



Title	マーケット・マイクロストラクチャー・ノイズの系列 相関の推定
Author(s)	大屋, 幸輔
Citation	大阪大学経済学. 2008, 57(4), p. 229-241
Version Type	VoR
URL	<a href="https://doi.org/10.18910/16371">https://doi.org/10.18910/16371</a>
rights	
Note	

*The University of Osaka Institutional Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

# マーケット・マイクロストラクチャー・ノイズの 系列相関の推定

大 屋 幸 輔

## 概 要

金融市場における資産価格の高頻度観測データが利用可能になり、資産価格形成メカニズムの解明がより一層進展することが期待されている。しかしながら経済理論的な解明が進展をとげる一方で、統計的分析においては大量データにもとづく漸近理論が必ずしもうまく適用できていない状況に直面している。高頻度観測においては、大量の法則が適用できる状況にあるにも関わらず実際には高頻度観測にもとづくデータによる推定量のばらつきは極めて大きく、安定的な推定値を得ることができない。このような状況が生じる理由を簡明に表現したものが、観測対象の価格系列はなんらかの観測誤差を含んで観測されている、とする考え方である。この観測誤差とは具体的には、約定価格がビッドで約定するかアスクで約定するかによって、不規則におこるノイズのように振舞うこと（ビッド・アスク・バウンスとよばれる）に起因するものと考えられている。マーケット・マイクロストラクチャー・ノイズは初期の研究においては、時間に関して独立、あるいは無相関であることが仮定されていたが、その性質上、必ずしも独立ではなく従属性があることが指摘されている。この従属性の構造が特定化できれば、投資家の発注行動の分析にも有用と考えられる。本稿ではマイクロストラクチャー・ノイズにおける系列相関を検証するために、その自己相関の推定方法を提案し、その推定量の漸近分布を導出した。さらにモンテカルロ実験により、推定量の経験分布をもとめ、漸近分布による近似の精度を確認した。その結果、推定量の分布は自己回帰係数が1に近い値のときを除いて、導出された漸近分布によって正確に近似できることが確認された。

JEL classification: C13; C22; D49

Keywords: 高頻度観測; マーケット・マイクロストラクチャー・ノイズ; 系列相関; 漸近分布

## 1 イントロダクション

金融市場における資産価格の高頻度観測データが利用可能になり、資産価格形成メカニズムの解明がより一層進展することが期待されている。しかしながら経済理論的な解明が進展をとげる一方で、統計的分析においては大量データにもとづく漸近理論が必ずしもうまく適用できていない状況に直面している。

高頻度観測においては、大量の法則が適用で

きる状況にあるにも関わらず実際には高頻度観測にもとづくデータによる推定量のばらつきは極めて大きく、安定的な推定値を得ることができない。このような状況がなぜ生じるのかを簡明な形で表現したものが、観測対象の価格系列はなんらかの観測誤差を含んで観測されている、とする考え方である。この観測誤差とは具体的には、約定価格がビッドで約定するかアスクで約定するかによって、不規則におこるノイズのように振舞うこと（ビッド・アスク・バウンス

とよばれる)に起因するものと考えられている。マーケット・マイクロストラクチャー・ノイズは初期の研究においては、時間に関して独立であることが仮定されていたが、その性質上、必ずしも独立ではなく従属性があることが指摘されている。その相関構造を明らかにすることができれば、投資家の発注行動を説明する手がかりとなることが期待される。

次節では観測価格と MMS ノイズに関して概説をあたえる。第 3 節では、MMS ノイズが存在しているとき、その自己相関を推定する方法を提案し、推定量の漸近分布を導出する。第 4 節では Monte Carlo 実験によって推定量の経験分布をもとめ、導出された漸近分布による近似精度を確認する。

## 2 観測価格とモデル

分析対象である資産の価格は次の連続確率過程に従うとする。

$$dP_t^* = \sigma_t dB_t \quad (1)$$

ドリフトはゼロ、 $\sigma_t$  は漸進的可測、 $B_t$  は標準ブラウン運動とする。また以下、特に断らない限り価格とは資産の対数価格をさすことにする。観測は時刻  $[0, T]$  の間で時刻  $t_i, (i = 0, 1, \dots, n)$  においておこなわれるものとする。これらの観測時刻は  $0 = t_0 < t_1 < t_2 < \dots < t_n = T$  であり、 $[0, T]$  を  $n$  分割するものである。また観測時点間隔を  $\Delta_i = t_i - t_{i-1}$  とあらわす。

時点  $t_i$  における価格は以下の形で観測されていると仮定する。

$$P_{t_i} = P_{t_i}^* + \varepsilon_{t_i} \quad (2)$$

ここで  $P_{t_i}^*$  は真の価格、 $P_{t_i}$  は観測価格、 $\varepsilon_{t_i}$  は観測誤差で、マーケット・マイクロストラクチャー・

ノイズ (以下、MMS ノイズ) とよばれる平均ゼロの確率変数である。

### 2.1 RV と MMS ノイズ

MMS ノイズに対して以下の仮定をおく。

#### 仮定 2.1.

(a)  $\varepsilon_{t_i} \sim i.i.d.(0, \sigma_\varepsilon^2)$

(b)  $P_{t_i}^*$  と  $\varepsilon_{t_i}$  は独立

(2) からあきらかなように真の価格を直接観測することはできない。また価格過程 (1) は連続であるのに対して、MMS ノイズは離散観測にもなっている。

Integrated Variance (IV) の推定量として代表的な Realized Volatility (RV) は

$$RV = \sum_{i=1}^n (\Delta P_{t_i})^2 = \sum_{i=1}^n (P_{t_i} - P_{t_{i-1}})^2 \quad (3)$$

と定義される。ここで高頻度観測の状況を考えるために観測区間  $[0, T]$  の分割数  $n$  を無限にする極限を考える。分割数  $n$  が無限に大きくなると観測時点間隔はゼロへ収束していく。実際、高頻度観測データ系列においては、非常に小さい時間間隔で大量のデータを観測していることになる。

観測された価格データが (2) に従っているとき、RV は漸近的に以下のように表現される。

$$\begin{aligned} RV &= \sum_{i=1}^n (\Delta P_{t_i}^* + \Delta \varepsilon_{t_i})^2 = \sum_{i=1}^n (\Delta P_{t_i}^*)^2 \\ &+ 2 \sum_{i=1}^n \Delta P_{t_i}^* \Delta \varepsilon_{t_i} + \sum_{i=1}^n (\Delta \varepsilon_{t_i})^2 \\ &\rightarrow \int_0^T \sigma_t^2 dt + 2 \lim_{n \rightarrow \infty} \sum_{i=1}^n \sigma_\varepsilon^2, \\ &\text{ただし } T \text{ は固定, } n \rightarrow \infty \end{aligned} \quad (4)$$

