モデル平均理論の新展開

劉　　慶　豊

Ⅰ　はじめに

データ解析を行う際に，利用可能なモデル群から最良と考えられるモデルを１個選ぶ手段として，モデル選択理論が研究されてきた。実証研究では，AIC（Akaike［1973］）や BIC（Schwarz［1978］）などの情報量基準を利用した選択法が広く利用されている。他方，モデル選択に伴う不確実性を考慮して，ベイズ的な立場から，ベイジアン・モデル平均法と理論が構築され，発展してきた。ここ数年は，非ベイズ的な研究者が，非ベイジアン・モデル平均理論の研究を始めた。本稿はモデル選択とモデル平均の理論を概観し，異なった考え方に基づく異なった手法の意味を明らかにする。さらに，この分野の未解決問題と，将来の発展方向を論じる。最後に例を示し，実証研究に適用するときの注意点を挙げる。

第Ⅱ章で，AIC や BIC を紹介する。そして，第Ⅲ章でモデル平均の意味合いを述べ，幾つかの例を挙げて，モデル平均推定量の性質をまとめる。第Ⅳ章で未解決な問題を説明し，第Ⅴ章ではモデル平均を利用した幾つかの研究例を紹介した上で，ファイナンスへの応用を示す。第Ⅵ章は結論である。

Ⅱ　情報量基準によるモデル選択

モデルを構築して，データの特性をできるだけ正確に説明しようとする際に，データを生成する真の確率分布を近似するための候補として，複数のモデルを考える。それらのモデルを評価するためには，さまざまな情報量基準が提案されているが，なかでも，AIC や BIC が最も広く利用されている。

1　赤池情報量基準 AIC

AIC の基本となっているものはカルバック―ライブラー距離である。確率変数 y の標本 $y=(y_1, y_2, \cdots, y_n)$ が観測されたとする。y の真の密度関数を $g(y)$ を，パラメットリックなモデル $f(y, \theta)$ で近似しよう。ただし，$\theta\in\Theta$ で $\Theta\subseteq\mathbb{R}^K$ である。$f(y, \theta)$ による近似の良さを計るための一つの基準として，Kullback and Leibler［1951］はカルバック―ライブラー距離

$$KL(g, f)=E_g\left(\log\frac{g(y)}{f(y, \theta)}\right)$$

$$=\int_{-\infty}^{\infty}\{\log g(y)-\log f(y, \theta)\}g(y)dy. \quad (1)$$

を提案した。$\log(\cdot)$ は自然対数を表す。この距離は，モデルで真の分布を近似するときの一種のリスク関数であると考えられる。

$g(y)$ はモデルに依存しないため，モデルの良さを評価する場合，$\log g(y)$ は無視できる。原理的には平均対数尤度と呼ばれる。

$$E_g(\log f(y, \theta))=\int_{-\infty}^{\infty}\log f(y, \theta)g(y)dy \quad (2)$$

を最大にするモデルが，最適なモデルとなる。大数の法則により

$$\lim\frac{1}{n}\sum_{i=1}^{n}\log f(y_i, \theta)=E_g(\log f(y, \theta)) \quad (3)$$

となる。そこで，対数尤度関数を

$$l_n(\theta)=\sum_{i=1}^{n}\log f(y_i, \theta) \quad (4)$$

とし，$\hat\theta$ を最尤推定量として，$\frac{1}{n}l_n(\hat\theta)$ を

$E_g(\log f(y,\hat{\theta}))$の推定量と考えることができるが，この推定量はバイアスを持っている。証明は小西・北川［2004］を参照されたい。そこで，バイアスが修正され，モデル選択の基準として

$$\mathrm{AIC} = -2l_n(\hat{\theta}) + 2p \tag{5}$$

が構築された。ただし，pはパラメーターθの次元で，平均対数尤度を推定するときのバイアスの補正として働く。$2p$は，モデルの複雑さに対するペナルティと解釈されることもある。AICを最小にするモデルが望ましいとされる。

2　他の情報量基準

ベイジアン情報量基準（BIC）も広く利用されている。他に，AICの修正版として竹内［1976］とStone［1977］によって考案された情報量基準TICもよく知られている。

BICの背景となっているのはベイジアン統計学で，事後確率$P(M|\mathrm{y})$が最大になるようなモデルMを選ぶというのがその発想である。導出の詳細はClaeskens and Hjort［2008］を参照されたい。事後確率の近似的な評価値として，

$$\mathrm{BIC} = -2l_n(\hat{\theta}) + (\log n)p \tag{6}$$

が定義される。AICと似た形になっているが，そのペナルティ$(\log n)p$は，nが大きくなるときAICのペナルティより大きくなる。nが大きくなるに連れAICはより複雑なモデルを選びがちであるが，BICは真のモデルよりパラメーター数の少ないモデルを選ぶ傾向がある。そして，真のモデルが候補となるモデル族の中に入っていれば，サンプル数nが無限大に近づくにつれ，BICが真のモデルを選ぶ確率は1に収束する。すなわち，BICはモデル選択の情報量基準として一致性を持つ。AICは一致性を持たない。

