統計数理(2006)
第 54 巻 第 2 号 481–510

特集「予測と発見」
[原著論文]

予測の平均二乗誤差を基準とするモデル選択について

片山 直也[†]

(受付 2005 年 12 月 27 日; 改訂 2006 年 3 月 24 日)

要　旨

本稿では，入れ子型の複数の線形時系列モデルにおいて，予測の平均二乗誤差(Prediction Mean Squared Error, PMSE)を基準にモデル選択を行う問題を考える．まず，標本の大きさに依存する局所的な母数空間において，入れ子構造になっている 2 つのモデルを考えた．そしてそれぞれ母数を推定したとき，2 つのモデルによる予測量の PMSE の意味での相対的な良さがある不等式で与えられ，その不等式はワルド統計量を用いた不等式で近似的に求められることを発見した．そこで簡単なモデルを帰無仮説，複雑なモデルを対立仮説とする検定でワルド統計量等の古典的検定統計量の漸近的性質を導き，さらに Hosoya (1984, 1986, 1989)の提唱する一般化尤度比検定を用いたモデル選択法を提案した．ただし棄却点は，自由度と非心度が同じ整数値の非心カイ 2 乗分布により決まる．このモデル選択法は，簡単なモデルの方が PMSE の意味で優れているのに複雑なモデルを選択する誤り，検定で言うところの第 1 種の過誤をコントロールしながらモデル選択を行える．またこのモデル選択法の枠組みから，すでに実用化されているモデル選択法である AIC (Akaike's Information Criterion)の解釈を与えた．

キーワード： 一般化尤度比検定，AIC，非心カイ二乗分布，線形時系列，PMSE，多重検定．

1. はじめに

本稿では，入れ子構造にある複数の線形時系列モデルにおいて，予測の平均二乗誤差(Prediction Mean Squared Error, PMSE)を基準にモデル選択を行う問題を考える．

予測問題とモデル選択問題は密接な関係にある．Mallows (1973)による C_p 基準や Akaike (1969)の FPE (Final Prediction Error)は，一期先予測の PMSE を不偏推定するというアイデアから生まれたものである．また，Akaike (1973)の赤池情報量規準(Akaike's Information Criterion, AIC)は，漸近的に FPE と一致することからこれも PMSE の観点からモデル選択をしていると解釈できる．これらの手法は，様々な分野で実用化されている．しかしながら，一致性の議論を除いて，これら手法で選択されたモデルが，実際に PMSE の意味で最良かどうかは，未だ理論的に解明されていない．つまり，PMSE の見地から最良のモデルを選択する確率がいかほどか分かっていない．

本稿は，モデル選択問題で時折みられる一致性については議論しない．なぜなら標本の大きさは有限であり，有限標本で真のモデルより簡単なモデルを採用するということは，PMSE の

[†] 一橋大学 経済学研究科：〒 186–8601 東京都国立市中 2–1

観点では必ずしも間違いではない．

例えば，真のモデルが定常な次数が2の自己回帰(AR(2))モデル$y_t=\phi_1 y_{t-1}+\phi_2 y_{t-2}+\varepsilon_t$であったとしよう．ここで$\varepsilon_t \sim iid\, N(0,\sigma^2)$，$\phi_1$と$\phi_2$は0ではないとする．もし，$\phi_2$が適当に大きな値であれば，AR(2)モデルで予測を行うのが最良であろう．しかしながら，ϕ_2が極めて0に近い場合，AR(1)モデルで予測を行うのは必ずしも間違いではない．なぜなら，標本の大きさに依存する標準誤差を伴いながらϕ_2を推定して予測を行うよりも，$\phi_2=0$と制約してAR(1)モデルとして予測を行うほうがPMSEの観点から優れている場合があるからである．似たケースはKatayama (2006)で紹介されている．また，AR(3)モデル$y_t=\phi_1 y_{t-1}+\phi_2 y_{t-2}+\phi_3 y_{t-3}+\varepsilon_t$で予測を行うよりもAR(2)モデルで予測を行う方がPMSEの観点から優れている．なぜなら，AR(2)モデルによる予測は，AR(3)モデルの母数$\phi_3=0$を正しく0と制約しているのに対し，AR(3)モデルによる予測は，標準誤差を伴いながら$\phi_3=0$を推定しているからである．これらの結果の一般化は2節で行われている．

しかし，我々は真のモデルがわからないのでAR(p)，$p=1,2,3,\ldots$の中でどうやってモデルを選択すればよいのかという疑問が残る．なぜならPMSEの比較は，標本の大きさ，推定量の標準誤差，母数の真の値に依存するが，後者2つは一般に未知であるし，既知であったとしても，何の選択基準でもってモデル選択をするのが最良かわかっていないからである．

これらの疑問は，次の3点に集約されるであろう．1)PMSEの観点からモデルを比較するための条件とは何か？，2)その条件が分かったとして，確認(推定)する手法は何か？，3)AIC等の既存のモデル選択法は1)の条件を満たすか？

本論文は，これらの疑問に答えようと書かれたものである．

まず2節では，論文を通して議論する線形時系列モデルを定義して，最良線形予測量(Best Linear Predictor, BLP)とそのPMSEを紹介する．また，母数空間が入れ子構造になっている2つのモデルを考え，それぞれ母数を推定し予測量を与えたとき，漸近的なPMSEを比較する．そして，相対的なPMSEの良さは，漸近的にある2次形式と母数の数との大小で決まることを示す．これは，Katayama (2006)を拡張するものである．この問題の過去の文献の紹介もKatayama (2006)にある．

3節では，2つのモデルを比較するワルド(Wald)統計量の漸近的性質を調べる．また尤度比(Likelihood Ratio, LR)統計量やスコア(Score)統計量がワルド統計量に漸近的に一致することを示す．

4節では，2節と3節の結果を利用して，入れ子構造にある3つ以上の線形時系列モデルの中からPMSEの意味で望ましいモデルを選択する方法を提案する．具体的には，Hosoya (1984, 1986, 1989)により提唱された一般化尤度比(Generalized Likelihood Ratio, GLR)検定を応用した方法である．ただし棄却点はχ^2分布ではなく，自由度と非心度が同じ整数値の非心χ^2分布により決定される．提案するモデル選択法は，簡単なモデルの方がPMSEの意味で優れているのに複雑なモデルを選択する誤り，検定で言うところの第1種の過誤をコントロールしながらモデル選択を行えるという利点を持つ．過去このようなモデル選択法は，佐和(1979)の7.1.7節等があるが，3つ以上のモデルから選択する問題はなされていない．またAICについて，このモデル選択法における解釈を与える．具体的には，竹内(1976)やHosoya (1984)により，AICをLR検定で解釈すると母数の数に依存するある整数値との大小で決まることがわかっているが，その整数値の解釈である．

5節では，4節までの議論に関係する非心χ^2分布のいくつかの結果と統計ソフトウェアS-PLUSによる実行方法を簡単に述べた．

論文を通して，数学的扱いの便宜上，母数が標本の大きさnに依存する局所的なケースで考えている．これは検定での局所対立仮説(local alternatives)の導入アイデアを利用して，真の

母数を n に依存させて，標本の大きさ，推定量の標準誤差，母数の真の値に依存する PMSE の相対的なよさの関係を簡単化しようとしたためである．

2 節から 4 節の証明は全て最後の節の補遺に記載した．最後に論文を通して用いる記号を記載する．L をラグオペレータ，$\|x\|=(\boldsymbol{x}'\boldsymbol{x})^{1/2}$ をベクトル x のユークリッドノルム，関数 f の微分を $\partial f(\boldsymbol{x})/\partial \boldsymbol{x}|_{\boldsymbol{x}=\boldsymbol{y}}=\partial f(\boldsymbol{y})/\partial \boldsymbol{x}$，$f^{(1)}(\boldsymbol{y})=\partial f(\boldsymbol{y})/\partial \boldsymbol{x}$，$f^{(2)}(\boldsymbol{y})=\partial^2 f(\boldsymbol{y})/\partial \boldsymbol{x}\partial \boldsymbol{x}'$，$\mathbb{I}$ は適当な次数の単位行列，$\mathbf{0}$ は全ての要素が 0 からなる適当な行と列の零行列とする．また確率変数 X_n と Y_n において，確率収束 $X_n-Y_n \xrightarrow{p} 0$ を $X_n \overset{\mathrm{a}}{=} Y_n$ とする．この記号は確率変数列ではなく数列の場合でも用いる．また，確率変数 X_n の確率変数 X への確率 1 収束を $X_n \xrightarrow{a.c.} X$ と表す．X_n と Y_n は連続型の分布を持つ確率変数で分布関数をそれぞれ F_n と G_n とおくとき，同じ漸近分布を持つこと，即ち $\lim_{n\to\infty} F_n(x)=\lim_{n\to\infty} G_n(x)$ を $X_n \overset{\mathrm{a}}{\sim} Y_n$ と表す．漸近的な全ての表現は，$n\to\infty$ とすることで得られる．故に混乱のない限りこの記述を略す．また，局所的な母数空間で漸近論を展開するために，記号の添字に n が必要とされるケースが多いが，混乱のない限りこの記述も略す．