上式の右辺第 2 項の  $\sigma_\varepsilon^2$  大きさは  $O(1)$  であり、第 2 項は  $n$  が大きくなると発散していく。従っ

て  $RV$  は右辺第 1 項の真の  $IV$  へ収束しない。すなわち高頻度観測の状況では、 $MMS$  ノイズからの変動によって、 $RV$  全体が支配されるようになり、 $RV$  が不安定な推定値を与えることになるのである。

## 2.2 系列相関をもつ $MMS$ ノイズ

これまで  $MMS$  ノイズ自体は独立、あるいは無相関な系列で  $E[\varepsilon_t \varepsilon_s] = 0, (t \neq s)$  と仮定しておこなわれる研究が多かった。しかし  $\varepsilon_t$  は単なる平均ゼロのノイズではなく、マーケット・マイクロストラクチャーという具体的発生理由をもったノイズである。外国為替市場や株式市場などでは売買されている対象の買値と売値が同一ではなく、その差であるビッド・アスク・スプレッドが存在している。従って観測している約定価格は買い約定と売り約定との間を振動しているようにみえる状況が見受けられる。また大量の売り注文を出す投資家は、一度に注文を出すことでマーケットインパクトを引き起こすことを避けるために、分割注文を出すことも考えられる。その場合、約定価格はある程度同じ価格で観測される、いわゆるクラスタリングが生じているようにも見える。これらは  $MMS$  ノイズ  $\varepsilon_t$  の変動パターンの例であるが、いずれにせよ、その変動に独立性を仮定するのは現実的ではない。

以下では  $\varepsilon_t$  に対して 1 階の自己回帰モデルを仮定し、そのモデルのパラメータを推定することで、 $MMS$  ノイズの従属性、自己相関について考察をおこなう。

### AR(1)- $MMS$ ノイズ

先に定義した観測時間間隔  $\Delta_i, (i = 1, \dots, n)$  において、最小時間間隔を  $h_m = \min_i \Delta_i$ 、さらに  $c_i = \Delta_i / h_m$  とする。このとき  $MMS$  ノイズ  $\varepsilon_{t_i}$  に対して以下の AR(1) を仮定する。

#### 仮定 2.2.

$$\begin{aligned} \varepsilon_{t_i} &= \rho \varepsilon_{t_i - h_m} + e_{t_i}, \quad e_{t_i} \sim i.i.d(0, \sigma_e^2), \\ \sigma_e &\neq 0, \quad E[e_{t_i}^4] = \eta \sigma_e^4, \quad E[e_{t_i}^6] < \infty \quad (5) \end{aligned}$$

ただし  $\varepsilon_{t_i}$  は  $P_{t_i}^*$  と独立であり、 $|\rho| < 1$  とする。 $e_{t_i}$  が正規分布に従う場合は  $\eta = 3$  である。

この仮定のもと、 $t_{i-1} = t_i - \Delta_i = t_i - c_i h_m$  であることから  $\varepsilon_{t_i}$  は

$$\begin{aligned} \varepsilon_{t_i} &= \rho \varepsilon_{t_i - h_m} + e_{t_i} = \rho^2 \varepsilon_{t_i - 2h_m} + \rho e_{t_i - h_m} + e_{t_i} \\ &= \rho^{c_i} \varepsilon_{t_i - c_i h_m} + \sum_{k=1}^{c_i} \rho^{c_i - k} e_{t_i - (c_i - k) h_m} \\ &= \rho^{c_i} \varepsilon_{t_{i-1}} + \sum_{k=1}^{c_i} \rho^{c_i - k} e_{t_i - (c_i - k) h_m} \end{aligned}$$

と表現され、確率変数  $\varepsilon_{t_i}$  の自己共分散関数を

$\gamma(s) = \text{Cov}[\varepsilon_{t_i} \varepsilon_{t_i - s h_m}]$  と定義すると (5) のもとで以下のようにあらわすことができる。

$$\gamma(s) = \text{Cov}[\varepsilon_{t_i} \varepsilon_{t_i - s h_m}] = \rho^s \gamma(0), \quad s \geq 0 \quad (6)$$

$MMS$  ノイズ  $\varepsilon_{t_i}$  が観測可能な変数であれば AR モデル (5) の係数の推定は容易だが、 $\varepsilon_{t_i}$  は観測できない潜在変数であり、通常の方法でのパラメータ推定は困難である。

次節では Ubukata and Oya (2007) でもちいられたアプローチを採用し、(5) のパラメータ推定の方法を提案する。

## 3 推定量と漸近分布

時点  $t_i$  でのリターンは、観測価格  $P_{t_i}$  の差分  $\Delta P_{t_i} = P_{t_i} - P_{t_{i-1}}$  で定義される。このリターンに対して、隣り合ったリターンの積  $\Delta P_{t_i} \Delta P_{t_{i-1}}$  は以下のように分解できる。

$$\begin{aligned} \Delta P_{t_i} \Delta P_{t_{i-1}} &= \Delta P_{t_i}^* \Delta P_{t_{i-1}}^* - \Delta P_{t_i}^* \Delta \varepsilon_{t_{i-1}} \\ &\quad - \Delta \varepsilon_{t_i} \Delta P_{t_{i-1}}^* + \Delta \varepsilon_{t_i} \Delta \varepsilon_{t_{i-1}} \quad (7) \end{aligned}$$

Ubukata and Oya (2007) では、真の価格系列の特性である独立増分性、真の価格系列と MMS ノイズの独立性から、複数の資産価格間の共分散推定量のバイアスに関する考察をおこなっている。ここでの 1 変量の観測価格系列に対しても同様な議論が展開でき、(7) の期待値は

$$\begin{aligned}
 E[\Delta P_{t_i} \Delta P_{t_{i-1}}] &= E[\Delta \varepsilon_{t_i} \Delta \varepsilon_{t_{i-1}}] \\
 &= E[\varepsilon_{t_i} \varepsilon_{t_{i-1}}] - E[\varepsilon_{t_i} \varepsilon_{t_{i-2}}] \\
 &\quad - E[\varepsilon_{t_{i-1}} \varepsilon_{t_{i-1}}] + E[\varepsilon_{t_{i-1}} \varepsilon_{t_{i-2}}] \\
 &= \gamma(c_i) - \gamma(c_i + c_{i-1}) \\
 &\quad - \gamma(0) + \gamma(c_{i-1}) \\
 &= -\gamma(0)(1 - \rho^{c_i})(1 - \rho^{c_{i-1}}) \quad (8)
 \end{aligned}$$