小西・北川［2004］によれば，AICを計算する際，想定したパラメトリックモデル族$\{f(y,\theta;\theta\in\Theta)\}$の中に，データを生成する確率分布$g(y)$が含まれていれば，$p$が平均対数尤度を推定するときのバイアス補正項になることが保障される。そうではないとき，pは正しいバイアス補正項とならない。竹内［1976］は正確な補正項を導出して，想定したモデル族の中に真の確率分布が含まれていない場合でも利用できる情報量基準（TIC）

$$\mathrm{TIC} = -2l_n(\hat{\theta}) + 2Tr(\hat{I}\hat{J}^{-1}) \tag{7}$$

を提案した。$\hat{I}$と$\hat{J}$はそれぞれ

$$I_0 = E_g\left[\frac{\partial \log f(y,\theta)}{\partial\theta}\frac{\partial \log f(y,\theta)}{\partial\theta'}\right] \tag{8}$$

$$J_0 = -E_g\left[\frac{\partial^2 \log f(y,\theta)}{\partial\theta\partial\theta'}\right] \tag{9}$$

の推定量で，一致推定量$\hat{\theta}$を，I_0とJ_0の標本統計量に代入することで得られる。

想定したモデル族の中に真の確率分布が含まれていないときは，AICは正確な補正項を使っていない。しかし，想定するモデルが真の確率分布に近い場合は，AICは近似的には厳密な選択基準である。

線形回帰式の情報量基準としては，Mallows［1973］がMallows' C_p

$$C_p = RSS + 2p\hat{\sigma}^2 - n\hat{\sigma}^2 \tag{10}$$

を提案した。C_pはリスク関数Mean Squared Error (MSE)の推定量となっており，それを最小にするように，説明変数を選ぶ。C_pは線形式にしか適用できないが，非線形式は，線形化してからの適用が可能である。

Ⅲ　モデル平均

AICやBICなどの情報量基準を利用してモデル選択を行うと，選ばれたモデルがそのまま真のモデルとして扱われ，さらなる推論が展開される。しかし，AICやBICなどの情報量基準によるモデル選択は必ずしも真のモデルを選ぶ保障はなく，不確実性をもたらす。その不確実性が推定量の性質に及ぼす影響を考慮しないと，誤った推論に繋がる。つまり，モデルから求まる信頼区間や検定結果が誤ったものにな

る。このような問題は，選択後推定量（Post Model Selection Estimator，PMSE）の問題と呼ばれる。モデル平均は，このような PMSE の問題を考慮する手法である。

モデル平均はモデル選択を基にした手法である。候補となるモデル族を $\mathcal{M}=\{M_1, M_2, \cdots, M_K\}$ とすると，母数 μ のモデル平均による推定量は

$$\hat{\mu}=\sum_{M_k\in\mathcal{M}} c(M_k)\hat{\mu}_{M_k} \tag{11}$$

である。$\hat{\mu}_{M_k}$ はモデル M_k のもとで得られる推定量で，$c(M_k)$ は各モデルに与えられるウエイトである。$c(M_k)$ は総和が 1 で，非確率的でも確率的でも良い。この表現は情報量基準でモデルを選択して得られる推定量を含んでいる。たとえば，AIC の場合，AIC で選んだモデルを M_{AIC} とする。そのモデルによって得られる推定量は

$$\hat{\mu}_{\text{AIC}}=\sum_{M_k\in\mathcal{M}} I(M_k=M_{\text{AIC}})\hat{\mu}_{M_k} \tag{12}$$

となる。$I(\cdot)$ は M_k が M_{AIC} に一致すれば 1，一致しない場合は 0 の値をとる定義関数である。AIC を基準にモデルを選ぶ際，選択は確率的に定まり，$I(M_k=M_{\text{AIC}})$ が確率変数となっていることに注意すべきである。この点は後ほど論じる。

モデル平均による推定量では，ウエイトはより一般的で，AIC の場合のウエイトを特殊ケースとして含む。そのため，リスクに関する最適なウエイトを使えば，モデル平均による推定量のリスクは，AIC のリスクに等しいか，より小さくなる。したがって，モデル平均を用いれば，AIC などの単純な情報量基準より優れた結果を得る可能性がある。他の伝統的なモデル選択の手法に関しても，同様のことが言える。

