

マーケット・マイクロストラクチャー・ノイズがある場合のボラティリティ推定に関する最近の研究*

Recent Studies on Estimation of Volatility in the Presence of Market Microstructure Noise

金谷太郎^{† ‡ §}
滋賀大学経済学部

2009年9月

概要

本稿は高頻度データを用いてボラティリティ推定する際に代表的な手法であるリアライズド・ボラティリティ (RV) に関するサーベイ論文である。特にその計測に大きな影響を与えるマーケット・マーケット・マイクロストラクチャー・ノイズ (MMN) を考慮に入れなければならない場合の対処法を中心に最近の研究を概観する。

*本稿の作成にあたり京都大学経済研究所・奥井亮氏から有益なコメントを多数頂いた。ここに記して感謝したい。また、日本学術振興会科学研究費 (研究課題番号: 21730174) による財政的支援にも感謝する。なお、本稿に含まれる誤りや漏れは全て筆者に属するものである。

[†] 〒 522-8522 滋賀県彦根市馬場 1-1-1 滋賀大学経済学部

[‡] HP: <http://www.biwako.shiga-u.ac.jp/sensei/t-kanatani/>

[§] Email: t-kanatani@biwako.shiga-u.ac.jp

I 序論

ボラティリティと呼ばれる量は、ファイナンスの分野で最も重要視される量で、特に計量ファイナンスの分野では中心的な研究対象である。それは金融資産の収益率の分散（あるいは標準偏差）で定義され、金融資産がもつリスクを計測する際に利用され、ときにはリスクそのものを示すものとしてとらえられる。定義上、ボラティリティは直接観測できず、それが時間を通じて変化していくような現象をモデル化するには、ARCH (Autoregressive Conditional Heteroscedasticity) 型モデルやSV (Stochastic Volatility) モデルなどで潜在変数として間接的にモデル化されるのが主流であった（この種のボラティリティ変動モデルに関しては、Ghysels, Harvey, and Renault [1996], Morimune [2007] など多くの優れたサーベイ論文がある）。

しかし、近年金融高頻度データの利用可能性が拡大したことによって事情が変わった。Andersen, Bollerslev, Diebold, and Labys [2003] などの一連の研究によって、本来潜在変数である日次ボラティリティの代理変数として高頻度データから算出したリアライズド・ボラティリティ (Realized Volatility, あるいは Realized Variance, 以下 RV) を用いて、ボラティリティを直接モデル化することの有用性が指摘された¹⁾。これ以降、RV を用いたボラティリティのモデル化に関する研究、一方で RV が真のボラティリティをいかに効率よく推定できるかに関する研究が飛躍的に増大した。前者の RV のモデル化を特に取り上げた日本語のサーベイ論文としては森棟 [2007] や渡辺 [2007] などがあるので、本稿では後者の RV やその修正推定量が真のボラティリティをどれだけ精度よく推定することができるかという研究に焦点をあてる。

この研究課題は古くから研究されてきた確率過程の2次変動の理論に基づけば一見非常に簡単な問題に見える。すなわち、RV の精度を上げるには、可能な限り高頻度なデータを使えばよいというわけである。しかし、実際の金融データはマーケット・マイクロストラクチャー・ノイズ (以下 MMN) を含むと考えられるため、あまりに高頻度なデータを使うことは避けるべきとされてきた。それゆえ、実証研究者は MMN の影響を受けないと考えられるデータの頻度を経験則や主観的な判断によって決定しなければならなかった

¹⁾冒頭で触れたとおり、ボラティリティは収益率の分散として定義される場合と、その平方根の標準偏差として定義される場合、またその両者として定義される場合がある。ボラティリティを収益率の標準偏差として分散とはっきり区別する場合、Realized Volatility は Realized Variance の平方根であるが、本稿では両者を分散として扱い区別せず、いずれにしても RV と略すことにする。

のである。

Zhou [1996] は「対数価格 + ノイズ」というモデルで為替レートの収益率系列の一次の負の相関を表現し、その結果として RV に上方バイアスが生じることを明らかにした。ノイズにやや強い仮定はあったものの、その仮定の下でバイアス修正推定量とその分散の振る舞いを明らかにした。この研究が、その後の「対数価格 + ノイズ」というタイプのモデルをつかった MMN 分析の先駆けとなった。そもそも MMN をどのようにモデル化するかは、議論の分かれるところであるが、本稿ではこの Zhou [1996] で提案された「対数価格 + ノイズ」という現在計量ファイナンスの分野で主流となっているモデルとその応用を中心にとりあげることにする。

Zhou [1996] 以降の研究の流れを簡単にまとめると次のようになる。Bandi and Russell [2008] は Zhou [1996] のモデルのノイズの仮定を緩め、その場合の RV の有限標本の平均二乗誤差 (MSE) を導出し、その MSE を最小にするデータの頻度を決定する方法を確立した。この研究によって、それまで研究者が主観的に決めざるをえなかった RV に使うデータの頻度の決定に理論的根拠を与えられた。また、その MSE を計算する際に必要なノイズのモーメントを計算する方法も示されており、その後の同種の研究に対して重要な貢献となっている。