と表現できる。さらに時点  $t_i$  のリターンと時点  $t_{i-2}$  のリターンの積  $\Delta P_{t_i} \Delta P_{t_{i-2}}$  の期待値は

$$\begin{aligned}
 E[\Delta P_{t_i} \Delta P_{t_{i-2}}] &= E[\Delta \varepsilon_{t_i} \Delta \varepsilon_{t_{i-2}}] \\
 &= \gamma(c_i + c_{i-1}) - \gamma(c_i + c_{i-1} \\
 &\quad + c_{i-2}) - \gamma(c_{i-1}) \\
 &\quad + \gamma(c_{i-1} + c_{i-2}) \\
 &= -\gamma(0)\rho^{c_{i-1}}(1 - \rho^{c_i})(1 - \rho^{c_{i-2}}) \quad (9)
 \end{aligned}$$

となる。観測時間間隔は一般には一様に同じではないため、(8), (9) 自体をその標本モーメントから求めることはできない。そこで同じ観測時間間隔幅の区間を集め、その区間に対するリターンの積による標本モーメントにより区間の長さ別に上記の期待値を求めるなどのアプローチが考えられるが、単純化のため、以下では観測時間間隔を一定 ( $\Delta t_i = h_m$ ) として分析をおこなう。このとき  $c_i = 1$  となっている。観測時間間隔を等間隔とすることは必ずしも一般的ではないが、現在、実際に利用できる東証一部上場企業の株価データのタイムスタンプは 1 分間隔が最小で

あり、取引が十分におこなわれている銘柄においては 1 分間隔でサンプリングされた観測系列に欠損値が生じることは少ないので分析目的上は十分であると考えられる。

あらためて先のリターンの積の期待値を  $T(1)$ ,  $T(2)$  とすると

$$\begin{aligned}
 T(1) &= E[\Delta P_{t_i} \Delta P_{t_{i-1}}] \\
 &= -\gamma(0)(1 - \rho)^2 \quad (10)
 \end{aligned}$$

$$\begin{aligned}
 T(2) &= E[\Delta P_{t_i} \Delta P_{t_{i-2}}] \\
 &= -\gamma(0)\rho(1 - \rho)^2 \quad (11)
 \end{aligned}$$

であり、 $\rho = T(2)/T(1)$  となっていることがわかる。

一般に  $E[\Delta P_{t_i} \Delta P_{t_{i-k}}]$  は

$$T(k) = E[\Delta P_{t_i} \Delta P_{t_{i-k}}] = -\gamma(0)\rho^{k-1}(1 - \rho)^2, \quad k \geq 1$$

となる。ただし  $T(0)$  に関しては便宜上、 $T(0) = E[(\Delta \varepsilon_{t_i})^2]$  とする。

(10), (11) の標本モーメントをそれぞれ

$$\widehat{T}(1) = \frac{1}{n} \sum_{i=2}^n \Delta P_{t_i} \Delta P_{t_{i-1}} \quad (12)$$

$$\widehat{T}(2) = \frac{1}{n} \sum_{i=3}^n \Delta P_{t_i} \Delta P_{t_{i-2}} \quad (13)$$

とする。このときこれらの標本モーメントの比によって定義された AR モデルの係数  $\rho$  の推定量に関して以下の定理が成り立つ。

**定理 3.1.** 観測価格  $P_{t_i}$  は (2)、真の価格  $P_{t_i}^*$  は (1)、MMS ノイズ  $\varepsilon_{t_i}$  は (5) に従っているとき、仮定 2.2 のもと AR モデル (5) の係数  $\rho$  の推定量を以下の  $\widehat{T}(1)$  と  $\widehat{T}(2)$  の比によって定義する。

$$\widehat{\rho} = \frac{\widehat{T}(2)}{\widehat{T}(1)} = \frac{\sum_{i=3}^n \Delta P_{t_i} \Delta P_{t_{i-2}}}{\sum_{i=2}^n \Delta P_{t_i} \Delta P_{t_{i-1}}} \quad (14)$$

このとき  $\widehat{\rho}$  は  $\rho$  の一致推定量であり、その漸近分布は

$$\begin{aligned} \sqrt{n}(\widehat{\rho} - \rho) &\rightarrow N(0, \sigma_\rho^2), \quad (n \rightarrow \infty) \\ \sigma_\rho^2 &= \frac{2(1+\rho)(3-\rho)}{(1-\rho)^2} \end{aligned} \quad (15)$$

となる。

**証明.** 仮定より  $T(1) \neq 0$  である。 $\widehat{\rho}$  に対して Taylor 展開を適用する。

$$\begin{aligned} \widehat{\rho} &= \frac{\widehat{T}(2)}{\widehat{T}(1)} = \frac{T(2)}{T(1)} \\ &+ \left( -\frac{T(2)}{T(1)^2} \quad \frac{1}{T(1)} \right) \begin{pmatrix} \widehat{T}(1) - T(1) \\ \widehat{T}(2) - T(2) \end{pmatrix} \\ &+ o_p(1) \\ &= \rho + \frac{1}{T(1)} \{ (\widehat{T}(2) - T(2)) \\ &- \rho(\widehat{T}(1) - T(1)) \} + o_p(1) \end{aligned} \quad (16)$$

$E[\widehat{T}(1)] = T(1) + o(1)$ ,  $E[\widehat{T}(2)] = T(2) + o(1)$   
より  $\widehat{\rho}$  の分散は

$$\begin{aligned} \text{Var}[\widehat{\rho}] &= E[(\widehat{\rho} - \rho)^2] \\ &= \frac{1}{T(1)^2} \{ \text{Var}[\widehat{T}(2)] \\ &- 2\rho \text{Cov}[\widehat{T}(2), \widehat{T}(1)] \\ &+ \rho^2 \text{Var}[\widehat{T}(1)] \} + o(n^{-1}) \end{aligned} \quad (17)$$

上式右辺の  $\{ \}$  の中は  $O(n^{-1})$  で  $n \rightarrow \infty$  のとき、 $\text{Var}[\widehat{\rho}]$  はゼロに収束する。従って  $\widehat{\rho}$  は  $\rho$  の一致推定量である。

$\sqrt{n}(\widehat{\rho} - \rho)$  の漸近正規性に関しては次のとおりである。仮定の下、Anderson (1971) の Theorem 8.4.2 より  $\widehat{T}(1)$  と  $\widehat{T}(2)$  に関する漸近正規性が成立する。

$$\begin{aligned} \mathbf{D} &= \begin{pmatrix} D_1 \\ D_2 \end{pmatrix} = \sqrt{n} \begin{pmatrix} \widehat{T}(1) - T(1) \\ \widehat{T}(2) - T(2) \end{pmatrix} \\ &\xrightarrow{d} N(\mathbf{0}, \mathbf{\Omega}), \quad \mathbf{\Omega} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix} \end{aligned}$$