以下，幾つかのモデル平均の手法を紹介し，次に，PMSE の問題を論じる。

1　ベイジアンモデル平均

近年，ベイズ的な枠組みを使ったモデル平均の手法が数多く発表されてきた（Draper［1995］，Hoeting et al.［1999］，Clyde and George［2004］）。ベイジアンでは，パラメーターを確率変数として考える。ここでは関心のある母数を μ とするが，これは各モデルのパラメーター θ_k の関数とする。分析では，事後分布 $\pi(\mu|y)$ や，条件付期待値 $\hat{\mu}=E(\mu|y)$ の推定を目的とする。

これまでの設定と同じく，候補となるモデル族を $\mathcal{M}=\{M_1, M_2, \cdots, M_K\}$ とする。ベイズの公式を利用して，モデル平均によって $\pi(\mu|y)$ と $E(\mu|y)$ が導出するが，そのためには，二つの事前分布が必要となる。一つは各モデルの事前分布で $P(M_k)$ と記す。もう一つは各モデルのパラメーター θ_k の，そのモデルのもとでの事前分布 $\pi(\theta_k|M_k)$ である。この二つの事前分布を所与とし，M_k のもとでの y の尤度関数を $L(y|M_k, \theta_k)$ とすると，M_k の事後分布は

$$P(M_k|y)=\frac{P(M_k)\lambda_k}{\sum_{k=1}^{K}P(M_k)\lambda_k} \tag{13}$$

となる。ただし

$$\lambda_k=\int L(y|M_k, \theta_k)\pi(\theta_k|M_k)d\theta_k \tag{14}$$

である。さらに μ の事後分布は

$$\pi(\mu|y)=\sum_{k=1}^{K}P(M_k|y)\pi(\mu|M_k, y) \tag{15}$$

となる，ただし，$\pi(\mu|M_k, y)$ は M_k が真であるという条件のもとでの μ の事後分布であり，事前分布 $\pi(\theta_k|M_k)$ を利用して，ベイズの公式により，解析的に，または MCMC などの方法で数値的に推定する。そして μ の事後期待値は

$$\hat{\mu}=E(\mu|y)=\sum_{k=1}^{K}P(M_k|y)E(\mu|M_k, y) \tag{16}$$

と計算される。

ベイジアンモデル平均はすでに多くの成果をあげているが，候補となるモデルに与える事前確率が主観に依存し場当たり的であることが多く，また，推定結果はこの事前確率に大きく影響を受けるという批判がある。

2 Smoothed-AIC，BIC

Buckland et al.［1997］はモデルの AIC および BIC の値を利用したモデル平均法を提案した。この方法はベイジアンモデル平均と密接な関連を持っている。

前述したように BIC は事後確率 $P(M_k|y)$ を評価している。Claeskens and Hjort［2008］によれば近似的に

$$\mathrm{BIC} \approx -2\log(\lambda_k) \tag{17}$$

が成立する。そこで $P(M_k)$ が k に関して均一であると仮定すれば，(13) 式から

$$P(M_k\,|\,y) \approx \frac{\exp(-BIC_k/2)}{\sum_{k=1}^{K}\exp(-BIC_k/2)} \tag{18}$$

となる。(16) 式は，モデル平均のウエイトとして

$$c_{BIC}(M_k) = \frac{\exp(-BIC_k/2)}{\sum_{k=1}^{K}\exp(-BIC_k/2)} \tag{19}$$

が一つの選択肢であることを示唆している。そのウエイトを利用すると，モデル平均の推定量は

$$\hat{\mu}_{MA-BIC} = \sum_{M_k\in\mathcal{M}} c_{BIC}(M_k)\hat{\mu}_{M_k} \tag{20}$$

となる。この推定量は，smoothed-BIC-based estimator と呼ばれる。

$\hat{\mu}_{MA-BIC}$ と同じ形を使って，smoothed-AIC-based estimator が

$$\hat{\mu}_{MA-\mathrm{AIC}} = \sum_{M_k\in\mathcal{M}} c_{\mathrm{AIC}}(M_k)\hat{\mu}_{M_k} \tag{21}$$

と定義される。ただし，

$$c_{\mathrm{AIC}}(M_k) = \frac{\exp(-\mathrm{AIC}_k/2)}{\sum_{k=1}^{K}\exp(-\mathrm{AIC}_k/2)} \tag{22}$$

である。モデル族 $\mathcal{M}$ に含まれる各モデルに最尤法を適用すれば，各モデルの推定量とウエイトが計算できるため，この二つの推定量の計算は簡単である。