2. モデルが未知の時の PMSE の漸近的性質

2.1 モデルの定義と仮定

次の線形時系列モデル $\{y_t\}$ について考える．

$$y_t=-\sum_{j=1}^{t-1}\pi_j(\boldsymbol{\theta}^0)y_{t-j}+\varepsilon_t=\sum_{j=1}^{t-1}\psi_j(\boldsymbol{\theta}^0)\varepsilon_{t-j}+\varepsilon_t,\quad t=2,3,\ldots, \tag{2.1}$$

ここで $\varepsilon_t\sim iid\,N(0,\sigma_0^2)$ で，$\boldsymbol{\theta}^0$ は真の母数ベクトルで，コンパクトかつ凸の母数空間 $\boldsymbol{\Theta}\subset\mathbb{R}^q$ の要素である．また，$\boldsymbol{\theta}\in\boldsymbol{\Theta}$ について一様に $\sum_{j=1}^{\infty}\pi_j(\boldsymbol{\theta})^2$ と $\sum_{j=1}^{\infty}\psi_j(\boldsymbol{\theta})^2$ は有界で，$\pi_0(\boldsymbol{\theta})=\psi_0(\boldsymbol{\theta})=1$ とする．さらに，理論展開の簡単化のため，$\{y_t=0, t\le 0\}$，$y_1=\varepsilon_1$ とする．

論文を通して，モデル(2.1)より $\{y_t\}_{t=1}^n$ を観測した場合で考える．

母数 $\boldsymbol{\theta}^0$ と σ_0^2 を最尤法により推定する．$e_t(\boldsymbol{\theta})=\sum_{j=0}^{t-1}\pi_j(\boldsymbol{\theta})y_{t-j}$ とするとき 2 階連続微分可能な対数尤度関数，

$$l(\boldsymbol{\theta},\sigma^2)=-\frac{n}{2}\log 2\pi-\frac{n}{2}\log\sigma^2-\frac{1}{2\sigma^2}\sum_{t=1}^{n}e_t(\boldsymbol{\theta})^2 \tag{2.2}$$

の尤度方程式を解いて得られる最尤推定量(Maximum Likelihood Estimator, MLE)を，それぞれ $\widehat{\boldsymbol{\theta}}$，$\widehat{\sigma}^2$ と置く．ここで $\widehat{\boldsymbol{\theta}}$ は $Q(\boldsymbol{\theta})=\sum_{t=1}^{n}e_t(\boldsymbol{\theta})^2/(2\sigma_0^2)$ を最小化し，$\widehat{\sigma}^2=\sum_{t=1}^{n}e_t(\widehat{\boldsymbol{\theta}})^2/n$ となる．

MLE の漸近的性質を保証するため仮定をおく．注意として，後に $\boldsymbol{\theta}^0$ は標本の大きさ n に依存させるが，以下の仮定は $\boldsymbol{\theta}^0$ が n に依存するケースを含め，論文を通して仮定する．

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n}\sum_{t=1}^{n}e_t(\boldsymbol{\theta})^2-\mathrm{E}[u_t(\boldsymbol{\theta})]^2\right|\xrightarrow{a.c.}\mathbf{0},\qquad \widehat{\boldsymbol{I}}(\boldsymbol{\theta}^0)=\frac{1}{n\sigma_0^2}\sum_{t=1}^{n}\frac{\partial e_t(\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}}\frac{\partial e_t(\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}'},\\
&\widehat{\boldsymbol{I}}(\boldsymbol{\theta}^0)-\boldsymbol{I}(\boldsymbol{\theta}^0)\xrightarrow{a.c.}\mathbf{0},\qquad \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|\frac{1}{n}\frac{\partial^2 Q(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}-\boldsymbol{I}(\boldsymbol{\theta})\right|\xrightarrow{a.c.}\mathbf{0},\\
&\frac{\partial e_t(\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}}=\sum_{j=1}^{t-1}\boldsymbol{d}_j(\boldsymbol{\theta}^0)\varepsilon_{t-j},\quad (t\ge 2),\qquad \boldsymbol{I}(\boldsymbol{\theta}^0)=\sum_{j=1}^{\infty}\boldsymbol{d}_j(\boldsymbol{\theta}^0)\boldsymbol{d}_j(\boldsymbol{\theta}^0)',\\
&\widehat{\boldsymbol{R}}(\boldsymbol{\theta}^*)=\frac{1}{n\sigma_0^2}\sum_{t=1}^{n}e_t(\boldsymbol{\theta}^0)\frac{\partial^2 e_t(\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}+\left(\frac{1}{n}\frac{\partial^2 Q(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}-\frac{1}{n}\frac{\partial^2 Q(\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right)\xrightarrow{a.c.}\mathbf{0},
\end{aligned}\tag{2.3}$$

ここで，$\mathrm{E}[u_t(\boldsymbol{\theta})]^2$ は $\boldsymbol{\Theta}$ 上有界かつ一様連続で $u_t(\boldsymbol{\theta})=\sum_{j=0}^{\infty}\pi_j(\boldsymbol{\theta})L^j\{\sum_{k=0}^{\infty}\psi_j(\boldsymbol{\theta}^0)\varepsilon_{t-k}\}$，$\boldsymbol{I}(\boldsymbol{\theta}^0)$ は $q\times q$ の正値定符号かつ対称な行列で $\boldsymbol{I}(\boldsymbol{\theta})$ は $\boldsymbol{\Theta}$ 上連続，$\boldsymbol{d}_j(\boldsymbol{\theta}^0)=\sum_{k=0}^{j}\{\partial\pi_k(\boldsymbol{\theta}^0)/\partial\boldsymbol{\theta}\}\psi_{j-k}(\boldsymbol{\theta}^0)$，

そして $\|\boldsymbol{\theta}^* - \boldsymbol{\theta}^0\| \le \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\|$ とする．(2.3)のこれらの条件は，Fuller (1996)の Lemma 5.5.2 や Theorem 5.5.1 等から考えられた強一致性や漸近正規性のための十分条件である．(2.3)の最初の式は，$\mathrm{E}[u_t(\boldsymbol{\theta})]^2$ が $\theta = \theta^0$ で一意に最小化されるため強一致性を保証する条件となっている．$\widehat{\theta}$ は，$Q(\theta)$ を最小化することから微分可能性と $\widehat{\theta} = \theta^0$ の Taylor 展開により，ある $n_0 > 0$ より大の全ての n で θ^* が存在し，

$$\frac{1}{\sqrt{n}}\frac{\partial Q(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \mathbf{0} = \frac{1}{\sqrt{n}}\frac{\partial Q(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} + \frac{1}{n}\frac{\partial^2 Q(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)$$

と確率 1 で展開できる．(2.3)の残りの式は，$\widehat{Q}^{(2)}(\boldsymbol{\theta}^*) = n^{-1}\partial^2 Q(\boldsymbol{\theta}^*)/(\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}')$ とすると $\widehat{\boldsymbol{Q}}^{(2)}(\boldsymbol{\theta}^*) = \widehat{\boldsymbol{I}}(\boldsymbol{\theta}^0) + \widehat{\boldsymbol{R}}(\boldsymbol{\theta}^*)$ で，$\widehat{Q}^{(2)}(\boldsymbol{\theta}^*) - \boldsymbol{I}(\boldsymbol{\theta}^0) \xrightarrow{a.c.} \mathbf{0}$ となる．また(2.3)の条件は，スコア関数が martingale 系列で，中心極限定理のためのモーメント条件を満たすため，Fuller (1996)の Theorem 5.5.1 の証明と同様にして次が成立する．

$$\boldsymbol{I}(\boldsymbol{\theta}^0)^{-1/2}\frac{1}{\sqrt{n}}\frac{\partial Q(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} = \boldsymbol{I}(\boldsymbol{\theta}^0)^{-1/2}\frac{1}{\sqrt{n}\sigma_0^2}\sum_{t=2}^{n}\left\{\sum_{j=1}^{t-1}\boldsymbol{d}_j(\boldsymbol{\theta}^0)\varepsilon_{t-j}\right\}\varepsilon_t \xrightarrow{d} N\big(\mathbf{0},\ \mathbb{I}\big),$$

ここで $\boldsymbol{I}(\boldsymbol{\theta}^0)^{-1/2}$ は $\boldsymbol{I}(\boldsymbol{\theta}^0) = \boldsymbol{I}(\boldsymbol{\theta}^0)^{1/2}\{\boldsymbol{I}(\boldsymbol{\theta}^0)^{1/2}\}'$ なる下三角行列 $\boldsymbol{I}(\boldsymbol{\theta}^0)^{1/2}$ の逆行列である．さらに Katayama (2006)の Corollary 2.1 を適用させて漸近正規性とモーメント収束が成立する．

$$\text{(2.4)}\qquad \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \overset{\text{a}}{\sim} N\big(\mathbf{0},\ \boldsymbol{I}(\boldsymbol{\theta}^0)^{-1}\big), \qquad \mathrm{E}\left\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right\|^r = O(n^{-r/2}) \qquad (\text{for } r \ge 2).$$