$\mathbf{\Omega}$  の各要素  $\omega_{11}$ ,  $\omega_{12} = \omega_{21}$ ,  $\omega_{22}$  は Appendix の (30), (31), (32) で与えられる。

この  $\widehat{T}(1)$  と  $\widehat{T}(2)$  の漸近正規性と (16) より、推定量  $\widehat{\rho}$  に関する漸近分布を得る。

$$\begin{aligned} \sqrt{n}(\widehat{\rho} - \rho) &= \frac{1}{T(1)} \{ \sqrt{n}(\widehat{T}(2) - T(2)) \\ &- \rho \sqrt{n}(\widehat{T}(1) - T(1)) \} \\ &= \frac{1}{T(1)} \begin{pmatrix} -\rho & 1 \end{pmatrix} \mathbf{D} \\ &\xrightarrow{d} N(0, \sigma_\rho^2) \end{aligned} \quad (18)$$

$$\begin{aligned} \sigma_\rho^2 &= \frac{1}{T(1)^2} (\rho^2 \omega_{11} - 2\rho \omega_{12} \\ &+ \omega_{22}) = \frac{2(1+\rho)(3-\rho)}{(1-\rho)^2} \end{aligned} \quad (19)$$

□

定理 3.1 で示された係数推定量  $\widehat{\rho}$  の一貫性および漸近正規性より、MMS ノイズに自己相関が存在しているかどうかは、帰無仮説  $\mathcal{H}_0 : \rho = 0$  を検定することで検証可能になる。

#### 4 Monte Carlo 実験

この節では前節で導出した MMS ノイズが従う AR モデルの係数  $\rho$  の推定量  $\widehat{\rho}$  の特性を Monte Carlo 実験により考察する。時点  $t_i$  での観測価格  $P_{t_i}$  は、真の価格  $P_{t_i}^*$  と MMS ノイズ  $\varepsilon_{t_i}$  の和として観測されているとする。

$$P_{t_i} = P_{t_i}^* + \varepsilon_{t_i}$$

真の価格系列と確率ボラティリティの系列の生成には Zhang, Mykland and Aït-Sahalia (2005) で

もちいられた Heston (1993) による確率ボラティリティモデル (20), (21) を採用する。

$$dP_t^* = (\mu - \sigma_t^2/2) dt + \sigma_t dW_t^{(1)} \quad (20)$$

$$d\sigma_t^2 = \kappa(\alpha - \sigma_t^2) dt + \gamma \sigma_t dW_t^{(2)} \quad (21)$$

$W_t^{(1)}$ ,  $W_t^{(2)}$  はそれぞれ標準ブラウン運動でそれらの相関は  $\rho_W$  とする。また条件  $2\kappa\alpha \geq \gamma^2$  を仮定する。パラメータ  $(\mu, \kappa, \alpha, \gamma, \rho_W)$  は Zhang, Mykland and Ait-Sahalia (2005) と同じ (0.05, 5, 0.04, 0.5, -0.5) とした。

MMS ノイズ  $\varepsilon_{t_i}$  の生成には以下の AR(1) をもちいる。

$$\begin{aligned} \varepsilon_{t_i} &= \rho \varepsilon_{t_i-h} + e_{t_i}, \quad e_{t_i} \\ &\sim i.i.d.N(0, \sigma_e^2), \quad E[e_{t_i}^4] = 3\sigma_e^4, \quad (22) \end{aligned}$$

(22) の  $h$  は観測時間間隔 (秒単位) をあらわしている。一日のボラティリティを推定する状況を想定し、日中の取引時間は東京証券取引所の取引時間 4 時間 30 分 (270 分 = 16200 秒) とした。真の価格系列は 1 秒ごとに (20) と (21) から生成し、その観測時間間隔は等間隔で  $h$  とする。

$RV$  への MMS ノイズの影響をみるために  $\varepsilon_{t_i}$  の標準偏差  $\sigma_e / \sqrt{(1-\rho^2)}$  を 0.05, 0.005 とする。この標準偏差の設定は実際の MMS ノイズの大きさを反映する必要がある。Hansen and Lunde (2006) によれば、分析の対象とされた銘柄に対して、Noise-to-Signal 比は (1 観測あたりの平均 MMS ノイズの分散)/(1 日平均  $IV$ ) = (min, max) = (0.0002, 0.006) であることが報告されている。ここで設定されたパラメータのもと  $IV = 2.75$  であり、 $\varepsilon_{t_i}$  の標準偏差を 0.05 としたときの Noise-to-Signal 比は 0.0009 となり現実的な MMS ノイズの大きさを表している。一方で標準偏差を 0.005 としたときは MMS ノイズの影響はほとんどない状況となっている。

図 1 は MMS ノイズに系列相関がない ( $\rho = 0$ ) 状態で  $RV$  をプロットした Volatility Signature

Plot である。横軸は観測間隔  $h$  で左に行くほど時間間隔が短くなっている。また図中の実線は真の  $IV$  を示している。 $\varepsilon_{t_i}$  の標準偏差の大きさが 0.05 のときは観測間隔  $h$  が短くなるほど  $RV$  の値が大きくなっており、観測間隔  $h$  を大きくするにつれて真の  $IV$  のまわりでばらついている様子がみてとれる。一方で  $\varepsilon_{t_i}$  の標準偏差が 0.005 のときにはノイズの影響は微小で観測間隔  $h$  が大きくなる (標本サイズが減少する) につれて  $RV$  のばらつきが顕著になることがわかる。ただしノイズの影響がある上の図と比較するとそのばらつきの程度は大きくはない。

次に MMS ノイズが従う AR モデルの係数  $\rho$  の推定量 (14) の性質をみていく。観測間隔  $h$  は 1 秒、真の AR 係数  $\rho$  は (-0.9, -0.6, -0.3, 0.0, 0.3, 0.6, 0.9) とした。MMS ノイズが  $RV$  の挙動に影響をあたえる状況を考察対象とするため、 $\varepsilon_{t_i}$  の標準偏差の大きさを 0.05 とした。このように設定された各値のもとでの 10000 回データを生成し、 $\rho$  の推定を (14) によっておこない推定値  $\hat{\rho}^{(i)}, (i = 1, \dots, 10000)$  を得た。表 1 は実験結果の要約である。表中の平均、標準偏差、MSE はそれぞれ実験によって得られた  $\hat{\rho}^{(i)}, (i = 1, \dots, 10000)$  の標本平均、標本標準偏差、標本 MSE である。また  $Var/\sigma_\rho^2$  は標本分散と漸近分散の比である。