3 Mallows' c_pを用いたモデル平均

Hansen［2007］は，Mallows' c_p を利用したモデル平均の方法を提案した。モデルは，$i=1, 2, \cdots, n$，について

$$y_i = \mu_i + \varepsilon_i \tag{23}$$

$$\mu_i = \sum_{j=1}^{\infty}\theta_j x_{ji} \tag{24}$$

$$E(\varepsilon_i|x_i) = 0 \tag{25}$$

$$E(\varepsilon_i^2|x_i) = \sigma^2 \tag{26}$$

と線形で表現されるとする。$(y_i, x_{1i}, x_{2i}, \cdots)$ は無作為標本であり，ε_i が撹乱項である。$\sum_{j=1}^{\infty}\theta_j x_{ji}$ は，非線形またはノンパラメトリックな関数の級数展開と理解することができるから，この方法の適用範囲は，線形モデルにとどまらない。

このモデルのもとで，x_i の最初の k 要素だけを説明変数とするモデル

$$\mu_i = \sum_{j=1}^{k}\theta_j x_{ji} \tag{27}$$

を M_k と表す。最小は M_1，最大は M_K とする。モデル平均の推定量は

$$\hat{\eta} = \sum_{k=1}^{K}\omega_k\begin{pmatrix}\hat{\eta}_k\\ 0\end{pmatrix} \tag{28}$$

となる。ただし，$\eta_k=(\theta_1, \theta_2, \cdots, \theta_k)'$ として，$\hat{\eta}_k=(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)'$ は，η_k の M_k のもとでの推定量，ω_k は M_k に与えるウエイトで，非負であり和は 1 である。K 個のウエイトをまとめて，$w=(\omega_1, \omega_2, \cdots, \omega_K)'$ と記そう。μ のモデル平均の推定量は，$x_i=(x_{1i}, x_{2i}, \cdots, x_{Ki})$ として，

$$\hat{\mu}_i(w) = x_i\hat{\eta}$$

となる。この論文では，損失関数を

$$L_n(w) = \sum_{i=1}^{n}(\hat{\mu}_i(w)-\mu_i)^2$$

として，Mallows'c_p が一定の条件のもとで，損失関数の期待値と定数の和の不偏推定量であることが示されている。ただし，Mallows'c_p は

$$c_n(w) = \sum_{i=1}^{n}(y_i - x_i\hat{\eta})^2 + 2\sigma^2 k^* \tag{29}$$

と定義される。x_i は最大モデルに関する説明変数のデータ行列，また，$k^*=\sum_{k=1}^{K}\omega_k k$ である。$c_n(w)$ を最小にするウエイトを

$$\hat{w}_n = \underset{w}{arg\,min}\, c_n(w) \tag{30}$$

と定義すると，一定の条件のもとで

$$\frac{L_n(\hat{w}_n)}{\inf_w L_n(w)} \xrightarrow{p} 1 \tag{31}$$

となり，$\hat{w}_n$ は漸近的に損失関数の値を最小にする最適ウエイトとなる。

4　PMSE 推定量の漸近分布

モデル選択の結果で選ばれたモデルに関して得られる推定量，ならびにモデル平均法により得られる推定量を PMSE と呼ぶ。通常の推定量は，モデルが真であるという前提の下で導かれていることに注意しよう。PMSE の漸近分布の導出を行った論文は，Hjort and Claeskens［2003］，Potscher［2006］，Leeb and Potscher［2008］などがある。ここでは，local misspecification のもとでの漸近分布の導出を行った Hjort and Claeskens［2003］を紹介する。

$(y_1, y_2, \cdots, y_n)$ を確率変数 y の標本とし，その密度関数を

$$f_{true}(y)=f_n(y)=f(y, \theta_0, \gamma),\ \gamma=\gamma_0+\frac{1}{\sqrt{n}}\delta \tag{32}$$

とする。θ_0 と γ_0 は，次元が p と K のパラメーターベクトルである。最小のモデルは δ に属するパラメーターを含まず，$f_{narr}(y, \theta)=f(y, \theta, \gamma_0)$，最大のモデルはすべての δ を含むようなモデルで，$f_{full}(y, \theta, \gamma)$ と表す。δ に属するパラメーターの一部を使うことによって，サブモデル M_{S_j} を構築することができる。ただし，S_j は M_{S_j} に入る δ の添え字の集合である。δ の要素のすべての組み合わせを考えると，最大と最小のモデルを含めて，サブモデルは 2^K 個ある。ある母数 μ を推定したい場合，そのモデル平均推定量は，$\hat{\mu}_{S_j}$ を M_{S_j} のもとでの推定量として，

$$\hat{\mu}=\sum_{j\in 2^K} c(M_{S_j})\hat{\mu}_{S_j} \tag{33}$$

と定義される。

このような設定のもとで，モデル平均推定量 $\hat{\mu}$ の漸近分布が導出された。一般的には，$c(M_{S_j})$ は確率変数となり，導出された分布は標準的な分布に従わず，混合正規分布になる。そして，モデル選択による推定量の漸近分布は，その特殊ケースになる。さらに，モデル選択により得られたモデルを真のモデルとして信頼空間を導出した場合，信頼空間が過小評価されるという結果も導かれている。異なったモデル平均推定量間の，リスクの比較も行われている。

