

# 経験類似度に基づくボラティリティ予測

森本 孝之<sup>1</sup>・川崎 能典<sup>2,3</sup>

(受付 2016 年 7 月 25 日；改訂 10 月 20 日；採択 10 月 28 日)

## 要 旨

事例ベース意思決定理論に基礎を置いた経験類似度(ES, Empirical Similarity)という概念を適用することにより、異なるモデルから生じるボラティリティ予測値を結合する。経験類似度の枠組みでは、意思決定者が予測モデルや予測値の尤もらしさに関する確率評価を行わずに、専ら類似性のみによって将来を予測することができる。具体的には、過去のモデル予測値と対応するボラティリティの実現値との距離を定量化することによって、予測の組合せの重みを決定する。そして、決定された重みを用い将来のボラティリティを予測する。本稿では、この経験類似度モデルから得られたボラティリティの予測値とその他時系列モデルの予測値とを実証的に比較する。モデルの予測力比較については、誤差関数に基づくモデル信頼集合(Model Confidence Set, MCS)を用いることにより、複数の銘柄と推定予測期間におけるモデルの予測力を順位付けし、最良モデルの累積頻度を分析し評価する。

キーワード：経験類似度、実現測度、HARQ, ESQ, モデル信頼集合。

## 1. はじめに

類推による推論は、過去の経験に基づき、未来の出来事を予測する基本的な方法の1つである(Gilboa et al., 2011)。帰納的推論の論理的妥当性に疑問を持ち、また類推的推論を議論したことで有名な Hume (1748)は、将来について過去から学ぶ基本思想である。一般的に、不確実あるいは無知な状況において、意思決定者は将来の確率を評価できない、あるいは評価したがらないが、将来について過去から学びその類似性に基づき思考することは可能である。より現代的に表現すれば、不確実性下の意思決定におけるフォンノイマン・モルゲンシュテルンの期待効用理論では、意思決定者は起こりうる状態をすべて列挙しつくした状態空間と、その上の確率分布から計算される期待効用を最大化すべく行動すると想定されている(尾山, 2012)。しかし、意思決定者が状態空間を完全に把握していると想定するにはあまりにも無理がある状況も多々あり、そのような状況での意思決定についての1つの考え方は、人々は過去の経験からの類推に基づいて現時点での行動を決めるであろう、というものである。これが Gilboa and Schmeidler (1995, 2001)の提唱する事例ベース意思決定理論である(尾山, 2012)。この類似性に基づく推論は、医学、法律、ビジネス、政治、あるいは人工知能における意思決定に幅広く応用されている(Gilboa and Schmeidler, 2001)。この事例ベース意思決定理論は、以前経験した過去の状況との類似性を考慮することによって、現状を評価するという類推的思考を意思決定者に仮定している

<sup>1</sup> 関西学院大学 理工学部：〒 669-1337 兵庫県三田市学園 2-1

<sup>2</sup> 統計数理研究所：〒 190-8562 東京都立川市緑町 10-3

<sup>3</sup> 総合研究大学院大学 複合科学研究科統計科学専攻：〒 190-8562 東京都立川市緑町 10-3

(Gilboa and Schmeidler, 2001). 現在の状況に類似した事例は、あまり類似していない事例と比較して、より大きい重みが与えられる。この考え方が事例ベース意思決定理論に基づく経験類似度 (ES, Empirical Similarity) の概念 (Gilboa et al., 2006, 2011) であり、Gilboa and Schmeidler (2012) によりデータから類似度関数を推定する計量経済学的枠組みが提供された。これにより、意思決定者によって認識される事例 (問題, 状況) 間の距離を計測することが可能となった。本稿では、非確率的な手法で異なるモデルから得られる予測値を組み合わせるために、Golosnoy et al. (2014) の提案した経験類似度の概念を利用する方法を用いる。ここでの設定では、競合するモデルから得られた異なる予測値は、現在観測された状態あるいは実現値にある程度類似している事例として評価される。直近でより正確な点予測値を与えるモデルには、その他のモデルと比較して、より大きい現在の重みを与える。Golosnoy et al. (2014) の核となるアイデアは、現在の観測値と、異なるモデルから得られる直近の 1 期先予測値との間の経験類似度距離を計測することである。この類似度距離により、次の期のモデルの重みが決定される。したがって、この経験類似度によるモデル組合せ手法は、予測モデルの組合せの重みを決定するために、異なるモデルの直近の予測力に関する情報を利用する。Golosnoy et al. (2014) によれば、その他の確率的な手法と比較して、この経験類似度によるモデル組合せ手法を用いる利点として、以下の 3 点が考えられる。

- (1) モデルの事後確率や予測値の平均二乗誤差 (Mean Squared Error, MSE) などを算出する必要がない。
- (2) 経済主体の選好に予測モデルの重みを関連づけられる。
- (3) 意思決定者が予測値と実現値間の類似度をどのように評価するかをデータから明らかにできる。

本稿の実証研究では、前述の Golosnoy et al. (2014) により提案された経験類似度によるモデル組合せ手法を、日次実現ボラティリティ過程をモデル化することにより分析する。この目的のために、先行研究同様、Corsi (2009) によって提案された、過去の異なる投資期間における推定結果をボラティリティの予測に反映できる HAR (Heterogeneous AutoRegressive) モデルの組合せに対する経験類似度を評価する。この実証研究において予測力評価に利用されるデータは、1999 年 1 月から 2013 年 12 月まで 15 年分の株価指数 6 銘柄と東京証券取引所 1 部上場の個別 24 銘柄の 1 分間隔高頻度データから得られた日次実現ボラティリティである。データの標本期間については、1999 年から 2013 年までの、インサンプル、アウトオブサンプルを含む計 225 通りの推定予測期間を分析対象とする。この 225 通りの組合せの内訳は、インサンプル 120 通り、アウトオブサンプル 105 通りである。これらインサンプルとアウトオブサンプルにおける予測を行うことにより、複数の一般的なボラティリティモデルに対して、この経験類似度によるモデル組合せ手法の予測力を比較する。予測力の比較に関しては、インサンプルとアウトオブサンプルにおいて得られた各モデルの誤差関数値を統計的仮説検定の枠組みで適切に予測力の評価を行うため、ここでは Hansen et al. (2011) が提案するモデル信頼集合 (Model Confidence Set, 以下 MCS) を用いる。MCS により、特定の真のモデルを仮定すること無しに、所与の有意水準での最良なモデル選択が可能となる。最後に、Mincer and Zarnowitz (1969) により提案された予測力を評価する一般的な手法の 1 つである Mincer-Zarnowitz (MZ) 回帰を、各モデルの予測値に対して実行し、得られた自由度調整済み決定係数の値を比較する。

本稿は、次のように構成される。第 2 節では、本稿の理論的背景となる経験類似度の統計モデルを詳細に解説する。第 3 節では、実証分析に用いるデータを説明した後、モデルの予測力比較を MCS および MZ 回帰により行う。第 4 節では、本稿における実証分析の結果を纏め、将来の研究に対する方向付けを示唆し結論とする。

## 2. 理論的背景

ここでは, Gilboa et al. (2011) および Golosnoy et al. (2014) に基づき, 本稿で用いられる経験類似度の理論的背景について解説する.

### 2.1 経験類似度

ある変数  $y_t$  の値を, 関連する変数の値  $x_t = (x_t^1, \dots, x_t^d)$  によって構成されるデータベースに基づき評価する. 例えば,  $y_t$  は, 家具の骨董品の価格であるとしよう. ここで  $x_t$  はそのスタイル, 製造年, 大きさなどといった特性値を表すとする.  $y_t$  を評価するために, 過去の観測値  $x_i$  と現在の値  $x_t$  とをいかにして結合すべきであろうか? もし Hume (1748) のアイデアに従うならば, 過去の条件  $x_i = (x_i^1, \dots, x_i^d)$  が  $x_t$  と似ているか, あるいは似ていないかを表す類似度の考えが必要となる. ここでは  $y_t$  の予測において, より似ていない条件の下で得られた観測値よりも, より似ている条件の下で得られた観測値により高い重みを与えたい. 上の例では, 最近売られた同様の骨董品の価格によって, この骨董品の価格を評価することは, 道理にかなっている. さらに, 「スタイル, 製造年, 大きさ, および売りだされた時期」に関して過去の観測値が現在の観測値により似ているほど, 現在の評価において, この観測値により大きい重みを置きたいと考える.

形式的に, 類似度関数  $s: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}_{++} = (0, \infty)$  を仮定し, データベース  $(x_i, y_i)_{i \leq n}$  と新しいデータ点  $x_t = (x_t^1, \dots, x_t^d) \in \mathbb{R}^d$  が与えられたならば,  $y_t$  の類似度予測子は

$$(2.1) \quad y_t^s = \frac{\sum_{i < t} s(x_i, x_t) y_i}{\sum_{i < t} s(x_i, x_t)}$$

と定式化できる. あるいは, もし  $(x_t, y_t)_{t \leq n}$  におけるデータ点の順序が任意ならば

$$(2.2) \quad y_t^s = \frac{\sum_{i \neq t} s(x_i, x_t) y_i}{\sum_{i \neq t} s(x_i, x_t)}$$

と定義することもできる. 類似度関数  $s$  については, いくつかの弱い仮定を満たすならば (Lieberman, 2010), 任意の関数形で表すことが可能である. 例えば Billot et al. (2008) は, 次の形をとる類似度関数と同値な類似度加重平均についての条件を与えている.

$$s(x, x') = \exp(-\|x - x'\|)$$

ここで  $\|\cdot\|$  は  $\mathbb{R}^d$  におけるノルムである. 具体的には, ここでは加重ユークリッド距離によって定義されたノルムの族に焦点を当てると

$$s_w(x, x') = \exp(-d_w(x, x'))$$

となる. ここで  $w \in \mathbb{R}_+^d$  は, 次式で与えられる 2 つのベクトル  $x, x' \in \mathbb{R}^d$  間の距離の加重ベクトルである.

$$(2.3) \quad d_w(x, x') = \sum_{j=1}^d w_j (x_j - x'_j)^2$$

従ってこの定式化では, 類似度関数は各予測子を含めたパラメータの  $d$ -次元ベクトルとなる.

統計的推論を実行し, 仮説検定により定性的結果を得るために, (2.1) 式と (2.2) 式は統計モデルに組み込むことができる. すなわち, それぞれ

$$(2.4) \quad y_t = \frac{\sum_{i < t} s_w(x_i, x_t) y_i}{\sum_{i < t} s_w(x_i, x_t)} + \varepsilon_t$$

および

$$(2.5) \quad y_t = \frac{\sum_{i \neq t} s_w(x_i, x_t) y_i}{\sum_{i \neq t} s_w(x_i, x_t)} + \varepsilon_t$$

というモデルを考える．ここで  $\{\varepsilon_t\}$  は  $\text{iid}(0, \sigma^2)$  である．(2.4) 式は，ある種の因果モデルとして解釈できる．例えば，経済主体による価格形成過程を考えてみると，この経済主体は，過去において既に価格が決定されている他の商品との類似度に応じて，不動産や美術品のような商品の価格を決定するだろう．したがって，(2.4) 式は，価格を決定する際の経済主体が関与する思考過程のモデルとして考えることができる．モデル (2.5) 式は，同様の方法で直接的に解釈することができない．各  $y_t$  の分布が他のすべての  $y_t$  に依存するので，(2.5) 式は，その過程の時間的進展を説明することができない．しかしながら，そのような相互依存性は，空間統計学の応用分野として一般的な地理学，社会学，あるいは政治学データでは自然に解釈できる．

## 2.2 経験類似度とカーネル推定量の関係

ここでは，説明を簡明にするため 1 次元，すなわち  $d = 1$  の変数として  $X$  が存在する場合を考える．ノンパラメトリック回帰モデルでは，通常次のようなデータ生成過程を仮定する．

$$y_i = m(x_i) + \varepsilon_i, \quad (i = 1, \dots, n), \quad \varepsilon_i \sim \text{iid}(0, \sigma^2)$$

ここで  $m: \mathbb{R} \rightarrow \mathbb{R}$  は  $x$  と  $y$  を関係づける未知の関数である．広く用いられている  $m(\cdot)$  のノンパラメトリック推定量は，ナダラヤ・ワトソン推定量であり

$$\hat{m}(x_t) = \frac{\sum_{i=1}^n K\left(\frac{x_i - x_t}{h}\right) y_i}{\sum_{i=1}^n K\left(\frac{x_i - x_t}{h}\right)}$$