Zhang, Mykland, and Ait-Sahalia [2005] は 2 種類の異なるデータ頻度で計算された RV を使ってノイズがある場合でも一致性をもつ推定量を提案した。この推定量は Two Scale Estimator (TS 推定量) と呼ばれ、高頻度データをバイアス修正、低頻度データをサブサンプリングとよばれる手法を介して分散の収束に寄与するように構成されている。また、Hansen and Lunde [2006] は Newey and West [1987] の長期分散推定量を基にして Kernel-Based Estimator とよばれる推定量を提唱している。この推定量は TS 推定量とほぼ同じであり、両者の漸近的な振る舞いは全く変わらない。また、Hansen and Lunde [2006] にはダウ・ジョーンズ工業平均株価指数に属する 30 種類の株価の MMN に関する実証研究がまとめられており、ノイズと株価の相関、あるいはノイズの系列相関など、独立ノイズの仮定では捉えきれないという結果が報告されている。

さらに、Barndorff-Nielsen, Hansen, Lunde, and Shephard [2008] は Hansen and Lunde [2006] の Kernel-Based Estimator の標本自己共分散に対する重み付けを一般化することによってより広いクラスの推定量を考察した。その一般化された推定量は Realized Kernel (以下 RK) と名付けられ、その漸近分

布が導出されている。それによって重み付けを決めるカーネル関数の形状と真のボラティリティへの収束速度の関係を整理している。Sun [2006] では RK よりも一般的なクラスの推定量を定義し、有限標本において最適な推定量を導出している。ただし、導出の過程で、かなり強いノイズの仮定や推定量のクラスの制限が必要である。

本稿の構成は以下の通りである。次章では RV の定義とその理論的根拠となる二次変動の理論を紹介する。第 III 章では「対数価格 + ノイズ」というタイプのモデルを導入し、RV に及ぼすノイズの影響をみる。第 IV 章では RV を修正しノイズの影響を回避するさまざまな手法とそれらの性質を概観する。第 V 章では結語として今後のこの分野の研究の展望などに簡単に触れる。

II 2次変動とRV

この節では導入として、RV の理論的根拠である 2 次変動の理論を概観するために、資産価格が MMN の影響を受けずに観測される場合を考える。 p_t を時刻 t における金融資産の真の対数価格とし、それがブラウン運動セミマルチンゲールというクラスの確率過程にしたがっているとす。簡単化のためにドリフト項は考慮しないことにする²⁾。すなわち、

$$p_t = p_0 + \int_0^t \sigma_s dW_s,$$

ただし、 σ_t は cadlag (左連続で右極限をもつ) 確率過程で、 W は標準ブラウン運動である。瞬間的な収益率の分散 $V(dp_t) = \sigma_t^2$ をスポット・ボラティリティとよび、これが興味の対象になる場合もあるが、ここでは区間 $[0, 1]$ における積分値を推定の対象にする。これを Integrated Volatility (あるいは Integrated Variance, 以下 IV) とよぶ。すなわち、 $IV = \int_0^1 \sigma_t^2 dt$ である。時間の単位は日であることが多く、このとき IV は日次ボラティリティなどとよばれる。

時刻 $0 = t_0, t_1, \dots, t_n = 1$ の対数価格の系列を $\{p_{t_i}\}_{i=0}^n$ とする。また、簡単化のために等間隔データの場合、すなわち $t_i - t_{i-1} = 1/n$ の場合を考える³⁾。

²⁾高頻度データを扱う文献ではしばしばこの簡単化が採用される。高頻度収益率を想定した微小区間ではドリフト項はブラウン運動の項に対して無視できるほど小さくなるからである。

³⁾等間隔データの仮定は特に有限標本での MSE の表現やノイズの系列相関の定義に影響を与える。本稿では 1 変数のボラティリティ推定を主に紹介するので、共分散推定の場合、

対数価格の一階差分（対数価格比） $x_i = p_{t_i} - p_{t_{i-1}}$ は期間 $[t_{i-1}, t_i]$ の収益率をあらわす。このとき，

$$p \lim_{n \rightarrow \infty} \underbrace{\sum_{i=1}^n x_i^2}_{RV_*^{(n)}} = \int_0^1 \sigma_t^2 dt, \quad (1)$$

が成立する。(1)の左辺を2次変動という。その有限和の部分（ $\sum_{i=1}^n x_i^2$ ）を Realized Volatility とよび， $IV(\int_0^1 \sigma_t^2 dt)$ の推定量として使う。このノイズを考慮しない場合の RV を $RV_*^{(n)} = \sum_{i=1}^n x_i^2$ と表記することにする。このとき RV の漸近分布が知られており，

$$\sqrt{n}(RV_*^{(n)} - IV) \xrightarrow{L} MN(0, 2IQ),$$

ただし， $IQ = \int_0^1 \sigma_t^4 dt$ ，となる⁴⁾。MN は混合正規分布をあらわし，分散が確率変数であることをあらわす（Barndorff-Nielsen and Shephard [2002]）。MMN を考慮しない場合，言い換えれば真の資産価格データが直接観測できる場合は，データ数を増やせば増やすほど RV の推定誤差分散を $1/n$ のオーダーで減らすことができる。