Ⅳ　未解決な問題

PMSE の漸近分布を用いて有限標本の分布を近似し，検定や信頼区間の導出を行うのは，一つの自然な考え方であろう。しかし，このような近似では，観測個数の大きさに関わらず，精度を高めることが不可能であることが，Potscher［2006］や Leeb and Potscher［2008］などによって示された。

以下では Leeb and Potscher［2008］の結果を説明しよう。モデルを行列表現で

$$y_i=x_i\beta+z_i\lambda+\varepsilon_i \tag{34}$$

と表す。x_i と z_i はそれぞれ $(1\times k)$ と $(1\times q)$ の非確率的な説明変数の行列，$\varepsilon_i \sim N(0, \sigma^2)$ を撹乱項，パラメーターをまとめて $\theta=(\beta', \lambda')'$ とする。x_i に含まれる説明変数は必ずモデルに入れるが，モデル選択により，z_i に含まれる説明変数のうち一部を残す。そして，選択されたモデルのもとで，β を含むパラメーターを推定する。

β の推定量 $\hat{\beta}$ の有限標本の分布を $G_{n,\theta,\sigma}(b)$ とする。論文は，任意の正値 δ に関して，$\hat{G}_{n}(b)$ が，

$$P_{n,\theta,\sigma}\left(|\hat{G}_n(b)-G_{n,\theta,\sigma}(b)|>\delta\right)\xrightarrow{n\to\infty}0 \tag{35}$$

を満たす，すなわち θ に関して，ポイントワイズに一致性を持つ推定量 $\hat{G}_n(b)$ が見つかることを証明した。しかし，他方で，(35) 式を満たす任意の一致推定量 $\hat{G}_n(b)$ は，正値の η と δ に関して

$$\sup_{\|\theta\|\eta} P_{n,\theta,\sigma}\left(|\hat{G}_n(b) - G_{n,\theta,\sigma}(b)| > \delta\right) \xrightarrow{n\to\infty} 1 \qquad (36)$$

を満たすことを示した。これは，$\hat{G}_n(b)$ が θ に関して一様な一致性を持たないことを意味する。一様に一致性を持つ推定量が存在しないという結論は，不可能性と名付けられている。不可能性はこのような回帰モデルの設定のもとだけで成立するのではなく，パラメトリックモデルとセミパラメトリックモデルに関しても成立する。さらに，不可能性は殆どのモデル選択の方法，そしてモデル平均に関しても成立することが示されている。

不可能性の結論により，観測個数を如何に大きくしても，PMSEの有限標本の分布を推定できないことが分かる。だから，有限標本分布の推定量のもとで，推論を進めることができない。Beran and Dumbgen［1998］と Beran［2000］の解決法は，まず，候補となるモデルに対応するリスクを推定し，リスクを最小にするモデルを選ぶ。次に，選んだモデルのリスクの推定量の漸近分布を導出し，この漸近分布のもとで，統計的推論を行う。彼らは，このアプローチに基づいて，推定量を中心とした母数の信頼集合を導いた。

V　モデル平均の応用

モデル選択の手法は，広く利用されている。それとは対照的に，モデル平均法の応用は，まだ数が少ない。この章では，モデル平均を利用した研究を紹介し，さらに，実際のデータを利用して，モデル平均による予測の例を示す。

1　既存研究

経済理論を実証的に分析する場合において重要な手法として，モーメント条件を利用した推定や推論があげられよう。この手法は，モーメント条件が有限個であれば容易に利用できる。しかし，モーメントが多く存在する場合，すべてのモーメント条件を利用すると推定量の性質が劣化するため，モーメント条件を選択して使う必要がある。この問題に関しては，Morimune［1983］は一つ重要な研究となっている。森棟［1985］，Bekker［1994］，Donald and Newey［2001］を参照されたい。そして，Kuersteiner and Okui［2009］はモーメントが複数存在するという問題の特殊ケースである操作変数選択の問題に，モデル平均法を応用した。二段階最小二乗法の第一段階で Hansen［2007］が提案したモデル平均の手法を利用して操作変数の選択を行い，次に，最小の損失関数の値をもたらす推定量を導出するのである。

