

先物市場の高頻度データ～立会時間延長の影響分析*

川崎能典[†] 吉田靖[‡]

概要

グローバルな取引市場間の競争が強まる中、投資家の利便性向上を目的として、特にわが国ではデリバティブ市場において、立会時間の延長が実施されている。本報告では、東京商品取引所に上場されている金先物および大阪証券取引所に上場されている日経平均先物の高頻度データを利用し、立会時間の制度変更により取引発生頻度の経時的変化を点過程モデルのあてはめを通じて観察する。特に、従来の国内の取引所は株式市場・商品市場共に、海外の取引所にはない昼休みが存在することが、特徴であった。この制度により、国内の高頻度データを使用した日内季節変動の分析では、昼休み前後の変動パターンが海外の実証結果と大きく異なる原因となっていた。本報告では、取引時間の延長により日中の変動パターンが変化していく様子を点過程モデルにより検証する。

キーワード：点過程，非定常ポアソン過程，金先物，日経平均先物

1 はじめに

1990年に大阪証券取引所の日経平均先物が取引代金ベースでCME (Chicago Mercantile Exchange) のS&P500先物を上回った翌年、Miller(1991)は米国の先物取引の国際競争力について分析している。その後も、取引所間の競争はデリバティブに限らず、2007年のニューヨーク証券取引所とユーロネクストの合併に代表されるように、クロスボーダー取引の発達によって、激しさを増している。国内でもかつては最大の先物取引所であった東京穀物取引所が2013年2月にすべての取引を終了し、東京工業品取引所は農産物が移管され、東京商品取引所に商号変更し、商品取引所は大阪堂島商品取引所との2取引所のみとなった。また、東京証券取引所と大阪証券取引所は2013年1月に経営統合するなど、統廃合が進んでいる。

取引所は、単に流動性を提供するだけでなく、価格の発見機能を通じて、効率的な資源配分に寄与し、また、金融センターの中心となる社会的インフラでもあり、国全体の競争力向上にとっても重要な存在である。さらに、投資家のニーズは、高度化・複雑化が著しく、取引システムの安定性、高速性、取引時間の延長から、決済制度の改善まで多岐に亘っている。具体的には、東京工業品取引所(当時)は、2009年5月から新取引システムを稼働させ、立会時間の延長や、処理速度の向上を実施し、東京証券取引所は、2010年1月からarrowheadを稼働させ高速注文への対応や、呼値の刻みの縮小など、高速性、信頼性、拡張性を向上させ、大阪証券取引所では、デリバティブ市場において2011年7月から早朝3時までのナイトセッションを開設するなど、積極的な利便性の向上が実施されている。

一方で学術的には、高頻度データによる分析が進展しており、特にリアライズド・ボラティリティの推計に代表される理論と実証については、Takahashi et al.(2009)などの多くの研究が発表され、生方・渡部(2011)にあるようにリスク管理などの分野でも大きな貢献が期待されている。また、Barclay,Hendershott(2003)やCooper et al.(2008)のように、夜間取引の分析をおこなっている論文も多数ある。しかしながら、前述の取引システム改善の効果の学術的な検証は必ずしも多くはなく、宇野・柴田(2012)や太田(2013)など

*2013年12月26日、科学研究費課題「経済リスクの統計学の新展開：稀な事象と再起的事象」(基盤研究(A) 研究代表者：国友直人 東京大学・経済学部)に基づく研究集会。

[†]統計数理研究所・モデリング研究系：〒190-8562 東京都立川市緑町10-3, e-mail address: kawasaki@ism.ac.jp

[‡]東京経済大学経営学部 〒185-8502 東京都国分寺市南町1-7-34, e-mail address: yyoshida@tku.ac.jp

の東証における取引高速化の分析に限定されており、同時に進められている取引時間変更の評価を行っている論文は非常に少ない。

このような背景の下、本研究は、大阪証券取引所の日経平均先物と東京商品取引所の金先物のティックデータを用いて、立会時間の延長が日中取引数量の推移パターンに与えた影響を検証し、新システムの導入が市場に与えた影響を評価することを目的とする。

従来、国内の取引所は株式市場・商品市場共に、海外の取引所にはない昼休みが存在することが特徴であった。この制度により、渡部(2004)が指摘するように国内の高頻度データを使用した日内季節変動の分析では、特に昼休み前後の変動パターンが海外の実証結果と大きく異なっていた。また、そのため渡部(2010)のようなりアライズド・ボラティリティの調整なども必要になっている。株式現物市場においても昼休みの廃止も議論されているものの、短縮が実施されたのみで現時点で昼休みは存在している。しかし、東京商品取引所と大阪証券取引所のデリバティブ市場においては、以前は昼休みが存在したが、現在は廃止されている。また、夜間の立会は、東京工業品取引所は午前4時まで、大阪証券取引所のデリバティブ市場は午前3時までと延長されている。現在では両取引所の分析をおこなうためのデータ蓄積も十分であり、さらに相互の比較も可能であるため、日内季節性の変動分析にとって、意義のある研究が可能である。

したがって、本研究は、大阪証券取引所および東京商品取引所の先物のティックデータを用いて、立会時間の延長が日中の価格変動性や取引数量に与えた影響を検証し、新制度の導入が市場の流動性に与えた影響を評価することを目的とし、取引システムの今後の改善やリスク管理の精度向上の基礎となることを目的とする。

つづく第2節では、分析対象である東京商品取引所の金先物と大阪証券取引所の日経平均先物の取引制度の変遷と分析に用いたデータについて述べる。第3節では、本稿で推計に用いた点過程モデルの推計について概説する。第4節は実証分析の結果である。最後に第5節として、まとめと今後の課題を述べる。

2 分析対象商品とデータ

東京商品取引所での分析対象は、商品先物の中で、最も流動性が高く、またグローバルな銘柄で、各地の取引所を利用することにより24時間の取引が可能でもあり、さらにETF(上場投資信託)の形での取引も多いことで証券市場との関連も高い金先物とする。東京商品取引所の制度では、金先物が2か月毎に6限月あるが、芹田他(2005)と同様にその中で最も取引量の多い先限の約定データを対象とする。分析にあたっては、約定データを1秒毎のデータに集約する。すなわち、取引量(枚数)は同一秒の中で合計し、価格は同一秒の中で最も遅い取引価格を採用する。

大阪証券取引所での分析対象は、最も取引量が多い日経平均先物とする。日経平均先物は、3か月毎に限月が存在するが、最も取引量の多い期近物とする。ただし、SQ前日において次の限月の取引量が多くなるのが通例であることから、SQ前日午前零時をもって、限月交代するものとして取り扱う。約定データを使用することと、秒データへの集約方法は、金先物と同様である。

季節的変動パターンモデルの構築にあたっては、既存研究では森保(2008)のようにスプライン関数を使用したモデルが多いが、周期性が残存している可能性があり、TIMSAC84(Akaike et al. 1985)のEPTRENを用いて条件付き強度関数によるモデリングを行ない比較する。

3 点過程モデルによる周期性解析

本節ではまず、さまざまな点過程データを解析するための一般的な枠組みを、条件付き強度関数¹(conditional intensity function)に基づき説明する。主として統計地震学を念頭に置いたものではあるが、点過程モデリングの優れた展望論文としてOgata(1999)を挙げておく。

¹Intensityの訳語として強度よりは生起度のほうが適切であるとする書もある。マクニール他(2008)を参照。

3.1 点過程モデルの最尤推定

正の実時間軸 $(0, \infty)$ 上でランダムに生起する系列事象 $\{t_i; 0 < t_1 < t_2 < \dots\}$ を考える．系列の記述にあたっては，まず隣り合う事象に関して生起時刻の差 $X_i = t_i - t_{i-1}$ を考えると，基本的に正の値を取る確率過程を考えることになる．(後に尤度関数の定義で問題が生じないように，生起時刻の差は厳密に正であるとすし，もし現実のデータに tie が存在する場合は，観測精度未満の時刻分だけいずれかをずらす操作を人為的に行う．) この生起間隔が独立同一分布に従う場合はよく知られた再生過程 (renewal process) であり，その周辺分布が指数分布であれば，系列事象は定常ポアソン過程に従う．

