preprint, 2008.
- [13] 深澤正彰. 実現ボラティリティの漸近分布について. *統計数理*, Vol. 57, pp. 3–16, 2009.

< マイクロ市場の計量-II >

8. 「高頻度金融データの分離情報最尤推定の頑健性」 国友直人・佐藤整尚
9. 「高頻度データの日内季節性調整について」 川崎能典
10. 「バイアス補正実現分散推定量の構築」 大屋幸輔

Robustness of the Separating Information Maximum Likelihood Estimation of Realized Volatility with Micro-Market Noise *

Naoto Kunitomo [†]

and

Seisho Sato [‡]

March 21, 2010

1. The SIML Estimation

Let y_{ij} be the i -th observation of the j -th (log-) price at t_i^n for $j = 1, \dots, p$; $0 = t_0^n \leq t_1^n \leq \dots \leq t_n^n = 1$. We set $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})'$ be a $p \times 1$ vector and $\mathbf{Y}_n = (\mathbf{y}_i')$ be an $n \times p$ matrix of observations. The underlying continuous process \mathbf{x}_i at t_i^n ($i = 1, \dots, n$) is not necessarily the same as the observed prices and let $\mathbf{v}_i' = (v_{i1}, \dots, v_{ip})$ be the vector of the micro-market noise at t_i^n . Then we assume

$$(1.1) \quad \mathbf{y}_i = \mathbf{x}_i + \mathbf{v}_i$$

and

$$(1.2) \quad \mathbf{x}_t = \mathbf{x}_0 + \int_0^t \mathbf{C}_x(s) d\mathbf{B}_s \quad (0 \leq t \leq 1),$$

where $\mathcal{E}(\mathbf{v}_i) = \mathbf{0}$, $\mathcal{E}(\mathbf{v}_i \mathbf{v}_i') = \boldsymbol{\Sigma}_v$, \mathbf{B}_s is a $q \times 1$ ($q \geq 1$) vector of the standard Brownian motions, $\mathbf{C}_x(s)$ is a $p \times q$ vector function adapted to the σ -field $\mathcal{F}(\mathbf{x}_r, \mathbf{B}_r, r \leq s)$, and we write the instantaneous covariance function $\boldsymbol{\Sigma}_x(s) = (\sigma_{ij}^{(x)}(s)) = \mathbf{C}_x(s) \mathbf{C}_x(s)'$ ($\sigma_{ij}^{(x)}(s)$ is the (i, j) -th element of $\boldsymbol{\Sigma}_x(s)$). The main statistical problem is to estimate the quadratic variations and co-variations

$$(1.3) \quad \boldsymbol{\Sigma}_x = (\sigma_{ij}^{(x)}) = \int_0^1 \boldsymbol{\Sigma}_x(s) ds$$

*KSIII-10-3-21. This is a summary version of Kunitomo and Sato (2010).

[†]Graduate School of Economics, University of Tokyo, Bunkyo-ku, Hongo 7-3-1 Tokyo, 113-0033, JAPAN, kunitomo@e.u-tokyo.ac.jp

[‡]Institute of Statistical Mathematics, Tachikawa-shi, Midori-cho 10-3, Tokyo 190-8562, JAPAN

of the underlying continuous process $\{\mathbf{x}_t\}$ and the variance-covariance $\Sigma_v = (\sigma_{ij}^{(v)})$ of the noise process from the observed \mathbf{y}_i ($i = 1, \dots, n$). We use the notation that $\sigma_{ij}^{(x)}$ and $\sigma_{ij}^{(v)}$ are the (i, j) -th element of $\Sigma_x(s)$ and Σ_v , respectively. Although we assume the Gaussian processes in order to derive the SIML estimation, the asymptotic results do not depend on the Gaussianity of the underlying processes. We first consider the standard situation when \mathbf{x}_t ($0 \leq t \leq 1$) and \mathbf{v}_i ($i = 1, \dots, n$) are independent with $\Sigma_x(s) = \Sigma_x$, and \mathbf{v}_i are independently, identically and normally distributed as $N_p(\mathbf{0}, \Sigma_v)$. Then given the initial condition \mathbf{y}_0 , we have

$$(1.4) \quad \mathbf{Y}_n \sim N_{n \times p} \left(\mathbf{1}_n \cdot \mathbf{y}'_0, \mathbf{I}_n \otimes \Sigma_v + \mathbf{C}_n \mathbf{C}'_n \otimes h_n \Sigma_x \right),$$

where $\mathbf{1}'_n = (1, \dots, 1)$, $h_n = 1/n$ ($= t_i^n - t_{i-1}^n$) and

$$(1.5) \quad \mathbf{C}_n = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ 1 & \dots & 1 & 1 & 0 \\ 1 & \dots & 1 & 1 & 1 \end{pmatrix}$$

We transform \mathbf{Y}_n to $\mathbf{Z}_n (= (\mathbf{z}'_k))$ by

$$(1.6) \quad \mathbf{Z}_n = h_n^{-1/2} \mathbf{P}'_n \mathbf{C}_n^{-1} (\mathbf{Y}_n - \bar{\mathbf{Y}}_0)$$

where

$$(1.7) \quad \bar{\mathbf{Y}}_0 = \mathbf{1}_n \cdot \mathbf{y}'_0.$$

Then the likelihood function under the Gaussian noises is given by

$$(1.8) \quad L_n^*(\boldsymbol{\theta}) = \left(\frac{1}{\sqrt{2\pi}} \right)^{np} \prod_{k=1}^n |a_{kn} \Sigma_v + \Sigma_x|^{-1/2} e^{\left\{ -\frac{1}{2} \mathbf{z}'_k (a_{kn} \Sigma_v + \Sigma_x)^{-1} \mathbf{z}_k \right\}},$$

where

$$(1.9) \quad a_{kn} = 4n \sin^2 \left[\frac{\pi}{2} \left(\frac{2k-1}{2n+1} \right) \right] \quad (k = 1, \dots, n).$$

Hence the maximum likelihood (ML) estimator can be defined as the solution of minimizing

$$(1.10) \quad L_n(\boldsymbol{\theta}) = - \sum_{k=1}^n \log |a_{kn} \Sigma_v + \Sigma_x|^{-1/2} + \frac{1}{2} \sum_{k=1}^n \mathbf{z}'_k [a_{kn} \Sigma_v + \Sigma_x]^{-1} \mathbf{z}_k.$$

From this representation we find that the ML estimator of unknown parameters is a rather complicated function of all observations in general because each a_{kn} terms depend on k as well as n . Let denote $a_{k_n, n}$ and then we can evaluate that $a_{k_n, n} \rightarrow 0$ as $n \rightarrow \infty$ when $k_n = O(n^\alpha)$ ($0 < \alpha < \frac{1}{2}$) since $\sin x \sim x$ as $x \rightarrow 0$. Also $a_{n+1-l_n, n} = O(n)$ when $l_n = O(n^\beta)$ ($0 < \beta < 1$).

When k_n is small, we expect that $a_{k_n, n}$ is small. Then we may approximate $2 \times L_n(\theta)$ by

$$(1.11) \quad L_{1n}(\theta) = m \log |\Sigma_x| + \sum_{k=1}^m \mathbf{z}'_k \Sigma_x^{-1} \mathbf{z}_k .$$

It is the standard likelihood function except the fact that we only use the first m_n terms and m_n depends on n . Then the SIML estimator of $\hat{\Sigma}_x$ is defined by

$$(1.12) \quad \hat{\Sigma}_x = \frac{1}{m_n} \sum_{k=1}^{m_n} \mathbf{z}_k \mathbf{z}'_k .$$

Similarly, we set and consider the terms when $a_{n+1-l_n, n}$ is large. We may approximate $2 \times L_n(\theta)$ by

$$(1.13) \quad L_{2n}(\theta) = \sum_{k=n+1-l}^n \log |a_{kn} \Sigma_v| + \sum_{k=n+1-l}^n \mathbf{z}'_k [a_{kn} \Sigma_v]^{-1} \mathbf{z}_k .$$

It is also the standard likelihood function approach except the fact that we only use the last l terms. Then the SIML estimator of $\hat{\Sigma}_v$ is defined by

$$(1.14) \quad \hat{\Sigma}_v = \frac{1}{l_n} \sum_{k=n+1-l_n}^n a_{kn}^{-1} \mathbf{z}_k \mathbf{z}'_k .$$

For both $\hat{\Sigma}_v$ and $\hat{\Sigma}_x$, the number of terms m_n and l_n should be dependent on n . Then we only need the order requirements that $m_n = O(n^\alpha)$ ($0 < \alpha < \frac{1}{2}$) and $l_n = O(n^\beta)$ ($0 < \beta < 1$) for Σ_x and Σ_v , respectively.

2. Asymptotic Robustness of the SIML Estimation

We shall investigate the effects of the serial correlations of noises on the asymptotic properties of the SIML estimator. Consider the case of $p = 1$ and we set

$$r_i = x_i - x_{i-1} = \int_{t_{i-1}}^{t_i} \sigma_s dB_s \quad (i = 1, \dots, n)$$

with $0 = t_0 \leq t_1 < \dots < t_n = 1$ ($i = 1, \dots, n$). For the simplicity, we take the equi-distance case as $t_i - t_{i-1} = 1/n$ and the volatility function σ_s ($0 \leq s \leq 1$) is non-stochastic. From Kunitomo and Sato (2008a,b), we shall investigate the asymptotic

distribution of $\hat{\sigma}_x^2 - \sigma_x^2$ and $\sigma_x^2 = \int_0^1 \sigma_s^2 ds$. From *Mathematical Appendix* of Kunitomo and Sato (2008a),

$$\begin{aligned}\sqrt{m} [\hat{\sigma}_x^2 - \sigma_x^2] &= \frac{1}{\sqrt{m}} \sum_{k=1}^m [x_{kn}^2 - \sigma_x^2] \\ &= \frac{1}{\sqrt{m}} \sum_{k=1}^m [x_{kn}^{(1)2} - \sigma_x^2] \\ &\quad + \frac{1}{\sqrt{m}} \sum_{k=1}^m \mathbf{E}[x_{kn}^{(2)2}] + \frac{1}{\sqrt{m}} \sum_{k=1}^m [x_{kn}^{(2)2} - \mathbf{E}[x_{kn}^{(2)2}]] + 2 \frac{1}{\sqrt{m}} \sum_{k=1}^m [x_{kn}^{(1)} x_{kn}^{(2)}]\end{aligned}$$

Then we want to find the conditions that three terms except the first one are $o_p(1)$.

Let $\mathbf{b}_k = \mathbf{e}'_k \mathbf{P}'_n \mathbf{C}_n^{-1} = (b_{kj})$, $\mathbf{e}'_k = (0, \dots, 1, 0, \dots)$ is an $n \times 1$ vectors. We write $x_{kn}^{(2)} = \sum_{j=1}^n b_{kj} v_j$ and note that $\sum_{j=1}^n b_{kj} b_{k'j} = \delta(k, k') a_{kn}$. (We shall set K_i being constants.)