のように定義される．ここで  $K(x)$  は，カーネル関数，すなわち，他の正則条件と同様に  $\int K(z) dz = 1$  を満たす非負の関数であり， $h$  は，バンド幅パラメータである．たとえば，もしガウシアンカーネルを選択すれば

$$(2.6) \quad \frac{1}{h} K\left(\frac{x_i - x_t}{h}\right) = (2\pi h^2)^{-1/2} \exp\left(-\frac{(x_i - x_t)^2}{2h^2}\right)$$

が得られる．分散とバイアス間のトレード・オフ関係があるので， $h$  の選択はノンパラメトリック統計学における重要な論点である．最適バンド幅の選択に対して最も一般的な基準の一つは，平均積分二乗誤差を最小化することである．すなわち，最適な  $h$  は

$$h^* = \arg \min_h E_{f_0} \int (\hat{m}(x) - m(x))^2 dx$$

を満たす．ここで期待値  $E_{f_0}$  は， $y$  の真の分布  $f_0$  の下での期待値を意味する．もし  $x$  が可算であり， $m(x)$  を  $y$  に置き換えるならば，誤差の二乗和の期待値を最小化するという基準で  $h^*$  を決めることになる．

いま，カーネルを基にした推定と経験類似度間関係を議論することにしよう．上で説明されたように，経験類似度法は

$$y_t = \frac{\sum_{i=1}^n s_w(x_i, x_t) y_i}{\sum_{i=1}^n s_w(x_i, x_t)}$$

によって  $y_t$  を予測することを提案している．ここで

$$s_w(x_i, x_t) = \exp(-dw) = (\pi/w)^{1/2} \left[ \frac{1}{(1/\sqrt{2w})} K\left(\frac{x_i - x_t}{1/\sqrt{2w}}\right) \right]$$

であり、 $d_w$  は(2.3)式において定義され、 $K$  は(2.6)式において与えられる。そして

$$\frac{\sum_{i=1}^n s_w(x_i, x_t) y_i}{\sum_{i=1}^n s_w(x_i, x_t)} = \frac{\sum_{i=1}^n K\left(\frac{x_i - x_t}{1/\sqrt{2w}}\right) y_i}{\sum_{i=1}^n K\left(\frac{x_i - x_t}{1/\sqrt{2w}}\right)}$$

であり、この設定では  $h = 1/\sqrt{2w}$  という結果となる。

### 2.3 モデル結合のための経験類似度

関心のある変数  $y_t$  を予測するために組み合わせられた  $d$  個のモデル  $x_t = (x_t^1, \dots, x_t^d)$  があると仮定する。Bates and Granger (1969)によれば予測の線形結合とは

$$(2.7) \quad \hat{y}_t = \sum_{j=1}^d a_{t-1}^j x_{t-1}^j$$

によって与えられる。ここで非負の  $a_t^j$  は、 $\sum_{j=1}^d a_t^j \equiv 1$  となる  $j$  番目のモデルの比率を表す。(2.7)式における重み  $a_t^j$  は、モデルや予測値の尤もらしさに関する(確率等の)定量的評価に関連づけた解釈が可能だが、何らかの目的関数を考えることによって、より適切な重み  $a_t^j$  を選択できるかもしれない。Elliott and Timmermann (2004)ではモデルから導かれる MSE の小ささが重み係数に対応している。この重み  $a_t^j$  を適切に選択するために、いくつかの接近法が提案されているが、どれも一般的な方法とまではなっていない。そこで Golosnoy et al. (2014)は、Gilboa et al. (2006)による ES 概念に基づいた予測の線形結合を、次のように定式化した。

$$y_t = \sum_{j=1}^p \phi[y_{t-1}, x_{t-2}^j] x_{t-1}^j + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma^2)$$

この定式化の特徴は、重み  $\phi[y_{t-1}, x_{t-2}^j]$  を得るために必要な一期前の予測値  $x_{t-2}^j$  と対応する実現値  $y_t$  間の距離を計測できることにある。そして、予測値  $x_t = (x_t^1, \dots, x_t^d)$  の加重和である予測の線形結合は

$$\hat{y}_t = \sum_{j=1}^d \phi[y_{t-1}, x_{t-2}^j] x_{t-1}^j$$

によって与えられる。

また、重み  $\phi[\cdot, \cdot]$  は、観測されたデータの過去の値に依存し、現在の実現値の代理変数と  $j$  番目のモデルの予測値間の距離は

$$\phi[y_t, x_{t-1}^j] = \frac{\theta[y_t, x_{t-1}^j]}{\sum_{k=1}^d \theta[y_t, x_{t-1}^k]}$$

のように算出される。重み  $\phi[y_t, x_{t-1}^j] \in [0, 1]$  は、 $\sum_{k=1}^d \phi[y_t, x_{t-1}^k] \equiv 1$  という性質を持つ正規化された相対経験類似度として解釈できる。右辺の  $\theta[y_t, x_{t-1}^j]$  は類似度関数であり、もし  $y_t$  と  $x_{t-1}^j$  間の距離が短ければ、 $\theta[y_t, x_{t-1}^j]$  は高い類似度を示唆する。本稿では前節で導入した次のような Billot et al. (2008)による指数型関数を類似度関数として用いる。

$$\theta[y_t, x_{t-1}^j] = \exp(-\omega_j (y_t - x_{t-1}^j)^2), \quad \omega_j \in \mathbb{R}$$

### 3. 実証分析

実証分析の目的は、東京証券取引所上場の株価指数および個別銘柄の日次ボラティリティの予測を通じて、前節で導入した経験類似度モデルの予測力を評価することにある。この目的の

表 1. 実証研究における分析対象銘柄 (30 銘柄).

株価指数	個別銘柄	
TOPIX	日本たばこ産業	本田技研工業
日経株価指数 300	信越化学工業	キヤノン
日経平均	武田薬品工業	任天堂
東証電気機器株価指数	アステラス製薬	三井物産
東証輸送用機器株価指数	富士フイルム HD	三菱商事
東証銀行業株価指数	新日本製鐵	野村 HD
	小松製作所	三菱地所
	日立製作所	東日本旅客鉄道
	松下電器産業	日本電信電話
	ソニー	KDDI
	日産自動車	NTT ドコモ
	トヨタ自動車	東京電力

ために、ここでは日経メディアマーケティング株式会社より提供された各株価指数と個別銘柄のティックデータを 1 分間隔の高頻度データに加工し用いている。標本期間は、1999 年 1 月 4 日から 2013 年 12 月 30 日までの 15 年分であり、使用した株価指数および個別銘柄は下記の通りである。まず株価指数に関しては、TOPIX、日経株価指数 300、日経平均、東証電気機器株価指数、東証輸送用機器株価指数、東証銀行業株価指数、の計 6 銘柄である。また個別銘柄に関しては、2009 年 4 月 1 日時点での TOPIX Core30 に含まれる銘柄から、1999 年から 2013 年まで継続的に市場で取引された計 24 銘柄を用いる。除外された 6 銘柄については、セブン & アイ HD (1999–2005 年)、JFE-HD (1999–2002 年)、三菱 UFJ-FG (1999–2001 年)、三井住友 FG (1999–2002 年)、みずほ FG (1999–2003 年)、東京海上 HD (1999–2002 年)、である (括弧内は欠損期間)。ここでの実証研究の分析対象に採用された株価指数 6 銘柄と個別 24 銘柄については、表 1 にまとめてある。

本研究における実証分析において用いられる日次時系列データは、株価原系列、その対数収益率、実現ボラティリティ (Realized Volatility,  $RV$ )、および Realized Quarticity ( $RQ$ ) である。なお、 $RV$  および  $RQ$  の対数および平方根を後述する HAR モデルで定式化し分析を行なったが、特に予測力に大きな変化は見られなかったため、これらの結果については割愛した。次に、これらのうち  $RV$  および  $RQ$  という 2 つの実現測度について解説する。

### 3.1 実現測度

金融市場における最も一般的なりスク指標の一つであるボラティリティは、対数収益率の分散あるいは標準偏差として定義される。これまで、ボラティリティを推定するために多くのモデルが提唱されたが、これらのモデルは、基本的にパラメトリックであり、日次、週次、月次といったボラティリティを同じ頻度で取られたデータを用いて推定するよう設計されている。しかし、近年、金融資産価格の日内データが広範に利用可能となり、その日の日次ボラティリティを事後的に算出するために、秒あるいは分刻みに取られた非常に高頻度なデータを利用することが可能となった。

そこで、ここでは日内高頻度データを利用した日次ボラティリティの推定手法の概略を以下に解説する。まず Bollerslev et al. (2016) に従い、ここでは確率微分方程式

$$d \log(P_t) = \mu_t dt + \sigma_t dW_t,$$

によって決定される金融資産価格過程  $P_t$  を考える。ここで  $\mu_t$  と  $\sigma_t$  は、それぞれドリフトと瞬

時的ボラティリティ過程を表し、 $W_t$  は標準ブラウン運動である。なお、ここでのモデルは、理解を容易にするために、ジャンプを含んでいないものとする。本論文の主要な目的は、潜在する日次ボラティリティ、つまり Integrated Variance (IV) を推定し予測することにある。具体的に、日次の IV は形式的に

$$IV_t = \int_{t-1}^t \sigma_s^2 ds$$

によって定義される。この IV は金融市場において直接的に観測できないが、次式で与えられる RV は日内高頻度収益率の二乗和として算出できる。時点  $t = 1, 2, \dots, T$  における株価を  $S_t$ 、その対数収益率を  $r_t = \log S_t - \log S_{t-1}$ 、そして実現ボラティリティは  $RV_t$  と表している。RV については、1 分間隔で標本抽出された日内収益率の 2 乗和として下記のように定義される (Andersen et al., 2001)。

$$RV_t = \sum_{i=1}^{n_t} r_{t,i}^2$$

ここで  $r_{t,i}$  は  $t$  日における  $i$  番目の観測された対数収益率であり、 $n_t$  は  $t$  日における標本数を表す。対数価格過程をセミマルチンゲールにおける連続マルチンゲール部分と考えることにより、この RV は IV の代理変数とみなすことができる。また RV は、IV の一致かつ不偏推定量であることが知られている (McAleer and Medeiros, 2008)。従って、日次 RV の推定には、24 時間の完全な高頻度データが必要となるが、東京証券取引所の取引時間は、前場 9:00–11:00 (2011 年 11 月 21 日以降は 11:30 まで) と後場 12:30–15:00 であり、昼休みと夜間の情報を無視した RV は「1 日分」のボラティリティ測度としては適切ではない。そこで、本稿では Hansen and Lunde (2005) の改良版である Masuda and Morimoto (2012) の手法を推定された RV に適用する。この手法は、前場、昼休み、後場、夜間それぞれに適した重み  $\lambda_1, \lambda_2, \lambda_3$  および  $\lambda_4$  を用い

$$RV_t^{weighted} = \lambda_1 Y_{t,1}^2 + \lambda_2 RV_{t,2} + \lambda_3 Y_{t,3}^2 + \lambda_4 RV_{t,4},$$

を定義する。ここで  $Y_{t,1}^2$  と  $Y_{t,3}^2$  は、それぞれ  $t$  日の夜間と昼休みにおける収益率の 2 乗を表し、 $RV_{t,2}$  と  $RV_{t,4}$  は、それぞれ  $t$  日の前場と後場における RV を表す。以上の修正は、モデルや予測法の構成にかかわらず必ず行う処置であるので、今後、表記の簡略化のため  $RV^{weighted}$  は RV と表す。

さらに、RV は 1 分間隔という比較的短い時点でサンプリングされた日内収益率を用いた場合、マイクロストラクチャーノイズと呼ばれる市場のミクロ構造に起因する観測誤差の影響を受けてしまうことが知られている。このバイアスを緩和するための方法としては、低頻度時間間隔の RV (例えば Andersen and Bollerslev (1997) や Bandi and Russell (2004) など)、サブサンプル法 (Zhang et al., 2005)、あるいはカーネル法を用いることなどが考えられているが、本稿では Hansen and Lunde (2005) に従い Bartlett カーネルを用いた Newey-West (NW) 推定量を用いる。この NW 推定量は前述のサブサンプル法とほぼ同一の推定量であることが Barndorff-Nielsen et al. (2008) によって示されている。

Bollerslev et al. (2016) によれば、上述の RV における推定誤差は、Barndorff-Nielsen and Shephard (2002) の  $\Delta \rightarrow 0$  における漸近分布理論により下記のように特徴づけられる。

$$(3.1) \quad RV_t = IV_t + \eta_t, \quad \eta_t \sim MN(0, 2\Delta IQ_t)$$

ここで  $IQ_t \equiv \int_{t-1}^t \sigma_s^4 ds$  は、Integrated Quarticity (IQ) を表し、MN は混合正規すなわち  $IQ_t$  の実現値に条件付けられた正規分布である。また、IV に対する RV と同様に考えることにより、Realized Quarticity (RQ) は