2.2 最良線形予測量とその平均二乗誤差

次にこのモデルの BLP とその PMSE を与える．

モデル(2.1)より $\{y_t\}_{t=1}^n$ を用いた y_{n+h} $(h = 1, 2, \ldots)$の BLP，$y_n(h, \boldsymbol{\theta}^0)$ は次で与えられる．

$$\text{(2.5)}\quad y_n(h,\boldsymbol{\theta}^0) = -\sum_{j=1}^{n+h-1}\pi_j(\boldsymbol{\theta}^0)y_n(h-j,\boldsymbol{\theta}^0) = \sum_{j=1}^{n}c_j(h,\boldsymbol{\theta}^0)y_{n+1-j} = \sum_{j=h}^{n+h-1}\psi_j(\boldsymbol{\theta}^0)e_{n+h-j}(\boldsymbol{\theta}^0),$$

ここで $y_n(h-j,\boldsymbol{\theta}^0) = y_{n+h-j}$ $(j \ge h)$，$e_{n+h-j}(\boldsymbol{\theta}^0) = \varepsilon_{n+h-j}$，そして $c_j(h,\boldsymbol{\theta})$ は Box and Jenkins (1976)の(A5.2.3)式より，$c_j(h,\boldsymbol{\theta}) = -\sum_{i=0}^{h-1}\psi_i(\boldsymbol{\theta})\pi_{j+h-i-1}(\boldsymbol{\theta})$ である．この PMSE，$\sigma_0^2(h)$ は次のようになる．

$$\text{(2.6)}\qquad \sigma_0^2(h) \equiv \mathrm{E}[y_{n+h} - y_n(h,\boldsymbol{\theta}^0)]^2 = \mathrm{E}\left[\sum_{j=0}^{h-1}\psi_j(\boldsymbol{\theta}^0)\varepsilon_{n+h-j}\right]^2 = \sigma_0^2\sum_{j=0}^{h-1}\psi_j(\boldsymbol{\theta}^0)^2.$$

2.3 モデルが未知のときの線形予測量とその平均二乗誤差

2.1 節の MLE はモデルが既知と仮定して得られた結果であるが，ここではモデルが未知で，2 つのモデルを仮定したときの MLE を用いて予測量を構成したときの PMSE の振る舞いを調べる．

まず，記号をいくつか追加する．$\boldsymbol{\theta} = (\boldsymbol{\delta}_1', \boldsymbol{\delta}_2', \boldsymbol{\delta}_3')'$ とし，$\boldsymbol{\delta}_i$ はそれぞれ p_i 次ベクトルで $q = p_1 + p_2 + p_3$ とする．対応させて真の母数を $\boldsymbol{\theta}^0 = ((\boldsymbol{\delta}_1^0)', (\boldsymbol{\delta}_2^0)', (\boldsymbol{\delta}_3^0)')'$ とする．ただし，$\boldsymbol{\delta}_2^0 = \boldsymbol{c}_2/\sqrt{n}$，$\boldsymbol{\delta}_3^0 = \boldsymbol{c}_3/\sqrt{n}$ とし，$\boldsymbol{c}_2$ と $\boldsymbol{c}_3$ はそれぞれ p_2 次と p_3 次の定数ベクトルとする．さて，$\boldsymbol{\delta}_2^0$ と $\boldsymbol{\delta}_3^0$ を $\mathbf{0}$ と制約して $\boldsymbol{\delta}_1^0$ を最尤推定したときの $\boldsymbol{\theta}^0$ と σ_0^2 の MLE を $\overline{\boldsymbol{\theta}} = (\overline{\boldsymbol{\delta}}_1', \mathbf{0}', \mathbf{0}')'$ と $\overline{\sigma}^2$，$\boldsymbol{\delta}_3^0$ を $\mathbf{0}$ と制約して $\boldsymbol{\delta}_1^0$ と $\boldsymbol{\delta}_2^0$ を最尤推定したときの $\boldsymbol{\theta}^0$ と σ_0^2 の MLE を $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\delta}}_1', \widetilde{\boldsymbol{\delta}}_2', \mathbf{0}')'$ と $\widetilde{\sigma}^2$ とする．仮定として，ある $n_0 > 0$ より大の全ての n で行列 $\boldsymbol{I}(\boldsymbol{\theta}^0)$ は正値定符号かつ対称とする．

そして式(2.5)の予測量 $y_n(h,\boldsymbol{\theta}^0)$ の $\boldsymbol{\theta}^0$ に $\widehat{\boldsymbol{\theta}}$，$\overline{\boldsymbol{\theta}}$，そして $\widetilde{\boldsymbol{\theta}}$ を各々代入して予測量を構成したとき，PMSE は次のようになる．

定理 2.1. **線形予測量** $y_n(h,\widehat{\theta})$，$y_n(h,\overline{\theta})$，**そして** $y_n(h,\widetilde{\theta})$ **の** PMSE **は，それぞれ，** $n\to\infty$ **で次のようになる．**

$$\mathrm{E}[y_{n+h}-y_n(h,\widehat{\boldsymbol{\theta}})]^2=\sigma_0^2(h)+\frac{\sigma_0^2}{n}\sum_{j=0}^{\infty}\boldsymbol{\varphi}_j(h,\boldsymbol{\theta}^0)'\boldsymbol{I}(\boldsymbol{\theta}^0)^{-1}\boldsymbol{\varphi}_j(h,\boldsymbol{\theta}^0)+o\left(\frac{1}{n}\right), \tag{2.7a}$$

$$\mathrm{E}[y_{n+h}-y_n(h,\overline{\boldsymbol{\theta}})]^2=\sigma_0^2(h)+\frac{\sigma_0^2}{n}\sum_{j=0}^{\infty}\boldsymbol{\varphi}_j(h,\boldsymbol{\theta}^0)'\overline{\boldsymbol{\Gamma}}(\boldsymbol{\theta}^0)\boldsymbol{\varphi}_j(h,\boldsymbol{\theta}^0)+o\left(\frac{1}{n}\right), \tag{2.7b}$$

$$\mathrm{E}[y_{n+h}-y_n(h,\widetilde{\boldsymbol{\theta}})]^2=\sigma_0^2(h)+\frac{\sigma_0^2}{n}\sum_{j=0}^{\infty}\boldsymbol{\varphi}_j(h,\boldsymbol{\theta}^0)'\widetilde{\boldsymbol{\Gamma}}(\boldsymbol{\theta}^0)\boldsymbol{\varphi}_j(h,\boldsymbol{\theta}^0)+o\left(\frac{1}{n}\right). \tag{2.7c}$$

ここで， $\sigma_0^2(h)$ **は**(2.6)，$\{\boldsymbol{d}_j(\boldsymbol{\theta}^0)\}$ **と** $\boldsymbol{I}(\boldsymbol{\theta}^0)$ **は**(2.3)**でそれぞれ定義され，**

$$\begin{aligned}
&\boldsymbol{\varphi}_j(h,\boldsymbol{\theta}^0)\equiv\sum_{k=0}^{j}\frac{\partial c_{k+1}(h,\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}}\psi_{j-k}(\boldsymbol{\theta}^0)=-\sum_{k=0}^{h-1}\psi_k(\boldsymbol{\theta}^0)\boldsymbol{d}_{j+h-k}(\boldsymbol{\theta}^0),\\
&\boldsymbol{I}(\boldsymbol{\theta}^0)=\begin{pmatrix}\boldsymbol{I}_{11}(\boldsymbol{\theta}^0) & \boldsymbol{I}_{12}(\boldsymbol{\theta}^0) & \boldsymbol{I}_{13}(\boldsymbol{\theta}^0)\\ \boldsymbol{I}_{12}(\boldsymbol{\theta}^0)' & \boldsymbol{I}_{22}(\boldsymbol{\theta}^0) & \boldsymbol{I}_{23}(\boldsymbol{\theta}^0)\\ \boldsymbol{I}_{13}(\boldsymbol{\theta}^0)' & \boldsymbol{I}_{23}(\boldsymbol{\theta}^0)' & \boldsymbol{I}_{33}(\boldsymbol{\theta}^0)\end{pmatrix},\\
&\overline{\boldsymbol{\Gamma}}(\boldsymbol{\theta}^0)=\begin{pmatrix}\boldsymbol{I}_{11}^{-1}+\boldsymbol{I}_{11}^{-1}\boldsymbol{I}_{1,23}\boldsymbol{c}_{23}\boldsymbol{c}_{23}'\boldsymbol{I}_{1,23}'\boldsymbol{I}_{11}^{-1} & -\boldsymbol{I}_{11}^{-1}\boldsymbol{I}_{1,23}\boldsymbol{c}_{23}\boldsymbol{c}_{23}'\\ -\boldsymbol{c}_{23}\boldsymbol{c}_{23}'\boldsymbol{I}_{1,23}'\boldsymbol{I}_{11}^{-1} & \boldsymbol{c}_{23}\boldsymbol{c}_{23}'\end{pmatrix},\\
&\widetilde{\boldsymbol{\Gamma}}(\boldsymbol{\theta}^0)=\begin{pmatrix}\boldsymbol{I}_{12,12}^{-1}+\boldsymbol{I}_{12,12}^{-1}\boldsymbol{I}_{12,3}\boldsymbol{c}_3\boldsymbol{c}_3'\boldsymbol{I}_{12,3}'\boldsymbol{I}_{12,12}^{-1} & -\boldsymbol{I}_{12,12}^{-1}\boldsymbol{I}_{12,3}\boldsymbol{c}_3\boldsymbol{c}_3'\\ -\boldsymbol{c}_3\boldsymbol{c}_3'\boldsymbol{I}_{12,3}'\boldsymbol{I}_{12,12}^{-1} & \boldsymbol{c}_3\boldsymbol{c}_3'\end{pmatrix}.
\end{aligned} \tag{2.8}$$