First we impose the condition

$$(I) \quad \mathbf{E}[v_i v_j] = c_1 \rho^{|i-j|} \quad (0 \leq \rho < 1),$$

where c_1 is a constant. Then by using the Cauchy-Schwartz inequality

$$\begin{aligned}\mathbf{E}[x_{kn}^{(2)}]^2 &= \mathbf{E}\left[\sum_{i=1}^n b_{ki} v_i \sum_{j=1}^n b_{kj} v_j\right] \\ &\leq \sum_{l=0}^n c_1 (1 + 2l) \rho^l \mathbf{E}\left[\sum_{i=1}^n b_{ki} b_{ki-l}\right] \\ &\leq K_1 \times a_{kn},\end{aligned}$$

provided that $\mathbf{E}[v_i^2]$ are bounded and we define $b_{kj} = 0$ ($j \leq 0$). Then the second term becomes

$$(2.1) \quad \frac{1}{\sqrt{m}} \sum_{k=1}^m \mathbf{E}[x_{kn}^{(2)}]^2 \leq K_1 \frac{1}{\sqrt{m}} \sum_{k=1}^m a_{kn} = O\left(\frac{m^{5/2}}{n}\right) \rightarrow 0$$

if $0 < \alpha < 0.4$.

For the fourth term,

$$\begin{aligned}(2.2) \quad \mathbf{E} \left[\frac{1}{\sqrt{m}} \sum_{j=1}^m x_{kn}^{(1)} x_{kn}^{(2)} \right]^2 &= \frac{1}{m} \sum_{k, k'=1}^m \mathbf{E} [x_{kn}^{(1)} x_{k'n}^{(1)} x_{kn}^{(2)} x_{k'n}^{(2)}] \\ &= \frac{1}{m} \sum_{k, k'=1}^m \mathbf{E} \left[2 \sum_{j, j'=1}^n s_{jk} s_{j'k'} \mathbf{E}(r_j r_{j'}) x_{kn}^{(2)} x_{k'n}^{(2)} \right] \\ &= \frac{1}{m} \sum_{k, k'=1}^m \mathbf{E} \left[2 \sum_{j=1}^n s_{jk} s_{jk'} \mathbf{E}(r_j^2) x_{kn}^{(2)} x_{k'n}^{(2)} \right]\end{aligned}$$

$$\begin{aligned}
&\leq K_2 \left[\left(\sup_{0 \leq s \leq 1} \sigma_s^2 \right) \frac{2}{n} \left(\frac{n}{2} + \frac{1}{4} \right) \right] \frac{1}{m} \sum_{k,k'=1}^m \sqrt{a_{kn}} \sqrt{a_{k'n}} \\
&\leq K_3 \sum_{k=1}^m a_{kn} = O\left(\frac{m^3}{n}\right)
\end{aligned}$$

by using the relations $\int_{t_{i-1}}^{t_i} \sigma_s^2 ds \leq (1/n)(\sup_{0 \leq s \leq 1} \sigma_s^2)$ and $|\sum_{j=1}^n s_{jk} s_{jk'}| \leq [\sum_{j=1}^n s_{jk}^2] = n/2 + 1/4$ for any $k \geq 1$.

Then we need the condition $0 < \alpha < 1/3$. If $\sigma_s = \sigma$ (i.e., the volatility is constant), (0.2) becomes $(1/m) \sum_{k=1}^m a_{kn} = O(m^2/n)$, which is satisfied if $0 < \alpha < .4$.

For the third term, we consider

$$x_{kn}^{(2)2} - \mathbf{E}[x_{kn}^{(2)2}] = \sum_{j,j'=1}^n b_{kj} b_{kj'} [v_j v_{j'} - \mathbf{E}(v_j v_{j'})]$$

and we try to evaluate the expectation of

$$[x_{kn}^{(2)2} - \mathbf{E}[x_{kn}^{(2)2}]] [x_{k'n}^{(2)2} - \mathbf{E}[x_{k'n}^{(2)2}]]$$

under an additional condition

$$(II) \quad \mathbf{E}[(v_i v_{i'} - \mathbf{E}(v_i v_{i'})) (v_{i''} v_{i'''} - \mathbf{E}(v_{i''} v_{i'''}))] = c_2 \rho^{\frac{1}{2}(|i-i'|+|i''-i'''|)} \quad (0 \leq \rho < 1),$$

where c_2 is a constant. The condition (II) is satisfied for the linear processes on $\{v_j\}$ with bounded 4th order moments. The calculations are straightforward, but there are many terms involved. We can use the fact that

$$\begin{aligned}
&\sum_{j,j'=1}^n b_{kj} b_{kj'} \rho^{|j-j'|/2} \sum_{j'',j'''=1}^n b_{k'j''} b_{k'j'''} \rho^{|j''-j'''|/2} \\
&\sim K_4 \left[\sum_{l,l'=1}^n \rho^{l+l'} l l' \right] \left[\sum_{j=1}^n b_{kj} b_{kj-l} \right] \left[\sum_{j=1}^n b_{k'j''} b_{k'j''-l'} \right] \\
&\sim K_5 \times a_{kn} a_{k'n}.
\end{aligned}$$

By applying these evaluations, we can obtain

$$\begin{aligned}
\mathbf{E} \left[\frac{1}{\sqrt{m}} \sum_{j=1}^m (x_{kn}^{(2)2} - \mathbf{E}[x_{kn}^{(2)2}]) \right]^2 &\leq \frac{1}{m} \sum_{k,k'=1}^m a_{kn} a_{k'n} \\
&= O\left(\frac{1}{m} \times \left(\frac{m^3}{n}\right)^2\right) = O\left(\frac{m^5}{n^2}\right)
\end{aligned}$$

since $\sum_{k=1}^m a_{kn} = O(m^3/n)$. Thus the third term is negligible if $0 < \alpha < .4$. We summarize our finding, but the conditions we have used can be further improved.

Theorem 1 : Assume Conditions (I) and (II) and set $0 < \alpha < 1/3$. Then the asymptotic distribution of $\sqrt{m_n} [\hat{\sigma}_x^2 - \sigma_x^2]$ is asymptotically $(m_n, n \rightarrow \infty)$ equivalent to that of $(1/\sqrt{m_n}) \sum_{k=1}^{m_n} [x_{kn}^{(1)2} - \sigma_x^2]$.

The only term involved on the correlations of noise and signal is the fourth term. Thus it is interesting to find the condition that they can be ignored for estimating the realized volatility and covariance. The second line of (0.2) can be written as

$$\begin{aligned} & \frac{1}{m} \sum_{k,k'=1}^m \mathbf{E} [x_{kn}^{(1)} x_{k'n}^{(1)} x_{kn}^{(2)} x_{k'n}^{(2)}] \\ &= \frac{1}{m} \sum_{k,k'=1}^m \mathbf{E} \left[\left(2 \sum_{j,j'=1}^n s_{jk} s_{j'k'} r_j r_{j'} \right) \left(\sum_{i=1}^n b_{ki} v_i \right) \left(\sum_{j=1}^n b_{k'j} v_j \right) \right] \end{aligned}$$

Thus the sufficient condition we need is

$$(I') \quad \mathbf{E}[v_i v_j | r_k, k = 1, \dots, n] = c'_1 \rho_1^{|i-j|} \quad (0 \leq \rho_1 < 1) \text{ a.s.}$$

We note that c'_1 may depend on $\int_0^1 \sigma_s^2 ds$, which is finite (a.s.). In that case we allow that the noises may depend on the volatility structure, but we need the condition that $c'_1 (\int_0^1 \sigma_s^2 ds) \int_0^1 \sigma_s^2 ds$ should be finite (a.s.) and integrable. By this argument, if both the correlations between signal and noise and the autocorrelations of noise are weak, the SIML estimator is consistent and it has the asymptotic normality.

Theorem 2 : Assume Conditions (I)' and (II) and set $0 < \alpha < 1/3$. Relax the independence assumption between the signal and noise terms. Then the result of Theorem 1 holds.

One nice feature of our approach is the fact that our arguments go through even when $p \geq 1$. When $p \geq 1$, we take an arbitrary constant vector \mathbf{c} and use $\mathbf{c}' r_i$ and $\mathbf{c}' v_i$ ($i = 1, \dots, n$). Then we can use the same arguments.

3. Simulations

We have done a large number of Monte Carlo experiments and we shall show only some results below.

Table 1 : Estimation of Realized Volatility (standard case, $v_t \sim i.i.d.Normal$)

n=300	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	2.05E-04	2.18E-06	1.40E-03	2.00E-04	3.82E-07	3.19E-04	2.01E-04	1.84E-07	2.01E-04
SD	9.61E-05	3.22E-07	1.35E-04	9.16E-05	5.61E-08	2.67E-05	9.29E-05	2.69E-08	1.61E-05
MSE	9.27E-09	1.36E-13		8.40E-09	3.61E-14		8.63E-09	3.39E-14	
AVAR	8.17E-09	8.34E-14		8.17E-09	8.34E-16		8.17E-09	8.34E-20	
n=5000	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	2.07E-04	2.01E-06	2.02E-02	2.01E-04	2.10E-07	2.20E-03	2.00E-04	1.23E-08	2.20E-04
SD	5.37E-05	9.44E-08	4.93E-04	5.12E-05	9.68E-09	5.22E-05	5.27E-05	5.68E-10	4.35E-06
MSE	2.92E-09	8.99E-15		2.62E-09	2.03E-16		2.78E-09	1.06E-16	
AVAR	2.65E-09	8.79E-15		2.65E-09	8.79E-17		2.65E-09	8.79E-21	
n=20000	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	2.05E-04	2.00E-06	8.02E-02	2.00E-04	2.03E-07	8.20E-03	2.00E-04	4.54E-09	2.80E-04
SD	4.10E-05	5.47E-08	9.82E-04	3.98E-05	5.43E-09	9.91E-05	3.90E-05	1.19E-10	2.87E-06
MSE	1.70E-09	3.00E-15		1.59E-09	3.59E-17		1.52E-09	6.46E-18	
AVAR	1.52E-09	2.90E-15		1.52E-09	2.90E-17		1.52E-09	2.90E-21	

