修士論文 要旨

拡散過程の非同期観測データに対する共変動の検定手法

数理情報学専攻 48216211　小田川拓利

指導教員　荻原 哲平 准教授

1　はじめに

日内の株価データを分析する際には，異なる株価が同時刻に観測されないことを考慮した拡散過程の非同期観測モデルが用いられる．二株式間の共変動は株価の連動性を測る上で重要な指標であり，2 つの拡散過程が非同期で離散時間のみ観測される場合の共変動を推定する問題においては，規則的な間隔にデータを補間し，従来の同期データに基づく共変動推定量を用いることがまず考えられる．しかし，規則的な間隔のサイズとデータ補間スキームの選択により，信頼性のない推定につながる可能性があった．[1] では元データの「同期」処理を一切行わず，従って，それに起因する偏りやその他の問題のない新しい推定量を提案し，一致性・漸近正規性を示した．一方，異なる株式間の共変動がゼロになるかどうかも重要な問題であるが，このような拡散過程の非同期観測モデルに対して，共変動がゼロかどうかの検定問題は今まで研究されていなかった．本研究では，[2] の漸近分散の推定量を構成し，共変動の有無に関する検定を可能とする検定統計量を構築し，データ数が増大する極限に於いて検定統計量が漸近理論として望ましい一致性をみたすことを確認する．また，数値実験において，ポアソン過程から生成される非同期観測時刻を用いた拡散過程のシミュレーションでこの検定統計量が帰無仮説に対する検出力を有することを確認する．

2　非同期観測下の共変動の推定

日内の株価データのモデルとして，拡散過程の非同期観測モデルを考える．本章ではまず [2] における拡散過程の共変動の推定量の先行研究を紹介する．以下では $T > 0$ を固定する．時刻 $t \in [0,T]$ における 2 つの価格・対数価格などの高頻度金融データを扱う．$l = 1,2$ に対して l 番目の株式の時刻 t での株価を P_t^l と書き，P_t^l が伊藤過程

$$dP_t^l = \sigma_t^l dW_t^l \qquad 0 \le t \le T, \quad l = 1,2$$

に従うとする．ただし，$W_t = (W_t^1, W_t^2)$ は二次元ブラウン運動, σ_t^l は t について Lipschitz 連続で, 非ランダムで Lipschitz 連続な $(\rho_t)_{t\in[0,T]}$ に対して，$\langle W^1, W^2\rangle_t = \int_0^t \rho_s ds$ とする．

これらをランダムな区間 $(I^i)_{i=1,2,\cdots}, (J^j)_{j=1,2,\cdots}$ ごとに観測することを考える．$T_i^1 := \inf I^{i+1}$, $T_j^2 := \inf J^{j+1}$, $K_{ij} := 1_{\{I^i \cap J^j \neq \varnothing\}}$ とおく．I^i, J^j は添え字 n にも依存し，n を与える毎に観測時刻列が定まるとして，$n \to \infty$ とすることでデータ数が増大する極限を表す．以下では，$\Delta X(I) := X_{\sup I} - X_{\inf I}$, $v(I) := \Delta\left\langle P^1, P^2\right\rangle(I)$, $v^l(I) := \Delta\left\langle P^l, P^l\right\rangle(I)$ $(l = 1,2)$ とおく．[1] では共変動の不偏推定量として，データの同期化を用いない以下の推定量（林・吉田推定量）を提案している．

$$\mathrm{HY}_n := \sum_{i,j} \Delta P^1(I^i)\Delta P^2(J^j)K_{ij}.$$

ここで $\Delta X(I) := X_{\sup I} - X_{\inf I}$ である．この推定量は次の仮定のもとで一致性や漸近正規性といった漸近理論における望ましい性質を満たす．具体的には,

[Condition (C1)]: (i) $(I^i), (J^i)$ は P^1, P^2 と独立 (ii) $\mathrm{E}[\max_i |I^i| \vee \max_j |J^j|] = o(1) \quad (n \to \infty)$

[Condition (C2)]: 数列 $(b_n) \subset (0,1)$ と $c \in \mathbb{R}$ で

$$b_n^{-1}\left(\sum_{ij} v^1(I^i)v^2(J^j)K_{ij} + \sum_i v(I^i)^2 + \sum_j v(J^j)^2 - \sum_{ij} v(I^i \cap J^j)^2\right) \xrightarrow{P} c$$

であり，かつある $q \in (0,1/4)$ で

$$\max_i |I^i| \vee \max_j |J^j| = o_P(b_n^{\frac{3}{4}+q}).$$

という仮定をおくことで，次の定理が成り立つ．

[Theorem 3.1 in [1], Theorem 2 in [2]] (C1) のもと $\mathrm{HY}_n \to \theta$ in L^2 が成り立つ．(C1)–(C3) のもと $b_n^{-1/2}(\mathrm{HY}_n - \theta) \xrightarrow{d} N(0,c)$ が成り立つ．

3　林・吉田推定量の漸近分散の推定量と検定手法

共変動 θ がゼロになるかどうかを非同期観測データから検定する問題を考える．$\theta = 0$ は対応する拡散過程の連動性がないことを意味するので，これが検定できれば株価データの相関関係を特定していくことができるため，重要な問題となる．