III マーケット・マイクロストラクチャー・ノイズ

Zhou [1996] は，収益率系列の一次の負の相関を表現し，その結果としての RV の上方バイアスの原因を明らかにするために，以下のような「対数価格 + ノイズ」というモデルを導入した。実際に観測される対数価格 \tilde{p}_t はノイズ ϵ_t を伴って観測されると考えられ，

$$\tilde{p}_t = p_t + \epsilon_t,$$

である。観測される収益率を $y_i = \tilde{p}_{t_i} - \tilde{p}_{t_{i-1}}$ ，また，ノイズの差分を $\eta_i = \epsilon_{t_i} - \epsilon_{t_{i-1}}$ と書くことにすると，

$$y_i = x_i + \eta_i$$

異なる金融資産が同時に観測されないことが原因となって生じる非同期バイアスを考慮する必要がないため，この簡単化を採用する。

⁴⁾IQ は Integrated Quarticity とよばれる量で， $(1/3) \sum_{i=1}^n x_i^4$ で一致推定できる。

である。

ここではひとまず Zhou [1996] で仮定された独立ノイズの仮定の下で何が起こるかみてみることにする。独立ノイズとは、ノイズ ϵ_{t_i} が対数価格 p_t と自らの系列 $\epsilon_{t_j} (j \neq i)$ の両方に対して独立なノイズのことをいう。また、ノイズの平均はゼロ、分散は ω^2 で一定であるとする。このとき $E(y_i y_{i-1}) = -\omega^2$ となり、観測される収益率は一次の負の系列相関をもつ。もし資産価格過程がセミマルチンゲールであれば、収益率には系列相関はないはずである。しかし、このような高頻度収益率データの負の相関の例は、Zhou [1996] の為替レートの例のほか、Dacorogna, Gencay, Muller, Olsen, and Pictet [2001] などで多数報告されている。

次に MMN があるときの RV のふるまいをみる。観測可能な $\{y_i\}_{i=1}^n$ を使って計算した RV を $RV^{(n)} = \sum_{i=1}^n y_i^2$ と書くことにする。このとき RV は

$$RV^{(n)} = RV_*^{(n)} + 2 \sum_{i=1}^n x_i \eta_i + \sum_{i=1}^n \eta_i^2 \quad (2)$$

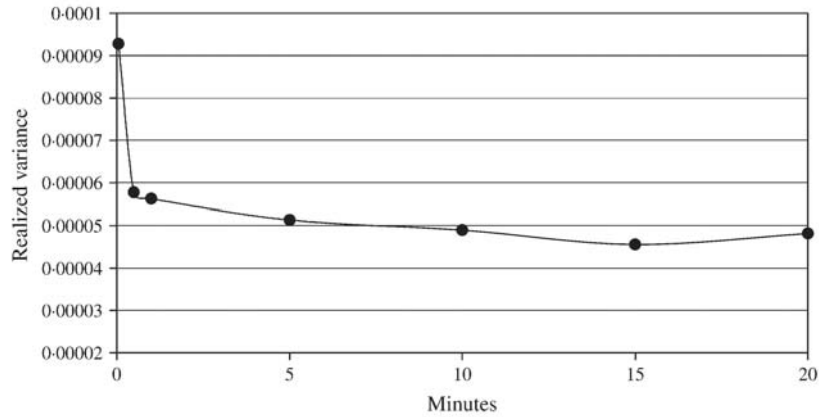
の3つの項に分けることができる。独立ノイズの場合、(2)の第三項の期待値である $2n\omega^2$ がバイアスとなる。これは一定区間内に観測されるデータ数 n に比例して RV の上方バイアスは大きくなることを意味する⁵⁾。図1は Bandi and Russell [2008] に掲載されている IBM 社の株価の Volatility signature plot (または Variance signature plot, 以下 VSP) である。VSP とは横軸に使うデータの頻度、縦軸に対応する頻度で計算された RV の平均をプロットしたもので、RV 関連の文献ではバイアスの程度を視覚的にとらえるためにしばしば用いられる⁶⁾。この図でもあまりに高頻度なデータを使ってしまうと RV の値が極端に大きくなっていく様子がみとれる。

しかし、Hansen and Lunde [2006] はダウ・ジョーンズ工業平均株価指数に属する30種類の株価に関して全ての銘柄の VSP が必ずしも図1のような形状にならないことを報告している。例えば、図1とは全く逆で、高頻度になるほど RV の値が小さくなるような VSP、すなわち下方バイアスを示す銘柄もあった。そのような現象を説明するには独立ノイズでは不十分でノイズの