16200 個のデータから得られた係数推定値は定理 3.1 によって示された漸近特性をほぼ満足するものとなった。四分位点は推定量の経験分布を  $\rho$  の真値と推定量の漸近分散の平方根 (標準誤差) で標準化したものに関して求めたものである。標準正規分布に関して、25%点、50%点、75%点のそれぞれは -0.674, 0.0, 0.674 であり、推定量の経験分布は漸近分布に比べて若干、裾が厚いが  $\rho = 0.9$  のとき以外は全般的に漸近分布での近似は良好であることがわかる。 $\rho = 0.9$  のときの分布はその右裾が厚くなっており、推定値が 1 を超える頻度が少なくないことが図 3 から確認できる。いわゆる Unit root の状況に近

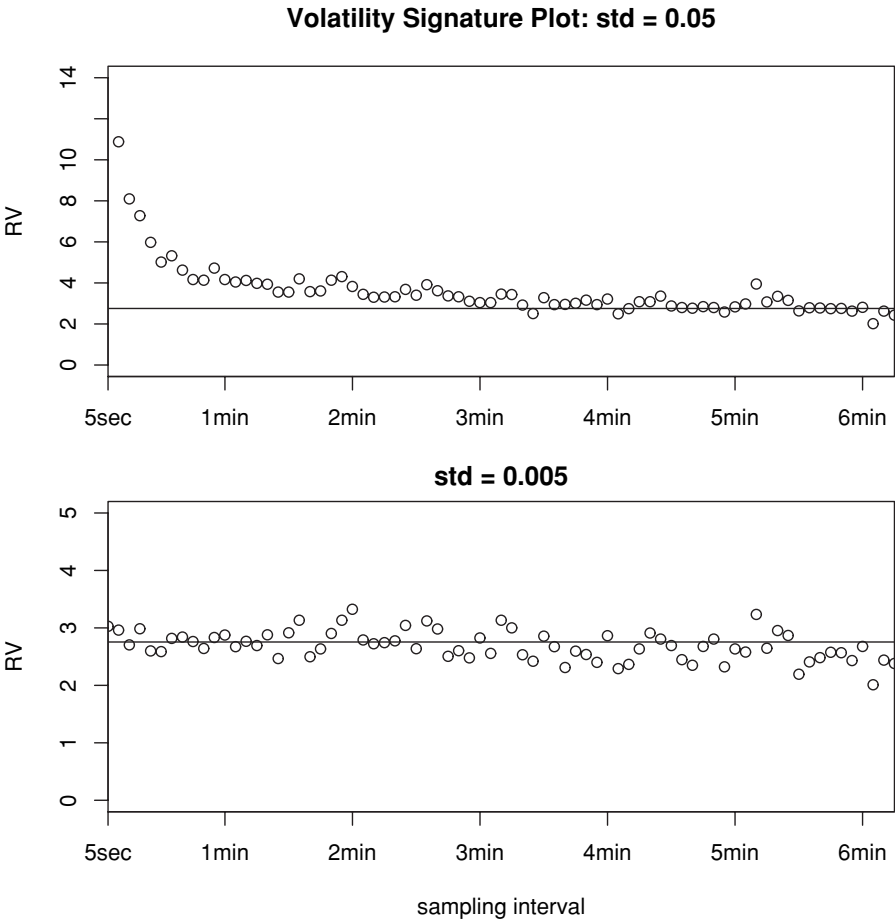


図 1: 観測時間間隔  $h$  と  $RV$  の挙動

表 1: 実験結果の要約

$SD(\varepsilon_t)$	$\rho$	平均	標準偏差	MSE	$\text{Var}/\sigma_\rho^2$	四分位点		
						25%	50%	75%
0.05	-0.9	-0.899833	0.003681	0.000014	1.0160	-0.6514	0.0231	0.7255
	-0.6	-0.599909	0.008311	0.000069	0.9940	-0.6614	-0.0058	0.6914
	-0.3	-0.299884	0.013287	0.000177	1.0460	-0.6906	-0.0051	0.6975
	0.0	-0.000020	0.019810	0.000392	1.0600	-0.6996	-0.0267	0.6873
	0.3	0.300343	0.030893	0.000954	1.0790	-0.6960	-0.0060	0.7076
	0.6	0.601105	0.058774	0.003456	1.1660	-0.7166	0.0025	0.7226
	0.9	0.952642	0.346757	0.123012	2.4410	-0.8201	0.0305	1.0138



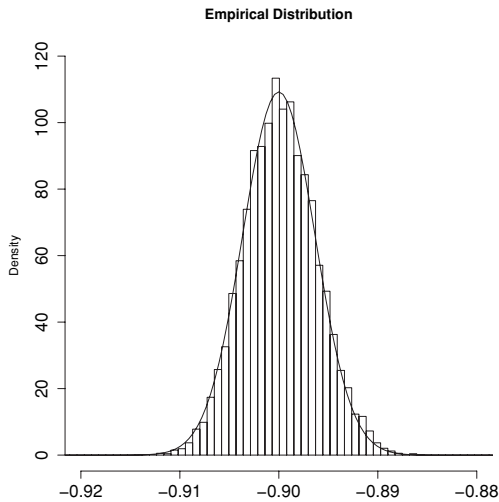


図 2:  $\rho = -0.9$

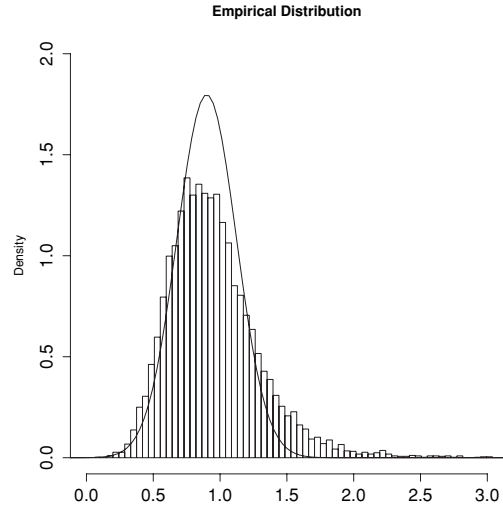


図 3:  $\rho = 0.9$

づいていることがその原因と考えられる。

## 5 結語

本稿では高頻度に観測される資産価格のデータを利用する際に、Integrated Volatility の代表的な推定量である Realized Volatility の漸近挙動が不安定になる主な原因と考えられているマーケット・マイクロストラクチャー・ノイズ (MMS ノイズ) に焦点をあてた。この MMS ノイズはビッド、アスクそれぞれの約定が不規則に観測されることで生じると考えられるが、それらの時系列構造 (時間的な従属性) が明らかになれば、投資家の注文メカニズムに何らかの規則性をみいだすことが可能になる。本来この MMS ノイズは観測されない潜在変数であり、その時間的な従属構造を明らかにすることは容易ではない。本稿では価格プロセスの独立増分性を使うことで、MMS ノイズの時間的な従属性を検討することが可能であることを示した。導出された推定量は MMS ノイズ自体の存在を前提としているが、ノイズ自体が存在しているかどうかは Volatility Signature Plot によって確認可能である。また MMS ノイズ系列が従うモデルと