Table 2 : Estimation of Realized Volatility (MA(1) noise, $\alpha = 0.5$)

n=300	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	2.03E-04	3.53E-06	1.87E-03	2.03E-04	5.18E-07	3.68E-04	1.99E-04	1.86E-07	2.01E-04
SD	9.62E-05	5.12E-07	1.99E-04	9.69E-05	7.45E-08	3.21E-05	9.35E-05	2.69E-08	1.64E-05
MSE	9.27E-09	2.61E-12		9.40E-09	1.07E-13		8.74E-09	3.44E-14	
AVAR	8.17E-09	8.34E-14		8.17E-09	8.34E-16		8.17E-09	8.34E-20	
n=5000	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	2.04E-04	3.52E-06	2.82E-02	2.00E-04	3.62E-07	3.00E-03	2.00E-04	1.38E-08	2.28E-04
SD	5.27E-05	1.63E-07	7.61E-04	5.12E-05	1.69E-08	7.94E-05	5.09E-05	6.45E-10	4.57E-06
MSE	2.79E-09	2.35E-12		2.62E-09	2.65E-14		2.59E-09	1.39E-16	
AVAR	2.65E-09	8.79E-15		2.65E-09	8.79E-17		2.65E-09	8.79E-21	
n=20000	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	2.05E-04	3.56E-06	1.12E-01	1.98E-04	3.57E-07	1.14E-02	2.00E-04	6.09E-09	3.12E-04
SD	3.95E-05	9.59E-08	1.53E-03	4.00E-05	9.58E-09	1.54E-04	3.92E-05	1.62E-10	3.30E-06
MSE	1.58E-09	2.43E-12		1.61E-09	2.48E-14		1.54E-09	1.67E-17	
AVAR	1.52E-09	2.90E-15		1.52E-09	2.90E-17		1.52E-09	2.90E-21	

Data generating process:

$$y_t = x_t + \sqrt{\sigma_v^2/(1+a^2)}v_t$$

$$x_t = x_{t-1} + \sqrt{\sigma_x^2/n}u_t$$

$$v_t = w_t - aw_{t-1}$$

$$u_t \sim i.i.d.N(0,1), w_t \sim i.i.d.N(0,1)$$

Table 4 : Estimation of Realized Volatility (Endogenous noise, $\rho = 0.5, l = 0$)

n=300	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	2.02E-04	1.76E-06	1.14E-03	1.98E-04	4.64E-07	3.68E-04	1.97E-04	2.02E-07	2.11E-04
SD	9.54E-05	2.54E-07	1.07E-04	9.31E-05	6.62E-08	3.14E-05	9.44E-05	2.98E-08	1.71E-05
MSE	9.11E-09	1.23E-13		8.67E-09	7.40E-14		8.93E-09	4.10E-14	
AVAR	8.17E-09	8.34E-14		8.17E-09	8.34E-16		8.17E-09	8.34E-20	
n=5000	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	2.02E-04	1.15E-06	1.16E-02	2.01E-04	1.55E-07	1.65E-03	1.99E-04	1.58E-08	2.55E-04
SD	5.27E-05	5.42E-08	2.82E-04	5.27E-05	7.22E-09	3.77E-05	5.13E-05	7.48E-10	5.19E-06
MSE	2.78E-09	7.20E-13		2.78E-09	2.08E-15		2.64E-09	1.90E-16	
AVAR	2.65E-09	8.79E-15		2.65E-09	8.79E-17		2.65E-09	8.79E-21	
n=20000	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	2.03E-04	1.07E-06	4.30E-02	2.00E-04	1.25E-07	5.09E-03	2.00E-04	5.78E-09	3.29E-04
SD	3.99E-05	2.87E-08	5.21E-04	3.91E-05	3.37E-09	6.12E-05	3.96E-05	1.57E-10	3.50E-06
MSE	1.60E-09	8.60E-13		1.53E-09	5.66E-15		1.57E-09	1.43E-17	
AVAR	1.52E-09	2.90E-15		1.52E-09	2.90E-17		1.52E-09	2.90E-21	

Data generating process:

$$y_t = x_t + \sqrt{\sigma_v^2} v_t$$

$$x_t = x_{t-1} + \sqrt{\sigma_x^2/n} u_t$$

$$v_t = (1 - \rho) w_t + \rho v_{t-1}$$

$$u_t \sim i.i.d.N(0, 1), w_t \sim i.i.d.N(0, 1)$$

Table 6 : Estimation of Realized Volatility (MA(1) and Endogenous noise, $a = 0.8, \rho = 0.9, l = 1$)

n=300	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	1.98E-04	1.71E-06	7.86E-04	2.00E-04	1.88E-08	6.91E-05	2.01E-04	1.39E-07	1.74E-04
SD	9.23E-05	2.47E-07	9.13E-05	9.30E-05	2.94E-09	7.78E-06	9.42E-05	2.02E-08	1.39E-05
MSE	8.53E-09	1.43E-13		8.65E-09	3.28E-14		8.88E-09	1.92E-14	
AVAR	8.17E-09	8.34E-14		8.17E-09	8.34E-16		8.17E-09	8.34E-20	
n=5000	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	2.00E-04	2.81E-06	2.10E-02	2.01E-04	2.13E-07	1.51E-03	2.00E-04	2.12E-09	1.11E-04
SD	5.31E-05	1.33E-07	6.02E-04	5.16E-05	1.02E-08	4.47E-05	5.15E-05	1.02E-10	2.51E-06
MSE	2.82E-09	6.67E-13		2.66E-09	2.61E-16		2.65E-09	2.51E-20	
AVAR	2.65E-09	8.79E-15		2.65E-09	8.79E-17		2.65E-09	8.79E-21	
n=20000	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	2.01E-04	3.01E-06	9.06E-02	2.01E-04	2.65E-07	7.70E-03	2.00E-04	7.30E-11	7.13E-05
SD	4.01E-05	8.08E-08	1.28E-03	4.05E-05	6.97E-09	1.08E-04	3.96E-05	1.96E-12	9.98E-07
MSE	1.61E-09	1.04E-12		1.64E-09	4.28E-15		1.56E-09	3.71E-18	
AVAR	1.52E-09	2.90E-15		1.52E-09	2.90E-17		1.52E-09	2.90E-21	

Data generating process: same as Table 5.

Table 8 : Estimation of Realized Volatility (MA(1) and Endogenous noise, $\alpha = 0.9, \rho = 0.9, l = -1$)

n=300	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	2.02E-04	2.49E-06	1.66E-03	1.99E-04	2.64E-07	3.46E-04	2.02E-04	1.63E-07	2.01E-04
SD	9.64E-05	3.67E-07	1.62E-04	9.41E-05	4.74E-08	3.21E-05	9.59E-05	2.42E-08	1.69E-05
MSE	9.29E-09	3.75E-13		8.86E-09	6.40E-15		9.21E-09	2.66E-14	
AVAR	8.17E-09	8.34E-14		8.17E-09	8.34E-16		8.17E-09	8.34E-20	
n=5000	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	2.00E-04	2.91E-06	2.48E-02	1.98E-04	2.38E-07	2.65E-03	1.99E-04	4.36E-09	2.24E-04
SD	5.14E-05	1.35E-07	6.68E-04	5.20E-05	1.10E-08	6.51E-05	5.30E-05	2.28E-10	4.78E-06
MSE	2.64E-09	8.45E-13		2.71E-09	1.53E-15		2.81E-09	5.64E-18	
AVAR	2.65E-09	8.79E-15		2.65E-09	8.79E-17		2.65E-09	8.79E-21	
n=20000	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	1.99E-04	3.07E-06	9.84E-02	2.00E-04	2.74E-07	1.00E-02	2.00E-04	7.45E-10	2.98E-04
SD	3.88E-05	8.40E-08	1.36E-03	3.91E-05	7.39E-09	1.29E-04	3.93E-05	2.59E-11	3.42E-06
MSE	1.50E-09	1.14E-12		1.53E-09	5.60E-15		1.54E-09	1.58E-18	
AVAR	1.52E-09	2.90E-15		1.52E-09	2.90E-17		1.52E-09	2.90E-21	

Data generating process: same as Table 5.

Table 10 : Estimation of Realized Volatility (MA(1) and Endogenous noise, $\alpha = 0.9, \rho = 0.9, l = 0$)

n=300	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	2.01E-04	4.46E-06	2.14E-03	2.00E-04	8.75E-07	4.91E-04	1.99E-04	2.24E-07	2.15E-04
SD	9.44E-05	6.43E-07	2.43E-04	9.61E-05	1.28E-07	4.94E-05	9.48E-05	3.31E-08	1.76E-05
MSE	8.90E-09	6.48E-12		9.24E-09	4.72E-13		8.99E-09	5.03E-14	
AVAR	8.17E-09	8.34E-14		8.17E-09	8.34E-16		8.17E-09	8.34E-20	
n=5000	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	2.00E-04	3.53E-06	2.66E-02	2.00E-04	4.37E-07	3.26E-03	1.99E-04	2.43E-08	2.84E-04
SD	5.20E-05	1.67E-07	7.58E-04	5.28E-05	2.08E-08	9.32E-05	5.01E-05	1.15E-09	6.17E-06
MSE	2.70E-09	2.38E-12		2.79E-09	5.65E-14		2.51E-09	4.98E-16	
AVAR	2.65E-09	8.79E-15		2.65E-09	8.79E-17		2.65E-09	8.79E-21	
n=20000	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol	σ_x^2	σ_v^2	H-vol
true-val	2.00E-04	2.00E-06		2.00E-04	2.00E-07		2.00E-04	2.00E-09	
mean	2.00E-04	3.40E-06	1.02E-01	2.01E-04	3.80E-07	1.12E-02	1.99E-04	1.13E-08	4.18E-04
SD	3.92E-05	9.07E-08	1.44E-03	3.91E-05	1.03E-08	1.59E-04	3.85E-05	3.03E-10	4.93E-06
MSE	1.54E-09	1.96E-12		1.53E-09	3.25E-14		1.49E-09	8.62E-17	
AVAR	1.52E-09	2.90E-15		1.52E-09	2.90E-17		1.52E-09	2.90E-21	

Data generating process: same as Table 5.

Table 11 : Comparing with ML (MA(1) noise, $a = 0.5, \alpha = 0.45$)

n=300	σ_x^2	σ_v^2	ML	σ_x^2	σ_v^2	ML	σ_x^2	σ_v^2	ML
true-val	5.00E-05	5.00E-07	5.00E-05	5.00E-05	5.00E-09	5.00E-05	5.00E-05	0	5.00E-05
mean	5.34E-05	2.10E-07	1.11E-04	4.91E-05	4.73E-08	5.10E-05	4.97E-05	4.54E-08	4.95E-05
SD	2.11E-05	3.05E-08	3.31E-05	1.96E-05	7.02E-09	7.23E-06	1.94E-05	6.83E-09	7.18E-06
n=5000	σ_x^2	σ_v^2	ML	σ_x^2	σ_v^2	ML	σ_x^2	σ_v^2	ML
true-val	5.00E-05	5.00E-07	5.00E-05	5.00E-05	5.00E-09	5.00E-05	5.00E-05	0	5.00E-05
mean	5.24E-05	1.24E-07	1.98E-04	5.00E-05	3.78E-09	6.66E-05	4.96E-05	2.57E-09	4.99E-05
SD	1.09E-05	5.74E-09	4.21E-05	1.05E-05	1.81E-10	2.78E-06	9.98E-06	1.19E-10	1.69E-06
n=20000	σ_x^2	σ_v^2	ML	σ_x^2	σ_v^2	ML	σ_x^2	σ_v^2	ML
true-val	5.00E-05	5.00E-07	5.00E-05	5.00E-05	5.00E-09	5.00E-05	5.00E-05	0	5.00E-05
mean	5.20E-05	1.13E-07	2.25E-04	4.99E-05	1.76E-09	9.69E-05	4.96E-05	6.35E-10	5.00E-05
SD	8.00E-06	3.02E-09	3.81E-05	7.55E-06	4.74E-11	2.85E-06	7.53E-06	1.76E-11	8.43E-07

Data generating process: same as Table 5.