ここで $\boldsymbol{I}_{ij}(\boldsymbol{\theta}^0)$ **は** $p_i\times p_j$ **行列で** $\boldsymbol{I}_{ij}=\boldsymbol{I}_{ij}(\boldsymbol{\theta}^0)$，$\boldsymbol{c}_{23}=(\boldsymbol{c}_2',\boldsymbol{c}_3')'$，$\boldsymbol{I}_{1,23}=\boldsymbol{I}_{1,23}(\boldsymbol{\theta}^0)=(\boldsymbol{I}_{12},\boldsymbol{I}_{13})$，$\boldsymbol{I}_{12,12}$ **は，** $\boldsymbol{I}(\boldsymbol{\theta}^0)$ **の最初の** p_1+p_2 **行** p_1+p_2 **列の小行列，** $\boldsymbol{I}_{12,3}=\boldsymbol{I}_{12,3}(\boldsymbol{\theta}^0)=(\boldsymbol{I}_{13}',\boldsymbol{I}_{23}')'$ **である．**

この漸近的な PMSE **の結果は，証明の**(A.2)**から** BLP **と母数を推定した予測量との差の** 2 **次モーメントが，** $O(1/n)$ **で** BLP **の** PMSE **より大きくなることを示し，** PMSE **の相対的な大きさが比較できることを示している．そこで次の記号を定義する．これは** Katayama (2006)**でも導入された記号である．**

定義 1. **確率過程** $\{x_t\}$ **より観測された** $\{x_t\}_{t=1}^n$ **を用いて** x_{n+h} **（**$h=1,2,\ldots$**）を予測する** 2 **つの予測量をそれぞれ，** $x_{1,n}(h)$ **と** $x_{2,n}(h)$ **とおく．このとき，漸近相対効率（**Asymptotic Relative Efficiency, ARE**），** $\mathrm{ARE}[x_{1,n}(h),\ x_{2,n}(h)]$ **は次の式で定義される．**

$$\mathrm{ARE}[x_{1,n}(h),\ x_{2,n}(h)]\equiv\lim_{n\to\infty}n(\mathrm{E}[x_{n+h}-x_{1,n}(h)]^2-\mathrm{E}[x_{n+h}-x_{2,n}(h)]^2).$$

ARE **は** 2 **つの予測量の** PMSE **を比較するので，その符号に興味がある．そこで，次の用語を定義する．**

定義 2. **定義** 1 **の用語と記号を用いる．** 2 **つの予測量** $x_{1,n}(h)$ **と** $x_{2,n}(h)$ **を構成する時系列モデルをそれぞれモデル** 1 **とモデル** 2 **とする．このとき** 2 **つの予測量の良さを次で言い表す．** $\mathrm{ARE}[x_{1,n}(h),\ x_{2,n}(h)]>0$ **ならば，** ARE **の意味で** $x_{2,n}(h)$ **（モデル** 2**）の方がよい．** $\mathrm{ARE}[x_{1,n}(h),\ x_{2,n}(h)]<0$ **ならば，** ARE **の意味で** $x_{1,n}(h)$ **（モデル** 1**）の方がよい．** $\mathrm{ARE}[x_{1,n}(h),\ x_{2,n}(h)]=0$ **ならば，** ARE **の意味で** $x_{1,n}(h)$ **と** $x_{2,n}(h)$ **（モデル** 1 **とモデル** 2**）は同程度である．**

次の結果は定理 2.1 の系である．

系 2.1.　1）1 期先予測における PMSE は $n\to\infty$ で

$$\mathrm{E}[y_{n+1}-y_n(1,\widehat{\boldsymbol{\theta}})]^2=\sigma_0^2\Big(1+\frac{q}{n}\Big)+o\Big(\frac{1}{n}\Big), \tag{2.9a}$$

$$\mathrm{E}[y_{n+1}-y_n(1,\overline{\boldsymbol{\theta}})]^2=\sigma_0^2+\frac{\sigma_0^2}{n}\big(p_1+\boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2+2\boldsymbol{c}_2'\boldsymbol{E}_{23}\boldsymbol{c}_3+\boldsymbol{c}_3'\boldsymbol{D}_{33}\boldsymbol{c}_3\big)+o\Big(\frac{1}{n}\Big), \tag{2.9b}$$

$$\mathrm{E}[y_{n+1}-y_n(1,\widetilde{\boldsymbol{\theta}})]^2=\sigma_0^2+\frac{\sigma_0^2}{n}\big(p_1+p_2+\boldsymbol{c}_3'\boldsymbol{D}_{33}\boldsymbol{c}_3+\boldsymbol{c}_3'\boldsymbol{E}_{23}'\boldsymbol{D}_{22}^{-1}\boldsymbol{E}_{23}\boldsymbol{c}_3\big)+o\Big(\frac{1}{n}\Big). \tag{2.9c}$$

2）$\boldsymbol{c}_3=\boldsymbol{\delta}_3^0=\boldsymbol{0}$ のとき，$n\to\infty$ で

$$\mathrm{ARE}[y_n(1,\overline{\boldsymbol{\theta}}),\ y_n(1,\widetilde{\boldsymbol{\theta}})]\overset{\mathrm{a}}{=}\sigma_0^2(\boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2-p_2), \tag{2.10a}$$

$$\mathrm{ARE}[y_n(h,\overline{\boldsymbol{\theta}}),\ y_n(h,\widetilde{\boldsymbol{\theta}})]\overset{\mathrm{a}}{=}\sigma_0^2\sum_{j=0}^{\infty}\boldsymbol{\kappa}_j(h,\boldsymbol{\theta}^0)'(\boldsymbol{c}_2\boldsymbol{c}_2'-\boldsymbol{D}_{22}^{-1})\boldsymbol{\kappa}_j(h,\boldsymbol{\theta}^0). \tag{2.10b}$$

3）$\boldsymbol{c}_3=\boldsymbol{\delta}_3^0=\boldsymbol{0}$ かつ $p_2=1$ のとき，$n\to\infty$ で

$$\mathrm{ARE}[y_n(h,\overline{\boldsymbol{\theta}}),\ y_n(h,\widetilde{\boldsymbol{\theta}})]\overset{\mathrm{a}}{=}\sigma_0^2(\boldsymbol{c}_2^2-\boldsymbol{D}_{22}^{-1})\sum_{j=0}^{\infty}\boldsymbol{\kappa}_j(h,\boldsymbol{\theta}^0)^2. \tag{2.11}$$

ここで $\boldsymbol{D}_{22}=\boldsymbol{I}_{22}-\boldsymbol{I}_{12}'\boldsymbol{I}_{11}^{-1}\boldsymbol{I}_{12}$，$\boldsymbol{D}_{33}=\boldsymbol{I}_{33}-\boldsymbol{I}_{13}'\boldsymbol{I}_{11}^{-1}\boldsymbol{I}_{13}$，$\boldsymbol{E}_{23}=\boldsymbol{I}_{23}-\boldsymbol{I}_{12}'\boldsymbol{I}_{11}^{-1}\boldsymbol{I}_{13}$，$\boldsymbol{\kappa}_j(h,\boldsymbol{\theta}^0)$ は p_2 次ベクトルで $\boldsymbol{\kappa}_j(h,\boldsymbol{\theta}^0)=(\boldsymbol{I}_{12}'\boldsymbol{I}_{11}^{-1}\quad -\mathbb{I}\quad \boldsymbol{0})\boldsymbol{\varphi}_j(h,\boldsymbol{\theta}^0)$，そして $\mathbb{I}$ は $p_2\times p_2$ の単位行列である．

仮定より真の母数は，$\theta^0\in\Theta\subset\mathbb{R}^q$ であるから，$\widehat{\theta}$ は非零の真の母数を全て推定しているが，$\overline{\theta}$ と $\widetilde{\theta}$ は必ずしもそうではない．しかしながら，系の ARE の結果は，$\widetilde{\theta}$ が非零の真の母数を全て推定しているときにあたる．また D_{22}^{-1} は，$c_3=\boldsymbol{0}$ のとき $\widetilde{\delta}_2$ の漸近分散にあたる．よって，$c_3=\boldsymbol{0}$ のときの $\mathrm{ARE}[y_n(h,\overline{\boldsymbol{\theta}}),\ y_n(h,\widetilde{\boldsymbol{\theta}})]$ の符号は，漸近的に真の母数 δ_2^0 を構成する $\boldsymbol{c}_2$ とその推定量 $\widetilde{\delta}_2$ の漸近分散との関係式で決まる．ただし，$c_3\neq\boldsymbol{0}$ のときの $\mathrm{ARE}[y_n(h,\overline{\boldsymbol{\theta}}),\ y_n(h,\widetilde{\boldsymbol{\theta}})]$ の符号は，(2.9b)と(2.9c)より分かるように関係式を導くのは一般に難しい．