$N(a, b)$ で，区間 (a, b) に入る点 (事象) の個数を表そう． N は 0 個であるかもしれないが負にはならず，非負整数値確率変数といえる．このとき，微小な時間区間に生起する事象の個数を予測する問題を考えよう．正の実時間軸 $(0, \infty)$ 上の点過程を想定し，それらを幅 δ の微少な区間に分割する．このとき， $\xi_k = N[(k-1)\delta, k\delta)$ を分割後の区間 $[(k-1)\delta, k\delta)$ 上の k 番目の確率変数とすると， $\{\xi_k\}$ はひとつの確率過程になっている．もし δ が十分小さければ， $\{\xi_k\}$ が二値確率過程 (binary process) であると見なして良い．もしいま考えている点過程が定常ポアソン過程なのであれば， $\{x_{i_k}\}$ 独立同一の分布に従うベルヌーイ系列である．しかし一般的には，系列事象の同時確率は条件付き確率 $P\{\xi_k = 1 | \xi_1, \dots, \xi_{k-1}\}$, $k = 1, 2, \dots$ の列，すなわち過去の履歴に依存してきまると考えるべきである．

このとき条件付き確率を時間に関して微分して，条件付き強度関数 $\lambda(t|\mathcal{F}_t)$ を

$$P\{N(t, t + \delta) = 1 | \mathcal{F}_t\} = \lambda(t|\mathcal{F}_t)\delta + o(\delta),$$

あるいは

$$\lambda(t|\mathcal{F}_t) = \lim_{\Delta \rightarrow 0} P\{\text{an event in } [t, t + \Delta] | \mathcal{F}_t\} / \Delta \quad (1)$$

で定義する．ここで \mathcal{F}_t は時刻 t までの観測値の集合が生成する情報セットであり，その中には事象生起時刻の履歴 $\mathcal{H}_t = \{t_i; t_i < t\}$ も含む．条件付き強度関数の定式化によって，対応する点過程は完全に特徴付けられることが知られている (Liptzer and Shiriyayev, 1978)．条件付き強度関数が定数であれば，それは明らかに定常ポアソン過程に帰着する．条件付き強度関数が過去の履歴とは独立だが生起時刻 t には依存している場合は， t に関する任意の非負関数 $\nu(t)$ を使って $\lambda(t|\mathcal{F}_t) = \nu(t)$ と書ける場合だが，この場合は非定常ポアソン過程と呼ばれる．本報告で用いるのはこのタイプのモデルである．

他にも興味深い点過程のクラスと，それに対応する条件付き強度関数がある．Hawkes の自己励起過程 (self-exciting process)

$$\lambda(t|\mathcal{F}_t) = \mu + \int_0^t g(t-s) dN_s = \mu + \sum_{t_i < t} g(t-t_i)$$

はその一例である (Hawkes, 1971, Hawkes and Oakes, 1974)．この関数形は，時系列解析における自己回帰モデルを想起させる．このモデルでは，事象が起きるかどうかの期待値が過去の生起事象の線形結合になっており，いわゆるインパルス応答関数 $g(\cdot)$ がその重みを与える．

区間 $[0, T]$ 上で事象の生起時刻が t_1, t_2, \dots, t_n と観測されたとして，そのときパラメトリックに表現された条件付き強度関数 $\lambda_\theta(t|\mathcal{F}_t)$ が与えられたとしよう．このとき尤度関数は以下の形を取る．

$$L_T(\theta | t_1, t_2, \dots, t_n; 0, T) = \left\{ \prod_{i=1}^n \lambda_\theta(t_i | \mathcal{F}_{t_i}) \right\} \exp \left\{ - \int_0^T \lambda_\theta(t | \mathcal{F}_t) dt \right\} \quad (2)$$

パラメータ θ の最尤推定値は，尤度関数の対数値

$$\log L_T(\theta | t_1, t_2, \dots, t_n; 0, T) = \sum_{i=1}^n \log \lambda_\theta(t_i | \mathcal{F}_{t_i}) - \int_0^T \lambda_\theta(t | \mathcal{F}_t) dt \quad (3)$$

を最大化することで得られる．

(3) 式右辺第 2 項が θ に関して解析的に表現されていれば，対数尤度関数の勾配微分は容易に得られる．そのような場合には，対数尤度関数の最大化は，標準的な非線形最適化の方法で実行できる．

複数の競合するモデルからひとつに絞らなければならないときには，赤池情報量規準 (Akaike, 1974)

$$\text{AIC} = -2 \times (\text{maximum log-likelihood}) + 2 \times (\text{number of parameters})$$

を最小化するモデルを選択する。

3.2 点過程の周期性モデリング

非定常ポアソン過程を想定し、以下のように条件付き強度関数を特定化する。

$$\lambda_{\theta}(t|\mathcal{F}_t) = a_0 + P_J(t) + C_K(t). \quad (4)$$

(4) 式の第 2 項は発展型のトレンドモデルとして

$$P_J(t) = \sum_{j=1}^J a_j \phi_j(t/T), \quad 0 < t < T \quad (5)$$

という定式化を与える。ここで T は観測の全区間であり、 $\phi_j(\cdot)$ は j 次の多項式である。(4) 式第 3 項は周期性をモデル化したものであり、フーリエ級数により

$$C_K(t) = \sum_{k=1}^K \{b_{2k-1} \cos(2k\pi t/T_0) + b_{2k} \sin(2k\pi t/T_0)\}, \quad (6)$$

という定式化を与える。ここで T_0 は分析に先立って特定化する、現象の周期であり、例えば一日の長さに相当しているとすれば、この周期性モデルは日内周期性を表現していることになる。

(4) 式の代わりに、これらのモデルを指数の肩に乗せたモデルを考えることがしばしば有益である。

$$\lambda_{\theta}(t|\mathcal{F}_t) = \exp\{a_0 + P_J(t) + C_K(t)\} \quad (7)$$

これにより、条件付き強度関数の正値性が保証される。条件付き強度関数を対数線形型にモデリングしていると考えて良い。この場合は、積分項 ((3) 式の右辺第 2 項) の解析的評価は、特殊な場合を除いて不可能である。例えば Lewis (1970) を参照。一般にはこの項を数値積分で評価することになる。非定常ポアソンで強度関数が例えば指数レートでしか緩慢変動しない場合には最尤推定は実行可能であることが知られており (MacLean 1974)、ソフトウェア実装としては Ogata and Katsura (1985) の EPTREN がある。

4 実証分析

4.1 金先物の結果

分析の対象期間は、2006 年 10 月 2 日から 2011 年 4 月 28 日までとし、これを立会時間の違いにより 4 分割し、

- 期間 A : 2006 年 10 月 2 日から 2008 年 1 月 4 日まで
 - 立会時間は 9:00-11:00 と 12:30-15:30
- 期間 B: 2008 年 1 月 7 日から 2009 年 5 月 1 日まで
 - 立会時間は 9:00-11:00 と 12:30-17:30
- 期間 C : 2009 年 5 月 7 日から 2010 年 9 月 17 日まで
 - 立会時間は 9:00-15:30 と 17:00-23:00
- 期間 D: 2010 年 9 月 21 日から 2011 年 4 月 28 日まで

- 立会時間は 9:00-15:30 と 17:00-翌日 4:00

とする。期間 A と期間 B の違いを EPTREN による推計結果 (図 1 のパネル A と B) を用いて比較すると、前場の寄り付きが最も強度の高い時間帯となり、次いで後場の引けとなっている。そして多くの先行研究で指摘されているように、昼休みの前後でも強度が高くなり、強度のグラフは W 字型になっている。後場の推移をより詳細に見ると、期間 A では 15 時前後が日中で最も強度の弱い時間帯となつてから急速に強度が引けまで上昇するが、期間 B では強度の最も低くなる時間帯はやや早まり、引けにかけて緩やかに上昇するパターンとなっている。

期間 C (図 1 のパネル C) では、それ以前と比べて昼休みがなくなるという大きな制度変更がなされているが、その結果、正午前後と夜間の 20 時前後に強度が低い時間帯が現れ、それまでと比べて、横長の W 字型となっている。そして、日中立会の寄り付きが最も強度が高い時間帯となることは変わらず、その次に強度が高くなる時間帯が日中の引けである点も同様であり、夜間立会は、一般的に日中立会よりも強度が低いので、W 字型よりも、V_v 型という表現がより正確である。