References

- [1] Kunitomo, N. and S. Sato (2008a), "Separating Information Maximum Likelihood Estimation of Realized Volatility and Covariance with Micro-Market Noise," Discussion Paper CIRJE-F-581, Graduate School of Economics, University of Tokyo, (<http://www.e.u-tokyo.ac.jp/cirje/research/dp/2008>).
- [2] Kunitomo, N. and S. Sato (2008b), "Realized Volatility, Covariance and Hedging Coefficient of Nikkei-225 Futures with Micro-Market Noise," Discussion Paper CIRJE-F-601, Graduate School of Economics, University of Tokyo, (<http://www.e.u-tokyo.ac.jp/cirje/research/dp/2008>).
- [3] Kunitomo, N. and S. Sato (2009), "The SIML Estimation of the Realized Volatility and Hedging Coefficient of Nikkei-225 Futures with Micro-Market Noise," forthcoming in *Mathematics and Simulations*, North-Holland.

高頻度データの日内季節性調整について*

Modeling Intraday Periodicity in High Frequent Financial Data

川崎能典^{1†}

Yoshinori Kawasaki¹

¹ 統計数理研究所 モデリング研究系

¹ The Institute of Statistical Mathematics, Dept. of Statistical Modeling

Abstract: This paper discusses modeling the intraday periodicity of high frequent financial data, especially of the quotes in yen-dollar exchange market. At first, we briefly review the most commonly used approach using spline smoothing. Then we propose a point process approach in which the intraday periodicity and trend (if any) and the cluster effect is simultaneously considered. It is confirmed that the cluster effect represented by Laguerre polynomial is essential in explaining the intensity of quote occurrence.

Keywords; High frequent data, autoregressive conditional duration model, point process, Laguerre polynomial

1 Introduction

Modeling with high frequent data in finance has become very popular since the middle of 1990's. One reason is due to the availability of such data itself. For example, a set of high frequent data was provided by a data service vendor to foster the empirical research in this field, and the conference and special issue of an academic journal were accompanied to the project. See Baillie and Dacorogna (1997), the preface of the special issue in *Journal of Empirical Finance*. Another reason is model development suitable for the analysis of high frequent data. Since the pioneering work of Engle and Russell (1997, 1998) which proposed a useful class of models called the Autoregressive Conditional Duration Models, it was followed by many empirical research and model ramifications such as Threshold ACD by Zhang, Russell and Tsay (2001) or Log-ACD by Bauwens and Giot (2000) among others.

The basic idea of the ACD model is to express the mean conditional duration time as a linear function of past duration time and mean conditional duration time. But we cannot fit the ACD model directly to the duration times calculated from original high frequent data because of intraday periodic pattern. As the

transactions or quotes in markets will be proportional to the actual economic activity and market time, the pattern of quote occurrence inevitably exhibits intraday periodicity.

Section 2 describes commonly employed two-step procedure to remove intraday periodicity prior to the fitting of the ACD model. The purpose of this paper is to propose simultaneous determination method for the models of intraday periodicity and of the cluster effects of financial transactions. Section 3 reports the framework and worked example of conditional intensity approach to point processes developed by Ogata (1983a, b), which has been applied in statistical modeling of seismicity analysis. In Section 4, we apply the method of Section 3 to detect the residual seasonality in the point process data adjusted by spline based two step method.

In closing this section we briefly state the data analyzed throughout this paper. The data is tick-by-tick data of yen-dollar exchange rate collected from July 10, 1997 to August 3, 1997. The data were captured from the screen of Telerate terminal, and then the obvious wrong orders and ticks with extremely high volatility were removed prior to the analysis. The intervals corresponding to weekends (from Friday 21:00 to Sunday 21:00 GST) are removed to make the data handling easy. One feature of foreign exchange rate data is that each data record is not always an actual transaction but just a quote. Secondly, the market is

*科学研究費補助金 (基礎研究 (A)) 「ファイナンス計量分析の展開と金融市場」研究集会。(2010年3月30日~31日, 於 東京大学経済学部 小島ホール)

[†]E-mail: kawasaki@ism.ac.jp Address: 〒190-8562 東京都立川市緑町 10-3

basically open 24 hours a day. But the transaction will be proportional to the actual economic activity and market time, the pattern of quote occurrence inevitably exhibits intraday periodicity.

2 Removing Periodicity in the ACD models

Since Engle and Russell (1998), a class of models called autoregressive conditional duration (ACD hereafter) model has received much attention, and applied to analyze the characteristics of financial markets using high frequent data. In this section we briefly review the basic framework of the ACD models, and subsequently the common recipes for removing intraday periodicity from high frequent financial data.

Suppose t_i stands for i -th transaction. It is assumed that $0 = t_0 < t_1 < \dots < t_n$ where n is the total number of transactions observed in the interval we consider. Then the duration time between trades is defined as $X_i = t_i - t_{i-1}$. Here we introduce the conditional mean duration time ψ_i as follows;

$$\psi_i = E[X_i | X_{i-1}, \dots, X_1; \theta_1]. \quad (1)$$

θ_1 is a set of unknown parameters that describe the structure of the conditional mean duration. The basic assumption in the ACD model is, if we correctly specify ψ_i , then i -th duration time divided by ψ_i will follow some distribution of mean 1 identically and independently, namely,

$$\frac{X_i}{\psi_i} = \epsilon_i \sim \mathcal{G}(\theta_2) \quad (2)$$

for some distribution \mathcal{G} with mean 1. θ_2 is a set of parameters of a distribution \mathcal{G} . The equation (1) shows a general framework, and the models used actually are the one that resembles the specification of volatility in ARCH and GARCH models. If ψ_i is defined as

$$\psi_i = a_0 + \sum_{m=1}^p a_m X_{i-m} + \sum_{n=1}^q b_n \psi_{i-n}, \quad (3)$$

then the duration time is said to follow $ACD(p, q)$ model.

Before the fitting of the ACD models, the removal of intraday periodicity is usually considered. The most commonly employed recipe seems the one described as follows. Divide one day into, say, the collection of five minutes bins. Every duration time is classified into

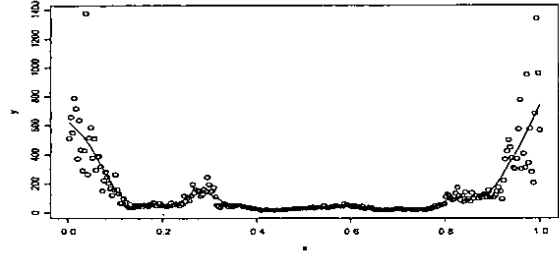


Figure 1: 5-minutes-bin-wise average duration time and its spline smoothing

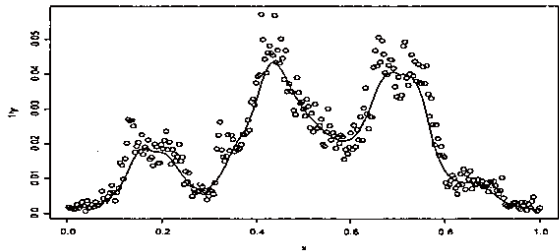


Figure 2: Reciprocal of 5-minutes-bin-wise average duration time and the reciprocal of smoothed curve estimated in Figure 1

anyone of the bins. Then the mean duration time is calculated by an sample average within a bin. Now we have 288 equally spaced data points. Then perform smoothing spline for the series of mean duration. If we divide each duration time by the corresponding value implied by the fitted spline curve, then we assume that the intraday periodicity is removed. See Engle and Russell (1997), Zhang, Russell and Tsay (2001) for example.

Figure 1 shows the results of applying the above stated procedure to our data. We used the function *supsmu* in S-PLUS. Both ends of the figure correspond to 21:00 GST. Figure 2 plots the reciprocal of duration time and the reciprocal of the estimated smoothing spline curve in Figure 1. Curve in Figure 2 clearly shows the intraday pattern of the quote intensity of yen-dollar exchange market. Three major humps correspond to the market time of Tokyo, London and New York respectively. Engle and Russell (1997) and following works in this field generally assert this two-stage approach works fine at least for the stock market in the United States. The aim of this paper is to examine such assertion by another modeling methodology, a class of parametric models for

point processes that will be estimated by the method of maximum likelihood.

3 Specification of Point Processes via Conditional Intensity

This section describes a general framework for analyzing various point process data by way of the conditional intensity function. An excellent review of modeling point processes mainly applied for seismicity analysis can be found in Ogata (1999).

Consider a series of events $\{t_i; 0 < t_1 < t_2 < \dots\}$ occurring at random on the real half line $(0, \infty)$. To describe this sequence we take time differences, $X_i = t_i - t_{i-1}$, between consecutive points, and we then consider as a positive valued stochastic process. When this is distributed independently and identically, we call this a renewal process, and if its marginal is exponential distribution, then it is the stationary Poisson process.

Let $N(a, b)$ be the number of points in an interval (a, b) on the real line such that this is a nonnegative integer valued random variable. Consider a prediction of an event occurring on a small time interval. Namely, assume a point process on the real half-line $(0, \infty)$, and divide it into small intervals of length δ . Then we get a stochastic process $\{\xi_k\}$, where $\xi_k = N[(k-1)\delta, k\delta)$ is k -th random variable on the subinterval $[(k-1)\delta, k\delta)$. If δ is small enough, we may assume that $\{\xi_i\}$ is a binary process. If the point process now considered is a stationary Poisson, then $\{x_{i_k}\}$ is identically and independently distributed Bernoulli series. But in general, the joint probability of the sequence is determined by a sequence of conditional probabilities $P\{\xi_k = 1 | \xi_1, \dots, \xi_{k-1}\}$, $k = 1, 2, \dots$, namely on the history of events.