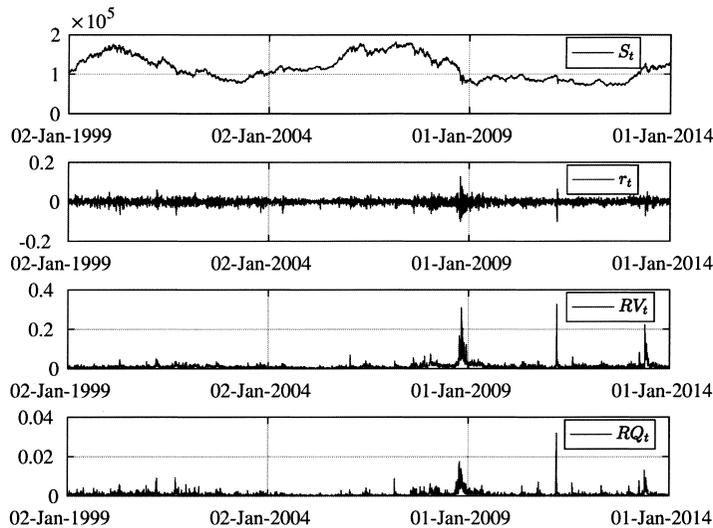


図 1. 株価, 収益率, 実現ボラティリティ, および  $RQ$  (TOPIX).

$$RQ_t \equiv \frac{M}{3} \sum_{i=1}^M r_{t,i}^4$$

によって与えられ, この  $RQ$  は,  $IQ$  の一致推定量となることが知られている. しかし,  $IQ$  の推定には, 測定誤差等多くのノイズを含む日内収益率の 4 次モーメントの推定を含んでいるため, ジャンプの大きさが小さくても  $RQ$  推定量は不安定にならざるを得ない. そこで, 例えば Andersen et al. (2012) は, ジャンプの影響を少しでも低減させるため, 隣接する各収益率の最小値あるいは中央値を使った 2 つの頑健な  $IQ$  の推定量として  $MinRQ$  および  $MedRQ$  を提案しているが, 本稿では  $RV$  と同様に Bartlett カーネルを用いた Newey-West (NW) 推定量を  $RQ$  の推定に用いている.

ここでは紙幅の関係上, 上述の 30 銘柄のうち株価指数と個別銘柄から各 1 銘柄に関する時系列データの特徴を概観する. まず, 図 1 および図 2 は, それぞれ TOPIX, 日立製作所の 1999 年 1 月から 2013 年 12 月までの 15 年間分の株価, 対数収益率, 実現ボラティリティ, および  $RQ$  を図示したものである. これらの図の特徴を見ると, 期間後半に 3 回の大きな変動が見られる. 特に実現ボラティリティの図を見ると, その山を捉えやすい. これらの大きな変動は, 古い方から順に 2008 年 9 月のリーマン・ショック, 2011 年 3 月に発生した東北地方太平洋沖地震, そして 2013 年 5 月 23 日の日経平均大暴落がそれぞれ対応している.

さらに, 表 2 および表 3 は, それぞれ TOPIX, 日立製作所の 1999 年 1 月から 2013 年 12 月までの 15 年間分の株価, 対数収益率, および実現ボラティリティから計算された基本統計量である. これらの表から分かることは, 両者とも対数収益率  $r_t$  の尖度が 3 以上となっており一般的な金融時系列の持つ特徴が表れている. 歪度は両者とも負の値を示しており, これは収益率分布の山が右側つまり正の方向に偏っていることになり, 興味深い結果である. また, 両者とも最大値と最小値は絶対値で 10% を超えており, 平均値はほぼ 0 に等しくなっている.

### 3.2 モデル

ここでは, 実証分析に用いる 15 個の時系列モデルを紹介する. まず, Corsi (2009) によって

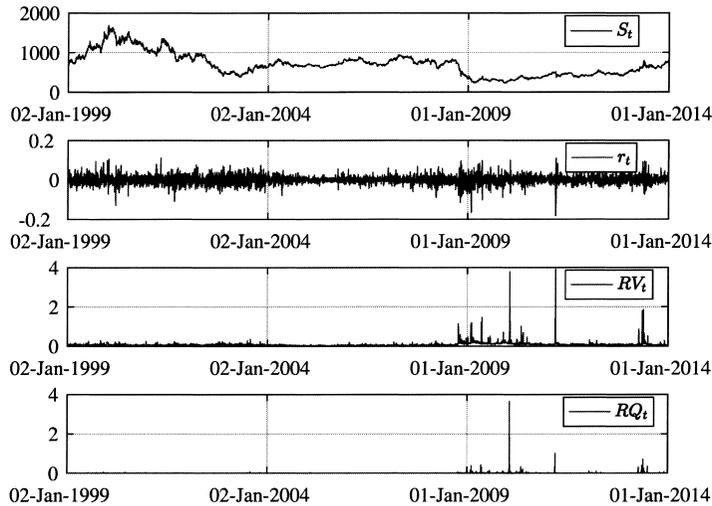


図 2. 株価, 収益率, 実現ボラティリティ, および RQ (日立製作所).

表 2. 基本統計量 (TOPIX) 1999–2013.

	$S_t$	$r_t$	$RV_t$	$\sqrt{RQ_t}$
平均	115815.4168	0.0001	0.0114	0.0008
中央値	111634.0000	0.0003	0.0080	0.0005
最大値	181697.0000	0.1286	0.3255	0.0318
最小値	69551.0000	-0.1001	0.0012	0.0000
標準偏差	30461.9413	0.0141	0.0155	0.0013
歪度	0.4294	-0.3543	9.7217	8.9328
尖度	1.9932	8.8653	144.9977	143.5033
標本数	3685	3684	3664	3664

表 3. 基本統計量 (日立製作所) 1999–2013.

	$S_t$	$r_t$	$RV_t$	$\sqrt{RQ_t}$
平均	701.3720	0.0000	0.0884	0.0095
中央値	683.0000	0.0000	0.0686	0.0051
最大値	1690.0000	0.1105	3.9003	3.6477
最小値	233.0000	-0.1827	0.0203	0.0015
標準偏差	286.7948	0.0233	0.1249	0.0668
歪度	0.7732	-0.1371	18.2170	45.8112
尖度	3.3460	6.7904	481.5358	2424.5612
標本数	3685	3684	3663	3663

提案された HAR (Heterogeneous AutoRegressive) モデルは, 単純な線形回帰の枠組みにおいて, 異なる周期において標本されたボラティリティ測度を組み合わせる. 日次ボラティリティ過程  $v_t$  に対する標準的 HAR モデルは

$$v_t = \alpha_0 + \omega_1 v_{t-1}^{(d)} + \omega_2 v_{t-1}^{(w)} + \omega_3 v_{t-1}^{(m)} + \varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$$

によって与えられる。ここで  $v_{t-1}^{(d)} = v_{t-1}$  は日次、 $v_{t-1}^{(w)}$  および  $v_{t-1}^{(m)}$  は、それぞれ週次平均と月次平均ボラティリティ測度である。これらは、 $v_t^{(w)} = 5^{-1} \sum_{i=1}^5 v_{t-i+1}$  and  $v_t^{(m)} = 22^{-1} \sum_{i=1}^{22} v_{t-i+1}$  として定義される。HAR モデルは、観測できない真の  $v_t$  を実現ボラティリティ  $rv_t$  で代用することにより、OLS 回帰の枠組みで推定できる。ボラティリティ成分に経済学的解釈を与えるとすれば、 $v^{(m)}$  はファンダメンタルなマクロ経済に起因する長期的な不確実性因子、 $v^{(w)}$  は新しくもたらされた情報が市場に浸透するまでの中期的な不確実性因子、 $v^{(d)}$  は投機的活動を反映する短期的な不確実性因子とそれぞれみなすことができるかもしれない (Golosnoy et al., 2014)。

次に、Golosnoy et al. (2014) において 1/3 モデルと呼ばれた HAR モデルの定数項を 0 とし、その他 3 つのパラメータを全て 1/3 の値に固定したモデル

$$v_t = \frac{1}{3}v_{t-1}^{(d)} + \frac{1}{3}v_{t-1}^{(w)} + \frac{1}{3}v_{t-1}^{(m)} + \varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$$

を、本稿内では ES0 モデルと言及する。この ES0 モデルは直接的に経験類似度を用いてはいないが、経験類似度 ES1 モデルのパラメータを  $\theta[v_{t-1}, v_{t-2}] = \theta[v_{t-1}, v_{t-2}^{(w)}] = \theta[v_{t-1}, v_{t-2}^{(m)}] = 1/3$  とした場合と考え ES の名前を冠している。ES0 モデルは、上述のようにパラメータが 1/3 と定数であるのでデータから値を推定する必要が無く、予測力評価の段階においてのみ登場する。

3 つ目のモデルは、本稿において中心的役割を果たす経験類似度モデル ES1 である。ES1 モデルは、経験類似度の概念により、直接的に観測された過去のデータに基づき、ボラティリティモデル成分の予測値に対する重みを推定する。モデル集合  $\mathcal{H}$  に属するモデル  $h$  から得られるボラティリティの予測値を  $v_t^{(h)}$  によって表すならば、ボラティリティ予測のための経験類似度モデルは

$$v_t = \sum_{h \in \mathcal{H}} \phi[v_{t-1}, v_{t-2}^{(h)}] \cdot v_{t-1}^{(h)} + \varepsilon_t = \frac{\sum_{h \in \mathcal{H}} \theta[v_{t-1}, v_{t-2}^{(h)}] \cdot v_{t-1}^{(h)}}{\sum_{h \in \mathcal{H}} \theta[v_{t-1}, v_{t-2}^{(h)}]} + \varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$$

と与えられる。ここで  $\sum_{h \in \mathcal{H}} \phi[v_{t-1}, v_{t-2}^{(h)}] \equiv 1$  であり、 $\theta[v_t, v_{t-1}^{(h)}] = e^{-w_h(v_t - v_{t-1}^{(h)})^2}$  によって定義される類似度関数は、現在のボラティリティの状態  $v_t$  と  $h$  番目のモデルの予測値  $v_{t-1}^{(h)}$  間の距離を計測する。こうして、重み  $\phi[v_t, v_{t-1}^{(h)}] \in [0, 1]$  とモデルの予測値  $v_{t-1}^{(h)}$  を用いることにより  $v_{t+1}$  が予測できる。

本稿では Golosnoy et al. (2014) と同様、ベンチマークとして HAR モデルを用いるので、その 3 つの成分と経験類似度モデルを組み合わせることに焦点を合わせる。ここでの目的は、現在のボラティリティと、異なる期間で標本されたボラティリティの加重和との相対距離が、どのように観測された過去のデータから決定されるかを評価することにある。つまり、異なる投資期間を持つ経済主体がこれらのボラティリティ過程の重みについて、どのように評価しているのか経験類似度を用いて分析したい。今後 ES1 モデルとして表される HAR 成分を持つ経験類似度モデルは

$$(3.2) \quad v_t = \frac{\theta[v_{t-1}, v_{t-2}]v_{t-1} + \theta[v_{t-1}, v_{t-2}^{(w)}]v_{t-1}^{(w)} + \theta[v_{t-1}, v_{t-2}^{(m)}]v_{t-1}^{(m)}}{\theta[v_{t-1}, v_{t-2}] + \theta[v_{t-1}, v_{t-2}^{(w)}] + \theta[v_{t-1}, v_{t-2}^{(m)}]} + \varepsilon_t, \quad \varepsilon_t \sim (0, \sigma^2)$$

と与えられる。ここで

$$\begin{aligned} \theta[v_{t-1}, v_{t-2}] &= \exp(-\omega_1(v_{t-1} - v_{t-2})^2) \\ \theta[v_{t-1}, v_{t-2}^{(w)}] &= \exp(-\omega_2(v_{t-1} - v_{t-2}^{(w)})^2) \\ \theta[v_{t-1}, v_{t-2}^{(m)}] &= \exp(-\omega_3(v_{t-1} - v_{t-2}^{(m)})^2) \end{aligned}$$

である。

ES1 モデルは、異なるボラティリティの単純な重み付き平均を仮定している予測モデルの組合せとして解釈できる。成分  $v_{t-1}$  は 1 日前のボラティリティから得られる予測値である一方、 $v_{t-1}^{(w)}$  と  $v_{t-1}^{(m)}$  直近 1 週間および 1 ヶ月における移動平均の予測値に他ならない。結局のところ、(3.2)式における日次ボラティリティ  $v_t$  は、過去の日次実現ボラティリティの加重平均として表される。(3.2)式より明らかなように、ES1 モデルは HAR モデルよりパラメータが 1 つ少ない、すなわち定数項が無いことに特徴がある。