モデル平均の手法を，ボラティリティモデルの推定や Value-at-Risk の分析に応用した実証研究もある。Brownlees and Gallo［2008］は Smoothed-AIC，BIC，FIC[1] の手法を用いてボラティリティモデルの変数選択を行った。また，Pesaran et al.［2009］はモデル平均の手法を利用して Value-at-Rsik の分析を行い，モデル平均の推定結果に利用できる診断テストを考案した。

2　モデル平均による予測

日経225のデータに関して，ARCH（Autoregressive Conditional Heteroscedasticity）型モデルを推定して，株価指数の実現ボラティリティの予測を行ってみよう。ARCH型モデルは，Engle［1982］によって時系列データの分散不均一性を捉えるためのモデルとして考案されてから，数多くのバラエティが提案されてきた。ここでは，その中の代表的なモデルを，モデル平均法の候補モデルとする。以下，それらの候

1）　FIC は Focused Information Criterion の略である。推定対象となる異なった母数の特質に配慮した情報量基準である。詳細に関しては Hjort and Claeskens［2003］または Claeskens and Hjort［2008］を参照されたい。

補モデルを紹介しよう。

2.1 ARCH 型モデル

Engle［1982］は時系列データの分散不均一性を捉えるために，ARCH（q）モデルを提唱した。ARCH（q）では，条件付分散の不均一性が仮定される。収益率 r_t は，

$$r_t = x_t\beta + \varepsilon_t, \tag{37}$$

$$\varepsilon_t = \sqrt{h_t z_t}, \quad z_t \sim iidn(0, 1), \tag{38}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2, \tag{39}$$

と表されるが，h_t は r_t の条件付分散で，ϕ_{t-1} を $t-1$ 期までの情報集合とすると，$r_t|\phi_{t-1} \sim N(x_t\beta, h_t)$ となる。h_t あるいは $\sqrt{h_t}$ は，ファイナンスにおけるボラティリティである。ARCH（q）モデルは，ボラティリティが，過去のショックの二乗に影響されることを表現している。h_t は条件付分散なので，非負でなければならず，$\alpha_0 > 0$，$\alpha_i \geq 0$ という制約が必要となる。

推定上は，ARCH（q）モデルを金融データに適用して，ボラティリティに対するショックを表現しようとすると，次数が長くなるという問題が生じる。これを回避するために提案されたのが，GARCH（p, q）モデル（Bollerslev［1986］）である。GARCH（p, q）モデルは，ARCH（q）モデルにボラティリティのラグ項を追加したもので

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{40}$$

となる。AR モデルに MA 項を追加することで，AR の次数を減らせることと同様に，GARCH モデルにボラティリティのラグ項を導入することで，短い次数で，データの特性を理解することができるようになる。

GARCH モデルにおいても，ボラティリティの非負性を保つため，ARCH における制約に加えて，$\beta_j \geq 0$，$j = 1, 2, \cdots p$ が必要となる。

実証研究が進展する中で，金融データのボラティリティに関する特徴が明らかになってきた。その一つとして，ボラティリティ変動の非対称性がある。非対称性とは，前日の株価変動の正負によって，ボラティリティの大きさが変わるという特性である。経験的には，前日の株価が下落した場合には，前日の株価が上昇した場合よりも，ボラティリティが大きくなる傾向が認められた。GARCH モデルでは，この非対称性を捉えることができない。そこで，GJR-GARCH モデル，EGARCH モデル，TGARCH モデルなどが提案されてきた。

Nelson［1991］によって提案された EGARCH（Exponential GARCH）モデルは，非説明変数の h_t を $\log(h_t)$ に置き換えることによって，ボラティリティ変動の非対称性を捉えることを可能にし，また GARCH モデルにおけるパラメーターの非負制約を除去することにも成功した。EGARCH（p, q）モデルは

$$\log(h_t) = \alpha_0 + \sum_{i=1}^{q} \alpha_i [\theta z_{t-i} + \gamma(|z_{t-i}| - E|z_{t-i}|)] + \sum_{j=1}^{q} \beta_j \log(h_{t-j}) \tag{41}$$

$$\varepsilon_t = \sqrt{h_t z_t}, \quad z_t \sim N[0, 1] i.i.d. \tag{42}$$

となる。ここでは，$\theta < 0$ であれば，ボラティリティ変動の非対称性が表現できる。$\log(h_t)$ は負値をとることもできるので，パラメーターに関する非負制約が不要となる。ただし，EGARCH は GARCH と異なり，ε_t の代わりに z_t でモデルを定式化するため，ε_t のボラティリティに与える影響が分かりにくくなるという欠点を持つ。

Liu and Morimune［2005］は，株価の連続的な上昇または下落がボラティリティに与える影響を捉えるため，OGARCH モデル（Overresponse GARCH モデル）を提案した。ボラティリティ式は

$$h_t = \alpha_0 + \alpha_1 \exp(\phi\gamma_{t-1})\varepsilon_{t-1}^2 + \beta h_{t-1}, \tag{43}$$