Then as a derivative of the conditional probability with respect to the time, the *conditional intensity function* $\lambda(t|\mathcal{F}_t)$ is defined by

$$P\{N(t, t + \delta) = 1 | \mathcal{F}_t\} = \lambda(t|\mathcal{F}_t)\delta + o(\delta),$$

or

$$\lambda(t|\mathcal{F}_t) = \lim_{\Delta \rightarrow 0} P\{\text{an event in } [t, t + \Delta] | \mathcal{F}_t\} / \Delta \quad (4)$$

where \mathcal{F}_t is a set of observations until time t , including the history of the occurrence times of the events $\mathcal{H}_t = \{t_i; t_i < t\}$. It is known that the conditional intensity completely characterizes the corresponding

point process (Liptzer and Shirayev, 1978). Clearly a constant conditional intensity provides a stationary Poisson process. If the conditional intensity is independent of the history but dependent only on the occurrence time t , like $\lambda(t|\mathcal{F}_t) = \nu(t)$ for any non-negative function $\nu(t)$ of t , then this means a non-stationary Poisson process which is employed in the subsection 3.1.

There are many interesting classes of point processes which are defined by certain conditional intensity functions. One of them is Hawkes' self-exciting process described by

$$\lambda(t|\mathcal{F}_t) = \mu + \int_0^t g(t-s) dN_s = \mu + \sum_{t_i < t} g(t-t_i).$$

(See Hawkes, 1971, Hawkes and Oakes, 1974.) This functional form resembles the autoregressive model in time series analysis. In this model the expectation of an event occurring is given by a linear combination of past occurrences, where the so-called impulse response function $g(\cdot)$ measures the weights of such combinations. This type of component is considered in the subsection 3.2.

Given a set of occurrence data t_1, t_2, \dots, t_n in an observed time interval $[0, T]$ and a parameterized conditional intensity $\lambda_\theta(t|\mathcal{F}_t)$, the likelihood is written in the form

$$L_T(\theta | t_1, t_2, \dots, t_n; 0, T) = \left\{ \prod_{i=1}^n \lambda_\theta(t_i | \mathcal{F}_{t_i}) \right\} \exp \left\{ - \int_0^T \lambda_\theta(t | \mathcal{F}_t) dt \right\}$$

The maximum likelihood estimate of θ is the value of the parameter vector which maximizes the logarithm of above,

$$\log L_T(\theta | t_1, t_2, \dots, t_n; 0, T) = \sum_{i=1}^n \log \lambda_\theta(t_i | \mathcal{F}_{t_i}) - \int_0^T \lambda_\theta(t | \mathcal{F}_t) dt. \quad (5)$$

If the second term in the right hand side of (5) can be expressed analytically in θ , then the gradients of the log-likelihood function can be easily obtained. In such a case the maximization of the function can be carried out by using a standard nonlinear optimization technique.

Assume that we have to choose the best model among proposed competing models. The Akaike Information Criterion (Akaike, 1974),

$$\text{AIC} = -2 \times (\text{maximum log-likelihood}) + 2 \times (\text{number of parameters})$$

is very suitable for such model comparisons. A model with a smaller AIC is considered to be a better fit.

3.1 Modeling with Cyclic Part

We start with a non-stationary Poisson model, and specify the conditional intensity function as follows.

$$\lambda_{\theta}(t|\mathcal{F}_t) = a_0 + P_J(t) + C_K(t). \quad (6)$$

The second term on the right-hand side of (6) represents the evolutionary trend where

$$P_J(t) = \sum_{j=1}^J a_j \phi_j(t/T), \quad 0 < t < T \quad (7)$$

T is the total length of the observed interval and $\phi_j(\cdot)$ is a polynomial of order j . If the data span of the high frequent data we consider is very long, then we may expect this term could capture the evolutionary change of the quote/trade intensity. The third term in (6) is the Fourier series

$$C_K(t) = \sum_{k=1}^K \{b_{2k-1} \cos(2k\pi t/T_0) + b_{2k} \sin(2k\pi t/T_0)\}, \quad (8)$$

which stands for cyclic effects with a given fixed cyclic length T_0 . If we set T_0 equal to the length of a day, then this cyclic term represents the intraday periodicity.

Sometimes it is useful to employ exponentiated version of trend-cycle model

$$\lambda_{\theta}(t|\mathcal{F}_t) = \exp\{a_0 + P_J(t) + C_K(t)\} \quad (9)$$

instead of (6) to guarantee the positivity of the conditional intensity function. This case can be viewed as the log-linear modeling of the conditional intensity function. In this case, the analytic calculation of the integral (the second part of the right-hand side in (5)) is not generally feasible except in special cases. See Lewis (1970) for example. However, for the slowly varying intensity model such as the exponential rate for a trend, numerical integration well approximates the integral term so that the maximum likelihood becomes feasible.

We fit the model (9) to the high frequent data from yen-dollar exchange market described in the section 2. Because the sample period covers only 16 days ($T = 16$), it is expected that any significant evolutionary movement cannot be conceived from this data.

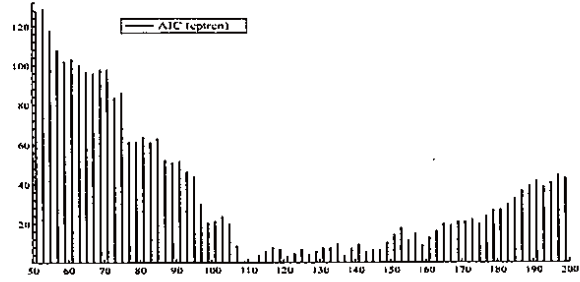


Figure 3: AIC difference vs. order of trigonometric function

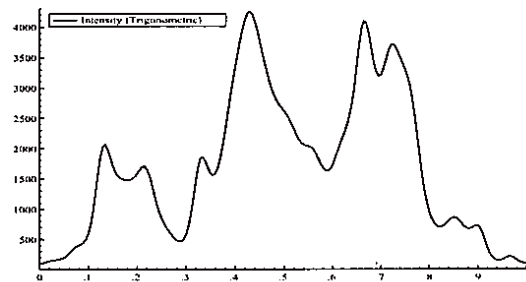


Figure 4: Intensity function induced by the minimum AIC model

Actually, the models with polynomial term of order one, two yields bigger value of AIC, hence those are not reported here. After restricting the model class to constant plus trigonometric functions, we search the best model up to the maximum order 1440. Figure 3 shows that minimum AIC is attained by the model order 111, which means a constant term and 55 pairs of trigonometric functions are required to describe the intensity of yen-dollar quote occurrence during July 1997. (The ordinates in the Figure 3 are the differences of AICs between each model and the minimum AIC model.) The highest frequency in the best model is approximately 26 minutes. Figure 4 shows the estimated cyclic intensity function. Both ends of horizontal axis are 21:00 GST, and the length of a day is set to one ($T_0 = 1$) in this analysis. Three major humps in the graph correspond to the market time of Tokyo, London and New York respectively.

M	K	J	AIC
2	0	0	-370461.21
2	6	0	-370667.19
2	10	0	-370857.12
2	20	0	-370910.87
2	30	0	-370921.32
2	34	0	-370924.39
2	40	0	-370918.18

表 1: AIC values for several configurations

3.2 Modeling with Cyclic and Cluster Part

The class of models we consider in this section follows;

$$\lambda_\theta(t|\mathcal{F}_t) = a_0 + P_J(t) + C_K(t) + \sum_{t_i < t} g_M(t - t_i). \quad (10)$$

The last term in (10) expresses the clustering effects such as quotes or trades invoked by precedent ones. The function $g_M(x)$ measures the increase in clustering due to a quote/trade. We may call this function a response function of an event, and parameterize it as

$$g_M(x) = \sum_{m=1}^M c_m x^{m-1} e^{-\alpha x}, \quad (11)$$

see Ogata and Akaike (1982). This model is used to examine the existence of each component by the comparison of AIC values among a possible set of configurations of (J, K, M) . Different from (9), trend and cyclic function is not exponentiated in (10). It is reasonable if we expect the most of intensity will be absorbed into the cluster effect component. Then the seasonal pattern should become smoother, which lead to the ordinary trigonometric function fitting rather than exponentially transformed one.

After trying various configurations we find that $M = 2$ and $J = 0$ generally lead to smaller values of AIC. After fixing M and J to the above mentioned value, we repeat the maximum likelihood estimation changing the order of the Fourier series, K . Table 1 and Figure 5 shows the model of $K = 34$ attains minimum AIC, -370924.39 . Estimated parameter values of the second order ($M = 2$) Laguerre polynomial in (11) are $\hat{\alpha} = 321.88$, $c_1 = 203.65$ and $c_2 = 45.36$, respectively. Because the cluster effect is almost concentrated around the origin, the intensity implied by the cluster effect part is drawn with the horizontal

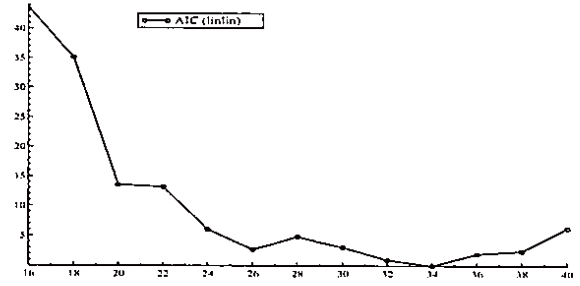


图 5: AIC difference vs. order of trigonometric function

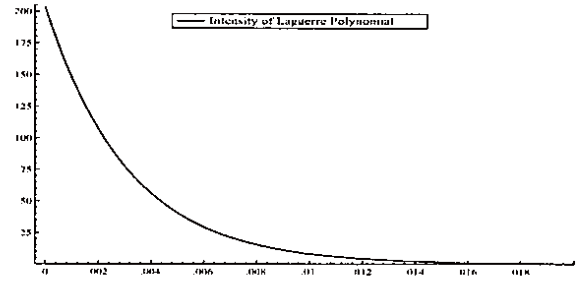


图 6: Intensity implied by the estimated Laguerre polynomial

axis ranges only over $[0, 0.2]$, which approximately corresponds to 4 hours and 50 minutes. Numerical integration of the function with the estimated parameters over $(0, 0.00485)$ produces almost $1/2$. Hence if a quote occurred right after the previous quote, the probability of the next quote occurrence is very high.

Figure 7 shows the shape of the cyclic part of the intensity function if we change the order of Fourier series K from 2 to 40. Model order increases horizontally, and the most bottom-right figure corresponds to $K = 40$ case. The cases of $K = 16, 24, 34, 40$ are dis-

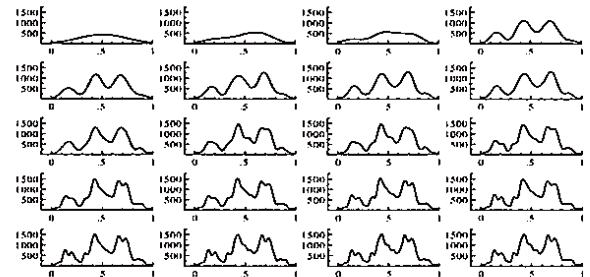


图 7: Cyclic part of the intensity function induced by the minimum AIC model. The order of trigonometric functions (K) varies from 2 to 40.