して本稿で採用したモデルは AR(1) であるが、MMS ノイズの自己相関がプラスなのか、マイナスなのか、それはどのような状況で生じるのか、といった分析目的からは十分であると考えられる。ただし推定量の導出のために、観測時間間隔は等間隔であると仮定したが、より広範囲な応用を念頭に置けばこの仮定を必要としない推定量とその漸近分布の導出が必要となる。この点に関しては、実際の高頻度観測の約定価格データの分析とあわせて別の機会に論じることにする。

(大阪大学大学院経済学研究科教授)

## 参考文献

- [1] Anderson, T.W. (1971), *The Statistical Analysis of Time Series*, John Wiley & Sons.
- [2] Fuller, W.A. (1976), *Introduction to Statistical Time Series*, John Wiley & Sons.

- [3] Hansen, P.R. and A.Lunde (2006), "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, vol.24, No.2, 127-161.
- [4] Heston, S.(1993), "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bounds and Currency Options," *Review of Financial Studies*, 6, 327-343.
- [5] Ubukata, M. and K.Oya (2007), "Test of Unbiasedness of the Integrated Covariance Estimation in the Presence of Noise," Discussion Papers in Economics and Business, 07-03, Osaka University.
- [6] Zhang, L., P.A. Mykland and Y. Aït-Sahalia (2005), "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, vol.100, No.472, 1394-1411.

## Appendix

ここでは定理の証明に必要な事項をまとめる。

(5) より確率変数  $\Delta \varepsilon_{t_i}$  は以下の線形過程として表現することができる。(ただし  $c_i = 1$ )

$$\begin{aligned}
 \Delta \varepsilon_{t_i} &= \varepsilon_{t_i} - \varepsilon_{t_{i-1}} = \varepsilon_{t_i} - \varepsilon_{t_i - c_i h_m} \\
 &= (1 - L^{c_i h_m}) \varepsilon_{t_i} = \frac{1 - L^{c_i h_m}}{1 - \rho L^{c_i h_m}} e_t \\
 &= (1 - L^{h_m}) \sum_{j=0}^{\infty} \rho^j e_{t-jh_m} = \sum_{j=0}^{\infty} a_j e_{t-jh_m}, \\
 a_j &= \begin{cases} 1 & \text{for } j = 0 \\ \rho^{j-1}(\rho - 1) & \text{for } j \geq 1 \end{cases}
 \end{aligned}$$

$|\rho| < 1$  より

$$\sum_{j=0}^{\infty} |a_j| = 1 + \frac{|\rho - 1|}{1 - |\rho|} < \infty$$

であり  $a_j$  は絶対総和可能な列である。ただし、 $\sum_{j=0}^{\infty} a_j = 0$  である。

$\Delta \varepsilon_{t_i}$  の自己共分散関数  $T(k)$

$$\begin{aligned}
 T(0) &= E[(\Delta \varepsilon_{t_i})^2] = E[\varepsilon_{t_i}^2] - 2E[\varepsilon_{t_i} \varepsilon_{t_{i-1}}] \\
 &\quad + E[\varepsilon_{t_{i-1}}^2] = 2(\gamma(0) - \gamma(1)) \\
 &= 2(1 - \rho)\gamma(0) \quad (23)
 \end{aligned}$$

$$\begin{aligned}
 T(k) &= E[\Delta \varepsilon_{t_i} \Delta \varepsilon_{t_{i-k}}] = \gamma(k) - \gamma(k+1) \\
 &\quad - \gamma(k-1) + \gamma(k) \\
 &= -(1 - \rho)^2 \rho^{k-1} \gamma(0), \quad k \geq 1 \quad (24)
 \end{aligned}$$

ただし  $\gamma(k)$  は  $\varepsilon_{t_i}$  の  $k$  次の自己共分散関数で、 $\gamma(k) = \rho^k \gamma(0)$  である。

以下、推定量  $\hat{\rho}$  の漸近分散の導出に必要な結果をまとめる。

$$\begin{aligned}
 \sum_{k=-\infty}^{\infty} T(k)^2 &= T(0) + 2 \sum_{k=1}^{\infty} T(k)^2 \\
 &= (1 - \rho)^2 \gamma(0)^2 \left\{ 4 + 2 \frac{1 - \rho}{1 + \rho} \right\} \quad (25)
 \end{aligned}$$

$$\begin{aligned}
 \sum_{k=-\infty}^{\infty} T(k+1)T(k-1) &= T(1)^2 + 2T(2)T(0) + 2 \sum_{k=1}^{\infty} T(k+2)T(k) \\
 &= (1 - \rho)^2 \gamma(0)^2 \left\{ (\rho - 1)(3\rho + 1) + 2 \frac{1 - \rho}{1 + \rho} \right\} \quad (26)
 \end{aligned}$$

$$\begin{aligned}
& \sum_{k=-\infty}^{\infty} T(k+2)T(k-2) \\
& = T(2)^2 + 2T(3)T(1) + 2T(4)T(0) \\
& \quad + 2 \sum_{k=1}^{\infty} T(k+4)T(k) \\
& = (1-\rho)^2\gamma(0)^2 \left\{ \rho^2(\rho-1)(5\rho-1) + 2\rho^2 \frac{1-\rho}{1+\rho} \right\} \\
& \quad (27)
\end{aligned}$$

$$\begin{aligned}
& \sum_{k=-\infty}^{\infty} T(k)T(k-1) \\
& = 2T(1)T(0) + 2 \sum_{k=1}^{\infty} T(k+1)T(k) \\
& = (1-\rho)^2\gamma(0)^2 \left\{ 4(\rho-1) + 2\rho \frac{1-\rho}{1+\rho} \right\} \quad (28)
\end{aligned}$$

$$\begin{aligned}
& \sum_{k=-\infty}^{\infty} T(k+1)T(k-2) \\
& = 2T(2)T(1) + 2T(3)T(0) + 2 \sum_{k=1}^{\infty} T(k+3)T(k) \\
& = (1-\rho)^2\gamma(0)^2 \left\{ 4\rho^2(\rho-1) + 2\rho \frac{1-\rho}{1+\rho} \right\} \quad (29)
\end{aligned}$$