次に、前節で導入した  $RQ$  を組み入れた 3 つのモデルを紹介する。まず、Bollerslev et al. (2016) は  $IV$  における動的依存関係を、次のようなラグ次数 1 の自己回帰モデルによって記述できると仮定した。

$$IV_t = \phi_0 + \phi_1 IV_{t-1} + u_t$$

ここでは、 $u_t \sim \text{iid}(0, \sigma_u^2)$  であり、 $IV_t$  の観測誤差  $\eta_t \sim \text{iid}(0, \sigma_\eta^2)$  とする。この観測誤差  $\eta_t$  を考慮した  $IV_t$  の AR(1) モデルは

$$(3.3) \quad IV_t + \eta_t = \beta_0 + \beta_1 (IV_{t-1} + \eta_{t-1}) + u_t$$

となる。 $RV$  に対して自己回帰モデルを適用することの形式的な理論的正当化は Andersen et al. (2003) によって与えられた。また Andersen et al. (2004) は、 $RV$  に対して連続時間に基づくモデルではなく単純な離散時間自己回帰モデルを用いる方が  $IV$  の予測力が顕著に向上することを示している。もし  $u_t$  と  $\eta_t$  が両方共 i.i.d. であると仮定するならば、 $\text{Cov}(RV_t, RV_{t-1}) = \phi_1 \text{Var}(IV_t)$  および  $\text{Var}(RV_t) = \text{Var}(IV_t) + 2\Delta IQ$  が成り立つので、 $\beta_1$  は

$$\beta_1 = \phi_1 \left( 1 + \frac{2\Delta IQ}{\text{Var}(IV_t)} \right)^{-1}$$

と表すことができる。したがって、いわゆる希釈バイアス (attenuation bias) により、 $RV$  の係数  $\beta_1$  は  $IV$  の係数  $\phi_1$  よりも小さい値となる。希釈バイアスの詳細については、例えば Wooldridge (2015) が詳しい。(3.2)式より、 $\beta_1$  は観測誤差の分散  $2\Delta IQ$  に依存して変化する。つまり、もし  $2\Delta IQ = 0$  ならば  $\beta_1 = \phi_1$  となるが、もし  $2\Delta IQ$  が大きければ  $\beta_1$  は 0 に近づく。一般的には、(3.2)式の  $\beta_1$  は観測誤差の分散が定数であるという仮定に基づいているが、実際には  $RV$  の推定誤差に関する分散は時間と共に変化する。 $IQ$  が小さい日には、 $RV$  は (3.1)式より  $IV$  に対して高い予測力を持ち、逆に  $IQ$  が大きい日では、 $RV$  は  $IV$  に対して相対的に弱い予測力しか持たない。よって、AR の係数が一定であると仮定するよりも、 $\beta_{1,t}$  といった時間と共に変化する自己回帰係数を仮定する方がより現実的である。

ここで (3.1)式の  $RV_t = IV_t + \eta_t$  という関係から、上の (3.3)式は、 $RV$  についての AR(1) モデル  $RV_t = \beta_0 + \beta_1 RV_{t-1} + u_t$  と考えることができる。この定式化において、 $\beta_1$  を  $IQ^{1/2}$  の推定値  $RQ_{t-1}^{1/2}$  に線形依存させることにより、 $RV$  の AR パラメータを時間と共に変化させることができる。

$$RV_t = \beta_0 + \underbrace{(\beta_1 + \beta_{1Q} RQ_{t-1}^{1/2})}_{\beta_{1,t}} RV_{t-1} + u_t$$

この定式化は、AutoRegressive Quarticity (ARQ) モデルと呼ばれる。このモデルは、標準的な最小二乗法を用いることにより容易に推定でき、自己回帰パラメータ  $\beta_{1,t}$  の値が、 $IV_t$  の観測誤差  $\eta$  の分散の推定値である  $RQ$  と共に変化する。もし  $RQ$  が定数であるならば、ARQ モデルは標準的な AR(1) モデルに帰着することは自明である。

しかし、(3.3)式のような AR(1) モデルでは単純過ぎて、大部分の  $RV$  系列に存在する長期依存構造を満足に記述することができない可能性がある。そこで、Bollerslev et al. (2016) は、

上で解説した Corsi (2009) の HAR モデルを用いて, ARQ モデルを次のように拡張した.

$$RV_t = \beta_0 + \underbrace{(\beta_1 + \beta_{1Q} RQ_{t-1}^{1/2})}_{\beta_{1,t}} RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + u_t$$

このモデルは, Heterogeneous AutoRegression Quarticity (HARQ) モデルと呼ばれる. ここでは, 日次の  $RV$  の係数のみ  $RQ^{1/2}$  の関数として時間と共に変化させている. 週次, 月次の  $RV$  の時変係数を含むモデルについては, Bollerslev et al. (2016) を参照していただきたい. さらに, 自然な拡張として上述の HARQ モデルの説明変数に  $RQ_{t-1|t-5}^{1/2}, RQ_{t-1|t-22}^{1/2}$  を加えることも考えられる. しかし, 先行研究 Bollerslev et al. (2016) において, HARQ-Full モデルと参照されているこのモデルは, 週次, 月次の観測誤差の分散を正確に推定することは実際上困難であるため, 上述の HARQ モデルと比較して一概に予測力が向上するわけではないという結果を報告している. その結果を踏まえ, 本論文では HARQ-Full モデルの分析は割愛した.

ところで, Golosnoy et al. (2014) によれば, ES モデルは, どんなボラティリティ予測子の組合せに対しても用いることができる. そこで, 過去の日次ボラティリティ  $v_{t-1}$ , HAR 予測子  $v_{t-1}^{(har)}$ , および上述の HARQ 予測値  $v_{t-1}^{(harq)}$  を組み合わせた新しいモデルを考えることも可能となる. このモデルをここでは ESQ モデルと呼ぶ. その表現は

$$v_t = \frac{\theta[v_{t-1}, v_{t-2}]v_{t-1} + \theta[v_{t-1}, v_{t-2}^{(har)}]v_{t-1}^{(har)} + \theta[v_{t-1}, v_{t-2}^{(harq)}]v_{t-1}^{(harq)}}{\theta[v_{t-1}, v_{t-2}] + \theta[v_{t-1}, v_{t-2}^{(har)}] + \theta[v_{t-1}, v_{t-2}^{(harq)}]} + \epsilon_t, \quad \epsilon_t \sim (0, \sigma^2).$$

として与えられる. 先に定義したように, ここでは

$$\begin{aligned} \theta[v_{t-1}, v_{t-2}] &= \exp(-\omega_1(v_{t-1} - v_{t-2})^2) \\ \theta[v_{t-1}, v_{t-2}^{(har)}] &= \exp(-\omega_2(v_{t-1} - v_{t-2}^{(har)})^2) \\ \theta[v_{t-1}, v_{t-2}^{(harq)}] &= \exp(-\omega_3(v_{t-1} - v_{t-2}^{(harq)})^2) \end{aligned}$$

である.

また, モデル比較の分析の一貫性の観点から, 日次ボラティリティ  $v_t$  と HAR 予測子  $v_t^{(har)}$  のみを組み合わせたモデル

$$v_t = \frac{\theta[v_{t-1}, v_{t-2}]v_{t-1} + \theta[v_{t-1}, v_{t-2}^{(har)}]v_{t-1}^{(har)}}{\theta[v_{t-1}, v_{t-2}] + \theta[v_{t-1}, v_{t-2}^{(har)}]} + \epsilon_t$$

および  $v_t, v_t^{(w)}, v_t^{(m)}, q_t = RQ_t^{1/2} RV_t$  をモデルを介さずそのまま組み合わせたモデル

$$v_t = \frac{\theta[v_{t-1}, v_{t-2}]v_{t-1} + \theta[v_{t-1}, v_{t-2}^{(w)}]v_{t-1}^{(w)} + \theta[v_{t-1}, v_{t-2}^{(m)}]v_{t-1}^{(m)} + \theta[v_{t-1}, q_{t-2}]q_{t-1}}{\theta[v_{t-1}, v_{t-2}] + \theta[v_{t-1}, v_{t-2}^{(w)}] + \theta[v_{t-1}, v_{t-2}^{(m)}] + \theta[v_{t-1}, q_{t-2}]} + \epsilon_t$$

ここで

$$\theta[v_{t-1}, q_{t-2}] = \exp(-\omega_4(v_{t-1} - q_{t-2})^2)$$

についても分析を行う. これらのモデルをここでは ES1a および ES1b モデルと呼ぶ.

また, 分析の中心となる上述の 5 つのモデルに加え, 予測力のベンチマークとして AR(1) モデル

$$v_t = \alpha_0 + \omega_1 v_{t-1} + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$$

を導入し, さらに GARCH 型 8 モデル, GARCH(1,1), GJR(1,1,1), EGARCH(1,1,1), IGARCH(1,1), AGARCH(1,1), NAGARCH(1,1), APARCH(1,1,1), ZARCH(1,1,1) を分析対象に加える. ただし, GARCH 型 8 モデルに関しては, 直接  $RV_t$  系列を用い推定予測を行うの

ではなく、株価の対数収益率  $r_t$  を用いての推定予測であるため、他の AR1, HAR, ES0, ES1, ES1a, ES1b, ARQ, HARQ, ESQ の各モデルに比べて圧倒的に不利な状況設定にある。そのため、ここでの GARCH 型 8 モデルと他の 9 モデルとの比較分析についての結果は、あくまで参考としていただきたいが、GARCH 型 8 モデル内での予測力比較は有益な情報となるであろう。

### 3.3 推定

本節では、前節で導入した標本期間 1999 年 1 月 4 日から 2013 年 12 月 30 日までの株価指数 6 銘柄、個別 24 銘柄の株価データを用い推定と予測を行う。まず、推定と予測に用いられた期間であるが、表 4 において示されているように、1999 年から 2013 年までのインサンプル、アウトオブサンプルを含む計 225 の推定予測期間を分析対象とする。ここでは紙幅の関係上、西暦の上 2 桁は省略して表記している。この 225 通りの組合せの内訳は、インサンプル 120 通り、アウトオブサンプル 105 通りである。さらに、インサンプルはそれぞれ 1 年間のインサンプル 15 通り、2 年以上 5 年未満のインサンプル 50 通り、5 年以上 15 年以下のインサンプル 55 通りに推定期間を分割している。アウトオブサンプルも同様に、推定期間 1 年のアウトオブサンプル 14 通り、推定期間 2 年以上 5 年未満のアウトオブサンプル 46 通り、推定期間 5 年以上 14 年以下のアウトオブサンプル 45 通りに推定期間を分割している。なお、アウトオブサンプルの予測期間に関しては、すべて推定期間の直近 1 年間のみとしている。

また、推定期間において外れ値が存在する場合、非線形モデルである ES1, ES1a, ES1b, ESQ モデルはその影響を受けやすい。そこで、ここでは RV 系列に Cook の距離を用いてその外れ値を検出し、モデルを推定する際にはその外れ値を除外したデータを用いている。外れ値が検出される頻度については、各銘柄ごとに差はあるものの概ね 5% 程度(最小値: 3.2%, 最大値: 6.6%)であった。Cook の距離の定義については例えば 竹内 他 (2000) が詳しい。

なお、今回のパラメータ推定に関して、ES1, ES1a, ES1b, HAR, AR1, ARQ, HARQ, ESQ の各モデルについては MATLAB の Statistics and Machine Learning Toolbox (<http://jp.mathworks.com/help/stats/>)を用い、GARCH 型 8 モデルについては Kevin Sheppard 氏が公開している MFE Toolbox ([https://www.kevinsheppard.com/MFE\\_Toolbox](https://www.kevinsheppard.com/MFE_Toolbox))を用いている。

### 3.4 予測

本節では、本稿の中心的貢献となる経験類似度モデルとその他時系列モデルとのボラティリティ予測力の比較分析を行う。手順としては、まず表 4 で示された推定期間すべてにおいて分析対象の 30 銘柄の時系列データを用い、ES0 モデルを除く 16 モデルのパラメータを推定する。そして、それらのパラメータを用い、表 4 で対応する予測期間において、以下で述べる誤差関数を用いモデル間の予測力の比較を行う。なお、ES1a および ES1b モデルについては、紙幅の関係上、後述する MCS によるモデル順位と MZ 回帰の自由度調整済み決定係数の結果のみ掲載している。

アウトオブサンプルの予測値については、例えば推定期間 99-99 で 00-00 の場合、1999 年 1 年間のデータを用いてモデルのパラメータ推定し、その推定値を用い 2000 年の RV に対して 1 日ずつ 1 期先予測を 1 年分繰り返している。ただし、ここではパラメータ推定を逐次的に 1 日ごとに行うローリングウィンドウ予測を行っていない。その理由としては、次の 2 点が挙げられる。(1) まず一点目としては、単純に計算量が大きすぎることである。本研究での目標は、複数(15 種類)のモデルに対する多様な(30 種類の)株式指数と個別銘柄の長短期(225 通りの推定予測期間)に対するボラティリティの予測力比較にある。そのため、単純計算で  $15 \times 30 \times 225 = 101250$  もの組合せに対して、1 日ごとにパラメータ推定を行い 1 日先予測を行うことは計算時間を考

表 4. 実証研究における全推定予測期間.