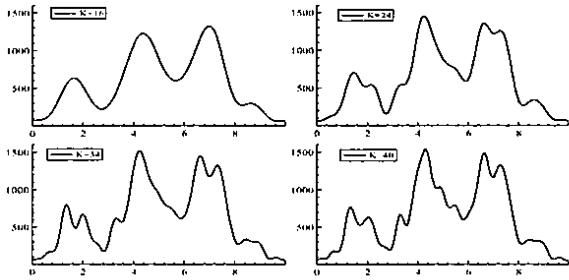


图 8: Cyclic part of the intensity function induced by the minimum AIC model.

M	K	J	AIC
0	110	0	-368454.04
2	0	0	-370461.21
2	34	0	-370924.39

表 2: Comparison of AIC between the best model and other extreme models

played in Figure 8. Compared to the Figure 4, we find the shape is almost alike but the scale of ordinates is very different. In table 2 we report the AIC values of the best model and two extreme cases. Introducing the second order Laguerre polynomial is far better than the Fourier series modeling where the difference of AIC is more than 2000. Adding periodicity to the pure cluster effect model results in improvement of AIC by 463.18.

4 Residual Seasonality Analysis

To demonstrate an advantage to employ direct seasonality modeling in a point process model, we check residual seasonality in a seasonally adjusted point process by spline smoothing. To begin with, we divide durations in a 5 minutes bin by the estimated value of the time of the day function. Then we rescale the whole time axis so that the terminal values of the original and adjusted point process data coincide. We call it the intra-day periodicity adjusted point process data.

To check the residual seasonality in the adjusted point process data, we start by fitting the baseline model of which conditional intensity function consists of constant term plus second order Laguerre polynomial. Alternative model considers *one* sine or cosine term with the baseline model. We change the fre-

quency from 24 hours to 36 minutes. If intra-day periodicity is completely removed, AIC of the baseline model should be smaller.

K	AIC Diff.	K	AIC Diff.
1	0.86	21	1.57
2	1.58	22	-0.02
3	1.88	23	1.66
4	1.66	24	0.74
5	-0.06	25	1.94
6	1.67	26	1.57
7	1.81	27	0.82
8	1.95	28	-3.01
9	-1.86	29	1.63
10	-1.07	30	0.78
11	-1.20	31	-1.86
12	1.88	32	1.29
13	0.41	33	-1.57
14	2.00	34	-0.69
15	-0.22	35	0.30
16	0.53	36	1.67
17	0.61	37	-2.22
18	-0.64	38	1.74
19	1.51	39	1.73
20	1.96	40	1.57

表 3: Differences in AIC between alternative and baseline model.

The results are summarized in Table 3. Numbers reported in the table are the differences in AIC, namely the AIC of the alternative model minus the AIC of the baseline model. Thus the negative value suggests that there may remain some seasonality in the adjusted point process data.

After investigating single separate component, we prepare two more alternative models with multiple sine and cosine functions. In Alternative 1, we add all the frequency components that showed negative AIC difference in Table 3. On the other hand, Alternative 2 includes terms with strong evidence only. So Alternative 1 has 12 sine/cosine terms while Alternative 2 has only 7 terms with all AIC difference less than -0.5.

The results are summarized in Table 4. In any case, the usual spline based custom to adjust intra-day periodicity might lead to some under adjustment, and there remains seasonality in the residual to some extent.

Model	AIC	Δ AIC
Baseline	-356839.58	—
Alternative 1	-356854.03	-14.45
Alternative 2	-356852.42	-12.84

表 4: Comparison of AIC between the best model and other extreme models

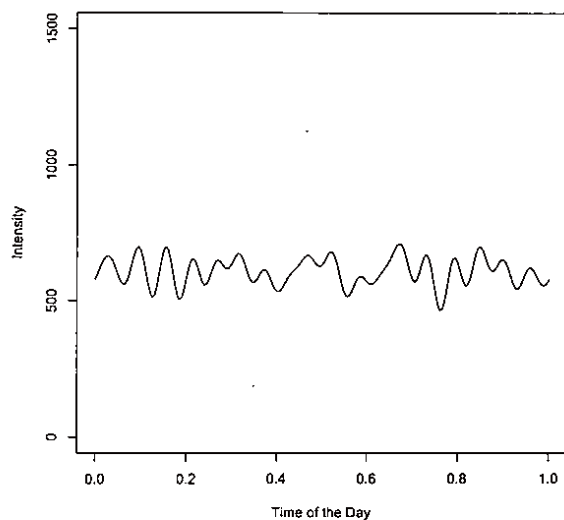


图 9: Estimated periodic intensity of time deformed process.

We can visually inspect the residual periodicity in Figure 9. Horizontal axis stands for the time of a day, and estimated periodic intensity of time deformed process is plotted.

5 Conclusion

A class of models for high frequent financial data is proposed. It turns out that incorporating cluster effect is essential to the modeling, which is parallel to the commonly used duration time based autoregressive models. The estimated shapes of the cyclic component of the intensity function shows the simple-minded application of built-in spline smoothing method may lead to smoother intraday periodicity pattern. Fitting conditional intensity model reveals the remaining seasonality in adjusted data by spline based two step method.

参考文献

- [1] Akaike, H., A new look at the statistical model identification, *IEEE Trans. Autom. Control*, **19**, 716-723.
- [2] Akaike, H. et al., TIMSAC-84 part 2, *Computer Science Monographs No. 23*, The institute of Statistical Mathematics, 1985.
- [3] Baillie, R. T. and M. M. Dacorogna, High frequent data in finance, *Journal of Empirical Finance*, **4**, 69-72, 1997.
- [4] Bauwens, L. and P. Giot, The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks, *Annales d'Economie et de Statistique*, **60**, 117-149, 2000.
- [5] Engle, R., Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**, pp. 987-1007, 1982.
- [6] Engle, R. and J. Russell, Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model, *J. of Empirical Finance*, **4**, 187-212, 1997.
- [7] Engle, R. and J. Russell, Autoregressive conditional duration: a new approach for irregularly spaced transaction data, *Econometrica*, **66**, 1127-1162, 1998.
- [8] Frankel, J., G. Galli and A. Giovannini (eds.), *The microstructure of foreign exchange markets*, University of Chicago Press, 1996.
- [9] Hawkes, A. G., Point spectra of some mutually exciting point processes, *J. Roy. Statist. Soc. (B)*, **33**, 438-443, 1971.
- [10] Hawkes, A. G. and D. A. Oakes, A cluster process representations of self-exciting process, *J. Appl. Probab.*, **11**, 493-503, 1974.
- [11] Kuwana, Y., M. Susai and Y. Kawasaki, How the scheduled macroeconomic announcements influence foreign exchange markets (*in Japanese*), *Proceedings of the Institute of Statistical Mathematics*, **48**, 213-227, 2000.

- [12] Lewis, P. A. W., Remarks on the theory, computation and application of the spectral analysis of series of events, *J. Sound. Vib.*, **12**, 353–375, 1970.
- [13] Liptzer, R. S. and A. N. Shiriyayev, *Statistics of Random Processes II: Applications*, Springer-Verlag, New York, 1978.
- [14] MacLean, C. J., Estimation and testing of an exponential polynomial rate function within the non-stationary Poisson process, *Biometrika*, **61**, pp. 81–86, 1974.
- [15] Ogata, Y., Estimation of the parameters in the Modified Omori Formula for Aftershock Frequencies by the Maximum Likelihood Procedure, *J. Phys. Earth*, **31**, 115–124, 1983a.
- [16] Ogata, Y., Likelihood Analysis of Point Processes and its Applications to Seismological Data, *Bull. Int. Statist. Inst.*, **50**, Book 2, 943–961, 1983b.
- [17] Ogata, Y. and K. Katsura, EPTREN in TIMSAC-84 part 2 (Akaike et al. eds.), *Computer Science Monographs No. 23*, The institute of Statistical Mathematics, pp. 187–197, 1985.
- [18] Ogata, Y. and K. Katsura, LINLIN in TIMSAC-84 part 2 (Akaike et al. eds.), *Computer Science Monographs No. 23*, The institute of Statistical Mathematics, pp. 198–211, 1985.
- [19] Ogata, Y. and H. Akaike, On linear intensity models for mixed doubly stochastic Poisson and self-exciting point process, *J. Royal Statist. Soc. B*, **44**, pp. 104 – 107, 1982.
- [20] Ogata, Y., Seismicity analysis through point-process modeling: a review, *Pure and Applied Geophysics*, **155**, 471–507, 1999.
- [21] Zhang, M. Y., J. R. Russell and R. T. Tsay, A nonlinear autoregressive conditional duration model with application to financial transaction data, *J. of Econometrics*, **104**, 179–207, 2001.

「バイアス補正実現分散推定量の構築」

大屋 幸輔

大阪大学 大学院経済学研究科 & CSFI
独立行政法人 科学技術振興機構, GREST

2010年3月30, 31日

「桜の季節に計量ファイナンス 2010」 東京大学経済学部

概要

先行研究と研究目的

価格過程

RV と関連する推定量

Lag (J, K) の選択

バイアス補正 RV

Simulation

Conclusion

References



K.Oya Bias 補正 RV	大阪大学
先行研究と研究目的	References
価格過程	Conclusion
Simulation	References
Conclusion	References
References	References

先行研究と研究目的

- そもそも問題
 - Realized Variance using High-frequency obs. is badly biased due to microstructure effect such as bid-ask bounce, price discreteness
- 先行研究：独立でないノイズ系列に対して
 - Extended TSRV: Ait-Sahalia, Mykland & Zhang (2005)
 - Realized kernel: B.-N., H., L. & S. (2008)

目標

- 1 ノイズ系列の従属性をどうやって測るか
- 2 Two time scales of RV の選択方法
- 3 ノイズの自己共分散を利用したバイアス補正法の提案



K.Oya Bias 補正 RV	大阪大学
先行研究と研究目的	References
価格過程	Conclusion
Simulation	References
Conclusion	References
References	References

価格過程 & Integrated variance

- 真の価格過程 & IV

$$dP^*(t) = \sigma(t)dW(t), t \in [0, T] \quad (1)$$

$$IV = \langle P^*, P^* \rangle_T = \int_0^T \sigma^2(s)ds, \sigma(t) > 0 \quad (2)$$

where $W(t)$ is standard B.M. and $\sigma_i(t)$ is a bounded measurable func.