3.　古典的検定統計量の漸近的性質

3.1　モデル候補が 2 つのときの検定統計量の漸近結果

この節では，2 節のモデルと記号を引き継いで，検定問題，

$$H_0\colon\ \boldsymbol{\delta}_2^0=\boldsymbol{c}_2=\boldsymbol{0},\ \boldsymbol{\delta}_3^0=\boldsymbol{c}_3=\boldsymbol{0}\qquad \text{vs}\qquad H_1\colon\ \boldsymbol{\delta}_2^0=\boldsymbol{c}_2/\sqrt{n}\neq\boldsymbol{0},\ \boldsymbol{\delta}_3^0=\boldsymbol{c}_3=\boldsymbol{0} \tag{3.1}$$

における古典的な検定統計量(Wald, LR, Score)の漸近分布を導く．ただし，通常の検定問題の議論と異なり，真のモデルは必ずしも $c_3=\boldsymbol{0}$ とは限らない．つまり対立仮説のモデルも誤っているケース($c_3\neq\boldsymbol{0}$)も考える．

Wald, LR, そして Score 検定統計量をそれぞれ，ξ_n^W，ξ_n^{LR}，ξ_n^S として次のようにおく．

$$\begin{aligned}
\xi_n^W&=n\widetilde{\boldsymbol{\delta}}_2'\boldsymbol{D}_{22}(\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{\delta}}_2,\\
\xi_n^{LR}&=-2\Big\{l(\overline{\boldsymbol{\theta}},\overline{\sigma}^2)-l(\widetilde{\boldsymbol{\theta}},\widetilde{\sigma}^2)\Big\},\\
\xi_n^S&=\frac{1}{n}\frac{\partial l(\overline{\boldsymbol{\theta}},\overline{\sigma}^2)}{\partial\boldsymbol{\delta}_2'}\boldsymbol{D}_{22}(\overline{\boldsymbol{\theta}})^{-1}\frac{\partial l(\overline{\boldsymbol{\theta}},\overline{\sigma}^2)}{\partial\boldsymbol{\delta}_2}.
\end{aligned} \tag{3.2}$$

ここで $D_{22}(\widetilde{\theta})$ と $D_{22}(\overline{\theta})$ は，それぞれ系 2.1 で定義された D_{22} の θ^0 に推定量 $\widetilde{\theta}$ と $\overline{\theta}$ を代入したものである．

このとき，次の定理が成立する．

定理 3.1. 統計量 ξ_n^W，ξ_n^{LR}，ξ_n^S について，$n\to\infty$ で次が成立する．

$$
\begin{aligned}
&\xi_n^W \overset{\mathrm{a}}{=} \xi_n^{LR} \overset{\mathrm{a}}{=} \xi_n^S,\\
&\xi_n^W \overset{\mathrm{a}}{\sim} \chi^2(p_2,\lambda), \quad \xi_n^{LR} \overset{\mathrm{a}}{\sim} \chi^2(p_2,\lambda), \quad \xi_n^S \overset{\mathrm{a}}{\sim} \chi^2(p_2,\lambda),\\
&\quad\lambda = (\boldsymbol{c}_2' + \boldsymbol{c}_3'\boldsymbol{E}_{23}'\boldsymbol{D}_{22}^{-1})\boldsymbol{D}_{22}(\boldsymbol{c}_2 + \boldsymbol{D}_{22}^{-1}\boldsymbol{E}_{23}\boldsymbol{c}_3)\\
&\qquad = \boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2 + 2\boldsymbol{c}_2'\boldsymbol{E}_{23}\boldsymbol{c}_3 + \boldsymbol{c}_3'\boldsymbol{E}_{23}'\boldsymbol{D}_{22}^{-1}\boldsymbol{E}_{23}\boldsymbol{c}_3.
\end{aligned}
\tag{3.3}
$$

ここで $\chi^2(\nu,\lambda)$ は，自由度 ν，非心度 λ の非心 χ^2 分布で 5.1 節で定義される．

対立仮説のモデルが真のモデルを含んでいない($\boldsymbol{c}_3\neq\boldsymbol{0}$)とき，より複雑なモデル H_1 が採択される，つまり H_0 は棄却されるのが望ましいだろう．しかし定理 3.1 の結果は $\boldsymbol{c}_3$ に関係する項が影響して，$\boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2$ が増加しても漸近的な検出力があがらないケースがあることを示している．一方，対立仮説のモデルが真のモデルを含んでいる($\boldsymbol{c}_3=\boldsymbol{0}$)とき，$\boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2$ が増加するにつれ検出力が上がる．

3.2 モデル候補が 3 つ以上あるときの検定統計量の漸近結果

ここでは，3.1 節の結果を拡張して 3 つ以上のモデル候補があるとき，古典的検定統計量の漸近的な同時確率分布を示す．そのために，いくつか記号と結果を追加する．

まず，真の母数 θ^0 を $\theta^0=(\theta_1^0,\theta_2^0,\ldots,\theta_p^0,(\boldsymbol{\delta}_3^0)')'=((\boldsymbol{\delta}_1^0)',(\boldsymbol{\delta}_2^0)',(\boldsymbol{\delta}_3^0)')'$ とおく．ここで，$p=p_1+p_2$，$\boldsymbol{\delta}_3^0=\boldsymbol{c}_3=\boldsymbol{0}$ である．また，$q-i$ 個の 0 制約を置いた $\boldsymbol{\theta}^0$ の MLE を $\widehat{\boldsymbol{\theta}}(i)=(\widehat{\theta}_{1i},\widehat{\theta}_{2i},\ldots,\widehat{\theta}_{ii},0,\ldots,0)'$ ($i=1,2,\ldots,q$)，そしてそのときの σ_0^2 の MLE を $\widehat{\sigma}^2(i)$ とおく．2 節の記号の定義から，$(\overline{\boldsymbol{\theta}},\widetilde{\boldsymbol{\theta}},\widehat{\boldsymbol{\theta}})=(\widehat{\boldsymbol{\theta}}(p_1),\widehat{\boldsymbol{\theta}}(p),\widehat{\boldsymbol{\theta}}(q))$ である．

一般に p は未知であるから，検定問題(3.1)を拡張して，検定問題，

$$
H_0^{(i)}\colon p=i \quad \text{vs} \quad H_1^{(j)}\colon p=j, \qquad (1\le i<j\le q)
\tag{3.4}
$$

の古典的検定統計量(Wald, LR，または Score 統計量のいずれか任意)を $\xi_n(i,j)$ とし，Wald, LR, Score 統計量をそれぞれ $\xi_n^W(i,j)$，$\xi_n^{LR}(i,j)$，$\xi_n^S(i,j)$ とする．検定問題(3.1)における古典的検定統計量(3.2)との対応は，それぞれ $(\xi_n^W,\xi_n^{LR},\xi_n^S)=(\xi_n^W(p_1,p)$，$\xi_n^{LR}(p_1,p)$，$\xi_n^S(p_1,p))$ である．以降，$p=i$ とする(真とは限らない)仮想の時系列モデルをモデル i と呼ぶことにする．

以下は，定理 3.1 の $\boldsymbol{c}_3=\boldsymbol{0}$ の場合を拡張しており，$\widetilde{\theta}$ が非零の真の母数を全て推定している場合の結果である．

定理 3.2. 上記の小節 3.2 の仮定の下，$n\to\infty$ で次が成立する．$j=p+1,p+2,\ldots,q$ のとき，

$$
\xi_n(p_1,j) \overset{\mathrm{a}}{=} \xi_n(p_1,p)+\xi_n(p,j) \overset{\mathrm{a}}{\sim} \chi^2(j-p_1,\boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2).
\tag{3.5}
$$

ここで，$\xi_n(p_1,p)$ と $\xi_n(p,j)$ は漸近的に独立で，

$$
\xi_n(p_1,p) \overset{\mathrm{a}}{\sim} \chi^2(p_2,\boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2), \qquad \xi_n(p,j) \overset{\mathrm{a}}{\sim} \chi^2(j-p,0).
\tag{3.6}
$$

定理 3.2 は，$p=p_1+p_2$ を真の母数の次数として扱っているが，階層的モデルで考えている

ので，$p_1+p_2\le p\le q$ でも同じように成立する．定理 3.1 より，$k=1,2,\dots,p_3$ において，

$$\xi_n(p_1,p+k)\overset{\mathrm{a}}{=}\xi_n^{LR}(p_1,p+k)=\xi_n^{LR}(p_1,p)+\sum_{i=1}^{k}\xi_n^{LR}(p+i-1,p+i).$$