### $\widehat{T}(1)$ のモーメント

(12) の定義より  $\widehat{T}(1)$  を以下のように分解する。

$$\begin{aligned}
\widehat{T}(1) &= \frac{1}{n} \sum_{i=2}^n \left\{ \Delta P_{t_i}^* \Delta P_{t_{i-1}}^* - \Delta P_{t_i}^* \Delta \varepsilon_{t_{i-1}} \right. \\
&\quad \left. - \Delta \varepsilon_{t_i} \Delta P_{t_{i-1}}^* + \Delta \varepsilon_{t_i} \Delta \varepsilon_{t_{i-1}} \right\} \\
&= A_0 + A_1 + A_2 + A_3
\end{aligned}$$

$$A_0 = \frac{1}{n} \sum_{i=2}^n \Delta \varepsilon_{t_i} \Delta \varepsilon_{t_{i-1}},$$

$$A_1 = \frac{1}{n} \sum_{i=2}^n \Delta P_{t_i}^* \Delta P_{t_{i-1}}^*,$$

$$A_2 = -\frac{1}{n} \sum_{i=2}^n \Delta P_{t_i}^* \Delta \varepsilon_{t_{i-1}},$$

$$A_3 = -\frac{1}{n} \sum_{i=2}^n \Delta \varepsilon_{t_i} \Delta P_{t_{i-1}}^*$$

### $\widehat{T}(1)$ の期待値

$A_0$  以外はその期待値はゼロであることにより以下を得る。

$$\begin{aligned}
E[\widehat{T}(1)] &= E[A_0] = \frac{1}{n} \sum_{i=2}^n \Delta \varepsilon_{t_i} \Delta \varepsilon_{t_{i-1}} \\
&= \frac{n-1}{n} \{-\gamma(0) + 2\gamma(1) - \gamma(2)\} \\
&= -\gamma(0)(1-\rho)^2 + O(n^{-1}) \\
&= T(1) + O(n^{-1})
\end{aligned}$$

### $\widehat{T}(1)$ の分散

$A_0$  以外のそれぞれの分散の大きさは以下のようになる。

$$\begin{aligned}
\text{Var}[A_1] &= \frac{1}{n^2} \sum_{i=2}^n \{E[(\Delta P_{t_i}^*)^2] E[(\Delta P_{t_{i-1}}^*)^2]\} \\
&= \frac{1}{n^2} \sum_{i=2}^n O(n^{-2}) = O(n^{-3}) \\
\text{Var}[A_2] &= \frac{1}{n^2} \sum_{i=2}^n \{E[(\Delta P_{t_i}^*)^2] E[(\Delta \varepsilon_{t_{i-1}})^2]\} \\
&= \frac{1}{n^2} \sum_{i=2}^n O(n^{-1}) = O(n^{-2}) \\
\text{Var}[A_3] &= O(n^{-2})
\end{aligned}$$

また  $\text{Cov}[A_2, A_3]$  以外の共分散はゼロで、 $\text{Cov}[A_2, A_3]$  については

$$\begin{aligned}
\text{Cov}[A_2, A_3] &= \frac{1}{n^2} \sum_{i=2}^n E[(\Delta P_{t_i}^* \Delta \varepsilon_{t_{i-1}})(\Delta \varepsilon_{t_{i+1}} \Delta P_{t_i}^*)] \\
&= \frac{1}{n^2} \sum_{i=2}^n E[(\Delta P_{t_i}^*)^2] E[\Delta \varepsilon_{t_{i-1}} \Delta \varepsilon_{t_{i+1}}] \\
&= O(n^{-2})
\end{aligned}$$

となる。

$A_0$  の分散に関しては以下のとおり（詳細は Fuller(1976) の Theorem 6.2.1.）。

$$\begin{aligned}
 \lim_{n \rightarrow \infty} n \operatorname{Var}[A_0] &= (\eta - 3)T(1)^2 \\
 &+ \sum_{k=-\infty}^{\infty} \{T(k)T(k) + T(k+1)T(k-1)\} \\
 &= (\eta - 3)T(1)^2 + \sum_{k=-\infty}^{\infty} \{T(k)^2 \\
 &\quad + T(k+1)T(k-1)\} \\
 &= (1 - \rho)^2 \gamma(0)^2 \left\{ (\eta - 3)(1 - \rho)^2 \right. \\
 &\quad \left. + (\rho - 1)(3\rho + 1) + \frac{8}{1 + \rho} \right\}
 \end{aligned}$$

以上より,  $\widehat{T}(1)$  の分散に関して次の (30) を得る。

$$\begin{aligned}
 \lim_{n \rightarrow \infty} n \operatorname{Var}[\widehat{T}(1)] &= (1 - \rho)^2 \gamma(0)^2 \left\{ (\eta - 3)(1 - \rho)^2 \right. \\
 &\quad \left. + (\rho - 1)(3\rho + 1) + \frac{8}{1 + \rho} \right\} \quad (30)
 \end{aligned}$$

## $\widehat{T}(2)$ のモーメント

$\widehat{T}(1)$  と同様に

$$\begin{aligned}
 \widehat{T}(2) &= \frac{1}{n} \sum_{i=3}^n \left\{ \Delta P_{t_i}^* \Delta P_{t_{i-2}}^* - \Delta P_{t_i}^* \Delta \varepsilon_{t_{i-2}} \right. \\
 &\quad \left. - \Delta \varepsilon_{t_i} \Delta P_{t_{i-2}}^* + \Delta \varepsilon_{t_i} \Delta \varepsilon_{t_{i-2}} \right\} \\
 &= B_0 + B_1 + B_2 + B_3
 \end{aligned}$$

$$B_0 = \frac{1}{n} \sum_{i=3}^n \Delta \varepsilon_{t_i} \Delta \varepsilon_{t_{i-2}}, \quad B_1 = \frac{1}{n} \sum_{i=3}^n \Delta P_{t_i}^* \Delta P_{t_{i-2}}^*,$$

$$B_2 = -\frac{1}{n} \sum_{i=3}^n \Delta P_{t_i}^* \Delta \varepsilon_{t_{i-2}}, \quad B_3 = -\frac{1}{n} \sum_{i=3}^n \Delta \varepsilon_{t_i} \Delta P_{t_{i-2}}^*$$

## $\widehat{T}(2)$ の期待値

$$\begin{aligned}
 E[\widehat{T}(2)] &= E[B_0] = \frac{n-2}{n} \{\gamma(2) - \gamma(3) \\
 &\quad - \gamma(1) + \gamma(2)\} \\
 &= -\gamma(0)(1 - \rho)^2 \rho + O(n^{-1}) \\
 &= T(2) + O(n^{-1})
 \end{aligned}$$