⁵⁾Delattre and Jacod [1997] は丸め誤差がある場合に RV の上方バイアスが無限に大きくなるという結果を得ている。丸め誤差は金融資産の場合、価格の離散性にあたり MMN のうちの一つとして考えられる。

⁶⁾Andersen, Bollerslev, Diebold, and Labys [2000] によって Volatility signature plot と名付けられて以降、計量ファイナンスの分野で定着したが、Fang [1996] にも同様の図がみられる。

図 1: IBM の Volatility signature plot



Bandi and Russell [2008] より転載。ただし，データ頻度ゼロ分は quote-to-quote データで計算された RV の平均という意味である。

仮定をより一般的にしなければならない。(2) の第三項の期待値は系列相関を考慮してもコーシー・シュワルツの不等式により非負であるので，この下方バイアスが生じるためには，第二項の真の収益率 y_i とノイズの差分 η_i に負の相関がなければならない。また，独立ノイズによるバイアスを修正する推定量（後述 (3) の $RV_{AC1}^{(n)}$ ）を使ってもバイアスが修正しきれていないという実証結果も報告されており，ノイズには系列相関があると結論付けられている。

この節の最後に Bandi and Russell [2008] の RV の最適頻度に関する研究に触れておく。上に述べたとおり MMN がバイアスを引き起こすことはよく知られており，高頻度データを全て使わずにある程度の頻度のデータを使うということは経験的に行われてきた。頻度を下げることによってバイアスは回避できるが，当然推定の分散は大きくなってしまふ。Bandi and Russell [2008] は，Zhou [1996] 以来の「対数価格 + ノイズ」というモデルの枠組みの中で RV に使うための最適頻度を導出した。バイアスと分散のトレードオフを考慮して，最適化の基準として有限標本の MSE が採用された。その MSE を最小にするデータ数はおおよそ $n^* \approx (IQ/E(\eta^2)^2)^{1/3}$ となる。つまり，最大限利用できるデータ数 n に対して最適な頻度は n/n^* 毎のデータを使って RV を計算することである。ところで，この最適頻度や MSE を評価するためにはノイズの

モーメントを知る必要がある。Bandi and Russell [2008] では、適当な条件の下で

$$\text{p} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n y_i^q = E(\eta^q), \quad q = 2, 3, 4$$

と一致推定できることが示されている。

IV さまざまな修正推定量

本節ではボラティリティ推定における MMN の影響を回避するためのさまざまな RV の修正法に関する研究を紹介する。

Zhou [1996] では上述のとおり独立ノイズが仮定されていたため、収益率には一次の自己相関があった。それゆえ、以下のように一次の標本自己共分散項を追加することによってバイスを修正することが考えられた。

$$RV_{AC1}^{(n)} = RV^{(n)} + \sum_{i=2}^n y_i y_{i-1} + \sum_{i=1}^{n-1} y_i y_{i+1} \quad (3)$$

独立ノイズなら、第二項・第三項の期待値はともに $-n\omega^2$ なので、 $RV^{(n)}$ のバイス $2n\omega^2$ と打ち消しあう。また、ボラティリティ一定の条件を追加して、有限標本での MSE とそれを最小にするデータ数（およそ $\sqrt{3}\sigma^2/(2\omega^2)$ ）も導出されている。この推定量も RV 同様一致性をもたないが、サブサンプリングという手法を使って分散が改善されることが示唆されている。例えば、1分収益率のデータをもっていたとして、MMN の影響を考えて 15分収益率のデータを使うことを考える。15分収益率のデータは 15通りでき、15分 RV 推定量も 15個計算できる。その平均をとるのがサブサンプリングである。

Zhang, Mykland, and Ait-Sahalia [2005] はそのサブサンプリングを使って一致推定量を提案した。上述のように低頻度収益率データ系列を K 通りつくることを考えたとき、その k 番目の系列を次のように定義する。

$$y_{j,k} = \sum_{i=1}^K y_{i+K(j-1)+k-1}, \quad j = 1, 2, \dots, m_k$$

ただし, $m_k = \lfloor n/K \rfloor$ で $[\cdot]$ はガウス記号である。 k 番目の低頻度データから計算された RV を

$$RV^{(K,k)} = \sum_{j=1}^{m_k} y_{j,k}^2,$$

とする。 $\{RV^{(K,k)}\}_{k=1}^K$ のうち $RV^{(K,1)}$ だけを使うのが, 従来の低頻度データ RV を使って MMN を避ける方法であった。しかし, RV のサブサンプリングとは以下のように K 個の RV の平均をすることをいう。すなわち,

$$SRV^{(K)} = \frac{1}{K} \sum_{k=1}^K RV^{(K,k)}, \quad (4)$$

であり, 低頻度データを使うことによってバイアスを軽減し, MSE を減少させる効果が期待できる⁷⁾。また, K がデータ数に対して適当な程度で大きくなれば分散を収束させることもできると考えられる。Zhang, Mykland, and Ait-Sahalia [2005] は (4) の低頻度データを使ったサブサンプリング RV のバイアスを高頻度データ RV を使って修正し, 以下の推定量を提案した。