さらに定理 3.1 と定理 3.2 を繰り返し用いて次が成立する．証明は略す．

系 3.1. 定理 3.2 の仮定の下，$n\to\infty$ で次が成立する．$k=1,2,\dots,p_3$ において，

$$\xi_n(p_1,p+k)\overset{\mathrm{a}}{=}\xi_n(p_1,p)+\sum_{i=1}^{k}\xi_n(p+i-1,p+i)\overset{\mathrm{a}}{\sim}\chi^2(p_2,\boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2)+\sum_{i=1}^{k}Z_i^2, \tag{3.7}$$

そして $\xi_n(p+i-1,p+i)\overset{\mathrm{a}}{\sim}Z_i^2\ \ (i=1,2,\dots,k)$．ここで $\{Z_i\}\sim iidN(0,1)$ で，各 Z_i と $\chi^2(p_2,\boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2)$ は互いに独立である．

定理 3.1，定理 3.2 そして系 3.1 は，条件 $\boldsymbol{c}_2\neq\boldsymbol{0}$ とは無関係に成立するため，$\boldsymbol{c}_2=\boldsymbol{0}$，つまり真の母数の非零の次数は高々$p_1$ としてもよい．次の系は，$\bar{\theta}$ が非零の真の母数を全て推定しているケースである．

系 3.2. 定理 3.2 の条件を仮定し，さらに $p_2=1$ かつ $\boldsymbol{c}_2=\boldsymbol{0}$ とするとき，$n\to\infty$ で次が成立する．$k=0,1,2,\dots,p_3$ において，

$$\xi_n(p_1,p_1+k+1)\overset{\mathrm{a}}{=}\sum_{i=0}^{k}\xi_n(p_1+i,p_1+i+1)\overset{\mathrm{a}}{\sim}\sum_{i=0}^{k}Z_i^2, \tag{3.8}$$

そして $\xi_n(p_1+i,p_1+i+1)\overset{\mathrm{a}}{\sim}Z_i^2\ \ (i=0,1,\dots,k)$．ここで $\{Z_i\}\sim iidN(0,1)$ である．

証明は，定理 3.1 と系 3.1 に条件 $p_2=1$，$\boldsymbol{c}_2=\boldsymbol{0}$，$\boldsymbol{c}_3=\boldsymbol{0}$ を代入すればよい．

系 3.2 は $\boldsymbol{c}_2=\boldsymbol{0}$ を仮定しているが，$\boldsymbol{c}_2\neq\boldsymbol{0}$ のときの同時分布はどうなるだろうか．次の定理はこの疑問に答えるもので定理 3.2 の拡張である．

定理 3.3. 定理 3.2 の仮定の下，$n\to\infty$ で次が成立する．

$$\begin{aligned}&(\xi_n(p_1,p_1+1),\xi_n(p_1,p_1+2),\dots,\xi_n(p_1,q))'\\&\quad\overset{\mathrm{a}}{=}\boldsymbol{U}\big(\xi_n(p_1,p_1+1),\xi_n(p_1+1,p_1+2),\dots,\xi_n(q-1,q)\big)'\\&\quad\overset{\mathrm{a}}{\sim}\boldsymbol{U}\big(V(p_1+1),V(p_1+2),\dots,V(q)\big)'.\end{aligned} \tag{3.9}$$

ここで $\boldsymbol{U}$ は，$i\ge j$ なる (i,j) 成分が全て 1，$i<j$ なる (i,j) 成分が全て 0 の $(q-p_1)\times(q-p_1)$ 下三角行列，$V(i)=\{Z_i+\mu(i)\}^2\ \ (i=p_1+1,p_1+2,\dots,q)$，そして $\{Z_i\}\sim iidN(0,1)$ である．また $\mu(i)$ について，$i=p_1+1,p_1+2,\dots,p$ のとき $\mu(i)$ は $\boldsymbol{c}_2$ と θ^0 に依存する定数，$i=p+1,p+2,\dots,q$ のとき $\mu(i)=0$，そして次の関係式が成立する．

$$\sum_{i=p_1+1}^{p}\mu(i)^2=\sum_{i=p_1+1}^{q}\mu(i)^2=\boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2. \tag{3.10}$$

4. 一般化尤度比検定の枠組みを用いたモデル選択

この節では，2 節と 3 節の結果を利用して漸近 PMSE を基準としたモデル選択を導入する．

モデルとその仮定ならびに記号は，2 節と 3 節を引き継ぐ．

4.1　予測問題と検定問題を結合させるアイデア

2 節と 3 節の漸近的に成立する結果をまとめる．

まず，2 節の PMSE の結果をまとめる．系 2.1 より $\boldsymbol{c}_3=\boldsymbol{\delta}_3^0=\boldsymbol{0}$ ならば，ARE の意味で $y_n(1,\overline{\boldsymbol{\theta}})$ より $y_n(1,\widetilde{\boldsymbol{\theta}})$ の方が同程度またはよくなる必要十分条件は，$\boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2\le p_2$ であり，さらに $p_2=1$ ならば，ARE の意味で $y_n(h,\overline{\boldsymbol{\theta}})$ より $y_n(h,\widetilde{\boldsymbol{\theta}})$ の方が同程度またはよくなる必要十分条件は，$c_2^2\boldsymbol{D}_{22}\le 1$ といえる．$y_n(1,\overline{\boldsymbol{\theta}})$ と $y_n(h,\overline{\boldsymbol{\theta}})$ の方が同程度またはよくなる必要十分条件は，それぞれ逆の不等号になる．これより，少なくとも 1 期先予測においては，2 次形式 $\boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2$ と母数の数 p_2 との大小が ARE の意味で予測量の良さを決めると言えよう．

次に，3 節の検定統計量の結果をまとめる．定理 3.1 より，$\boldsymbol{c}_3=\boldsymbol{\delta}_3^0=\boldsymbol{0}$ ならば，$\boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2$ が大きくなるにつれ，簡単なモデル H_0: $\boldsymbol{\delta}_2^0=\boldsymbol{c}_2=\boldsymbol{0}$ を棄却し，複雑なモデル H_1: $\boldsymbol{\delta}_2^0=\boldsymbol{c}_2/\sqrt{n}\neq\boldsymbol{0}$ を採択する確率は増加する．また，$\boldsymbol{\delta}_2^0=\boldsymbol{c}_2/\sqrt{n}$ なので，

$$\boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2=n(\boldsymbol{\delta}_2^0)'\boldsymbol{D}_{22}\boldsymbol{\delta}_2^0 \tag{4.1}$$

と書ける．対して Wald 統計量 $\xi_n^W=n\widetilde{\boldsymbol{\delta}}_2'\boldsymbol{D}_{22}(\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{\delta}}_2$ に着目すると，$\widetilde{\boldsymbol{\delta}}_2$ と $\boldsymbol{D}_{22}(\widetilde{\boldsymbol{\theta}})$ は，それぞれ $\boldsymbol{\delta}_2^0$ と $\boldsymbol{D}_{22}$ の強一致推定量であることから，Wald 統計量 ξ_n^W は $\boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2$ を推定しているといえる．そして，ξ_n^W と他の古典的検定統計量 ξ_n^{LR} と ξ_n^S との漸近分布は一致する．

これから，$\boldsymbol{c}_3=\boldsymbol{\delta}_3^0=\boldsymbol{0}$ のとき，検定問題，

$$H_0^{(P)}\colon \boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2\le p_2 \quad \text{vs} \quad H_1^{(P)}\colon \boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2> p_2 \tag{4.2}$$

が予測の意味でのモデル選択問題となる．そして，Wald 統計量等の古典的統計量がそれを判断する統計量の候補となる．

4.2　簡単なケース

ここでは，この小節に限り，$p_2=1$，$\boldsymbol{\delta}_3^0=\boldsymbol{0}$，$c_2\ge 0$ のケースに議論を簡単化して，2 つの推定量による予測量の候補 $y_n(h,\overline{\boldsymbol{\theta}})$ と $y_n(h,\widetilde{\boldsymbol{\theta}})$ から予測量を選択する方法について考える．片側検定の Wald 統計量を $W_n=\{n\boldsymbol{D}_{22}(\widetilde{\boldsymbol{\theta}})\}^{1/2}\widetilde{\delta}_2$ とすると，定理 3.1 の証明（A.8）より，$W_n\overset{\mathrm{a}}{\sim}N(\sqrt{\boldsymbol{D}_{22}}\boldsymbol{c}_2,\,1)$ である．一方，系 2.1 の（2.11）より，ARE の意味で $y_n(h,\overline{\boldsymbol{\theta}})$ より $y_n(h,\widetilde{\boldsymbol{\theta}})$ の方が同程度またはよくなる必要十分条件は，$c_2\sqrt{\boldsymbol{D}_{22}}\le 1$ である．

これより，$p_2=1$ を基準としたモデル選択：$W_n\le 1$ ならば，$y_n(h,\overline{\boldsymbol{\theta}})$ を選択し，$W_n>1$ ならば，$y_n(h,\widetilde{\boldsymbol{\theta}})$ を選択する，というモデル選択法について次が成立する．証明は略す．