推定	予測								
99-99	99-99	00-08	00-08	02-06	07-07	04-09	04-09	07-09	07-09
99-99	00-00	00-08	09-09	02-07	02-07	04-09	10-10	07-09	10-10
99-00	99-00	00-09	00-09	02-07	08-08	04-10	04-10	07-10	07-10
99-00	01-01	00-09	10-10	02-08	02-08	04-10	11-11	07-10	11-11
99-01	99-01	00-10	00-10	02-08	09-09	04-11	04-11	07-11	07-11
99-01	02-02	00-10	11-11	02-09	02-09	04-11	12-12	07-11	12-12
99-02	99-02	00-11	00-11	02-09	10-10	04-12	04-12	07-12	07-12
99-02	03-03	00-11	12-12	02-10	02-10	04-12	13-13	07-12	13-13
99-03	99-03	00-12	00-12	02-10	11-11	04-13	04-13	07-13	07-13
99-03	04-04	00-12	13-13	02-11	02-11	05-05	05-05	08-08	08-08
99-04	99-04	00-13	00-13	02-11	12-12	05-05	06-06	08-08	09-09
99-04	05-05	01-01	01-01	02-12	02-12	05-06	05-06	08-09	08-09
99-05	99-05	01-01	02-02	02-12	13-13	05-06	07-07	08-09	10-10
99-05	06-06	01-02	01-02	02-13	02-13	05-07	05-07	08-10	08-10
99-06	99-06	01-02	03-03	03-03	03-03	05-07	08-08	08-10	11-11
99-06	07-07	01-03	01-03	03-03	04-04	05-08	05-08	08-11	08-11
99-07	99-07	01-03	04-04	03-04	03-04	05-08	09-09	08-11	12-12
99-07	08-08	01-04	01-04	03-04	05-05	05-09	05-09	08-12	08-12
99-08	99-08	01-04	05-05	03-05	03-05	05-09	10-10	08-12	13-13
99-08	09-09	01-05	01-05	03-05	06-06	05-10	05-10	08-13	08-13
99-09	99-09	01-05	06-06	03-06	03-06	05-10	11-11	09-09	09-09
99-09	10-10	01-06	01-06	03-06	07-07	05-11	05-11	09-09	10-10
99-10	99-10	01-06	07-07	03-07	03-07	05-11	12-12	09-10	09-10
99-10	11-11	01-07	01-07	03-07	08-08	05-12	05-12	09-10	11-11
99-11	99-11	01-07	08-08	03-08	03-08	05-12	13-13	09-11	09-11
99-11	12-12	01-08	01-08	03-08	09-09	05-13	05-13	09-11	12-12
99-12	99-12	01-08	09-09	03-09	03-09	06-06	06-06	09-12	09-12
99-12	13-13	01-09	01-09	03-09	10-10	06-06	07-07	09-12	13-13
99-13	99-13	01-09	10-10	03-10	03-10	06-07	06-07	09-13	09-13
00-00	00-00	01-10	01-10	03-10	11-11	06-07	08-08	10-10	10-10
00-00	01-01	01-10	11-11	03-11	03-11	06-08	06-08	10-10	11-11
00-01	00-01	01-11	01-11	03-11	12-12	06-08	09-09	10-11	10-11
00-01	02-02	01-11	12-12	03-12	03-12	06-09	06-09	10-11	12-12
00-02	00-02	01-12	01-12	03-12	13-13	06-09	10-10	10-12	10-12
00-02	03-03	01-12	13-13	03-13	03-13	06-10	06-10	10-12	13-13
00-03	00-03	01-13	01-13	04-04	04-04	06-10	11-11	10-13	10-13
00-03	04-04	02-02	02-02	04-04	05-05	06-11	06-11	11-11	11-11
00-04	00-04	02-02	03-03	04-05	04-05	06-11	12-12	11-11	12-12
00-04	05-05	02-03	02-03	04-05	06-06	06-12	06-12	11-12	11-12
00-05	00-05	02-03	04-04	04-06	04-06	06-12	13-13	11-12	13-13
00-05	06-06	02-04	02-04	04-06	07-07	06-13	06-13	11-13	11-13
00-06	00-06	02-04	05-05	04-07	04-07	07-07	07-07	12-12	12-12
00-06	07-07	02-05	02-05	04-07	08-08	07-07	08-08	12-12	13-13
00-07	00-07	02-05	06-06	04-08	04-08	07-08	07-08	12-13	12-13
00-07	08-08	02-06	02-06	04-08	09-09	07-08	09-09	13-13	13-13

えると現実的ではない。(2)また一点目と関連するが、ここでの目標は2, 3のモデル間における厳密な予測力の比較ではなく、多種多様なデータと推定予測期間に対するモデルの順位付けである。よって、膨大な計算時間を費やしローリングウィンドウ予測を実行し結果を出すより

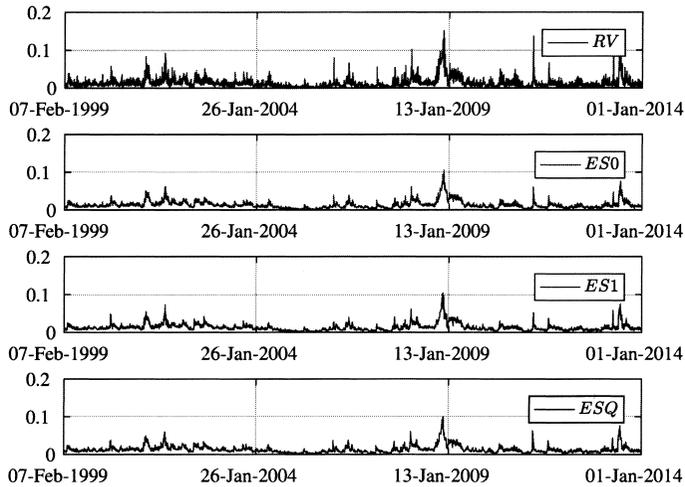


図 3.  $RV$  と各モデルの予測値(日経平均).

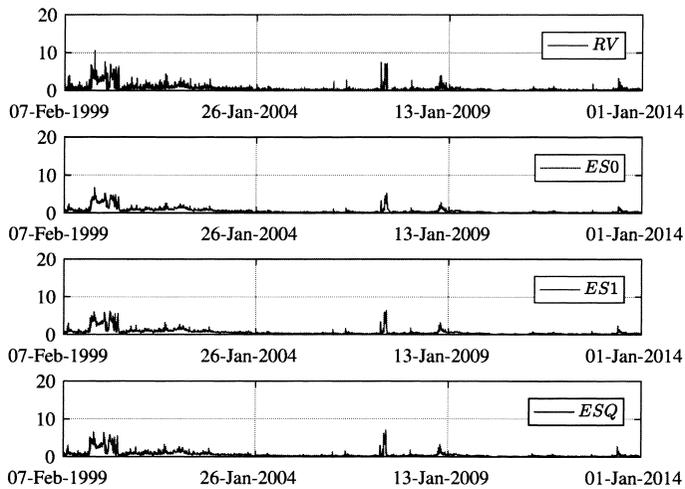


図 4.  $RV$  と各モデルの予測値(KDDI).

も、パラメータ推定値は固定し 1 期先予測を行なった方が費用対効果で考えるならば効率的な分析といえる。

図 3 および図 4 は、日経平均と KDDI の 1999 年 2 月から 2013 年 12 月までのインサンプルにおける  $RV$  と  $ES0$ ,  $ES1$ ,  $HAR$ ,  $HARQ$ ,  $ESQ$  の各モデルから計算されたボラティリティの予測値である。期間の開始が 1999 年 2 月からとなっているのは、 $HAR$  モデルが直近 22 日間のボラティリティの平均値を予測に必要としているため、最初の 22 日間の予測値は除外しているからである。まず、図 3 を見ると、2008 年後半におけるリーマン・ショックによる大きな変動が目立つ。この変動における  $RV$  と各モデルによる予測値を比較すると、 $ES0$ ,  $ES1$ ,  $HAR$ ,  $HARQ$ ,  $ESQ$  共にボラティリティを過小評価していることが分かる。全体的に見ても、リーマン・ショック以外の期間においても  $RV$  より各モデルによる予測値が低い値となっていること

が分かる。次に、図 4 を見ると、ボラティリティの動きが日経平均と KDDI では大きく異なっていることが分かる。特に 1999 年から 2000 年にかけての長期間のボラティリティクラスタリングが目立つ。これは、1999 年 12 月の KDDI と DDI と IDO が合併を正式発表してから、2000 年 10 月の実際の実際の合併までの期間がほぼ該当する。個別銘柄の動きは、各銘柄固有の事象がその動きに影響を及ぼすので、株価指数とは異なる動きになることは興味深い。RV と各モデルによる予測値を比較すると、日経平均のグラフと同様、ES0, ES1, HAR, HARQ, ESQ 共に、全体的にボラティリティを過小評価していることが分かる。日経平均のグラフにおける RV の y 軸の目盛が 0.2 までだったことを考えると、株価指数に比べて個別銘柄のボラティリティは非常に高い値となる場合があることが分かる。

### 3.4.1 誤差関数による基本統計量の分析

では、上で見た各モデルによる予測値の中で、どの予測値が最も高い予測力を持っているだろうか。この疑問に答えるため、ここでは Patton (2011) によって提案された誤差関数のクラスを利用し、予測力比較を行う。この誤差関数は、実現ボラティリティの代理変数におけるノイズの存在に頑健であり、他の予測モデルの順位付けに利用することができる。ある実数  $b$  によってパラメータ化することにより、この誤差関数のクラスは

$$\mathcal{L}(rv, \hat{v}, b) = \begin{cases} \frac{1}{(b+1)(b+2)}(rv^{b+2} - \hat{v}^{b+2}) - \frac{1}{b+1}\hat{v}^{b+1}(rv - \hat{v}) & \text{for } b \notin \{-1, -2\} \\ \hat{v} - rv + rv \cdot \log(rv/\hat{v}) & \text{for } b = -1 \\ \frac{rv}{\hat{v}} - \log \frac{rv}{\hat{v}} - 1 & \text{for } b = -2 \end{cases}$$

のように定義される。ここで  $rv$  は、ボラティリティ測度であり、 $\hat{v}$  は対応する予測値である。この誤差関数は  $b = -2$  のとき 擬似尤度誤差 (Quasi-LIKelihood, QLIKE) に対応する一方、 $b = 0$  のときは平均自乗誤差 (Mean Squared Error, MSE) 測度に対応している。Patton and Sheppard (2009) によれば、尤度に基づく誤差関数である QLIKE は、ノイズに対して頑健であるため、MSE と比較してボラティリティの予測力を比較する上でより好ましい誤差関数である。また、大きい正値の  $b$  に対しては、この誤差関数は真値の過大推定により重く罰則を与える一方、負値の  $b$  に対しては、真値を過小評価するほど誤差関数の値が大きくなる (Patton, 2011)。

本研究では、 $b \in \{1, 0, -1, -2\}$  という 4 種類の値を用いる。実際の計算された誤差関数の値は、表 4 における 225 通りの推定予測期間に銘柄数 30 と上述の 4 種類の  $b$  を組合せた数  $225 \times 30 \times 4 = 27000$  に及ぶ。表 5 は、これらの誤差関数のうち  $b = -2$  つまり QLIKE を用いて算出された平均誤差の値の一部である。この表を見ると異なる推定予測期間では、結果が変わっていることが分かる。また、27000 通りにも及ぶ誤差関数の結果に対して、これらの値を一つ一つ比較しモデルを評価することは困難である。

そこで、ここでは推定予測期間をインサンプルとアウトオブサンプルに大別し、それらの各基本統計量を見ることにより全体的な誤差関数値の特徴を捉えることにする。表 6 はインサンプル、表 7 はアウトオブサンプルにおける各誤差関数の基本統計量である。まず平均値に注目すると、インサンプルの表 6 から、 $b = 1$  では ES1、それ以外の  $b$  では HARQ が最も低い値となっており、逆に最も高い値は  $b = 0$  では ARQ、それ以外では AR1 となっている。この表から、インサンプルにおける誤差の平均値は、全体的な傾向として  $\text{HARQ} < \text{ES1} < \text{HAR} \approx \text{ESQ} < \text{ES0} < \text{ARQ} < \text{AR1}$  となっていることが分かる。また、アウトオブサンプルの表 6 を見ると、 $b = 0$  では ES1、それ以外の  $b$  では ES0 が最も低い値となっており、逆に最も高い値は  $b = 1$  では ARQ、それ以外では AR1 となっている。この表から、アウトオブサンプルにおける誤差の平均値は、全体的な傾向として  $\text{ES0} < \text{ES1} < \text{HAR} \approx \text{HARQ} \approx \text{ESQ} < \text{ARQ} < \text{AR1}$  となっていることが分かる。次に最大値に着目すると、表 6 および表 7 共に、 $b$  の値によって結果がまちまちであるが、一

表 5. 誤差関数 ( $b = -2$ ) の結果 (抜粋).