となる。γ_{t-1} は，第 t 期まで同じ符号のショックが連続的に現れた日数であり，

$$\gamma_{t-1} \equiv i, \; if \; sign(\varepsilon_{t-1}) = \cdots = sign(\varepsilon_{t-i})$$

$$= -sign(\varepsilon_{t-(i+1)}) \tag{44}$$

となる。α_0，α_1，β，ϕ は未知パラメーターで，$a_0 \geq 0$，$a_1 > 0$，$\beta > 0$ という制約が課せられ，ϕ も正値であることが期待される。したがって，株価が同じ方向に連続的に動くと，ボラティリティは大きくなる可能性が高いと考えられている。

2.2 実現ボラティリティの予測

モデル平均を応用するが，GARCH（1,1），OGARCH（1,1）および EGARCH（1,1）を利用する。モデル平均法としては，Smoothed BIC を採用する。データは1984年1月9日から2007年12月25日までの日経225の日次および週次データで[2)]，サンプルサイズは各々5902と1242である。日次データから週次の実現ボラティリティ（Realized Volatility，RV）を計算する。他方，週次データを用いて上記 ARCH 型モデルを推定し，RV を予測する。

週次の株価を y_t，$t=1,2,\cdots n$ とすると，対数収益率は

$$r_t = 100[\log(y_t) - \log(y_{t-1})] \tag{45}$$

である。週次の RV を

$$RV_t = \sum_{i=1}^{n_t} r_{t_i}^2 \tag{46}$$

と定義する。ただし，n_t は第 t 週の営業日数，r_{t_i} は第 t 週第 i 日の対数収益率である。ローリング・ウインドの幅を442とする。442週のデータを利用してモデルを推定し，1週先の RV を予測する。このようにしてローリング・ウインドを移動しながら，残りの800週間の RV を予測するが，予測の良さを計るため，平均予測誤差二乗和

$$MSPE = \frac{1}{n}\sum_{t=1}^{n}(\widehat{RV}_t - RV_t)^2 \tag{47}$$

を利用する。$\widehat{RV}_t$ は RV_t の推定値である。

Smoothed BIC によるモデル平均は2種類計算して，予測を行う。一つは GARCH（1,1）と OGARCH（1,1）の二つのモデルの平均，もう一つは GARCH（1,1），OGARCH（1,1），EGARCH（1,1）の三つのモデルによる平均である。モデル平均による予測値は

$$\widetilde{RV}_t = \sum_{m=1}^{M} c_{BIC}(m)\widehat{RV}_{m,t} \tag{48}$$

と定義する。M はモデル平均に使われるモデル数で，$c_{BIC}(m)$ はモデル m に対応する BIC によって計算されるウエイトである。結果を表1にまとめた。

Best-BIC-I は，ローリング予測の中で，GARCH（1,1）と OGARCH（1,1）の中の BIC が小さい方で予測した場合の結果を表す。Best-BIC-II は3モデルの場合である。そして，2モデルと3モデルは，二つのモデルと三つのモデルによるモデル平均を表す。単独の GARCH や OGARCH と比べて，単独の EGARCH の $MSPE$ が一番大きい。そして，二つのモデルによるモデル平均の結果は Best-BIC-I よりも良く，最も良い結果を出している。三つのモデル平均の結果は，EGARCH のパフォーマンスに影響され悪い結果となっているが，それでも Best-BIC-II よりは良い。

以上は，モデル平均が単純な BIC によるモデル選択よりも多少優れていることを示しているが，実証研究を行う際，この結果はいつも成立するわけではない。この例についても，ローリング・ウインドの幅を変えたり，予測期間を長くしたり，候補となるモデル群を変えたりする

表1．予測の結果

	GARCH	OGARCH	EGARCH	Best-BIC-I	Best-BIC-II	2モデル	3モデル
MSPE	95.91	96.26	100.05	96.26	98.77	95.83	97.87

2） データの出所：Morningstar, Inc.

ことで，結果が異なりうる。モデル平均は理論的に優れているが，それが実証研究に反映されるには，モデル平均以外のさらなる工夫が必要である。

Ⅵ　結論

本稿では，モデル選択の拡張であるモデル平均に関して，最近の研究成果を概観した。さらに，その未解決問題を説明し，応用例を挙げた。モデル平均は，リスクの観点で比較すると，単純なモデル選択より優れている。また，モデル平均の漸近分布が導出されており，モデル平均による推定量のリスクの評価法も確立されている。しかし，漸近分布が有限標本分布の良い近似とはなっておらず，統計的推論は難しい。この問題を解決するために，Beran and Dumbgen［1998］と Beran［2000］の方法を始めとして，様々な解決法が試みられる必要がある。

謝辞

詳細なコメントをいただいた森棟公夫先生に深く感謝します。

参考文献

Akaike, H.［1973］“Information Theory and an Extension of the Maximum Likelihood Principle,” in N., P. B. and C. F. eds. *Proc. of the 2nd Int. Symp. on Information Theory*, pp. 267–281.