- 観測価格

obs.	true	noise
$P(t_i)$	$P^*(t_i)$	$\eta(t_i)$

- $P(t_i)$: 時刻 t_i での執行価格
- $\eta(t_i)$: microstructure noise
- $(t_{i-1}, t_i]$: 第 i 番目の執行間隔
- $\tau_i := P(t_i) - P(t_{i-1})$: 第 i 番目の return



Microstructure Noise に対する仮定

Assumption 1: $\eta(t)$

(1a) $E[\eta(t)] = 0$

(1b) stationarity with autocovariance function, which has finite dependence in the sense that:

$$\gamma(\ell) = E[\eta(t)\eta(t-\ell)] = 0, \text{ for all } |\ell| > m.$$

m is a finite positive number. $\gamma(\ell)$ is a decreasing func. of $|\ell|$.

(1c) $E|\eta(t)\eta(s)|^{4\beta} < \infty$ for some $\beta > 1$ and all t, s .

(1d) $P \perp \perp \eta(t)$.

- 仮定 (1d) が成立していなくても、高頻度 (執行間隔が非常に短い) のときは影響は軽微 \Leftarrow Hansen & Lunde (2006)

Two Scale RV: Zhang, Mykland & Ait-Sahalia (2005)

- original grid of observation time $\mathcal{G} = \{t_0, t_1, \dots, t_n\}$
- K non-overlapping subgrids $\mathcal{G}_K^{(j)}, j = 1, \dots, K, \mathcal{G} = \cup_{j=1}^K \mathcal{G}_K^{(j)}$
 $\mathcal{G}_K^{(j)} = \{t_{j-1}, t_{j-1+K}, t_{j-1+2K}, \dots, t_{j-1+nK}\}$

Two Scale RV

RV for subgrid $\mathcal{G}_K^{(j)}$ $RV_K^{(j)} = \sum_{i=1}^{n_j} (P(t_{(j-1)+iK}) - P(t_{(j-1)+(i-1)K}))^2$ (5)

averaging $RV_K^{(avg)} = \frac{1}{K} \sum_{j=1}^K RV_K^{(j)}$ (6)

bias correction $RV_K = RV_K^{(avg)} - \frac{\bar{\eta}_K}{n} RV^{(alt)}$ (7)

- The second term of (7) is a bias correction.
- This bias correction is valid if η_{t_i} s are uncorrelated.

Extended TSRV: Ait-Sahalia, Mykland & Zhang (2006)

with dependent microstructure noise

- the bias of TSRV is not removed completely
 $bias = 2\bar{\eta}_K \{ \gamma(1) - \gamma(K) \}, \bar{\eta}_K = (n - K + 1)/K$
- Since K is large ($K = O(n^{2/3})$), $\gamma(K)$ is negligible but $\gamma(1)$ remains.
- Extended TSRV uses the fast time scale J instead of $RV^{(alt)}$.

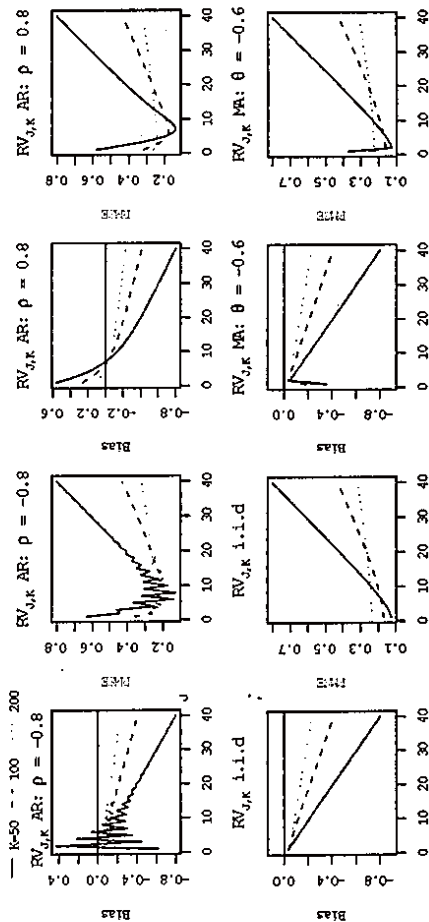
Extended TSRV

$$RV_{j,K} = \underbrace{RV_K^{(avg)}}_{\text{slow time scale}} - \frac{\bar{\eta}_K}{\bar{\eta}_j} \times \underbrace{RV_j^{(avg)}}_{\text{fast time scale}} \quad (8)$$

$$1 \leq j < K \leq n, K = o(n), RV_{1,K} = TSRV$$

- Bias is $2\bar{\eta}_K \{ \gamma(1) - \gamma(K) \}, \gamma(1)$ and $\gamma(K)$ decay as J being larger.
- However, we must find appropriate K and J in practice.

- 横軸は J で、縦軸は $\text{Bias} = (\text{estimate} - IV)/IV$ と RMSE
- 相対 Bias と RMSE はラグ J に強く依存
- 適切に J を選択できれば、 K の選択の影響は少ない



従属性の閾値 m の決定: $\gamma(m+1) = 0, \gamma(m) \neq 0$

二つの収益率の積の期待値

$$\begin{aligned}
 E_i[Z_{\ell,ij}] &= E_i \left[\underbrace{(P^*(t_i) - P^*(t_{i-1}))}_{\text{zero}} (P^*(t_j) - P^*(t_{j-1})) \right] \\
 &+ E_i \left[\{\eta(t_i) - \eta(t_{i-1})\} \{\eta(t_j) - \eta(t_{j-1})\} \right] \\
 &= \gamma(\ell + \Delta t_i) - \gamma(\ell + \Delta t_i + \Delta t_j) - \gamma(\ell) + \gamma(\ell + \Delta t_j)
 \end{aligned}$$

- 十分大きな L を考える。ただし $\gamma(L) = 0$ and $\gamma(i) = 0$ for $i > L$
- $s^* = \min_s \{s \mid \gamma(L-s) \neq 0, s > 0\}$ とすると
 $E_i[Z_{L,ij}] = \dots = E_i[Z_{L-(s^*-1),ij}] = 0$ and $E_i[Z_{L-s^*,ij}] \neq 0$.
- 閾値 $m = L - s^*$ は $E_i[Z_{\ell,ij}] = 0$ かどうかを検定することで特定する



どのようにして従属性の程度を測るか

資産が時点 $t_{i-1}, t_i, \dots, t_{j-1}, t_j, \dots$ で約定 (価格が観測)

- 二つの重なりの無い区間 $(t_{i-1}, t_i]$ and $(t_{j-1}, t_j]$, $i < j$
- この二つの区間の距離を $\ell = t_{j-1} - t_i$ で定義
- この距離 ℓ の二つの区間上での収益率の積を $Z_{\ell,ij}$ と定義

$$Z_{\ell,ij} = (P(t_i) - P(t_{i-1})) (P(t_j) - P(t_{j-1}))$$

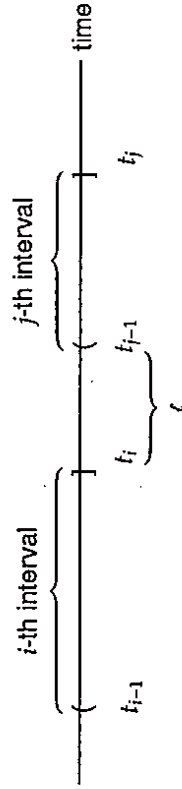


Figure: 距離が ℓ の重なりの無い区間



ノイズの従属性に関する検定

For given ℓ

- $H_0: E_i[Z_{\ell,k}] = 0$ for all k , H_1 : all possible deviations from H_0

Theorem 1: Ubukata & Oya (2009)

Suppose Assumption 1 and 2 hold. As N_ℓ goes to infinity we have

$$\tau(\ell) := \frac{\sqrt{N_\ell} Z_{\ell, N_\ell}}{\hat{\sigma}_\ell} \xrightarrow{d} N(0, 1) \text{ under } H_0. \quad (9)$$

Under H_1 , $\tau(\ell)$ diverges.

- Let $\{Z_{\ell,k}\}_{k=1}^{N_\ell}$ be a sequence that arranges $Z_{\ell,ij}$ satisfying $\ell = t_{j-1} - t_i \geq 0$ for given ℓ in ascending order of index i .
- \bar{Z}_{ℓ, N_ℓ} is a sample mean of $\{Z_{\ell,k}\}$.
- $\hat{\sigma}_\ell^2$ is a subsampling est. of $\lim_{N_\ell \rightarrow \infty} N_\ell E[(\bar{Z}_{\ell, N_\ell} - E[\bar{Z}_{\ell, N_\ell}])^2]$.



観測時定化の検定手続き

- ① 十分に大きな L に関して $E_l[Z_{L,k}] = 0$ for all k を検定
- ② $H_0: E_l[Z_{L,k}] = 0$ が棄却されなければ, $H_0: E_l[Z_{L-1,k}] = 0$ を検定.
- ③ 帰無仮説が棄却されるまで, 検定をつづける
- ④ はじめに帰無仮説が棄却された距離を閾値 m の推定値とする
 $\gamma(m) \neq 0$ and $\gamma(\ell) = 0$ for all $\ell > m$
- ⑤ この閾値 m の推定値を利用しノイズの自己共分散の不偏推定量を構築
- ⑥ ノイズの自己共分散推定に際して, m の過大推定は支障は無いが, 過小推定の場合には自己共分散推定量の不偏性が保証されない.