最後に期間 D (図 1 のパネル D) の推定結果を見ると、強度が最大となる点が 15:30 の日中立会の引けの時刻となる点が、それ以前の A から C の全期間と大きく異なる点である。夜間立会に関しては、20 時前後の強度が最も低くなる点は期間 C と同様で、その後 23 時頃まで強度が上昇する点も同様であるが、それ以降は 4 時にかけて徐々に強度が低くなる点も、それまで引けにかけては強度が上昇していたことと比べて大きな違いがある。

以上のように、日内季節変動のパターンは、特に期間 D において大きく変動したことがわかった。

なお、ここでは推定された強度の縦軸は基準化していない。従って、強度関数の面積は当該標本期間に生じた事象の総数に等しい。グラフは制度改変に対応して 4 つの期間に分けられているが、それぞれ事象総数は A:743,447, B:1,257,597, C:2,032,120, D:987,151 である。

4.2 日経平均先物の結果

日経平均先物に関しては、立会時間の延長 (の有無) によって以下の 6 期間に分けて分析している。金先物同様、強度を表す縦軸は事象数などで基準化はしていない。 n は各期間での事象数である。

- 期間 A: 2006 年 9 月 7 日 ~ 2007 年 9 月 14 日
 - イブニングセッション導入以前, $n = 1,168,040$
- 期間 B: 2007 年 9 月 18 日 ~ 2008 年 10 月 10 日
 - 16:30-19:00 のイブニングセッション始まる, $n = 1,560,363$
- 期間 C: 2008 年 10 月 14 日 ~ 2010 年 7 月 16 日
 - 16:30-20:00 にイブニングセッション延長, $n = 2,008,698$
- 期間 D: 2010 年 7 月 20 日 ~ 2011 年 2 月 10 日
 - 16:30-23:30 にイブニングセッション延長, $n = 487,672$
- 期間 E: 2011 年 2 月 14 日 ~ 2011 年 7 月 15 日
 - 日中立会昼休み廃止, $n = 399,835$
- 期間 F: 2011 年 7 月 19 日 ~ 2012 年 12 月 12 日
 - 16:30-03:00 のナイトセッションに, $n = 1,006,938$

結果は図 2 の 6 枚のパネルに示されている。期間 4 の取引が前場の終わりに集中しており、ここだけ突出して強度が高い。

5 まとめ

実証分析の結果, 条件付き強度関数のモデリングである EPTREN を用いて, 先物市場での取引頻度の日中の推移を検証できることが確認された。夜間取引そのものの取引量は多くないが朝方の寄り付きパターンには影響を与えている可能性があることと, 昼休みの廃止が大きく影響していることが示唆された。リプライズド・ボラティリティの影響などは今後の課題である。

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] Barclay Michael J., and Hendershott, Terrence, Price Discovery and Trading After Hours, *Review of Financial Studies*, 16(4), 1041–1071, 2003.
- [4] Breedon, Francis and Ranaldo, Angelo, Intraday Patterns in FX Returns and Order Flow (April 3, 2012). Queen Mary, University of London, School of Economics and Finance WP 694. Available at SSRN: <http://ssrn.com/abstract=2099321> or <http://dx.doi.org/10.2139/ssrn.2099321>
- [5] Cooper, Michael J., Cliff, Michael T. and Gulen, Huseyin, Return Differences between Trading and Non-Trading Hours: Like Night and Day (September 26, 2008). Available at SSRN: <http://ssrn.com/abstract=1004081> or <http://dx.doi.org/10.2139/ssrn.1004081>
- [6] Hawkes, A. G., Point spectra of some mutually exciting point processes, *J. Roy. Statist. Soc. (B)*, **33**, 438–443, 1971.
- [7] Hawkes, A. G. and D. A. Oakes, A cluster process representations of self-exciting process, *J. Appl. Probab.*, **11**, 493–503, 1974.
- [8] 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.
- [9] Liptzer, R. S. and A. N. Shiriyayev, *Statistics of Random Processes II : Applications*, Springer-Verlag, New York, 1978.
- [10] Lopez de Prado, Marcos, Advances in High Frequency Strategies (December 1, 2011). Doctoral Dissertation, Complutense University, Madrid, 2011. Available at SSRN: <http://ssrn.com/abstract=2106117> or <http://dx.doi.org/10.2139/ssrn.2106117>
- [11] MacLean, C. J., Estimation and testing of an exponential polynomial rate function within the non-stationary Poisson process, *Biometrika*, **61**, pp. 81–86, 1974.
- [12] マクニール・A, フライ・R, エンブレヒツ・P, 「定量的リスク管理-基礎概念と数理技法-」(塚原英敦・小林俊・三浦良造・川崎能典・中川秀敏 訳), 共立出版, 2008.
- [13] Merton H. Miller, International Competitiveness of U.S. Futures Exchanges, *Journal of Applied Corporate Finance*, Volume 3, Issue 4, Winter 1991, Pages: 6–19, DOI: 10.1111/j.1745-6622.1991.tb00559.x

- [14] 森保洋, 金先物市場のマイクロストラクチャー-ティック・データを利用した実証分析-, 日本商品先物振興協会 先物取引研究, 第 11 巻第 1 号 No.15, pp. 1-19, 2008.
- [15] 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.
- [16] Ogata, Y., Seismicity analysis through point-process modeling: a review, *Pure and Applied Geophysics*, **155**, 471-507, 1999.
- [17] 太田亘, 取引システム高速化の流動性に対する長期的影響, 日本ファイナンス学会 第 21 回大会予稿集, 2013.
- [18] 芹田敏夫, 濱田隆道, 荒木浩介, 坂本智幸, 金先物市場の日中の価格形成と流動性: 取引データに基づく実証研究, 東京工業品取引所 市場構造研究所ワーキングペーパーシリーズ, 05(1), 2005.
- [19] Takahashi, Makoto, Yasuhiro Omori, and Toshiaki Watanabe, Estimating Stochastic Volatility Models Using Daily Returns and Realized Volatility Simultaneously, *Computational Statistics and Data Analysis*, **53**, 2404-2426, 2009.
- [20] Treepongkaruna, Sirimon, Brooks, Robert Darren and Gray, Stephen, Do Trading Hours Affect Volatility Links in the Foreign Exchange Market? (April 4, 2012). *Australian Journal of Management*, Vol. 37, No. 1, 2012. Available at SSRN: <http://ssrn.com/abstract=2050440>
- [21] 生方雅人, 渡部敏明, 実現ボラティリティ - ボラティリティの計測方法の発展とリスクマネジメントへの応用可能性 -, 証券アナリストジャーナル, **49**(8), 16-26, 2011.
- [22] 宇野淳, 柴田舞, 取引の高速化と流動性へのインパクト: 東証アローヘッドのケース. *現代ファイナンス* **31**, 87-107, 2012.
- [23] 渡部敏明, 日経 225 先物の価格および取引高の日中の変動パターン, 先物・オプションレポート, 大阪証券取引所, **16**(12), 2004.
- [24] 渡部敏明, 日経 225 の Realized Volatility - マイクロストラクチャ・ノイズと夜間・昼休みの調整 -, 先物・オプションレポート, 大阪証券取引所, **22**(2), 2010.