Bekker, P. A.［1994］“Alternative Approximations to the Distributions of Instrumental Variable Estimators,” *Econometrica*, Vol. 62, No. 3, pp. 657–81, May.

Beran, R.［2000］“REACT Scatterplot Smoothers : Superefficiency through Basis Economy,” *Journal of the American Statistical Association*, Vol. 95, No. 449, pp. 155–171.

Beran, R. and L. Dumbgen［1998］“Modulation of Estimators and Confidence Sets,” *Annals of Statistics*, Vol. 26, No. 5, pp. 1826–1856.

Bollerslev, T.［1986］“Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, Vol. 31, No. 3, pp. 307–327, April.

Brownlees, C. T. and G. M. Gallo［2008］“On Variable Selection for Volatility Forecasting : The Role of Focused Selection Criteria,” *Journal of Financial Econometrics*, Vol. 6, No. 4, pp. 513–539.

Buckland, S. T., C. Burnham, K. P. Burnham, and N. H. Augustin［1997］“Model Selection : An Integral Part of Inference,” *Biometrics*, Vol. 53, pp. 603–618.

Claeskens, G. and N. L. Hjort［2008］*Model Selection and Model Averaging* : Cambridge University Press.

Clyde, M. and E. I. George［2004］“Model Uncertainty,” *Statistical Science*, Vol. 19, No. 1, pp. 81–94.

Donald, S. G. and W. K. Newey［2001］“Choosing the Number of Instruments,” *Econometrica*, Vol. 69, No. 5, pp. 1161–1191, September.

Draper, D.［1995］“Assessment and Propagation of Model Uncertainty,” *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 57, No. 1, pp. 45–97.

Engle, R. F.［1982］“Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation,” *Econometrica*, Vol. 50, No. 4, pp. 987–1007, July.

Hansen, B. E.［2007］“Least Squares Model Averaging,” *Econometrica*, Vol. 75, No. 4, pp. 1175–1189, 07.

Hjort, N. and G. Claeskens［2003］“Frequentist Model Average Estimators,” *Journal of the American Statistical Association*, Vol. 98, pp. 879–899, January.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky［1999］“Bayesian Model Averaging : A Tutorial,” *Statistical Science*, Vol. 14, No. 4, pp. 382–417. with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors.

Kuersteiner, G. and R. Okui［2009］“Instrument Selection by First Stage Prediction Averaging.” Unpublished manuscript.

Kullback, S. and R. A. Leibler［1951］“On Information and Sufficiency,” *The Annals of Mathematical Statistics*, Vol. 22, No. 1, pp. 79–86.

Leeb, H. and B. M. Potscher [2008] "Can One Estimate The Unconditional Distribution of Post-Model-Selection Estimators ?" *Econometric Theory*, Vol. 24, No. 3, pp. 338–376.

Liu, Q. and K. Morimune [2005] "A Modified GARCH Model with Spells of Shocks," *Asia*-Pacific Financial Markets, Vol. 12, No. 1, pp. 29–44.

Mallows, C. L. [1973] "Some Comments on c_p," *Technometrics*, Vol. 15, pp. 661–675.

Morimune, K. [1983] "Approximate Distributions of k-Class Estimators When the Degree of Overidentifiability Is Large Compared with the Sample Size," *Econometrica*, Vol. 51, No. 3, pp. 821–841, May.

Nelson, D. B. [1991] "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, Vol. 59, No. 2, pp. 347–70, March.

Pesaran, M. H., C. Schleicher, and P. Zaffaroni [2009] "Model Averaging in Risk Management with An Application to Future Markets," *Journal of Empirical Finance*, Vol. 16, No. 2, pp. 280–305.

Potscher, B. M. [2006] "The Distribution of Model Averaging Estimators and an Impossibility Result Regarding Its Estimation," MPRA Paper 73, University Library of Munich, Germany.

Schwarz, G. [1978] "Estimating the Dimension of a Model," *The Annals of Statistics*, Vol. 6, No. 2, pp. 461–464.

Stone, M. [1977] "An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1, pp. 44–47.

小西貞則・北川源四郎［2004］『情報量基準』朝倉書店。

竹内啓［1976］「情報統計量の分布とモデルの適切さの基準」『数理科学』第 153 巻，12–18 ページ。

森棟公夫［1985］『経済モデルの推定と検定』共立出版。