林・吉田推定量の推定誤差は漸近的に正規分布に従うため，その分散 c の一致推定量：$U_n \xrightarrow{P} c$ を構築できれば，その推定量のルートで林・吉田推定量の推定誤差を割ることで

$$\frac{\mathrm{HY}_n - \theta}{\sqrt{b_n U_n}} \xrightarrow{d} N(0,1)$$

を得る．その為，帰無仮説「$\theta = 0$」に対して，検定統計量 $T_n := \mathrm{HY}_n/\sqrt{b_n U_n}$ を定めることで，帰無仮説の下，$T_n \xrightarrow{d} N(0,1)$ が成立する．よって，有意水準を $\alpha \in (0,1)$ として，標準正規分布の上側 100α %点 z_α に対して，$P(|T_n| > z_{\alpha/2}) \to \alpha$ が成立するので，$|T_n| > z_{\alpha/2}$ の時に帰無仮説を棄却することができる．

そのため，上の検定手法を正当化するために林・吉田推定量の漸近分散 c の一致推定量 U_n を構築する．具体的には，Condition (C2) の第 1 式左辺の統計量は c に確率収束するので，(C2) 第 1 式左辺の推定量を構築する．推定量は次のようになる．ただし，(C1), (C2) に加え，次の仮定が必要となる．

Condtion (C3): σ_t^1, σ_t^2, ρ_t は t に依存しないとする．

論文では (C3) の代わりに $b_n^{-1}\sum_{ij}|I^i \cap J^j|^2 = O_P(1)$ を仮定した場合についても示しているが具体的な式は省略する．

(C1) – (C3) が満たされる時，次が成り立つ．

$$U_n := b_n^{-1}\left(\sum_{ij}(Y_i^1)^2(Y_j^2)^2K_{ij} + \frac{1}{3}\sum_i (Y_i^1)^4 + \frac{1}{3}\sum_j (Y_j^2)^4 - \frac{3}{T^2}\left(\sum_{ij} Y_i^1 Y_j^2 K_{ij}\right)^2 \sum_{ij}|I^i \cap J^j|^2\right) \xrightarrow{P} c.$$

ここで，$Y_i^1 = \Delta P^1(I^i)$, $Y_j^2 = \Delta P^2(J^j)$ である．したがって，この U_n を用いることで，前述の方法での検定が可能となる．

4 数値実験

以下では特に $b_n = 1/n$, $T = 1$ とおく．拡散過程 P_t^1, P_t^2 が定数 θ に対して，$P_0^1 = 0, P_0^2 = 0$ a.s.

$$dP_t^l = \sqrt{1 + \left(\frac{\theta}{2}\right)^2}\, dW_t^l, \quad 0 \le t \le 1, l = 1, 2$$

に従うとする．ただし，$\rho = \theta/\left(1 + \left(\frac{\theta}{2}\right)^2\right)$ であり，$W_t = (W_t^1, W_t^2)$ は $\langle W^1, W^2 \rangle_t = \rho t$ を満たす二次元ブラウン運動とする．このとき，$\langle P^1, P^2 \rangle_T = \theta$ となる．

N_t^1, N_t^2 をポアソン過程として，強度パラメータをそれぞれ $\lambda_1 = 0.2n$, $\lambda_2 = 0.3n$ とする．P_t^l の観測時刻 T_i^j は N_t^l から以下のように生成されるとする：

$$T_i^l = \inf\{t \ge 0; N_t^l \ge i\}.$$

ポアソン過程から生成される観測時刻に対して，(C1)–(C3) が成立することを確認できる．

各 n, θ について 1000 回シミュレーションを行い，検定を行った結果棄却された回数を示したのが表 1 である．n が大きくなるにつれて θ が大きな値をとる場合にほぼすべてのケースで帰無仮説 H_0 が棄却され，逆に $\theta = 0$ では棄却される回数が有意水準 5% 程度に収まっているため，提案手法による仮説検定がうまくいっていることが確認できる．また，$n = 10000, \theta = 0$ として，1000 回の実験結果の $\sqrt{n}\,\mathrm{HY}_n/\sqrt{U_n}$ をヒストグラムと QQ-plot で表したものが，図 1 である．漸近理論の結果と整合的になっていることがわかる．

表 1. 各 n, θ について 1000 回の仮説検定の結果，帰無仮説 H_0 が棄却された回数.

$n \setminus \theta$	0.00	0.10	0.20	0.30	0.40	0.50
500	54	83	178	331	498	640
1000	55	112	304	571	803	920
3000	31	268	743	976	999	1000
5000	54	412	923	999	1000	1000
10000	46	701	996	1000	1000	1000

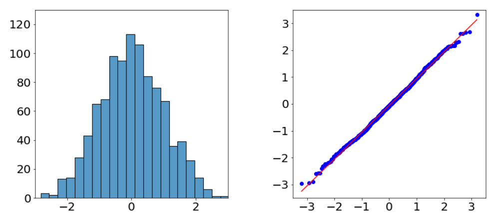

図 1. $n = 10000, \theta = 0$ について 1000 回実験を行った際の $\sqrt{n}\,\mathrm{HY}_n/\sqrt{U_n}$ のヒストグラムと QQ-plot.

参考文献

[1] Takaki Hayashi and Nakahiro Yoshida. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, Vol. 11, No. 2, pp. 359–379, 2005.

[2] Takaki Hayashi and Nakahiro Yoshida. Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes. *Annals of the Institute of Statistical Mathematics*, Vol. 60, No. 2, pp. 367–406, 2008.