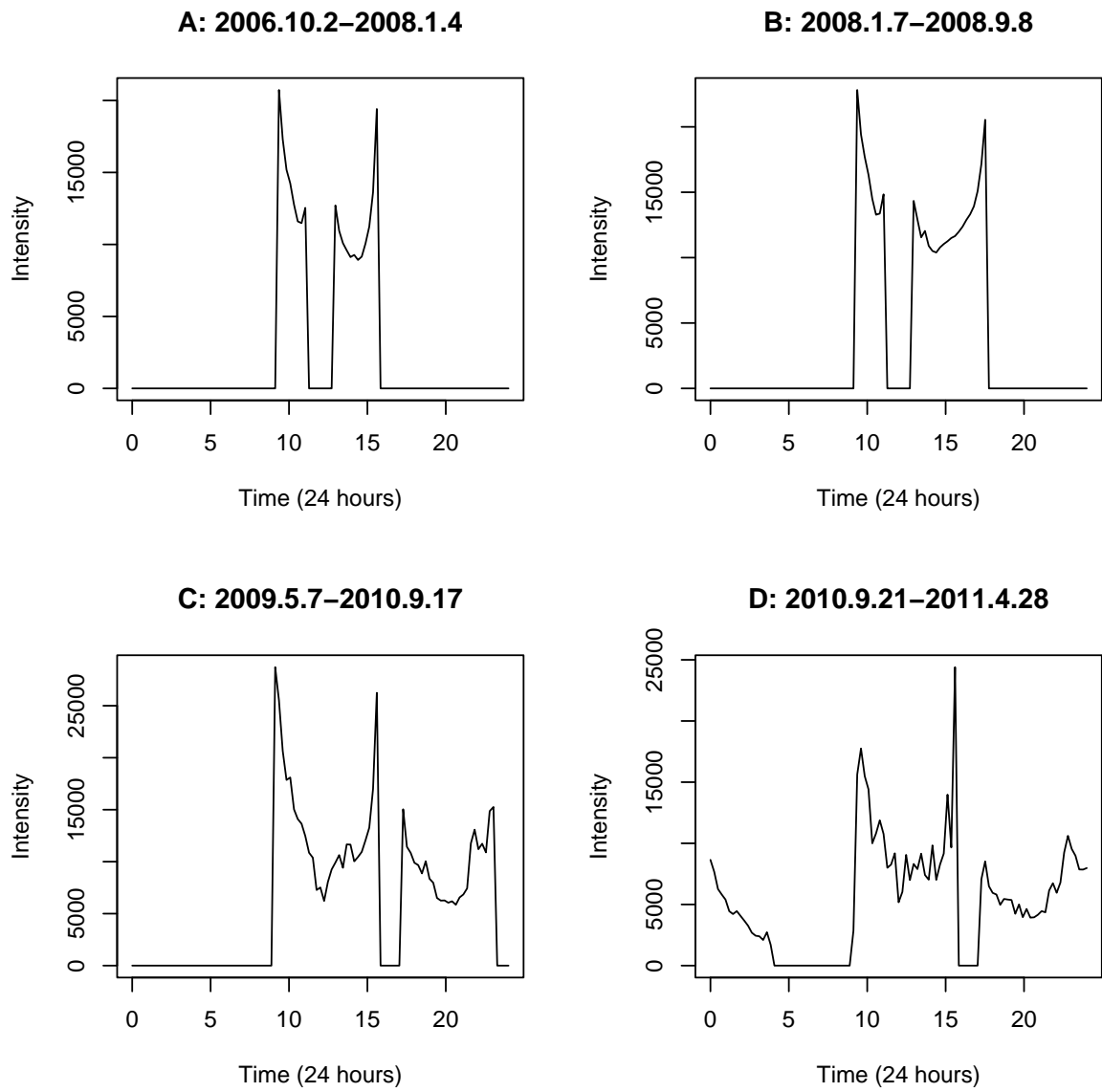


図 1: 点過程モデルから推定した金先物の取引強度

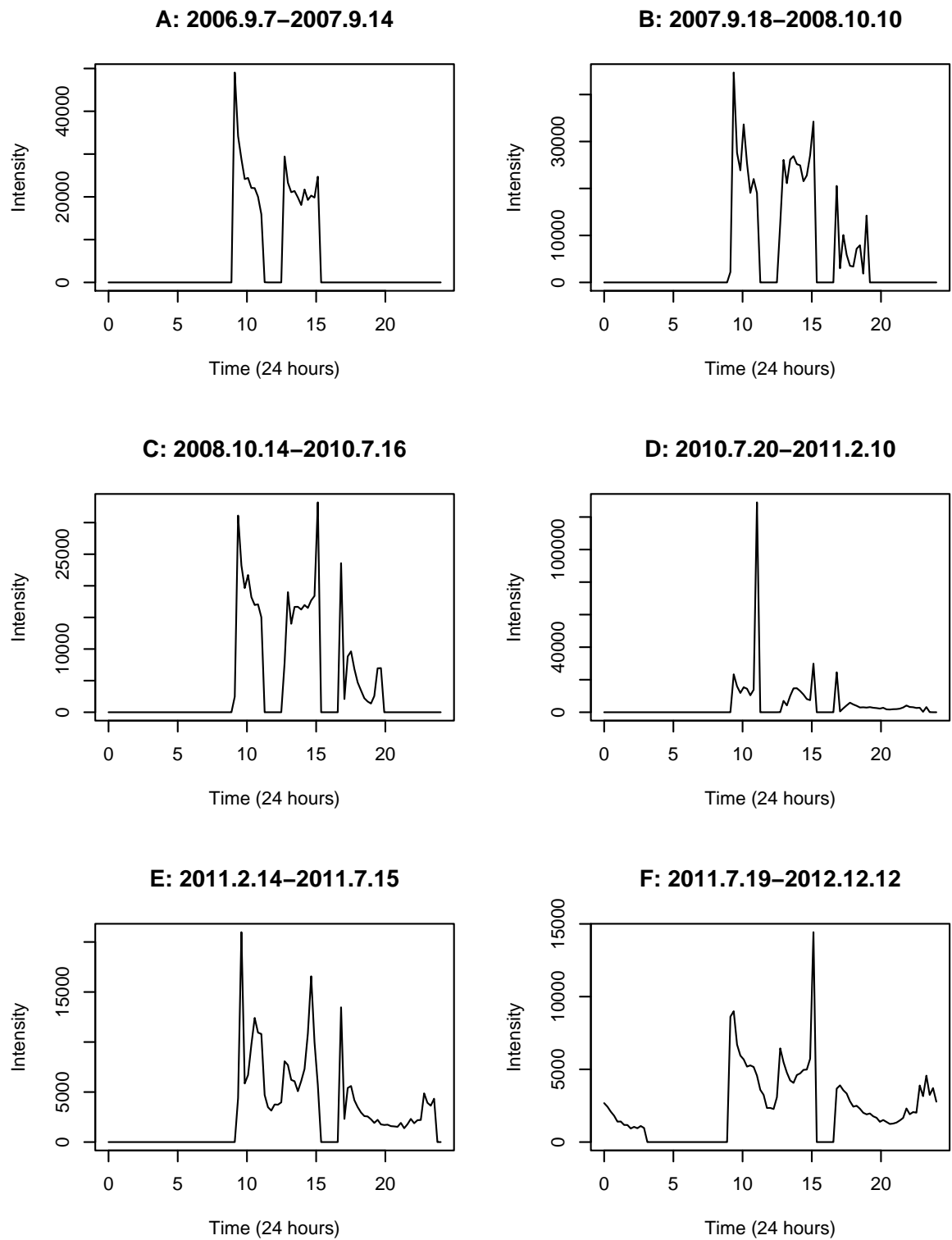


図 2: 点過程モデルから推定した日経平均先物の取引強度

Improving the Separating Information Maximum Likelihood Method for Continuous Diffusion Processes with Micro-Market Noise *

Naoto Kunitomo [†]

December 10, 2013

Abstract

For estimating the integrated covariances of continuous time diffusion process with micro-market noise, Kunitomo and Sato (2008, 2013) have proposed the Separating Information Maximum Likelihood (SIML) method by using high frequency financial data. We can improve the SIML method such that the modified SIML (MSIML) method is asymptotically optimal in a sense while it has the asymptotic robustness when the sample size is large. We investigate the effects of market adjustments (autocorrelated noises), round-off errors, and random sampling. We find that the MSIML estimator has reasonable finite sample properties and thus it would be useful for practice.

Key Words

Continuous Diffusion Process, Micro-Market Noise, High-Frequency Data, Separating Information Maximum Likelihood (SIML), Modified SIML (MSIML), Optimal Rate.

*This is a very preliminary manuscript. I thank Seisho Sato for providing some numerical examples.

[†]Graduate School of Economics, University of Tokyo, Bunkyo-ku, Hongo 7-3-1, Tokyo 113-0033, JAPAN, kunitomo@e.u-tokyo.ac.jp, phone: +81-3-5841-5614.