日経平均株価	推定	予測	ES0	ES1	HAR	AR1	ARQ	HARQ	ESQ
	99-11	99-11	0.156	0.156	0.154	0.186	0.180	0.153	0.153
	99-11	12-12	0.186	0.185	0.179	0.201	0.203	0.183	0.173
	99-12	99-12	0.159	0.159	0.156	0.187	0.183	0.156	0.155
	99-12	13-13	0.205	0.207	0.203	0.220	0.230	0.205	0.202
	99-13	99-13	0.162	0.162	0.160	0.190	0.186	0.160	0.159
アステラス製薬	推定	予測	ES0	ES1	HAR	AR1	ARQ	HARQ	ESQ
	99-11	99-11	0.025	0.025	0.025	0.032	0.032	0.025	0.025
	99-11	12-12	0.009	0.009	0.011	0.025	0.024	0.011	0.010
	99-12	99-12	0.024	0.024	0.024	0.031	0.030	0.024	0.024
	99-12	13-13	0.028	0.028	0.028	0.031	0.030	0.028	0.028
	99-13	99-13	0.024	0.024	0.023	0.030	0.029	0.023	0.024
コマツ	推定	予測	ES0	ES1	HAR	AR1	ARQ	HARQ	ESQ
	99-11	99-11	0.176	0.175	0.173	0.223	0.211	0.171	0.171
	99-11	12-12	0.107	0.107	0.116	0.212	0.182	0.114	0.107
	99-12	99-12	0.174	0.173	0.170	0.219	0.208	0.168	0.169
	99-12	13-13	0.240	0.246	0.263	0.373	0.335	0.257	0.255
	99-13	99-13	0.180	0.179	0.177	0.227	0.214	0.175	0.176
日立製作所	推定	予測	ES0	ES1	HAR	AR1	ARQ	HARQ	ESQ
	99-11	99-11	0.129	0.128	0.132	0.187	0.159	0.128	0.128
	99-11	12-12	0.030	0.030	0.030	0.030	0.032	0.029	0.031
	99-12	99-12	0.123	0.123	0.126	0.177	0.151	0.122	0.123
	99-12	13-13	0.212	0.207	0.212	0.230	0.211	0.205	0.212
	99-13	99-13	0.127	0.127	0.130	0.179	0.153	0.126	0.126
トヨタ自動車	推定	予測	ES0	ES1	HAR	AR1	ARQ	HARQ	ESQ
	99-11	99-11	0.079	0.079	0.079	0.101	0.094	0.078	0.079
	99-11	12-12	0.079	0.079	0.077	0.100	0.090	0.077	0.080
	99-12	99-12	0.083	0.083	0.083	0.105	0.098	0.082	0.083
	99-12	13-13	0.102	0.102	0.105	0.114	0.088	0.109	0.097
	99-13	99-13	0.100	0.099	0.097	0.117	0.108	0.096	0.097

つ言えることとしては AR1 でも他のモデルに比べて違いがそれ程極端に大きくはないことが分かる。

最後に標準偏差に関して見ると、インサンプルの表 6 では、すべての  $b$  において ARQ または AR1 が最も高い値となっているが、他の ES0, ES1, HAR, HARQ, ESQ については各  $b$  においてそれ程大きな違いは見られない。そして、アウトオブサンプルの表 6 では驚くべきことに、ES0 がすべての  $b$  において最も低い標準偏差の値となっている。前述した通り、ES0 はパラメータを推定せず、HAR の定数項以外のパラメータを  $1/3$  に固定したモデルであるが、最も分散が少なく安定的にアウトオブサンプルにおいて予測できているということである。この ES0 が、アウトオブサンプル予測において威力を発揮することは、先行研究において示唆された結果と整合的である。この ES0, すなわち  $1/3$  重み付けモデルがアウトオブサンプルにおいて良好な結果を示すことは、最適ポートフォリオの構成を選択する際に、無情報重み付け  $1/N$  手法を凌駕することが難しい (DeMiguel et al., 2009) という結果を想起させる。つまり、この結果は、無情報の意思決定者が、直近の日次、週次、月次ボラティリティの総体に均等な重み付けをすることにより、次の日のボラティリティを予測する傾向にあるという直観的洞察に対応している。

表 6. 誤差関数 ( $b = 1, 0, -1, -2$ ) の基本統計量 (インサンプル).

$b = 1$	ES0	ES1	HAR	AR1	ARQ	HARQ	ESQ
平均	0.0145	0.0142	0.0144	0.0162	0.0156	0.0147	0.0146
中央値	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
最大値	0.8704	0.9900	0.9717	0.9532	0.9082	0.9982	0.9041
標準偏差	0.0743	0.0734	0.0747	0.0838	0.0801	0.0766	0.0758
歪度	6.6617	6.9163	6.9393	6.7303	6.7412	7.0681	6.7948
尖度	51.19	57.44	57.50	52.10	52.44	60.28	54.26
標本数	3324	3324	3324	3324	3324	3324	3324
$b = 0$	ES0	ES1	HAR	AR1	ARQ	HARQ	ESQ
平均	0.0083	0.0080	0.0080	0.0092	0.0093	0.0079	0.0080
中央値	0.0004	0.0004	0.0004	0.0005	0.0005	0.0004	0.0004
最大値	0.9818	0.9703	0.9486	0.6279	0.9609	0.9093	0.9434
標準偏差	0.0397	0.0384	0.0381	0.0411	0.0434	0.0377	0.0384
歪度	10.2256	10.4631	10.2925	7.6756	9.1738	9.9895	10.1998
尖度	163.47	172.57	166.33	77.96	125.25	154.31	161.72
標本数	3324	3324	3324	3324	3324	3324	3324
$b = -1$	ES0	ES1	HAR	AR1	ARQ	HARQ	ESQ
平均	0.0110	0.0107	0.0107	0.0129	0.0125	0.0106	0.0107
中央値	0.0042	0.0041	0.0041	0.0050	0.0049	0.0041	0.0041
最大値	0.7353	0.7070	0.6926	0.7427	0.7270	0.6898	0.6713
標準偏差	0.0272	0.0262	0.0261	0.0307	0.0303	0.0259	0.0261
歪度	10.1021	10.0234	9.7591	8.4654	8.3573	9.7409	9.4669
尖度	188.98	186.43	177.16	129.45	125.12	176.75	162.61
標本数	3326	3326	3326	3326	3326	3326	3326
$b = -2$	ES0	ES1	HAR	AR1	ARQ	HARQ	ESQ
平均	0.1161	0.1153	0.1148	0.1375	0.1317	0.1138	0.1139
中央値	0.1090	0.1082	0.1077	0.1287	0.1229	0.1082	0.1074
最大値	0.6328	0.6067	0.9892	0.9444	0.9738	0.9847	0.9598
標準偏差	0.0747	0.0750	0.0757	0.0857	0.0839	0.0750	0.0739
歪度	1.0967	1.1221	1.7024	0.9060	1.2516	1.7417	1.4951
尖度	5.8412	5.9414	13.2327	5.9489	9.2506	13.8119	11.3362
標本数	3326	3326	3326	3326	3326	3326	3326

### 3.4.2 モデル信頼集合に基づく比較

上で見たインサンプルとアウトオブサンプルにおける誤差関数値の基本統計量を目視することにより、その結果を判断することはあくまで便宜的であり場当たりのなやり方である。得られた誤差関数の結果を統計的仮説検定の枠組みの上で適切に判断することにより、各モデルの予測力の比較を行うため、ここでは Hansen et al. (2011) が提案するモデル信頼集合 (Model Confidence Set, MCS) の考え方を導入する。MCS により、特定のモデルを前提とせず、所与の有意水準での最良なモデルを選択することが可能となる。MCS の一般的な手順は、Hamid and Heiden (2015) に従い下記の通り解説する。まず、候補となるモデルの組  $\mathcal{M}_0 = \{1, \dots, m_0\}$  を用意する。本研究では  $m_0 = 15$  となる。すべてのモデルの組に対して、各モデルから得られた誤差関数  $L$  の差分に基づきモデルの優劣を評価する。つまり、モデル  $i$  とモデル  $j$  ( $i, j = 1, \dots, m_0$ ) および、すべての時点  $t = 1, \dots, T$  に対して

$$d_{i,j,t} = L(rv_{it}, \hat{r}\hat{v}_{it}) - L(rv_{jt}, \hat{r}\hat{v}_{jt})$$

を評価する。そして、各  $d_{i,j,t}$  において、帰無仮説

表 7. 誤差関数 ( $b = 1, 0, -1, -2$ ) の基本統計量 (アウトオブサンプル).

$b = 1$	ES0	ES1	HAR	AR1	ARQ	HARQ	ESQ
平均	0.0026	0.0028	0.0027	0.0035	0.0040	0.0029	0.0029
中央値	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
最大値	0.4941	0.4660	0.5605	0.6676	0.9178	0.4867	0.5220
標準偏差	0.0175	0.0179	0.0185	0.0228	0.0276	0.0190	0.0192
歪度	14.1394	12.9544	15.3115	14.7657	17.6783	12.7591	13.4480
尖度	286.85	232.25	349.62	322.32	472.63	223.54	256.69
標本数	2840	2840	2840	2840	2840	2840	2840
$b = 0$	ES0	ES1	HAR	AR1	ARQ	HARQ	ESQ
平均	0.0050	0.0049	0.0056	0.0063	0.0058	0.0052	0.0054
中央値	0.0002	0.0002	0.0002	0.0003	0.0003	0.0002	0.0002
最大値	0.5086	0.7682	0.8326	0.8262	0.8362	0.6840	0.8029
標準偏差	0.0344	0.0363	0.0422	0.0432	0.0388	0.0375	0.0399
歪度	12.3288	15.1551	13.2819	13.7566	15.3827	14.0863	13.9322
尖度	166.77	264.39	193.23	212.51	272.53	220.87	218.47
標本数	2840	2840	2840	2840	2840	2840	2840
$b = -1$	ES0	ES1	HAR	AR1	ARQ	HARQ	ESQ
平均	0.0091	0.0096	0.0098	0.0130	0.0129	0.0101	0.0098
中央値	0.0035	0.0037	0.0039	0.0054	0.0051	0.0037	0.0037
最大値	0.3381	0.5879	0.4847	0.5723	0.9602	0.6940	0.5902
標準偏差	0.0224	0.0286	0.0264	0.0318	0.0375	0.0302	0.0280
歪度	10.0364	13.3357	10.7361	9.3700	12.9680	13.2420	11.8349
尖度	131.31	225.72	147.35	118.62	239.53	238.43	181.30
標本数	2841	2841	2841	2841	2841	2841	2841
$b = -2$	ES0	ES1	HAR	AR1	ARQ	HARQ	ESQ
平均	0.1225	0.1243	0.1275	0.1697	0.1624	0.1276	0.1245
中央値	0.1003	0.1016	0.1035	0.1351	0.1291	0.1016	0.1016
最大値	0.5506	0.7121	0.5686	0.9904	0.9495	0.8868	0.7665
標準偏差	0.1004	0.1028	0.1042	0.1415	0.1367	0.1071	0.1026
歪度	1.2581	1.3651	1.2641	1.4917	1.5307	1.5091	1.3682
尖度	4.4922	5.2335	4.4802	5.6011	6.0364	6.4795	5.3569
標本数	2841	2841	2841	2841	2841	2841	2841

$$H_0 : E[d_{ij,t}] = 0, \quad \forall i, j \in \mathcal{M}, \quad i > j$$

を、モデル集合  $\mathcal{M} \subset \mathcal{M}_0$  に対して検定する。初期値は  $\mathcal{M} = \mathcal{M}_0$  である。もし帰無仮説  $H_0$  が所与の有意水準 (たとえば 10%) で棄却されるならば、予測力が最も低いモデルがこのモデル集合から除外される。この方法を、 $H_0$  が棄却できなくなるまで継続する。ここでは Hansen et al. (2011) に従い、 $H_0$  を評価するために以下のような範囲統計量を用いる。

$$T_{R,k} = \max_{i,j \in \mathcal{M}} |t_{ij}| = \max_{i,j \in \mathcal{M}} \frac{|\bar{d}_{ij}|}{\sqrt{\widehat{\text{var}}(\bar{d}_{ij})}}$$

ここで  $\bar{d}_{ij} = \frac{1}{T} \sum_{t=1}^T d_{ij}$  および  $\widehat{\text{var}}(\bar{d}_{ij})$  は、ブロック・ブートストラップ法を用いることにより求められる。モデル集合  $\mathcal{M}$  から除外される最も良くない予測力を持つモデル  $i^*$  は

$$i^* = \arg \max_{i \in \mathcal{M}} \frac{\bar{d}_i}{\sqrt{\widehat{\text{var}}(\bar{d}_i)}}$$

によって選ばれる。ここで  $\bar{d}_i = \frac{1}{m-1} \sum_{j \in \mathcal{M}} \bar{d}_{ij}$  であり、 $m$  はモデル集合  $\mathcal{M}$  に含まれるモデル

表 8. MCS によるモデル順位(日経平均, 抜粋).