## $\widehat{T}(2)$ の分散

$\widehat{T}(1)$  と同様に  $\operatorname{Var}[B_1] = O(n^{-3})$ ,  $\operatorname{Var}[B_2] = O(n^{-2})$ ,  $\operatorname{Var}[B_3] = O(n^{-2})$ , また  $\operatorname{Cov}[B_2, B_3]$  以外の共分散はゼロで,  $\operatorname{Cov}[B_2, B_3]$  については

$$\begin{aligned}
 \operatorname{Cov}[B_2, B_3] &= \frac{1}{n^2} \sum_{i=3}^{n-2} E[(\Delta P_{t_i}^* \Delta \varepsilon_{t_{i-2}})(\Delta \varepsilon_{t_{i+2}} \Delta P_{t_i}^*)] \\
 &= \frac{1}{n^2} \sum_{i=3}^{n-2} E[(\Delta P_{t_i}^*)^2] E[\Delta \varepsilon_{t_{i-2}} \Delta \varepsilon_{t_{i+2}}] \\
 &= O(n^{-2})
 \end{aligned}$$

となる。さらに  $B_0$  の分散に関しては下記のとおりでである。

$$\begin{aligned}
 \lim_{n \rightarrow \infty} n \operatorname{Var}[B_0] &= (\eta - 3)T(2)^2 \\
 &+ \sum_{k=-\infty}^{\infty} \{T(k)^2 + T(k+2)T(k-2)\} \\
 &= (1 - \rho)^2 \gamma(0)^2 \left\{ (\eta - 3)(1 - \rho)^2 \right. \\
 &\quad \left. + \rho^2(\rho - 1)(5\rho - 1) - 2(\rho - 1)^2 + \frac{8}{1 + \rho^2} \right\}
 \end{aligned}$$

となる。以上より  $\widehat{T}(2)$  の分散に関して (31) を得る。

$$\begin{aligned}
 \lim_{n \rightarrow \infty} n \operatorname{Var}[\widehat{T}(2)] &= (1 - \rho)^2 \gamma(0)^2 \left\{ (\eta - 3)(1 - \rho)^2 \right. \\
 &\quad \left. + \rho^2(\rho - 1)(5\rho - 1) - 2(\rho - 1)^2 + \frac{8}{1 + \rho^2} \right\} \quad (31)
 \end{aligned}$$

$\widehat{T}(1)$  と  $\widehat{T}(2)$  の共分散

共分散  $\text{Cov}[A_i, B_j]$ ,  $(i, j = 0, \dots, 3)$  でゼロとならないのは,  $\text{Cov}[A_0, B_0]$ ,  $\text{Cov}[A_2, B_2]$ ,  $\text{Cov}[A_2, B_3]$ ,  $\text{Cov}[A_3, B_3]$  であり, それぞれは以下のとおりである。

$$\begin{aligned} \lim_{n \rightarrow \infty} n \text{Cov}[A_0, B_0] &= (\eta - 3)T(2)T(1) \\ &+ \sum_{k=-\infty}^{\infty} \{T(k)T(k-1) + T(k+1)T(k-2)\} \\ &= (1-\rho)^2\gamma(0)^2 \left\{ (\eta-3)(1-\rho)^2\rho \right. \\ &\quad \left. + 4(\rho-1)(1+\rho^2) + 4(2-\rho) - \frac{8}{1+\rho} \right\} \end{aligned}$$

$$\begin{aligned} \text{Cov}[A_2, B_2] &= \frac{1}{n^2} \sum_{i=3}^n E[(\Delta P_{t_i}^* \Delta \varepsilon_{t_{i-1}})(\Delta P_{t_i}^* \Delta \varepsilon_{t_{i-2}})] \\ &= \frac{1}{n^2} \sum_{i=3}^n E[(\Delta P_{t_i}^*)^2] E[\Delta \varepsilon_{t_{i-1}} \Delta \varepsilon_{t_{i-2}}] \\ &= O(n^{-2}) \\ \text{Cov}[A_2, B_3] &= \frac{1}{n^2} \sum_{i=2}^{n-2} E[(\Delta P_{t_i}^* \Delta \varepsilon_{t_{i-1}})(\Delta \varepsilon_{t_{i+2}} \Delta P_{t_i}^*)] \\ &= \frac{1}{n^2} \sum_{i=2}^{n-2} E[(\Delta P_{t_i}^*)^2] E[\Delta \varepsilon_{t_{i-1}} \Delta \varepsilon_{t_{i+2}}] \\ &= O(n^{-2}) \\ \text{Cov}[A_3, B_3] &= \frac{1}{n^2} \sum_{i=2}^{n-1} E[(\Delta \varepsilon_{t_i} \Delta P_{t_{i-1}}^*)(\Delta \varepsilon_{t_{i+1}} \Delta P_{t_{i-1}}^*)] \\ &= \frac{1}{n^2} \sum_{i=2}^{n-1} E[(\Delta P_{t_{i-1}}^*)^2] E[\Delta \varepsilon_{t_i} \Delta \varepsilon_{t_{i+1}}] \\ &= O(n^{-2}) \end{aligned}$$

以上より,  $\widehat{T}(1)$  と  $\widehat{T}(2)$  の共分散に関して (32) を得る。

$$\begin{aligned} \lim_{n \rightarrow \infty} n \text{Cov}[\widehat{T}(2), \widehat{T}(1)] \\ &= (1-\rho)^2\gamma(0)^2 \left\{ (\eta-3)(1-\rho)^2\rho \right. \\ &\quad \left. + 4(\rho-1)(1+\rho^2) + 4(2-\rho) - \frac{8}{1+\rho} \right\} \quad (32) \end{aligned}$$

## Estimation of Autocorrelation of Market Microstructure Noise

Kosuke Oya

Accurate estimation of a volatility of financial asset is much important since a variety of financial tradings require it. The standard estimator of the integrated volatility of the asset price is a realized volatility. We expect that the realized volatility will be accurate one when we use a high frequency data since the sample size is quite large. However, the realized volatility does not give a reliable estimate due to the market microstructure noise in practice. Although this microstructure noise is considered as an independently distributed random variable in an early stage of the research, it is well known that the microstructure noise is time-dependent. The structure of such time-dependence is helpful for the analysis of the investors trading strategy. In this paper, we suppose that a simple autoregressive model for the microstructure noise and propose to estimate the parameter of the autoregressive model of the noise process. We also show the asymptotic normality of the estimator. We conduct a series of Monte Carlo experiments to see the properties of the estimator. It is confirmed that the empirical distribution of the estimator is well approximated by the asymptotic distribution derived in this paper except the case of the autoregressive parameter is close to unity.

JEL Classification: C13; C22; D49

Key words: Hight Frequency Data; Market Microstructure Noise; Serial Correlation; Asymptotic Distribution