1. Introduction

Recently a considerable interest has been paid on the estimation problem of the continuous time diffusion processes and their relationships. It is partly because there are many theoretical use of diffusion processes in the area of mathematical finance. Since it is possible to use a large number of high-frequency data in financial markets, the estimation of the continuous time diffusion processes and their relationships have potentially many applications in practice. Although there were some discussion on the estimation of continuous stochastic processes in the statistical literature, the earlier studies often had ignored the presence of micro-market noises in financial markets when they tried to estimate the underlying stochastic processes. Because there are several reasons why the micro-market noises are important in high-frequency financial data both in economic theory and in statistical measurement, several new statistical estimation methods have been developed.

The main purpose of this paper is to propose a way to improve the Separating Information Maximum Likelihood (SIML) method for estimating the continuous time diffusion process by using high frequency data in the presence of possible micro-market noise. We shall call the resulting one as the MSIML (modified SIML) method because the SIML method was originally proposed by Kunitomo and Sato (2008, 2011) and its properties has been investigated by Kunitomo and Sato

(2013), Kunitomo and Misaki (2013). In this paper we shall show that the MSIML method estimator has reasonable asymptotic properties in the sense that it is asymptotically optimal when the sample size is large and it improves the finite sample properties of the SIML method under quite general situations.

The main motivation of our study is two-fold. First, the SIML method has reasonable asymptotic properties, but generally it does not attain the optimal rate in the ideal situation. Hence it may be important to improve the SIML and also understand the underlying main reason why the original SIML estimator does not attain the optimal convergence rate. Secondly, the variance-covariances of micro-market noise are also important because they cause important effects and thus the estimation of their magnitude gives key information on the underlying process. Because it is difficult to observe the micro-market noises, the assumption of i.i.d. random variables may be often too strong in the view of micro-market structure. In this paper we shall investigate the effects for estimating the variance and covariances of noises when they are weakly auto-correlated.

2. The MSIML estimation of the diffusion process with micro-market noise

2.1 The statistical models in continuous time and discrete time

Let $y_s(t_i^s)$ be the i -th observation of the (log-) price of the first asset at t_i^s for $0 = t_0^s < t_1^s < \dots < t_{n_s^*}^s \leq t_{n_s}^s = 1$ and $y_f(t_j^f)$ be the j -th observation of the (log-) price of the second asset at t_j^f for $0 = t_0^f < t_1^f < \dots < t_{n_f^*}^f \leq t_{n_f}^f = 1$. Let $t_{n_s^*}^s = \max_{i \geq 1, t_i^s \leq 1} t_i^s$, $t_{n_f^*}^f = \max_{i \geq 1, t_i^f \leq 1} t_i^f$ and we denote n as a constant index and n^* as a stochastic index. We consider the situation that the high-frequency data are observed at random times t_i^a ($a = s$ or f) under some conditions on random sampling.

Assumption 2.1 : There exist positive constants c_a ($a = s$ or f) such that

$$(2.1) \quad t_n^a \longrightarrow 1, \quad \frac{n_a^*}{n} \xrightarrow{p} c_a$$

and

$$(2.2) \quad \mathcal{E} [|t_i^a - t_{i-1}^a|] = O(n^{-1})$$

as $n \rightarrow \infty$, where $a = s$ or f .

These conditions imply that n^{-1} corresponds to the average duration of observations of the intervals in $[0, 1]$ when n is relatively large. Without loss of generality we take $c_s = c_f = 1$.

A typical example is the Poisson Process Sampling on t_i^s and t_i^f with the intensity functions $\lambda_n^{(s)} = nc_s$ and $\lambda_n^{(f)} = nc_f$. In this case the sequence of random variables τ_i^a ($a = s, f$) are exponentially distributed with $\mathcal{E}(\tau_i^a) = 1/n$ ($\tau_i^a = t_i^a - t_{i-1}^a$) if we take $c_s = c_f = 1$. In this

paper we make a further assumption on the independence of $X(t)$ and t_i^a ($i \geq 1$).

Assumption 2.2 : The stochastic process $X(t)$ ($0 \leq t \leq 1$) is independent of the random sequences t_i^s and t_j^f ($i, j \geq 1$).

The underlying two-dimensional continuous process $\mathbf{X}(t) = (X_s(t), X_f(t))'$ ($0 \leq t \leq 1$) is not necessarily the same as the observed (log-)price at t_i^s and t_j^f ($i, j \geq 1$) and

$$(2.3) \quad \mathbf{X}(t) = \mathbf{X}(0) + \int_0^t \mathbf{C}_x(s) d\mathbf{B}(s) \quad (0 \leq t \leq 1),$$

where $\mathbf{B}(s)$ is the two-dimensional Brownian motion, $\mathbf{C}_x(s)$ is the 2×2 instantaneous volatility matrix adapted to the σ -field $\mathcal{F}(\mathbf{x}(r), \mathbf{B}(r), r \leq s)$. The main statistical objective is to estimate the quadratic variation or the integrated volatility matrix

$$(2.4) \quad \Sigma_x = \int_0^1 \Sigma_x(s) ds = \begin{pmatrix} \sigma_{ss}^{(x)} & \sigma_{sf}^{(x)} \\ \sigma_{sf}^{(x)} & \sigma_{ff}^{(x)} \end{pmatrix}$$

($\Sigma_x(s) = \mathbf{C}_x(s)\mathbf{C}_x'(s)$) of the underlying continuous process $\mathbf{X}(t)$ ($0 \leq t \leq 1$) from the set of discrete observations on $(y_s(t_i^s), y_f(t_j^f))$ with the condition that $\Sigma_x(s)$ is a progressively measurable matrix and $\sup_{0 \leq s \leq 1} \Sigma_x(s) < \infty$ (*a.s.*).

We also consider the situation that the observed (log-)prices $y_s(t_i^s)$ and $y_f(t_j^f)$ are the sequence of discrete stochastic processes generated by

$$(2.5) \quad y_s(t_i^s) = h_s(\mathbf{X}(t_i^s), y_s(t_{i-1}^s), u_s(t_i^s))$$

and

$$(2.6) \quad y_f(t_j^f) = h_f \left(\mathbf{X}(t_j^f), y_f(t_{j-1}^f), u_f(t_j^f) \right) ,$$

where $h_s(\cdot)$ and $h_f(\cdot)$ are measurable functions, the (unobservable) continuous martingale process $\mathbf{X}(t)$ ($0 \leq t \leq 1$) is defined by (2.3) and the micro-market noises $u_s(t_i^s)$ and $u_f(t_j^f)$ are the discrete stochastic processes. In particular, we assume that $u_s(t_i^s)$ and $u_f(t_j^f)$ are a sequence of independently and identically distributed random variables with $\mathcal{E}(u_s(t_i^s)) = 0$, $\mathcal{E}(u_f(t_j^f)) = 0$ and $\mathcal{E}(u_s(t_i^s)^2) = \sigma_{ss}^{(u)}$, $\mathcal{E}(u_f(t_j^f)^2) = \sigma_{ff}^{(u)}$, $\mathcal{E}(u_s(t_i^s)u_f(t_j^f)) = \delta(t_i^s, t_j^f)\sigma_{sf}^{(u)}$.

There are special cases of (2.3), (2.5) and (2.6), which reflect the important aspects on modeling financial markets and the high frequency financial data. The basic (high-frequency) financial model with micro market noises can be represented by

$$(2.7) \quad y_s(t_i^s) = X_s(t_i^s) + u_s(t_i^s) , \quad y_f(t_j^f) = X_f(t_j^f) + u_f(t_j^f) ,$$

where the underlying process $\mathbf{X}(t) = (X_s(t), X_f(t))'$ is given by (2.3). The synchronous sampling means $t_i^s = t_i^f$ and the fixed grid observation means $t_i^a - t_{i-1}^a = n^{-1}$. We shall consider the more general situations, that is, we have the non-synchronous observations as well as the random sampling.

The most important statistical aspect of (2.7) is the fact that it is an additive (signal-plus-noise) measurement error model. However, there are some reasons why the standard situation as (2.7) is not

enough for applications. For instance, the high frequency financial models for micro-market price adjustments and the round-off-errors models for financial prices are not in the form of (2.7), but they can be represented as special cases of (2.3), (2.5) and (2.6). Sato and Kunitomo (2012) have discussed several important examples of (2.5) and (2.6) when the state variable is one dimension.