系 4.1.　定理 3.1 の仮定と上記の小節 4.2 の仮定の下，$n\to\infty$ で次が成立する．

$$\Pr(W_n\le 1\,|\,\boldsymbol{c}_2\sqrt{\boldsymbol{D}_{22}}\le 1)\overset{\mathrm{a}}{=}\Pr(Z\le 1-\boldsymbol{c}_2\sqrt{\boldsymbol{D}_{22}}\,|\,\boldsymbol{c}_2\sqrt{\boldsymbol{D}_{22}}\le 1)\ge 1/2. \tag{4.3}$$

$$\Pr(W_n> 1\,|\,\boldsymbol{c}_2\sqrt{\boldsymbol{D}_{22}}> 1)\overset{\mathrm{a}}{=}\Pr(Z> 1-\boldsymbol{c}_2\sqrt{\boldsymbol{D}_{22}}\,|\,\boldsymbol{c}_2\sqrt{\boldsymbol{D}_{22}}> 1)\ge 1/2. \tag{4.4}$$

ここで，$Z\sim N(0,1)$ である．

この結果は，漸近的に W_n と $p_2=1$ によるモデル選択が，佐和（1979）7.1.7 節のいう不偏な決定方式になることを示している．なお，$\Pr(Z>1)$ はおおよそ 0.16 であるから，このモデル選択法は近似的に有意水準 16%程度の正規検定とも言える．また，結果は検定の右片側検定にあたるが，左片側検定（$c_2<0$）に対応する W_n と -1 との大小によって決めるモデル選択法も同様の結果を導く．

それでもなお，(4.2)の検定に戻ると，2 つの疑問点が残る．1 つは，(4.2)の検定に古典的統計量を使うとして，どのように棄却域を構成するかである．一般には，$p_2 \geq 1$ であり，両側検定で考えることが必要とされるため，単純に p_2 との大小でモデル選択を行おうとしても，漸近的に，佐和(1979)の不偏な決定方式は導けない．もう 1 つは真の母数の次数が未知のため，3 つ以上のモデルの候補の中でどのように(4.2)の検定を実行するかである．

4.3 一般化尤度比検定を応用したモデル選択基準

ここでは，前述の疑問点に対し，3.2 節の条件を引き継ぎ，Hosoya (1984, 1986, 1989)による GLR 検定を応用したモデル選択法を提案する．

定理 3.3 の $\mu(i)$ による，$\boldsymbol{\mu} = (\mu(p_1+1), \mu(p_1+2), \ldots, \mu(q))' \in \mathbb{R}^{q-p_1}$ の集合を $\mathcal{M}_p = \{\boldsymbol{\mu} |\ \sum_{i=p_1+1}^{j} \mu(i)^2 \leq j - p_1,\ j = p, \ldots, q\}$，$\mathcal{A}_p = \{\boldsymbol{\mu} | \mu(i) = 0,\ i = p+1, \ldots, q\}$ とおくと，(4.2)の検定は，「$H_0^{(P)}$: $\mathcal{M}_p \bigcap \mathcal{A}_p$ vs $H_1^{(P)}$: $\mathcal{M}_p^c \bigcap \mathcal{A}_p$」と書ける．ただし，$p$ は未知であるから，検定を行うためにはなるべく p に依存しない形にする必要がある．

まず，$\mathcal{M}_p$ の代わりとして，$\mathcal{M}_{G1} = \{\boldsymbol{\mu} |\ \sum_{i=p_1+1}^{j} \mu(i)^2 \leq j - p_1,\ j = p_1+1, \ldots, q\} = \{\boldsymbol{\mu} | \mu(i)^2 \leq 1,\ i = p_1+1, \ldots, q\}$ をおき，次の検定を提案する．

$$H_0^{(G1)}:\ \mathcal{M}_{G1} \bigcap \mathcal{A}_p \quad \text{vs} \quad H_1^{(G1)}:\ \mathcal{M}_{G1}^c \bigcap \mathcal{A}_p. \tag{4.5}$$

これは，(3.4)の検定の記号を使うと，「$H_0^{(p_1)}$: $p = p_1$ vs $H_1^{(j)}$: $p = j$ $(j = p_1+1, p_1+2, \ldots, q)$」に対応している．これは GLR 検定と呼ばれるが，従来の GLR 検定と異なり，0 制約を置いたモデルが真のモデルかどうか検討するのではなく，ARE の意味でよいモデルかどうか検討している．検定統計量は，$\xi_n(p_1, j)$ を用いて，

$$\xi_n(p_1, j) > d(p_1, j) \quad (\text{for some } j = p_1+1, p_1+2, \ldots, q) \tag{4.6}$$

であれば，$H_0^{(G1)}$ を棄却，つまりモデル p_1 は対立仮説のモデル $p_1+1, p_1+2, \ldots, q$ と比較して，ARE の意味でよいモデルである保証が得られていないと判断する．また，そうでないときは，モデル p_1 をモデル $p_1+1, \ldots, q$ と比較して，ARE の意味でよいモデルとして採択する．

$H_0^{(G1)}$ を棄却するとき，モデル p_1 よりモデル $p_1+1, \ldots, q$ のいずれかの方が ARE の意味でよいモデルかどうかの解釈は曖昧である．なぜなら，$p > p_1+1$ では，定理 3.3 の(3.10)より $H_0^{(P)} \supset H_0^{(G1)}$，$H_1^{(P)} \subset H_1^{(G1)}$ だからである．よって $H_0^{(G1)}$ 下でモデル p_1 を棄却するとき，$H_0^{(P)}$ と $H_1^{(P)}$ の真偽の解釈は不明であるが，複雑なモデル p_1+1 について再び検定を行うこととなる．これは，$\mathcal{M}_{G1}$ が p に依存しないことに起因する．

この包含関係による不利益を軽減する検定法として次の方法が現実的かもしれない．$p_1 < k < q$ で，

$$H_0^{(Gk)}:\ \mathcal{M}_{Gk} \bigcap \mathcal{A}_p \quad \text{vs} \quad H_1^{(Gk)}:\ \mathcal{M}_{Gk}^c \bigcap \mathcal{A}_p. \tag{4.7}$$

ここで，$\mathcal{M}_{Gk} = \{\boldsymbol{\mu} |\ \sum_{i=p_1+1}^{j} \mu(i)^2 \leq j - p_1,\ j = k, k+1, \ldots, q\} = \{\boldsymbol{\mu} |\ \sum_{i=p_1+1}^{k} \mu(i)^2 \leq k - p_1$ かつ $\mu(i)^2 \leq 1,\ i = k+1, \ldots, q\}$ である．この検定は，(3.4)の記号を使うと，ARE の観点で比較する検定「$H_0^{(p_1)}$: $p = p_1$ vs $H_1^{(j)}$: $p = j$ $(j = k, k+1, \ldots, q)$」に対応しており，$k = p_1+1$ のとき，検定(4.5)に一致する．そして検定統計量は，$\xi_n(p_1, j)$ を用いて，

$$\xi_n(p_1, j) > d(p_1, j) \quad (\text{for some } j = k, k+1, \ldots, q) \tag{4.8}$$

であれば，モデル p_1 は，対立仮説のモデル $k, k+1, \ldots, q$ と比較して，ARE の意味でよいモデルである保証が得られていないものとして $H_0^{(Gk)}$ を棄却する．

検定(4.7)の手続きは，適当な検定の有意水準と棄却域 $d(p_1, j)$ を所与として，下降して $\xi_n(p_1, q), \xi_n(p_1, q-1), \ldots, \xi_n(p_1, k)$ の順に(4.8)を評価して，$\xi_n(p_1, j) > d(p_1, j)$ となる $j = k$ で

検定を止めて，$j<k$ の検定を行わず，モデル p_1 を ARE の観点から棄却しようというものである．棄却されると，$p_1=p_1+1$ として(4.7)の検定を再び行い，棄却されるたびこれを繰り返す．そして棄却されない最小の p_1 で検定をやめる．つまり全ての $j=p_1+1,p_1+2,\dots,q$ で $\xi_n(p_1,j)\le d(p_1,j)$ となるとき $H_0^{(G1)}$ を採択，換言すればモデル $p_1+1,\dots,q$ と比較して，モデル p_1 は ARE の意味でよいモデルとして p_1 を採択する．

この方法であれば，$H_0^{(Gk)}$ 下で帰無仮説のモデル p_1 が採択されるとき，検定問題(4.5)と(4.7)は同じで $H_0^{(P)}\supset H_0^{(G1)}=H_0^{(Gk)}$ であるから，ARE の意味でも帰無仮説のモデル p_1 の選択は正しい．一方，$H_0^{(Gk)}$ 下で帰無仮説のモデル p_1 が棄却されるとする．$p\le k$ であれば，$H_0^{(P)}\subset H_0^{(Gk)}$，$H_1^{(P)}\supset H_1^{(Gk)}$ で，$H_0^{(Gk)}$ は，モデル p_1 と真のモデル $k,k+1,\dots,q$ とを比較して ARE の意味でモデル p_1 の方がよいことを示している．即ち，帰無仮説のモデル p_1 は，対立仮説のモデル $k,k+1,\dots,q$ と比較して ARE の意味でよいモデルであるが棄却されることになり，有意水準の意味合いが明確となる．ただし $p>k$ で棄却されるとき，(4.5)のときと同じ理由で ARE の意味でよいモデルかどうかの解釈は曖昧である．