推定期間	予測期間	1st	2nd	3rd	4th	5th	6th
1999-2004	2005-2005	ES0 (1.000)	ES1 (0.945)	ES1b (0.572)	ESQ (0.141)	ES1a (0.053)	HARQ (0.024)
2001-2004	2005-2005	ES0 (1.000)	ES1 (0.066)	ES1b (0.047)	ESQ (0.047)	ES1a (0.047)	HARQ (0.047)
2002-2003	2004-2004	ES0 (1.000)	ES1b (0.226)	ES1 (0.226)	ESQ (0.203)	ES1a (0.160)	HAR (0.057)
推定期間	予測期間	1st	2nd	3rd	4th	5th	6th
2000-2004	2005-2005	ES1 (1.000)	ES0 (0.850)	ESQ (0.142)	ES1a (0.090)	ES1b (0.090)	HARQ (0.029)
2001-2002	2003-2003	ES1 (1.000)	ES1b (0.179)	ES0 (0.179)	ESQ (0.179)	ES1a (0.176)	HAR (0.097)
2008-2008	2009-2009	ES1 (1.000)	ES0 (0.816)	ES1b (0.313)	ESQ (0.313)	HAR (0.003)	ES1a (0.003)
推定期間	予測期間	1st	2nd	3rd	4th	5th	6th
1999-2009	2010-2010	HAR (1.000)	ES1b (0.939)	ES1 (0.250)	ESQ (0.250)	ES0 (0.225)	HARQ (0.017)
2002-2009	2010-2010	HAR (1.000)	ESQ (0.713)	ES0 (0.672)	HARQ (0.007)	ES1b (0.007)	ES1 (0.007)
2009-2011	2012-2012	HAR (1.000)	ESQ (0.498)	HARQ (0.438)	ES1b (0.438)	ES1 (0.135)	ES1a (0.135)
推定期間	予測期間	1st	2nd	3rd	4th	5th	6th
1999-2006	2007-2007	HARQ (1.000)	ESQ (0.924)	HAR (0.257)	ES1 (0.104)	ES0 (0.104)	ES1a (0.104)
2001-2006	2007-2007	HARQ (1.000)	ESQ (0.835)	HAR (0.388)	ES1a (0.388)	ES1b (0.388)	ES1 (0.388)
2001-2012	2013-2013	HARQ (1.000)	ESQ (0.963)	ES1a (0.963)	ARQ (0.963)	HAR (0.555)	ES1b (0.555)
推定期間	予測期間	1st	2nd	3rd	4th	5th	6th
1999-2007	1999-2007	ESQ (1.000)	HARQ (0.506)	HAR (0.391)	ES1 (0.034)	ES0 (0.028)	ES1a (0.001)
2000-2007	2000-2007	ESQ (1.000)	HARQ (0.340)	HAR (0.288)	ES1 (0.067)	ES1a (0.004)	ES0 (0.004)
2001-2004	2001-2004	ESQ (1.000)	HARQ (0.692)	HAR (0.692)	ES1a (0.044)	ES0 (0.044)	ES1b (0.044)

の数である。本研究では、ブロックの長さは 15 とした 10000 回の繰り返しのブロック・ブートストラップ法を実行している。有意水準については 90% に設定し、実際の分析には前述の MFE Toolbox を用いている。

では、実際のデータを用いた MCS の結果を見ていくことにしよう。MCS は、誤差関数同様、225 通りの推定予測期間に銘柄数 30 と 4 種類の  $b$  を組合せた数  $225 \times 30 \times 4 = 27000$  通り実行できる。そこで、ここでも実際に得られた結果の一部を抜粋して表 8 に紹介する。公平を期するために、ES0, ES1, HAR, HARQ, ESQ それぞれが最良であった期間を選んで掲載している。まず、この表 8 の最も左の列から順に、MCS によるモデルの予測力順位を 5 位まで掲載している。各モデル名の下にある括弧は P 値を表している。つまり、ここでは P 値が 1.00 であるモデルが最良であり、その後 P 値の高さによってモデルの予測力を順位付けしている。

次に、上述の MCS によって得られたモデルの予測力順位で P 値が 1.00 だったモデル、つまり MCS 基準で予測力が最良だったモデルの累積頻度を図 5 から図 7 に示す。まず、図 5 は、

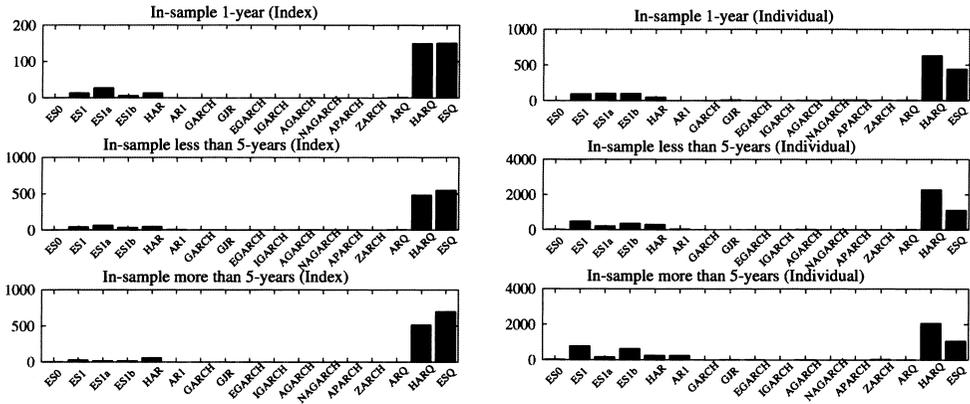


図 5. MCS 最良モデルの累積頻度(推定期間別, インサンプル).

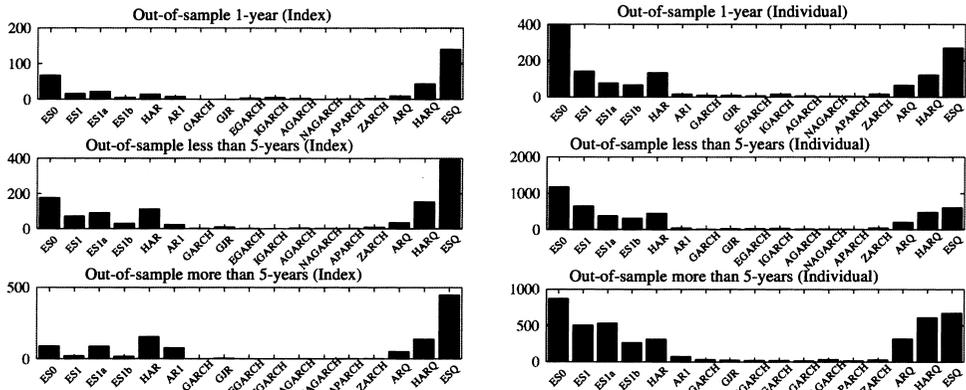


図 6. MCS 最良モデルの累積頻度(推定期間別, アウトオブサンプル).

列方向に左側が株式指数, 右側が個別銘柄, 行方向に上から推定期間が1年, 5年未満, 5年以上のデータセットから得られたインサンプルでのMCS最良モデルの累積頻度である。一見して分かることとして, 左側の株式指数では推定期間が1年ではESQ, その他ではHARQが最良モデルとしての頻度が最も高くなっており, いずれにせよHARQとESQが他のモデルを圧倒していることが分かる。ただし, 右側の個別銘柄では, 株式指数と同様の傾向ではあるが, 5年以上の推定期間ではES1が比較的健闘していることが分かる。

図6は, 図5と同様の設定でアウトオブサンプルでのMCS最良モデルの累積頻度である。ここで特徴的なこととして, 表7における誤差関数の結果により示唆されたように, アウトオブサンプルにおける個別銘柄のすべての推定期間においてES0の頻度が最も高くなっている。株式指数では, 推定期間が5年以上のデータセットではHAR, それ以外ではESQがそれぞれ最良モデルとしての頻度が最も高くなっている。

図7は, 図5および図6をインサンプル, アウトオブサンプル別と株式指数, 個別銘柄別にまとめたものである。列方向に左側が株式指数, 右側が個別銘柄, 行方向に上側がインサンプル, 下側がアウトオブサンプルの各データセットから得られたMCS最良モデルの累積頻度である。この図から分かることは, 左側の株式指数においてインサンプルではHARQ, アウトオ

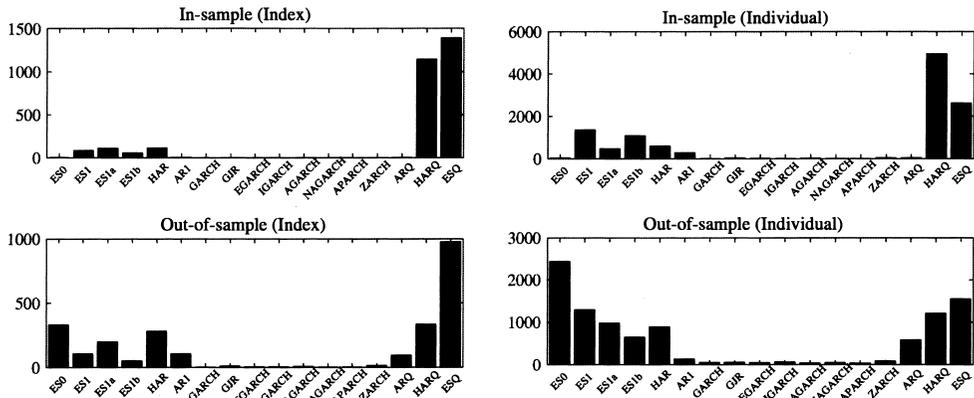


図 7. MCS 最良モデルの累積頻度(全推定期間).

ブサンプルでは ESQ が最良モデルとしての頻度が最も高くなっている。右側の個別銘柄については、インサンプルでは HARQ、アウトオブサンプルでは ESQ がそれぞれ最良モデルとしての頻度が最も高くなっている。なお、前述した通り、GARCH 型 8 モデルに関しては、直接  $RV_t$  系列を用い推定予測を行うのではなく、株価の対数収益率  $r_t$  を用いての推定予測であるため、他の ES0, ES1, HAR, HARQ, ESQ の各モデルと予測力を直接比較することは適当でないことに留意すべきであろう。

### 3.4.3 Mincer-Zarnowitz 回帰に基づく比較

最後に、各モデルの予測値に対して Mincer-Zarnowitz (MZ) 回帰を行なった結果を報告する。Patton and Sheppard (2009)によれば、MZ 回帰は Mincer and Zarnowitz (1969)により提案された、予測力を評価する一般的な手法の 1 つである。この MZ 回帰は

$$\hat{\sigma}_t^2 = \alpha + \beta h_t + e_t$$

により定式化される。ここで、 $\hat{\sigma}_t^2$  はボラティリティの代理変数、 $\alpha$  および  $\beta$  はパラメータ、 $h_t$  は各モデルから得られたボラティリティの予測値、そして  $e_t$  は誤差項である。 $\hat{\sigma}_t^2$  に関して、本研究では一貫して  $RV$  をボラティリティの代理変数として用いている。表 9 は、この MZ 回帰を株式指数および個別銘柄の各インサンプル、アウトオブサンプルにおける推定期間別のデータセットに対して実行した際の自由度調整済み決定係数  $R^2$  の結果である。この表中の下線を引いている値が、各行における最大の  $R^2$  の値である。この表から分かることは、株式指数におけるインサンプルではすべての推定期間において ESQ の  $R^2$  が最大であり、アウトオブサンプルでは推定期間 5 年以上を除けば、すべての推定期間において ES0 の  $R^2$  が最大となっている。また、個別銘柄については、インサンプルでは推定期間 1 年を除いては ES1 の  $R^2$  が最大であり、アウトオブサンプルではすべての推定期間において ES0 の  $R^2$  が最大となっている。つまり、アウトオブサンプルにおいて、株式指数の推定期間 5 年以上を除けば、すべて ES0 の  $R^2$  が最大となっている。このことは、先の誤差関数、MCS の結果とも整合的である。

MZ 回帰の結果に基づいて全体的傾向を縮約すれば、インサンプルでは 4 次モーメントの情報を使った方が予測が良いものの、アウトオブサンプルではそのような効果は認められない。収益率データにはしばしばボラティリティクラスタリングが観察される通り、荒れている時期もあれば風いでいる時期もある。インサンプル期間におけるそのような違いを説明するのに、RQ は重要な説明要素たりうるが、大きな変動性が将来的にいつ訪れるかに関しては、今回分

表 9. MZ 回帰による各モデルの自由度調整済み決定係数  $R^2$ .