$$TS^{(K)} = SRV^{(K)} - \frac{\bar{m}}{n} RV^{(n)}, \quad (5)$$

ただし, $\bar{m} = (1/K) \sum_{k=1}^K m_k$ である。この推定量は高頻度・低頻度の2つの異なる頻度のデータを使ってつくられているので Two Scale 推定量とよばれている。独立ノイズというやや強い仮定となっているが, MMN の影響下でも一貫性をもつ初の RV 推定量と認識されている。漸近分布は本稿では省略するが, 漸近正規性をもち, $n^{1/6}$ のオーダーで真のボラティリティ, すなわち IV に収束する。MSE 基準の最適な \bar{m} の目安は $(IQ/(3\omega^2))^{1/3}$ である。Zhang [2006] はさらに異なる頻度のデータを使うことで TS 推定量より効率を改善することができることを示している。この推定量は Multi-Scale 推定量 (以下 MS 推定量) とよばれ, 収束のオーダーは $n^{1/4}$ である。これは Ait-Sahalia, Mykland, and Zhang [2005] がパラメトリックな設定の下で導出した最尤推定量と同じオーダーである。

Hansen and Lunde [2006] は前節で触れたように独立ノイズの仮定では彼らの実証結果を捉えきれないことを示した。そこで, Newey and West [1987] の

⁷⁾ただし, あくまでバイアスの「軽減」であり, バイアスはいくらか残るので (5) の第二項による修正が必要である。

自己相関に対してロバストな長期分散推定量を応用して、(3)の一期ラグ項に加えて、 H 期までのラグをバートレット・ウェイトで重み付けて追加した。

$$HL^{(H)} = \gamma_0 + \sum_{h=1}^H \left(1 - \frac{h-1}{H}\right) (\gamma_h + \gamma_{-h}), \quad (6)$$

ただし、 $\gamma_q = \sum_{i=1}^n y_i y_{i-q}$ であり、 $\gamma_0 = RV^{(n)}$ 、また $H = 1$ のときは $HL^{(1)} = RV_{AC1}^{(n)}$ である。この推定量は Kernel-Based Estimator とよばれ、TS 推定量とは異なるモチベーションで考案されたが、両者はほぼ同じ推定量であることがわかり、漸近的性質は同一である。

Barndorff-Nielsen, Hansen, Lunde, and Shephard [2008] は (6) の自己共分散項への重み付けを一般化した以下の推定量を分析した。

$$RK^{(H)} = \gamma_0 + \sum_{h=1}^H k\left(\frac{h-1}{H}\right) (\gamma_h + \gamma_{-h}), \quad (7)$$

ただし、 $k(x)$ は定義域が $[0, 1]$ で、確率変数ではない。この推定量を Realized Kernel といい、既出の推定量を全て含んでいる。漸近分布はカーネル関数 $k(x)$ の形状によって異なる。

表 1 に主なカーネル関数と RK の収束のオーダー、効率（漸近分散の相対的な大きさ）をまとめた。上段の 3 つのカーネルは $k(0) = 1$ 、 $k(1) = 0$ を満たす。 $H = c_0 n^{2/3}$ (c_0 は推定可能な定数) としたとき、RK が IV に $n^{1/6}$ のオーダーで収束する。Bartlett カーネルは上述の通り、Zhang, Mykland, and Ait-Sahalia [2005] の TS 推定量と Hansen and Lunde [2006] の Kernel-Based Estimator と同じである。

中段のカーネルには $k'(0) = k'(1) = 0$ という条件が加わる。このクラスの推定量は $H = c_0 n^{1/2}$ としたとき、 $n^{1/4}$ のオーダーで収束する。Ait-Sahalia, Mykland, and Zhang [2005] がボラティリティー一定、正規独立ノイズなどのパラメトリックな設定の下で導出した最尤推定量と同じオーダーを達成している。Turkey-Hanning カーネルの次数を増やしていくと、最尤推定量の効率 8 に近づく。なお、Zhang [2006] の MS 推定量は RK の形で書くと Cubic カーネルを持つ事がわかっている。

下段は定義域に $[0, \infty)$ をとるカーネル、すなわち無限のラグの自己共分散項 $\gamma_h + \gamma_{-h}$ をとることが可能なカーネルである。これらの推定量は $H = (\omega/IQ)n^{1/2}$ としたとき、 $n^{1/4}$ のオーダーで収束する。特に Optimal カーネ