More generally, it is straightforward to extend our analysis to the cases when the observations are the p -dimensional vector value process $y_j(t_i^{(j)})$, $j = 1, \dots, p$. The model we have discussed has been the case when $p = 2$.

2.2 The MSIML estimation

We consider the situation when \mathbf{x}_i and \mathbf{v}_i ($i = 1, \dots, n$) are independent with $\Sigma_x(s) = \Sigma_x$ ($0 \leq s \leq 1$), and \mathbf{v}_i are independently, identically and normally distributed as $N_p(\mathbf{0}, \Sigma_v)$. We use an $n \times p$ matrix $\mathbf{Y} = (\mathbf{y}'_i)$ and consider the distribution of $np \times 1$ random vector $(\mathbf{y}'_1, \dots, \mathbf{y}'_n)'$. Given the initial condition \mathbf{y}_0 , we have

$$(2.8) \quad \mathbf{Y}_n \sim N_{n \times p}(\mathbf{1}_n \cdot \mathbf{y}'_0, \mathbf{I}_n \otimes \Sigma_v + \mathbf{C}_n \mathbf{C}'_n \otimes h_n \Sigma_x) ,$$

where $\mathbf{1}'_n = (1, \dots, 1)$, $h_n = 1/n (= t_i^n - t_{i-1}^n)$ and

$$(2.9) \quad \mathbf{C}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ 1 & \cdots & 1 & 1 & 0 \\ 1 & \cdots & 1 & 1 & 1 \end{pmatrix}.$$

We transform \mathbf{Y}_n to $\mathbf{Z}_n (= (\mathbf{z}'_k))$ by

$$(2.10) \quad \mathbf{Z}_n = h_n^{-1/2} \mathbf{P}_n \mathbf{C}_n^{-1} (\mathbf{Y}_n - \bar{\mathbf{Y}}_0)$$

where

$$(2.11) \quad \mathbf{C}_n^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix},$$

$$(2.12) \quad \mathbf{P}_n = (p_{jk}), \quad p_{jk} = \sqrt{\frac{2}{n + \frac{1}{2}}} \cos \left[\frac{2\pi}{2n + 1} \left(k - \frac{1}{2} \right) \left(j - \frac{1}{2} \right) \right]$$

and $\bar{\mathbf{Y}}_0 = \mathbf{1}_n \cdot \mathbf{y}'_0$. We have the spectral decomposition $\mathbf{C}_n^{-1} \mathbf{C}_n'^{-1} = \mathbf{P}_n \mathbf{D}_n \mathbf{P}_n'$ and \mathbf{D}_n is a diagonal matrix with the k-th element

$$(2.13) \quad a_{kn} = 2 \left[1 - \cos \left(\pi \left(\frac{2k-1}{2n+1} \right) \right) \right] = 4n \sin^2 \left[\frac{\pi}{2} \left(\frac{2k-1}{2n+1} \right) \right].$$

Then given the initial condition \mathbf{y}_0 the likelihood function can be defined as

$$(2.14) L_n = \sum_{k=1}^n \log |a_{kn} \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_x|^{-1/2} - \frac{1}{2} \sum_{k=1}^n \mathbf{z}'_k [a_{kn} \boldsymbol{\Sigma}_v + \boldsymbol{\Sigma}_x]^{-1} \mathbf{z}_k.$$

Since $a_{k_n, n} \rightarrow 0$ as $n \rightarrow \infty$ when $k_n = O(n^\alpha)$ ($0 < \alpha < \frac{1}{2}$) and $a_{n+1-l_n, n} = O(n)$ when $l_n = O(n^\beta)$ ($0 < \beta < 1$). Then we may approximate $2 \times L_n$ by $2 \times L_{m_n}$ and the separating information maximum likelihood (SIML) estimator of $\hat{\Sigma}_x$ is defined by

$$(2.15) \quad \hat{\Sigma}_x = \frac{1}{m_n} \sum_{k=1}^{m_n} \mathbf{z}_k \mathbf{z}_k'$$

and the SIML estimator of Σ_v is defined by

$$(2.16) \quad \hat{\Sigma}_v = \frac{1}{l_n} \sum_{k=n+1-l_n}^n a_{kn}^{-1} \mathbf{z}_k \mathbf{z}_k'.$$

For both $\hat{\Sigma}_v$ and $\hat{\Sigma}_x$, the number of terms m_n and l_n are dependent on n and we have 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.

In order to improve the SIML method, we notice that the (1st order) asymptotic bias term of $\hat{\Sigma}_x$ is given by

$$(2.17) \quad \text{ABIAS} = \left(\frac{1}{m_n} \sum_{k=1}^{m_n} a_{kn} \right) \Sigma_v + o_p\left(\frac{m_n^2}{n}\right).$$

In this paper we propose to modify the SIML method and to use the MSIML estimator of Σ_v by

$$(2.18) \quad \hat{\Sigma}_v = \frac{1}{l_n} \sum_{k=n_1+1}^{n_2} [a_{kn}^{-1} \mathbf{z}_k \mathbf{z}_k']$$

where $l_n = n_2 - n_1$ and $n_1 = n^{\beta_1}, n_2 = n^{\beta_2}$ ($\frac{1}{2} < \beta_1 < \beta_2 < 1$).

Then given (2.14) the modified SIML (MSIML) estimator of $\hat{\Sigma}_x$ is defined by

$$(2.19) \quad \hat{\Sigma}_{x,m} = \left[\frac{1}{m_n} \sum_{k=1}^{m_n} \mathbf{z}_k \mathbf{z}_k' \right] - \left[\frac{1}{m_n} \sum_{k=1}^{m_n} a_{kn} \right] \hat{\Sigma}_v.$$

2.3 Asymptotic Properties of the MSIML estimator

The asymptotic properties of the SIML estimator have been investigated by Kunitomo and Sato (2008, 2011) on the estimation problem of the integrated volatility and integrated covariances. For the simplicity, we take the case of $p = 1$ and we consider the estimation problem of the variance of micro-market noise. It seems that there have not been any clear statement on the asymptotic properties of alternative estimators for the noise variance.

We have the following result.

Theorem 2.1 : We assume that $X(t_i^n)$ and $v_i = v(t_i^n)$ ($i = 1, \dots, n^*$) are given by (2.1) and (2.2) with $\sup_{0 \leq s \leq 1} \sigma_x^2(s) < \infty$.

For $1/2 < \beta_1, \beta_2 < 1$ and $0 < \alpha < 0.5$, as $n \rightarrow \infty$,

$$(2.20) \quad \sqrt{l_n} [\hat{\sigma}_v^2 - \sigma_v^2] \xrightarrow{d} N[0, V] ,$$

where

$$(2.21) \quad V = 2\sigma^4 .$$

For the deterministic time varying volatility case the asymptotic properties of the MSIML estimator can be summarized as the next proposition.

Theorem 2.2 : We assume that $X(t_i^n)$ and $v_i = v(t_i^n)$ ($i = 1, \dots, n^*$) are given by (2.1) and (2.2) with $\sup_{0 \leq s \leq 1} \sigma_x^2(s) < \infty$ and $\sigma_x^2 = \int_0^1 \sigma_x^2(s) ds$ is a positive constant (or deterministic). Define the MSIML estimator of σ_x^2 by (2.19).

For $m_n = n^\alpha$ and $0 < \alpha < 0.5$, as $n \rightarrow \infty$

$$(2.22) \quad \sqrt{m_n} [\hat{\sigma}_{x,m}^2 - \sigma_x^2] \xrightarrow{d} N[0, V] ,$$

where

$$(2.23) \quad V = 2 \int_0^1 [\sigma_x^2(s)]^2 ds .$$

When σ_x^2 is a random variable, we need the concept of *stable convergence in law* because the limiting distribution of the SIML estimator is the mixed-Gaussian distribution. In order to discuss the stable convergence in law we extend the probability space (Ω, \mathcal{F}, P) to the extended probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P})$ as explained by Chapter VIII of Jacod and Syriyaev (2003) or Jacod (2007). We say that a sequence of random variables Z_n with an index n stably converges in law if

$$(2.24) \quad \mathbf{E}[Y f(Z_n)] \rightarrow \tilde{\mathbf{E}}[Y f(Z)]$$

for all bounded continuous functions $f(\cdot)$ and all bounded random variables Y , and $\tilde{\mathbf{E}}[\cdot]$ is the expectation operator with respect to the extended probability space. We write this convergence as

$$(2.25) \quad Z_n \xrightarrow{\mathcal{L}\text{-}s} Z .$$

(See Jacod and Shriyaev (2003) and Jacod (2007) for the details.)