株式指数	ES0	ES1	ES1a	ES1b	HAR	AR1	ARQ	HARQ	ESQ
インサンプル									
推定期間 1 年	0.293	0.320	0.323	0.313	0.312	0.250	0.269	0.322	<u>0.330</u>
推定期間 5 年未満	0.436	0.451	0.450	0.449	0.444	0.368	0.385	0.450	<u>0.457</u>
推定期間 5 年以上	0.547	0.554	0.554	0.552	0.552	0.468	0.485	0.557	<u>0.560</u>
アウトオブサンプル									
推定期間 1 年	<u>0.302</u>	0.254	0.270	0.259	0.291	0.258	0.191	0.247	0.263
推定期間 5 年未満	<u>0.317</u>	0.300	0.305	0.304	0.317	0.272	0.241	0.290	0.300
推定期間 5 年以上	0.324	0.308	<u>0.326</u>	0.312	0.325	0.294	0.277	0.316	0.325
個別銘柄									
インサンプル									
推定期間 1 年	0.284	0.308	0.305	0.304	0.299	0.240	0.259	<u>0.309</u>	0.308
推定期間 5 年未満	0.432	<u>0.447</u>	0.438	0.444	0.440	0.364	0.385	0.446	0.443
推定期間 5 年以上	0.538	<u>0.548</u>	0.540	0.546	0.543	0.456	0.480	0.547	0.545
アウトオブサンプル									
推定期間 1 年	<u>0.289</u>	0.265	0.261	0.268	0.275	0.245	0.208	0.244	0.261
推定期間 5 年未満	<u>0.295</u>	0.280	0.282	0.284	0.289	0.252	0.231	0.273	0.280
推定期間 5 年以上	<u>0.286</u>	0.275	0.280	0.278	0.281	0.249	0.236	0.273	0.277

析したモデル群は殆ど構造を持っていない。

外挿型の点予測を行う状況では、4 次モーメントを取り込んで広めに構えていても、ナイーブな方式で予測を形成しても大きな違いはなく、fat tail を想定しておいたメリットが活きるケースは低頻度と考えられる。予測ストレッチが短ければ、たまたまそういう想定が奏功して予測評価の尺度は上がるかもしれないが、ストレッチが長くなるほどそういうケースは率として低下してゆく。それが、アウトオブサンプルで ES0 が平均的に優越しているということの、ひとつの解釈たり得るのではないだろうか。

#### 4. 結論

本稿では、事例ベース意思決定理論 (Gilboa and Schmeidler, 1995, 2001) に基礎を置いた Gilboa et al. (2006) を端緒とする経験類似度の枠組みに着目し、Golosnoy et al. (2014) の提案した経験類似度に基づく時系列モデルを用いボラティリティ予測の実証分析を行なった。

ここでの実証研究の中心的貢献は、経験類似度モデル ES0, ES1, ESQ とその他の時系列モデルとの予測力比較にある。モデルの予測力比較については、最初に 4 つの誤差関数による MCS を用いることにより、複数の銘柄と推定予測期間におけるモデルの予測力を順位付けし、最良モデルの累積頻度を分析した。分析結果としては、インサンプルでは株式指数、個別銘柄共に HARQ、アウトオブサンプルでは株式指数は ESQ、個別銘柄 ES0 がそれぞれ最良モデルとしての頻度が最も高いという結果となった。次に、複数の銘柄と推定予測期間におけるモデルの予測力を比較するために、MZ 回帰を実行した。この MZ 回帰の自由度調整済み決定係数  $R^2$  に基づく分析結果によれば、インサンプルでは株式指数は ESQ、個別銘柄は ES1、アウトオブサンプルでは株式指数、個別銘柄共に ES0 が最良という結果となった。MCS および MZ 回帰の結果において、アウトオブサンプルにおける ES0 の予測力が他のモデルと比較して高くなる傾向は、先行研究 Golosnoy et al. (2014) の結果と整合的である。また、HARQ, ESQ といった高頻度データから算出された実現測度 RQ を含むモデルが誤差関数、MCS, MZ 回帰において全般的に好評価であったことは、高頻度金融データの集約がもたらす情報量の多さを改めて確認さ

せられる結果となった。

また、本稿ではボラティリティの非対称性について特に考慮に入れていないが、非対称性はボラティリティを予測する上で非常に重要な因子であるため、この点については今後の研究課題としたい。

最後に、経験類似度モデルを日本の株式市場におけるボラティリティ分析に適用し、さらに、HARQ, ESQ といった実現測度 RQ を含むボラティリティモデルを体系的に扱った文献は、我々の知る限り、本研究が本邦初である。従って、経験類似度モデルの金融分野への適用および実現測度 RQ を含むボラティリティモデルの応用は始まったばかりであり、今後の研究の発展に期待したい。

## 謝 辞

本研究は、統計数理研究所公募型共同利用(課題番号 28-共研-2012)に基づくと同時に、科学研究費基盤研究(A)25245033, (A)15H01943, (B)15H03337, (C)16K00067, (C)26380279, (C)15K03406 による援助を受けています。また、改稿にあたって有益なコメントをくださった 2 人の査読者および Union Investment の Moritz Heiden 博士に感謝いたします。

## 参 考 文 献

- Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets, *Journal of Empirical Finance*, **4**, 115–158.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001). The distribution of realized stock return volatility, *Journal of Financial Economics*, **61**, 43–76.
- Andersen, T. G., Bollerslev, T., Diebold, F.X. and Labys, P. (2003). Modeling and forecasting realized volatility, *Econometrica*, **71**, 579–625.
- Andersen, T. G., Bollerslev, T. and Meddahi, N. (2004). Analytical evaluation of volatility forecasts, *International Economic Review*, **45**, 1079–1110.
- Andersen, T. G., Dobrev, D. and Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation, *Journal of Econometrics*, **169**, 75–93.
- Bandi, F. M. and Russell, J. R. (2004). Microstructure noise, realized variance, and optimal sampling, *Review of Economic Studies*, **75**, 339–369.
- Barndorff-Nielsen, O. E. 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 (Statistical Methodology)*, **64**, 253–280.
- Barndorff-Nielsen, O. E., 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**, 1481–1536.
- Bates, J. M. and Granger, C. W. J. (1969). The combination of forecasts, *Operation Research Quarterly*, **20**, 451–468.
- Billot, A., Gilboa, I. and Schmeidler, D. (2008). An axiomatization of an exponential similarity function, *Mathematical Social Sciences*, **55**, 107–115.
- Bollerslev, T., Patton, A. J. and Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting, *Journal of Econometrics*, **192**, 1–18.
- Corsi, F. (2009). A simple long memory model of realized volatility, *Journal of Financial Econometrics*, **7**, 174–196.
- DeMiguel, V., Garlappi, L. and Uppal, R. (2009). Optimal versus naïve diversification: How inefficient

- is the  $1/N$  portfolio strategy?, *Review of Financial Studies*, **22**, 1915–1953.
- Elliott, G. and Timmermann, A. (2004). Optimal forecast combinations under general loss functions and forecast error distributions, *Journal of Econometrics*, **122**, 47–79.
- Gilboa, I. and Schmeidler, D. (1995). Case-based decision theory, *Quarterly Journal of Economics*, **110**, 605–639.
- Gilboa, I. and Schmeidler, D. (2001). *A Theory of Cased-based Decisions*, Cambridge University Press, Cambridge. (浅野貴央, 尾山大輔, 松井彰彦 共訳 (2005). 『決め方の科学—事例ベース意思決定理論』, 勁草書房, 東京.)
- Gilboa, I. and Schmeidler, D. (2012). *Case-based Predictions*, World Scientific, Singapore.
- Gilboa, I., Lieberman, O. and Schmeidler, D. (2006). Empirical similarity, *Review of Economics and Statistics*, **88**, 433–444.
- Gilboa, I., Lieberman, O. and Schmeidler, D. (2011). A similarity-based approach to prediction, *Journal of Econometrics*, **162**, 124–131.
- Golosnoy, V., Hamid, A. and Okhrin, Y. (2014). The empirical similarity approach for volatility prediction, *Journal of Banking & Finance*, **40**, 321–329.
- Hamid, A. and Heiden, M. (2015). Forecasting volatility with empirical similarity and Google trends, *Journal of Economic Behavior & Organization*, **117**, 62–81.
- Hansen, P. R. and Lunde, A. (2005). A realized variance for the whole day based on intermittent high-frequency data, *Journal of Financial Econometrics*, **3**, 525–554.
- Hansen, P. R., Lunde, A. and Nason, J. M. (2011). The model confidence set, *Econometrica*, **79**, 453–497.
- Hume, D. (1748). *Enquiry into the Human Understanding*, Clarendon Press, Oxford.
- Lieberman, O. (2010). Asymptotic theory for empirical similarity models, *Econometric Theory*, **26**, 1032–1059.
- Masuda, H. and Morimoto, T. (2012). Optimal weight for realized variance based on intermittent high-frequency data, *Japanese Economic Review*, **63**, 497–527.
- McAleer, M. and Medeiros, M. C. (2008). Realized volatility: A review, *Econometric Reviews*, **27**, 10–45.
- Mincer, J. A. and Zarnowitz, V. (1969). The evaluation of economic forecasts, *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance* (ed. J. A. Mincer), 3–46, National Bureau of Economic Research, New York.
- 尾山大輔 (2012). 事例ベース意思決定理論とは, オペレーションズ・リサーチ: 経営の科学, **57**, 557–559.
- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies, *Journal of Econometrics*, **160**, 246–256.
- Patton, A. J. and Sheppard, K. (2009). Evaluating volatility and correlation forecasts, *Handbook of Financial Time Series* (eds. T. G. Andersen, R. A. Davis, J. P. Kreiss and T. Mikosch), 801–838, Springer, Berlin.
- 竹内秀一, 近河拓也, 篠崎信雄 (2000). 複数個の外れ値を検出するときの Cook の距離の検出力, 応用統計学, **29**, 83–99.
- Wooldridge, J. M. (2015). *Introductory Econometrics: A Modern Approach*, 6th ed., Cengage Learning, Boston.
- Zhang, L., Mykland, P. A. and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high frequency data, *Journal of the American Statistical Association*, **100**, 1394–1411.

## Volatility Forecasting with Empirical Similarity: Japanese Stock Market Case

Takayuki Morimoto<sup>1</sup> and Yoshinori Kawasaki<sup>2,3</sup>

<sup>1</sup>School of Science and Technology, Kwansai Gakuin University

<sup>2</sup>The Institute of Statistical Mathematics

<sup>3</sup>Department of Statistical Science, School of Multidisciplinary Sciences,  
The Graduate University for Advanced Studies

In this research, we compare the forecasting ability of various volatility models through within-sample and out-of-sample forecasting simulations. Models considered here are heterogeneous auto regression models (HAR), a 1/3 model where the weight coefficients are all set to 1/3 in the HAR model (ES0), and an HAR model in which weight coefficients are determined by empirical similarity. We also try AR(1), ARCH/GARCH and their variants, and models incorporating the Realized Quarticity (RQ), which are referred to as ARQ, HARQ and ESQ. As stock data, we pick 6 index series from the Tokyo Stock Exchange, and 24 individual stock series all of which had enough liquidity from April 1st 1999 to December 30th 2013. Minute-by-minute data were created based on high-frequency data. Forecasting evaluation depends on what kind of evaluation function we employ. We make use of Patton's error function. Changing the length of the estimation period and the forecasting period, and also the parameter of Patton's error function, we try 27,000 patterns of forecasting simulations. We find that ESQ and HARQ are almost comparative in within-sample forecasting, whereas ES0 is outstanding in out-of-sample forecasting experiments. We also tried model comparison based on the pair-wise testing procedure proposed by Hansen et al. We see similar results but the details are a little bit different in index series and in individual stock series.