表 1: カーネルの種類とその漸近分散の比較

| カーネル名 | $k(x)$ | オーダー | 効率 |
|-------------------------|--|-----------|------|
| Bartlett(TS) | $1 - x$ | $n^{1/6}$ | 0.76 |
| Epanechnikov | $1 - x^2$ | $n^{1/6}$ | 1.31 |
| 2nd order | $1 - 2x + x^2$ | $n^{1/6}$ | 0.68 |
| Cubic(MS) | $1 - 3x^2 + 2x^3$ | $n^{1/4}$ | 9.04 |
| 5th order | $1 - 10x^3 + 15x^4 - 6x^5$ | $n^{1/4}$ | 10.2 |
| 6th order | $1 - 15x^4 + 24x^5 - 10x^6$ | $n^{1/4}$ | 12.1 |
| 7th order | $1 - 21x^5 + 35x^6 - 15x^7$ | $n^{1/4}$ | 13.9 |
| 8th order | $1 - 28x^6 + 48x^7 - 21x^8$ | $n^{1/4}$ | 15.7 |
| Parzen | $1 - 6x^2 + 6x^3$ ($0 \leq x \leq 1/2$) $2(1 - x)^3$ ($1/2 < x \leq 1$) | $n^{1/4}$ | 8.54 |
| Turkey-Hanning1 | $\sin^2(\pi(1 - x)/2)$ | $n^{1/4}$ | 9.18 |
| Turkey-Hanning2 | $\sin^2(\pi(1 - x)^2/2)$ | $n^{1/4}$ | 8.29 |
| Turkey-Hanning5 | $\sin^2(\pi(1 - x)^5/2)$ | $n^{1/4}$ | 8.07 |
| Turkey-Hanning16 | $\sin^2(\pi(1 - x)^{16}/2)$ | $n^{1/4}$ | 8.02 |
| Turkey-Hanning ∞ | $\sin^2(\pi \exp(-x)/2)$ | $n^{1/4}$ | 8.01 |
| Quadratic spectral | $(3/x^2)((\sin(x)/x) - \cos x)$ | $n^{1/4}$ | 9.38 |
| Dirichlet(Fourier) | $\sin(x)/x$ | $n^{1/4}$ | 11.7 |
| Fejer | $(\sin(x)/x)^2$ | $n^{1/4}$ | 8.89 |
| Optimal | $(1 + x)e^{-x}$ | $n^{1/4}$ | 8.00 |

効率（各カーネル関数を使ったときの RK の漸近分散の相対的な大きさ）は同じ収束のオーダー間で比較可能である。Ait-Sahalia, Mykland, and Zhang [2005] の最尤推定量はオーダー $n^{1/4}$ で効率 8 となり、下限となる。

ルを使ったときその漸近分散の1次の項は Ait-Sahalia, Mykland, and Zhang [2005] の最尤推定量のそれと同じである。しかし, 無限ラグの自己共分散項を実用する際, 有限個で打ち切ることによる近似の問題や計算量の問題もあり, 中段の有限の Turkey-Hanning 等の漸近効率が十分下限の8に近いので, あえて下段のカーネルを使うことに必ずしもメリットを感じられない。

また, 表1下段のクラスの Dirichlet カーネルを使った RK は, Malliavin and Mancino [2002] のフーリエ推定量の RK 表現となっている。フーリエ推定量は他の時間領域でのボラティリティ推定量とは異なり, 周波数領域での分析により導出された新しい推定量であると認識され, MMN に対してロバストであるというシュミレーションあるいは実証研究の結果が Oya [2006] や Nielsen and Frederiksen [2008] などで報告されてきた。また, Kanatani [2004b] や Mancino and Sanfelici [2007] によって RK の形に書き換えることができることが示され, ノイズに由来するバイアスを緩和する理由が理論的に明らかになった。しかし, 表1にもあるように, 他のカーネルに比べて効率において劣っているので実用において魅力がある推定量であるとは言いがたい。

RK に関してはそのサブサンプリング版 (Barndorff-Nielsen, Hansen, Lunde, and Shephard [2007]) や実用法 (Barndorff-Nielsen, Hansen, Lunde, and Shephard [2009]) が議論されつつあり, 今後ボラティリティ推定において標準的な手法になることが予想される。

最後に有限標本での最適化に取り組んでいる Sun [2006] の研究を紹介する。以下のような二次形式は既存の推定量全てを含む⁸⁾。

$$WRV = \sum_{i=1}^n \sum_{j=1}^n w_{ij} y_i y_j,$$

例えば, この表現において RK は $|i - j| = h$ 毎に同じ重み付けをした特殊形となる。Sun [2006] は正規独立ノイズ, ボラティリティ一定, さらに推定量のクラスを不偏推定量に限って, 有限標本 MSE を最小化するウェイト付けが以下のように明示的に書けることを示した。

$$w_{ij}^* = \frac{\lambda}{\sqrt{\lambda(\lambda + 4)}} \frac{\cosh(\alpha(n + 1 - |i - j|)) - \cosh(\alpha(n + 1 - i - j))}{\sinh(\alpha(n + 1))},$$

⁸⁾この推定量のクラスは Kanatani [2004a] において定義され, Weighted Realized Volatility と名付けられている。Sun [2006] ではこの推定量を Quadratic (Form) Estimator とよんでいる。

ただし,

$$\lambda = \frac{IV}{n\omega^2}, \quad \alpha = \log \left(\frac{1}{2}\lambda + \frac{1}{2}\sqrt{4\lambda + \lambda^2 + 1} \right),$$