As a typical stochastic volatility case in the continuous time, we consider that the volatility function $\sigma_x(t)$ is a strong solution of the stochastic differential equation (SDE)

$$(2.26) \quad \begin{aligned} \sigma_x(t) = & \sigma_x(0) + \int_0^t \mu_\sigma(s, \sigma_x(s)) ds + \int_0^t \gamma_\sigma(s, \sigma_x(s)) dB(s) \\ & + \int_0^t \gamma_\sigma^*(s, \sigma_x(s)) dB^*(s) , \end{aligned}$$

where the coefficients $\mu_\sigma(s), \gamma_\sigma(s)$ and $\gamma_\sigma^*(s)$ are in the class of \mathcal{A}^1 (extensively measurable, continuous and bounded), and $B^*(s)$ is a Brownian motion which is orthogonal to $B(s)$. Here we set $\mathbf{B}(t) = (B(t), B(t)^*)'$ as the vector of Brownian motions. Then there exists a strong solution such that $\sup_{0 \leq s \leq 1} \mathcal{E}[\sigma_x^4(s)] < \infty$. (There can be weaker conditions on the coefficients which give the existence of a strong solution and the moment conditions. See Chapter III of Ikeda and Watanabe (1989) for the notations and the details.)

The asymptotic properties of the MSIML estimator in the stochastic volatility cases can be summarized as Theorem 2.3.

Theorem 2.3 : We assume that $X(t_i^n)$ and $v_i = v(t_i^n)$ ($i = 1, \dots, n^*$) are given by (2.1) and (2.2) with $\sup_{0 \leq s \leq 1} \sigma_x^2(s) < \infty$ and $\sigma_x^2 = \int_0^1 \sigma_x^2(s) ds (> 0)$ is finite (a.s.). We assume that $\mathcal{E}[v(t_i^n)^4] < \infty$. For $m_n = n^\alpha$ and $0 < \alpha < 0.5$, as $n \rightarrow \infty$ we have the weak convergence

$$(2.27) \quad Z_{n^*} = \sqrt{m_n} [\hat{\sigma}_{x,m}^2 - \sigma_x^2] \xrightarrow{w} Z^* ,$$

where the characteristic function $g_n(t) = \mathcal{E}[\exp(itZ_{n^*})]$ of Z_{n^*} converges to the characteristic function of Z^* , which is written as

$$(2.28) \quad g(t) = \mathcal{E}[e^{-\frac{Vt^2}{2}}] ,$$

where

$$(2.29) \quad V = 2 \int_0^1 [\sigma_x^2(s)]^2 ds .$$

2.4 Generalizations

It is straightforward to extend our analysis in the previous section to the p dimension cases ($p \geq 1$). Another direction to extend our analysis would be to assume that the $p \times 1$ vector noise process $\{v_i\}$ is stationary process which can be represented as

$$(2.30) \quad v_i = \sum_{s=-\infty}^{\infty} \gamma_s w_{i-s} \quad (i = \dots, -1, 0, 1, \dots),$$

where $w_j = (w_{ij})$ are the vector sequence of independent random variables with $\mathcal{E}(w_{ij}) = 0$, $\mathcal{E}(w_{ij}^2) = \sigma_{jj}$, $\mathcal{E}(w_{li}w_{kj}) = 0$ ($i \neq j$), $\mathcal{E}(w_{ij}^4) < \infty$ and $\sum_{-\infty}^{\infty} \|\gamma_s\|^2 < \infty$.

3. Simulations

We have investigated the robustness properties of the MSIML estimator for the integrated volatility based on a set of simulations and the number of replications is 1,000. We have taken the sample size $n = 20,000$, and we have chosen $\alpha = 0.5$ and $n_1 = n^{c_1}$, $n_2 = n - n^{c_2}$ ($c_1 = 0.85$, $c_2 = 0.66$). The other details of the simulation procedure are similar to the corresponding ones reported by Kunitomo and Sato (2008, 2011).

In our simulations we consider several cases when the observations are generated by (2.3) and (2.7) as the basic case. The volatility function ($\sigma_x^2(s)$) is given by

$$(3.1) \quad \sigma_x^2(s) = \sigma(0)^2 [a_0 + a_1s + a_2s^2],$$

where a_i ($i = 0, 1, 2$) are constants and we have some restrictions such that $\sigma_x(s)^2 > 0$ for $s \in [0, 1]$. It is a typical time varying (but deterministic) case and the integrated volatility σ_x^2 is given by

$$(3.2) \quad \sigma_x^2 = \int_0^1 \sigma_x(s)^2 ds = \sigma_x(0)^2 \left[a_0 + \frac{a_1}{2} + \frac{a_2}{3} \right] .$$

In this example we have taken several intra-day volatility patterns including the flat (or constant) volatility, the monotone (decreasing or increasing) movements and the U-shaped movements.

In our Monte-Carlo simulations, we also investigate the situation that the observed (log-)price $y(t_i^n)$ is a sequence of discrete stochastic process generated by

$$(3.3) \quad y(t_i^n) = h(X(t_i^n), y(t_{i-1}^n), v(t_i^n)) ,$$

where $h(\cdot)$ is a measurable function and the (unobservable) continuous martingale process $X(t)$ ($0 \leq t \leq 1$) is defined by (2.3) and $v(t_i^n)$ is the micro-market noise process. In Appendix we give some results and each model corresponds to the cases when we take $h(\cdot, \cdot, \cdot)$ as

$$\text{Model 1} \quad h_1(x, y, u) = y + g(x - y) + u \quad (g : \text{a constant}) ,$$

$$\text{Model 2} \quad h_2(x, y, u) = y + g_\eta(x - y + u) \quad (g_\eta(\cdot) \text{ is (2.7)}) ,$$

$$\text{Model 3} \quad h_3(x, y, u) = y + g_\eta(x - y) + u \quad (g_\eta(\cdot) \text{ is (2.7)}) ,$$

$$\text{Model 4} \quad h_4(x, y, u) = y + u + \begin{cases} g_1(x - y) & \text{if } y \geq 0 \quad (g_1 : \text{a constant}) \\ g_2(x - y) & \text{if } y < 0 \quad (g_2 : \text{a constant}) \end{cases} ,$$

$$\text{Model 5} \quad h_5(x, y, u) = y + [g_1 + g_2 \exp(-\gamma|x - y|^2)] (x - y)$$

$$(g_1, g_2 : \text{constants}) ,$$

respectively.

Model 1 is the basic additive model when $g = 1$. When $0 < g < 2$, Model 1 corresponds to the linear price adjustment model with the micro-market noise. Model 2 and Model 3 are the micro-market models with the round-off errors. Model 2 is the basic round-off errors model and Model 3 has a more complicated nonlinearity. Model 4 and Model 5 are the SSAR model and the exponential AR model, which have been known as nonlinear (discrete) time series models.

For a comparison we have calculated the historical volatility (HI) estimates. Overall the estimates of the MSIML method are quite stable and robust against the possible values of the variance ratio even in the nonlinear transformations we have considered.

For Model-1, the estimates obtained by historical-volatility (H-vol) are badly-biased, which have been known in the analysis of high frequency financial data. Actually, the values of H-vol are badly-biased in all cases of our simulations.

By examining these results of our simulations we conclude that we can estimate the integrated volatility of the hidden martingale part reasonably by the MSIML estimation method despite of the possible non-linear transformations. It may be surprising to find that the MSIML method gives reasonable estimates even when we have non-linear transformations of the original unobservable security (intrinsic) values. We have conducted a number of further simulations, but the results are quite similar as we have reported in this section.