自己共分散推定量と検定統計量 Ubukata & Oya (2009)

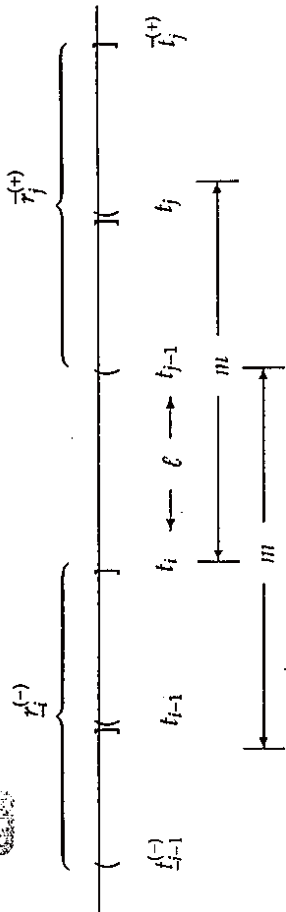
- Autocovariance estimator of noise

$$\hat{\gamma}(\ell) = -\frac{1}{N_\ell} \sum_{k=1}^{N_\ell} Z_{\ell,k}^{(+)} Z_{\ell,k}^{(-)}, \quad N_\ell^{1/2}(\hat{\gamma}(\ell) - \gamma(\ell)) \xrightarrow{d} N(0, \omega_\ell^2)$$
 - $\{Z_{\ell,k}^{(\pm)}\}_{k=1}^{N_\ell}$ arranged $\{Z_{\ell,j}^{(\pm)}\}$ in ascending order of index i
 - $E_j[Z_{\ell,k}^{(\pm)}] = -\gamma(\ell)$
- Test statistic for Autocovariance of noise
 - $H_0: \gamma(\ell) = 0, H_1: \gamma(\ell) \neq 0$, for $\ell > 0$
$$\tau^*(\ell) = \frac{\sqrt{N_\ell} \hat{\gamma}(\ell)}{\hat{\omega}_\ell} \xrightarrow{d} N(0, 1) \text{ as } N_\ell \rightarrow \infty \text{ under } H_0.$$

$\tau^*(\ell)$ を Extended TSRV の J の選択に利用する

 $\gamma(\ell)$ の不偏推定量の構築

- 下図のように区間を拡大し二つの収益率の積から $\gamma(\ell)$ 以外の影響を排除



- $t_j^{(+)}$ は t_j 以降で $t_j^{(+)} - t_j > m$ を満たす最初の約定時刻
- $t_{j-1}^{(-)}$ は t_{j-1} 以前で $t_{j-1} - t_{j-1}^{(-)} > m$ を満たす最後の約定時刻
- $Z_{\ell,j}^{(\pm)} = Z_{\ell}^{(-)}(t_j^{(-)}) (P(t_j^{(-)})) - P(t_{j-1}^{(-)})$

First time scale J と K の選択

- Selection J
 - ① $\gamma(1) = 0, \gamma(2) = 0, \dots$ を $\tau^*(\ell)$ によって選択
 - ② $H_0: \gamma(\ell) = 0$ が棄却されなかった最初の ℓ をノイズの従属性の閾値の推定値 j とする
- Selection K
 - *i.i.d.* noise case, Optimal choice of K は Zhang et al.(2005)
 - *dependent* noise case
 - ① original grid of observation times $\mathcal{G} = \{t_0, t_1, \dots, t_n\}$ に対して, j をもちいて new subgrid $\mathcal{G}_j = \{t_0, t_j, t_{2j}, \dots, t_{\lfloor n/j \rfloor}\}$ を定義
 - ② ノイズ系列 $\{\eta(t_{ij})\}_{i=0}^{\lfloor n/j \rfloor}$ を無相関系列と見なし, 観測価格系列 $\{P(t_{ij})\}_{i=0}^{\lfloor n/j \rfloor}$ に対して, Zhang et al.(2005) の方法を適用

Alternative way to remove bias

- Bias of avg. subsample RV is $2\bar{n}_K\gamma(0)$
- $\gamma(0)$ can be estimated by proposed method in UO(2009).

バイアス補正 RV (with appropriate choices of K and J)

$$RV_K^{(bc)} = \underbrace{RV_K^{(avg)}}_{\text{avg. of subsample RV}} - \underbrace{2\bar{n}_K \gamma(0)}_{\text{bias correction}} \xrightarrow{J} IV$$

Comparison of Estimators

$$RV_{j,K} = RV_K^{(avg)} - \frac{\bar{n}_K}{\bar{n}_j} \times RV_j^{(avg)}$$

$$RV_{j,K}^{(adj)} = \left(1 - \frac{\bar{n}_K}{\bar{n}_j}\right)^{-1} RV_{j,K} : \text{small sample correction}$$

$$RV_K^{(bc)} = RV_K^{(avg)} - 2\bar{n}_K \gamma(0)$$

実験デザイン: Zhang, Mykland & Ait-Sahalia (2005)

$$dP^*(t) = (0.05 - v(t)/2)dt + \sigma(t)dB(t)$$

$$dv(t) = 5(0.04 - v(t))dt + 0.5v(t)^{1/2}dW(t), \quad v(t) = \sigma^2(t)$$

- parameter の値は Z.M.A. (2005) と同じ
- $\Delta t = 1 \text{ sec.}$, $T = 1 \text{ day} = 6.5 \text{ hours}(23400 \text{ sec.})$ of open trading
- 観測価格 $P(t_i) = P^*(t_i) + \eta(t_i)$
- ノイズの従属性:

$$AR(1): \eta(t_i) = \rho \eta(t_{i-1}) + \varepsilon(t_i), \quad \rho = -0.8, -0.4, 0.0, 0.4, 0.8$$

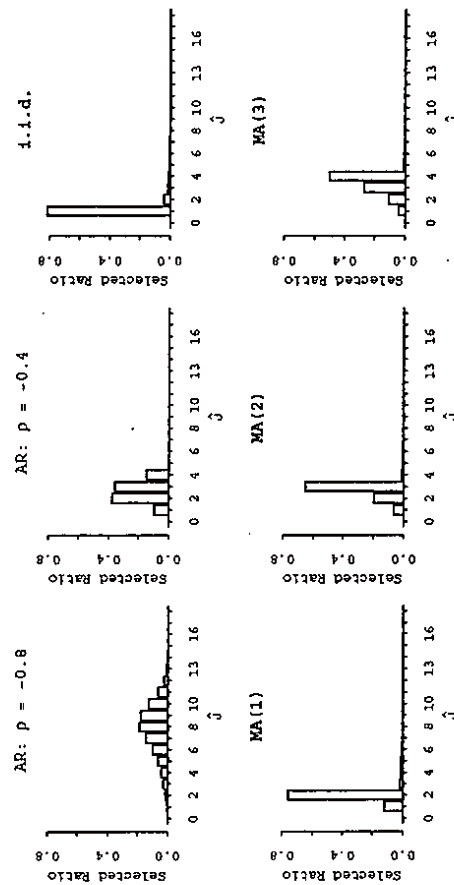
$$MA(3): \eta(t_i) = \varepsilon(t_i) + \theta_1 \varepsilon(t_{i-1}) + \theta_2 \varepsilon(t_{i-2}) + \theta_3 \varepsilon(t_{i-3}),$$

$$(\theta_1, \theta_2, \theta_3) = (-0.6, 0, 0), (0.6, 0, 0), (0.6, 0.5, 0), (0.6, 0.5, 0.4).$$

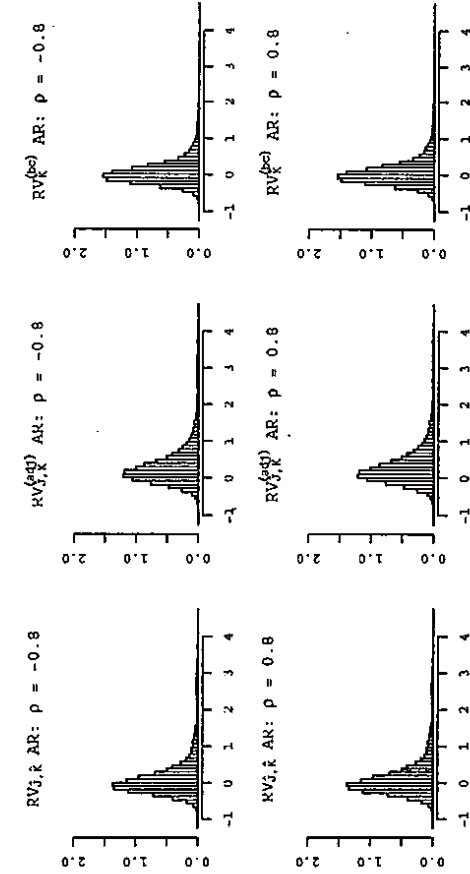
- noise to signal ratio: $\sigma_{\eta}^2/IV = 0.004$

- Hansen & Lunde (2006): $\sigma_{\eta}^2/IV = 0.0002 \sim 0.006$ in NYSE and NASDAQ

Performance of selection J



各推定値の相対誤差の分布









Summary of simulation




Table 1: Relative Bias and RMSE of estimators

Noise Type	Bias			RMSE		
	$RV_{j,k}$	$RV_{j,k}^{(adj)}$	$RV_k^{(bc)}$	$RV_{j,k}$	$RV_{j,k}^{(adj)}$	$RV_k^{(bc)}$
AR: $\rho = -0.8$	-0.131	0.035	0.008	0.326	0.366	0.259
AR: $\rho = -0.4$	-0.064	0.009	-0.006	0.208	0.218	0.214
AR: $\rho = 0.4$	0.097	0.181	0.041	0.336	0.374	0.295
AR: $\rho = 0.8$	0.123	0.344	0.077	0.543	0.688	0.385
i.i.d.	-0.074	-0.019	-0.005	0.160	0.141	0.213
MA(1)	0.016	0.081	0.025	0.334	0.358	0.292
MA(2)	0.058	0.146	0.032	0.400	0.446	0.314
MA(3)	0.088	0.196	0.037	0.441	0.506	0.336

References I

 Ait-Sahalia, Y., P. Mykland and L. Zhang (2006), "Ultra high frequency volatility estimation with dependent microstructure noise," NBER Working Paper No. 11380.
 Bandi, F. M. and J. R. Russell (2008a), "Microstructure noise, realized variance, and optimal sampling," *Review of Economic Studies*, 75 (2), 339-369.
 Bandi, F. M. and J. R. Russell (2009), "Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations," *Journal of Econometrics*, forthcoming.
 Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2008), "Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise," *Econometrica*, vol.76, No.6, 1481-1536.
 Hansen P.R. and A. Lunde (2006), "Realized variance and market microstructure noise (with discussions)," *Journal of Business and Economic Statistics*, 24, 127-218.
 McAleer, M. and M. C. Medeiros (2008), "Realized Volatility: A Review," *Econometric Reviews*, 27, 10-45.

References II

 Ubukata, M. and K. Oya (2009a), "Estimation and Testing for Dependence in Market Microstructure Noise," *Journal of Financial Econometrics*, 7, 106-151.
 Ubukata, M. and K. Oya (2009b), "Statistical Properties of Covariance Estimator of Market Microstructure Noise: Dependence, Rare Jumps and Endogeneity," to appear in *Recent Advance in Financial Engineering*, World Scientific.
 Zhang, L., P. Mykland, and Y. Ait-Sahalia (2005), "A tale of two time scales: Determining integrated volatility with noisy high frequency data," *Journal of the American Statistical Association*, 100, 1394-1411.

Conclusion & Future Research

- This study provides
 - Extended TSRV の two time scales (J, K) の選択方法を提案
 - Microstructure noise の分散推定量をもちいた Extended TSRV のバイアス補正法を提案
 - Monte Carlo 実験の結果から
 - Extended TSRV は適切な J の選択を要求
 - 提案した選択方法の結果は良好
 - 提案したバイアス補正 RV は相対的に小さなバイアスと RMSE
 - Future topics
 - Subsampling realized kernel 法 (two time scale タイプよりも効率的) との比較
- Bandi and Russell(2009) によれば実証分析で典型的な標準サイズでは必ずしも漸近理論が示すとおりではない
- Small sample correction