である。この最適ウェイトを使った推定量は最良二次不偏推定量 (Best Quadratic Unbiased Estimator, 以下 BQU 推定量) と名付けられている。最適ウェイト w_{ij}^* には推定すべき対象の IV が入っており (ありえない設定だが) もし IV が既知ならば, 収束のオーダーは $n^{1/2}$ であり, TS 推定量などで計算した推定値を使った場合には, 先に述べた RK や最尤推定量同様, BQU 推定量は $n^{1/4}$ で収束することも証明されている。

V 結語

以上, 高頻度データを用いたボラティリティの推定問題を RV とその修正推定量を中心に概観した。まず, 近年の高頻度データの利用可能性の拡大が RV という古典的で簡便な手法に光を当てた。それを受けて, 高頻度データを扱う上で避けることができない MMN の問題に「対数価格 + ノイズ」という Zhou [1996] 以降のモデルを用いて, 多くの研究者がこの問題に取り組んでいる。本論でみたように, 簡単なバイアス修正, 最適頻度の導出, 一致推定量の構成, ノイズの仮定の拡張, さらに推定の効率をあげる手法の開発, そしてそらの推定量を一般化した統一理論といった研究がここ数年の間に飛躍的に進んだ。しかし, 「対数価格 + ノイズ」というモデルの下で RV 推定量の精度をあげる方向には研究は進んでいるが, Hansen and Lunde [2006] のようなノイズの性質を十分に議論するような実証研究はまだ少ない。このことは, その利用可能性が拡大したとはいえ, 十分な高頻度データが一部の欧米の研究機関を除いて必ずしも容易に利用できないことにも原因がある。特に欧米に比べて日本の金融高頻度データは利用が難しいのが現状である。しかし, 今後高頻度データの利用可能性がより高まり実証研究が進めば, 広範な市場と金融資産の MMN の統計的性質が明らかになるはずである。現在は「対数価格 + ノイズ」というデータ記述型のモデルも, 既存のマーケット・マイクロストラクチャーの理論 (O'Hara [1995] など) との整合性や解釈が議論され, ファイナンスの理論研究においても大きく貢献していくと期待される。

本稿では主に分散推定についての研究を紹介したが, 最後に共分散推定について少し触れておく。恒等式 $4 \cdot Cov(X, Y) = V(X + Y) - V(X - Y)$ から

共分散に関しては分散の推定がわかれば十分のように考えられがちである。しかし、取引データのような金融高頻度データを扱う場合は、 X と Y が同時観測されず、 $X + Y$ や $X - Y$ といったポートフォリオが観測不可能であるので、共分散推定には分散推定にはない固有の問題が含まれることになる⁹⁾。

Hayashi and Yoshida [2005] は同時観測されない2つの系列の共分散の不偏推定量を提案して、その一致性を示した。この推定量は非同期バイアスの問題を解決する手法として認識され、後に Cumulative Covariance Estimator (CC 推定量) と名付けられた。Griffin and Oomen [2009] は、上述の Bandi and Russell [2008] が RV の最適頻度を導出したのと同様に、CC 推定量の有限標本の MSE を分析し、CC 推定量を計算する際に最適なデータの頻度を求めた。Voev and Lunde [2008] では、より一般的な状況で CC 推定量のバイアスあるいは分散が考察され、そのサブサンプリング版の有用性が指摘されている。有限標本では Kanatani [2008] がカーネルの選択法、Kanatani [2009] が CC 推定量のサブサンプリングについて研究している。しかし、全体としては共分散の場合は観測の非同期性を考慮しなければならないため、特に漸近論において、分散の場合に比べて研究は進んでおらず、その重要性を鑑みれば今後の研究の進展が期待される。

参考文献

- 森棟公夫 [2007]. “実現ボラティリティ,” 現代経済学の潮流 2007, pp. 3–32.
- 渡部敏明 [2007]. “Realized volatility – サーベイと日本の株式市場への応用,” 経済研究 58(4), pp. 352–373.
- Ait-Sahalia, Y., Mykland, P., and Zhang, L. [2005]. “How often to sample a continuous-time process in the presence of market microstructure noise,” *Review of Financial Studies* 18, pp. 351–416.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. [2000]. “Great realizations,” *Risk* 13(3), pp. 105–108.