4. Concluding Remarks

In the present study we propose a way to improve the statistical estimation method of the integrated volatility and covariances in the presence micro-market noises. We extend the Separating Information Maximum Likelihood (SIML) method proposed by Kunitomo and Sato (2008, 2011). We have shown that the modified SIML (MSIML) method has reasonable asymptotic properties; it is consistent and it has the asymptotic normality (or the stable convergence in the general case) and it is asymptotically optimal in a sense when the sample size is large and the data frequency interval is small under reasonable conditions. The MSIML estimator has reasonable finite sample properties and also it has the asymptotic robustness properties.

The MSIML estimator is so simple that it can be practically used not only for the integrated volatility but also the integrated covariances of the multivariate high frequency financial series and the hedging ratios. Further developments of applications will be discussed in other occasions.

References

- [1] Gloter, A. and J. Jacod (2001), “Diffusions with Measurement Errors. II : Optimal Estimators,” *ESAIM : Probability and Statistics*, Vol. 5, 243-260.

- [2] Ikeda, N. and S. Watanabe (1989), *Stochastic Differential Equations and Diffusion Processes*, 2nd Edition, North-Holland.
- [3] Jacod, J. (2007), "Statistics and high frequency data," Unpublished Manuscript, Paris.
- [4] Jacod, J. and A. N. Shiryaev (2003), *Limit Theorems for Stochastic Processes*, Springer.
- [5] Kunitomo, N. and S. Sato (2008), "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>), forthcoming in *North American Journal of Economics and Finance*.
- [6] Kunitomo, N. and S. Sato (2011), "The SIML Estimation of Realized Volatility of Nikkei-225 Futures and Hedging Coefficient with Micro-Market Noise," *Mathematics and Computers in Simulation*, Vol.81, 1272-1289.
- [7] Kunitomo, N. and H. Misaki (2013), "The SIML Estimation of Integrated Covariance and Hedging Coefficients with Micro-market noise and Random Sampling," Unpublished Manuscript.
- [8] Sato, S. and N. Kunitomo (2012), "A Robust Estimation of Integrated Volatility and covariance with Micro-market Adjustments and Round-off Errors," Unpublished Manuscript.

APPENDIX : TABLES

In Tables the estimates of the variance (σ_x^2) are calculated by the MSIML method while H-vol are calculated by the historical volatility estimation. The true-val means the true parameter value in simulations and mean, SD and MSE correspond to the sample mean, the sample standard deviation and the sample mean squared error of each estimators, respectively.

B-1 : Estimation of integrated volatility (Model-1)

$$(a_0 = 1, a_1 = 0, a_2 = 0; \sigma_v^2 = 1.00\text{E} - 04, g = 0.2)$$

n=20000	σ_x^2	σ_v^2	H-vol	$\sigma_{x,m}^2$
true-val	1	1.00E-04	1	1
mean	1.000648	6.93E-05	2.333708	1.000422
SD	0.118420	9.85E-07	0.023389	0.118420
MSE	0.014024	9.44E-10	1.779325	0.014024

B-2 : Estimation of integrated volatility (Model-1)

$$(a_0 = 1, a_1 = 0, a_2 = 0; \sigma_v^2 = 0.0, g = 0.2)$$

n=20000	σ_x^2	σ_v^2	H-vol	$\sigma_{x,m}^2$
true-val	1	0.00E+00	1	1
mean	0.992748	1.16E-06	0.1110989	0.992744
SD	0.111878	2.38E-08	0.0024040	0.111878
MSE	0.012569	1.35E-12	0.7901510	0.012569

B-3 : Estimation of integrated volatility (Model-1)

$$(a_0 = 1, a_1 = 0, a_2 = 0; \sigma_v^2 = 0.0, g = 1.5)$$

n=20000	σ_x^2	σ_v^2	H-vol	$\sigma_{x,m}^2$
true-val	1	0.00E+00	1	1
mean	1.003343	6.53E-05	2.99923	1.00313
SD	0.12221	7.87E-07	0.036817	0.12221
MSE	0.014946	4.27E-09	3.998269	0.01494

B-4 : Estimation of integrated volatility (Model-1)

$$(a_0 = 1, a_1 = 0, a_2 = 0; \sigma_v^2 = 1.00\text{E} - 05, g = 1.0)$$

n=20000	σ_x^2	σ_v^2	H-vol	$\sigma_{x,m}^2$
true-val	1	1.00E-05	1	1
mean	1.004956	3.84E-05	1.3997	1.00483
SD	0.121406	4.88E-07	0.013967	0.121406
MSE	0.014764	8.06E-10	0.159962	0.014763

B-5 : Estimation of integrated volatility (Model-1)

$$(a_0 = 1, a_1 = 0, a_2 = 0; \sigma_v^2 = 1.00\text{E} - 06, g = 0.01)$$

n=20000	σ_x^2	σ_v^2	H-vol	$\sigma_{x,m}^2$
true-val	1	1.00E-06	1	1
mean	0.523326	5.77E-07	0.0250976	0.523324
SD	0.071447	8.16E-09	0.0005492	0.071447
MSE	0.232323	1.79E-13	0.9504351	0.232325

B-6 : Estimation of integrated volatility (Model-2)

$$(a_0 = 7, a_1 = -12, a_2 = 6; \sigma_v^2 = 2.00\text{E} - 02, \eta = 0.5)$$

n=20000	σ_x^2	σ_v^2	H-vol	$\sigma_{x,m}^2$
true-val	45	0.02	45	45
mean	46.59003	0.004002	136.3733	46.57694
SD	6.399936	0.00019	6.166522	6.399906
MSE	43.48737	0.000256	8387.097	43.44555

B-7 : Estimation of integrated volatility (Model-3)

$$(a_0 = 7, a_1 = -12, a_2 = 6; \sigma_v^2 = 1.00\text{E} - 02, \eta = 0.5)$$

n=20000	σ_x^2	σ_v^2	H-vol	$\sigma_{x,m}^2$
true-val	45	1.00E-02	45	45
mean	47.185	1.17E-02	394.7923	47.14678
SD	6.548271	2.29E-04	7.1138	6.548203
MSE	47.6541	2.91E-06	122405.23	47.48761

B-8 : Estimation of integrated volatility (Model-3)

$$(a_0 = 1, a_1 = 0, a_2 = 0; \sigma_v^2 = 0.0, \eta = 0.005)$$

n=20000	σ_x^2	σ_v^2	H-vol	$\sigma_{x,m}^2$
true-val	1	0.00E+00	1	1
mean	1.002935	1.94E-05	0.68542	1.002872
SD	0.117878	3.32E-07	0.008776	0.117878
MSE	0.013904	3.78E-10	0.099037	0.013903

B-9 : Estimation of integrated volatility (Model-4)

$$(a_0 = 1, a_1 = 0, a_2 = 0; \sigma_v^2 = 0.0, g_1 = 0.2, g_2 = 5)$$

n=20000	σ_x^2	σ_v^2	H-vol	$\sigma_{x,m}^2$
true-val	1	0.00E+00	1	1
mean	1.001114	6.76E-05	2.221652	1.000893
SD	0.11993	2.22E-06	0.068095	0.11993
MSE	0.014384	4.57E-09	1.497069	0.014384

B-10 : Estimation of integrated volatility (Model-4)

$$(a_0 = 1, a_1 = 0, a_2 = 0; \sigma_v^2 = 1.00E - 03, g_1 = 0.2, g_2 = 5)$$

n=20000	σ_x^2	σ_v^2	H-vol	$\sigma_{x,m}^2$
true-val	1	1.00E-03	1	1
mean	1.029582	1.99E-03	66.6307	1.02307
SD	0.122565	4.86E-05	1.547753	0.122566
MSE	0.015897	9.86E-07	4309.795	0.015555

B-11 : Estimation of integrated volatility (Model-5)

$$(a_0 = 1, a_1 = 0, a_2 = 0; \sigma_v^2 = 0.0, g_1 = 1.9, g_2 = -1.7, \gamma = 10000)$$

n=20000	σ_x^2	σ_v^2	H-vol	$\sigma_{x,m}^2$
true-val	1	0.00E+00	1	1
mean	0.996518	9.72E-05	6.375448	0.9962
SD	0.121183	3.21E-06	0.367669	0.12118
MSE	0.014697	9.46E-09	29.03062	0.01470