⁹⁾データの同時観測が仮定されている場合で、かつノイズを考えない場合の Realized Covariance (RC) Matrix については Barndorff-Nielsen and Shephard [2004], ノイズを考える場合の RC の最適頻度については Bandi and Russell [2005] などがある。

- Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. [2003]. “Modeling and forecasting realized volatility,” *Econometrica* **71**(2), pp. 579–625.
- Bandi, F. M. and Russell, J. R. [2005]. “Realized covariation, realized beta and microstructure noise,” Working Paper, University of Chicago.
- Bandi, F. M. and Russell, J. R. [2008]. “Microstructure noise, realized variance, and optimal sampling,” *The Review of Economic Studies* **75**, pp. 339–369.
- Barndorff-Nielsen, O., Hansen, P. R., Lunde, A., and Shephard, N. [2007]. “Subsampling realized kernels,” *Journal of Econometrics*, forthcoming.
- Barndorff-Nielsen, O., Hansen, P. R., Lunde, A., and Shephard, N. [2008]. “Designing realized kernels to measure the ex post variation of equity prices in the presence of noise,” *Econometrica* **76**(6), pp. 1481–1536.
- Barndorff-Nielsen, O., Hansen, P. R., Lunde, A., and Shephard, N. [2009]. “Realized kernels in Practice: Trades and Quotes,” *Econometrics Journal*, forthcoming.
- Barndorff-Nielsen, O. and Shephard, N. [2002]. “Econometric analysis of realized volatility and its use in estimating stochastic volatility models,” *Journal of the Royal Statistical Society, Series B* **64**, pp. 253–280.
- Barndorff-Nielsen, O. and Shephard, N. [2004]. “Econometric analysis of realized covariation: high frequency covariance, regression and correlation in financial economics,” *Econometrica* **72**, pp. 885–925.
- Dacorogna, M. M., Gencay, R., Muller, U., Olsen, R. B., and Pictet, O. V. [2001]. *An Introduction to High-Frequency Finance*, Academic Press, London.
- Delattre, S. and Jacod, J. [1997]. “A central limit theorem for normalized functions of the increments of a diffusion process in the presence of round off errors,” *Bernoulli* **3**, pp. 1–28.

- Fang, Y. [1996]. “Volatility modeling and estimation of high-frequency data with Gaussian noise,” Ph. d. thesis, MIT, Sloan School of Management.
- Ghysels, E., Harvey, A. C., and Renault, E. [1996]. “Stochastic volatility,” in *Handbook of Econometrics, Vol. 14, Statistical Methods in Finance*, ed. by G. S. Maddala and C. R. Rao. Amsterdam: North-Holland, 119–191.
- Griffin, J. E. and Oomen, R. C. A. [2009]. “Covariance measurement in the presence of non-synchronous trading and market microstructure noise,” Working Paper, University of Warwick.
- Hansen, P. R. and Lunde, A. [2006]. “Realized variance and market microstructure noise” (with discussions), *Journal of Business & Economic Statistics* **24**, pp. 127–161.
- Hayashi, T. and Yoshida, N. [2005]. “On covariance estimation of non-synchronously observed diffusion processes,” *Bernoulli* **11**, pp. 359–379.
- Kanatani, T. [2004a]. “High frequency data and realized volatility,” Ph. d. thesis, Graduate School of Economics, Kyoto University.
- Kanatani, T. [2004b]. “Integrated volatility measuring from unvenly sampled observations,” *Economics Bulletin* **3**(36), pp. 1–8.
- Kanatani, T. [2008]. “Finite sample analysis of weighted realized covariance with noisy asynchronous observations,” Working Paper, Kyoto University.
- Kanatani, T. [2009]. “Subsampling cumulative covariance estimator,” Working Paper, Shiga University.
- Malliavin, P. and Mancino, M. E. [2002]. “Fourier series method for measurement of multivariate volatilities,” *Finance and Stochastics* **6**, pp. 49–61.
- Mancino, M. E. and Sanfelici, S. [2007]. “Robustness of Fourier estimator of in the presence of microstructure noise,” *Computational Statistics and Data Analysis* **52**, pp. 2966–2989.

- Morimune, K. [2007]. “Volatility models,” *The Japanese Economic Review* **58**(1), pp. 1–23.
- Nielsen, M. O. and Frederiksen, P. [2008]. “Finite sample accuracy and choice of sampling frequency in integrated volatility estimation,” *Journal of Empirical Finance* **15**, pp. 265–286.
- Newey, W. and West, K. [1987]. “A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix,” *Econometrica* **55**, pp. 703–708.
- O’Hara, M. [1995]. *Market microstructure theory*, Oxford: Blackwell Publishers.
- Oya, K. [2005]. “Measurement of volatility of diffusion processes with noisy high frequency data,” *Proceeding of MODSIM05*, pp. 940–945.
- Sun, Y. [2006]. “Best quadratic unbiased estimators of integrated variance in the presence of market microstructure noise,” Working Paper, University of California, San Diego.
- Voev, V. and Lunde, A. [2007]. “Integrated covariance estimation using high-frequency data in the presence of the noise,” *Journal of Financial Econometrics* **5**, pp. 68–104.
- Zhang, L. [2006]. “Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach,” *Bernoulli* **12**, pp. 1019–1043.
- Zhang, L., Mykland, P., and Ait-Sahalia, Y. [2005]. “A tale of two time scales: determining integrated volatility with noisy high-frequency data,” *Journal of the American Statistical Association* **100**, pp. 1394–1411.
- Zhou, B. [1996]. “High-frequency data and volatility in foreign exchange rates,” *Journal of Business & Economic Statistics* **14**(1), pp. 45–52.