この検定の有意水準は，未知の p に依存させないために，帰無仮説の集合 $\mathcal{M}_{Gk}$ を条件付きとする確率よりその上界が求められる．具体的には，次の定理である．

定理 4.1. 検定(4.7)において，$n\to\infty$ で次が成立する．定数 $d_k<d_{k+1}<\cdots<d_q$ と $p_1<k<q$ において，

$$\Pr\left[\xi_n(p_1,j)>d_j\,\middle|\,H_0^{(Gk)}\right]\overset{\mathrm{a}}{=}1-\Pr\left[\sum_{i=p_1+1}^{j}V(i)\le d_j\,\middle|\,H_0^{(Gk)}\right]\le 1-F(d_j;j-p_1,j-p_1)\qquad(j=k,k+1,\dots,q),\tag{4.9}$$

$$\Pr\left[\bigcup_{j=k}^{q}\{\xi_n(p_1,j)>d_j\}\,\middle|\,H_0^{(Gk)}\right]\overset{\mathrm{a}}{=}1-\Pr\left[\bigcap_{j=k}^{q}\left\{\sum_{i=p_1+1}^{j}V(i)\le d_j\right\}\middle|\,H_0^{(Gk)}\right]\le 1-C_k^q(d_k,d_{k+1},\dots,d_q).\tag{4.10}$$

ここで，$V(i)$ は定理 3.3 により定義され，

$$\begin{aligned}C_k^q(d_k,d_{k+1},\dots,d_q)&=\Pr\left[\bigcap_{j=k}^{q}\left\{\sum_{i=p_1+1}^{j}(Z_i+1)^2\le d_j\right\}\right]\\&=\int_0^{d_k}p(x_k;k-p_1,k-p_1)\int_0^{d_{k+1}-x_k}p(x_{k+1};1,1)\\&\quad\times\cdots\times\int_0^{d_q-x_{q-1}-\cdots-x_k}p(x_q;1,1)dx_qdx_{q-1}\cdots dx_{k+1}dx_k,\end{aligned}\tag{4.11}$$

$\{Z_i\}\sim iidN(0,1)$，$F(x;\nu,\lambda)$ と $p(x;\nu,\lambda)$ は，それぞれ自由度 ν，非心度 λ の非心 χ^2 分布の分布関数と確率密度関数で(5.2)と(5.1)で定義される．

この定理は，モデル p_1 と，モデル $k,k+1,\dots,q$ とを ARE の観点で比較する GLR 検定(4.7)の有意水準を評価している．

$p\le k$ であれば，ARE の意味でモデル p_1 が，モデル j $(j=k,k+1,\dots,q)$ よりよい状況で，帰無仮説のモデル p_1 を棄却する確率が高々どの程度かを示していると解釈できる．(4.9)は，モデル j と比較して，ARE の意味でよいモデルでないものとしてモデル p_1 を棄却する確率

の上界が $1-F(d_j; j-p_1, j-p_1)$ となることを示している．(4.10)は，モデル p_1 をモデル j $(j=k, k+1, \ldots, q)$ のそれぞれと比較したとき，少なくとも 1 つのモデルが，モデル p_1 より ARE の意味でよいとして棄却される確率，つまり(4.8)の確率の上界が $1-C_k^q(d_k, d_{k+1}, \ldots, d_q)$ となることを示している．

一方，$p>k$ のときは，ARE による解釈は曖昧であるが，帰無仮説 $H_0^{(Gk)}$ が正しいにも関わらず，帰無仮説のモデル p_1 を棄却する確率の上界を示している．

定理の結果より，(4.8)における漸近的な棄却域と有意水準の決定に，次の 2 通りが考えられる．

1 つは，$j=k+1, \ldots, q$ で

$$F(d(p_1,k); k-p_1, k-p_1) = F(d(p_1,j); j-p_1, j-p_1) = 1-\beta \tag{4.12}$$

とし，各検定の有意水準 $\Pr[\xi_n(p_1,j) > d(p_1,j) | H_0^{(Gk)}]$ を β 以下と与えて，$d(p_1,j)$ を求め，

$$C_k^q(d(p_1,k), d(p_1,k+1), \ldots, d(p_1,q)) = 1-\alpha(k,q) \tag{4.13}$$

とし，検定全体の有意水準を $\alpha(k,q)$ 以下と求めるやり方である．

もう 1 つは，

$$C_k^q(d(p_1,k), d(p_1,k+1), \ldots, d(p_1,q)) = 1-\alpha \tag{4.14}$$

とし，検定全体の有意水準を α 以下と与え，$j=k+1, \ldots, q$ で

$$F(d(p_1,k); k-p_1, k-p_1) = F(d(p_1,j); j-p_1, j-p_1) = 1-\beta(p_1,j) \tag{4.15}$$

とし，各検定の有意水準が等しくなるよう $\beta(p_1,j)$ と $d(p_1,j)$ を求めるやり方である．

4.4 AIC の解釈

ここでは，よく知られたモデル選択基準である AIC を前節のモデル選択問題に適用してみる．母数を k 個推定したとき AIC は，

$$AIC(k) = -2l(\widehat{\boldsymbol{\theta}}(k), \widehat{\sigma}^2(k)) + 2k$$

と書ける．よってモデル p_1 とモデル p_1+p_0 ($p_0 = 1, 2, \ldots, q-p_1$)とを比較するときは，$AIC(p_1)$ と $AIC(p_1+p_0)$ の小さい方を選択する．つまり，

$$\begin{aligned} AIC(p_1) - AIC(p_1+p_0) &= -2\Big\{ l(\widehat{\boldsymbol{\theta}}(p_1), \widehat{\sigma}^2(p_1)) - l(\widehat{\boldsymbol{\theta}}(p_1+p_0), \widehat{\sigma}^2(p_1+p_0)) \Big\} - 2p_0 \\ &= \xi_n^{LR}(p_1, p_1+p_0) - 2p_0 \end{aligned} \tag{4.16}$$

が負(正)の値をとれば，p_1 (p_1+p_0)を採用する．即ち $\xi_n^{LR}(p_1, p_1+p_0)$ が $2p_0$ より小(大)であれば，p_1 (p_1+p_0)を採用する．最小の $AIC(k)$ でモデルを採用することは次と同等である：p_1 を適当に最小のモデルから開始し，ある p_0 で $\xi_n^{LR}(p_1, p_1+p_0) > 2p_0$ となれば，$p_1 = p_1+1$ として $\xi_n^{LR}(p_1, p_1+p_0)$ と $2p_0$ との大小を改めて評価する．そして大であれば再び $p_1 = p_1+1$ として繰り返す．そして全ての p_0 で $\xi_n^{LR}(p_1, p_1+p_0) \le 2p_0$ であれば，その p_1 を採用する．

AIC についてここまではすでに知られている事実であるが，以降検定問題(4.5)における(4.16)の $2p_0$ の解釈を与える．(4.9)より，適当な $d_0 > 0$ で

$$\begin{aligned} \Pr\Big[\xi_n^{LR}(p_1, p_1+p_0) > d_0 \Big| H_0^{(G1)}\Big] &\overset{\mathrm{a}}{=} \Pr\left[\sum_{i=p_1+1}^{p_1+p_0} V(i) > d_0 \middle| H_0^{(G1)}\right] \\ &\le \Pr\big[\chi^2(p_0, p_0) > d_0\big]. \end{aligned}$$

ここで $\chi^2(p_0,p_0)$ は自由度と非心度が p_0 の非心 χ^2 分布に従う確率変数である．さて，AIC では d_0 を $2p_0$ としていることと，$\mathrm{E}[\chi^2(p_0,p_0)]=2p_0$ であることを思い出せば，(4.12)–(4.15) は，$\chi^2(p_0,p_0)$ のパーセント点で d_0 を決めるのに対して，AIC では $\chi^2(p_0,p_0)$ の平均で d_0 を決めていると解釈できる．Sen (1989) より $\chi^2(p_0,p_0)$ のメディアンは平均 $2p_0$ より小さいため $\Pr[\chi^2(p_0,p_0)>2p_0]$ は 0.5 より小さく，$p_0=1,2,3,\ldots$ を調べると，中心極限定理が働いて，0.35, 0.39, 0.41, … と単調増加し 0.5 に近づいていく．つまり，ARE の意味で帰無仮説のモデルと対立仮説のモデルが同等であるとき，帰無仮説のモデルを選択する確率は，0.5 より大きいが p_0 が大きくなるにつれ 0.5 に近づき佐和(1979)の不偏な決定方式となる．さらに，c_2 と c_3 が 0 のときは，系 3.2 より $\Pr[\xi_n^{LR}(p_1,p_1+p_0)>2p_0]\stackrel{\mathrm{a}}{=}\Pr(\sum_{i=1}^{p_0}Z_i^2>2p_0)$ であるが，p_0 が大きくなるにつれ，右辺の確率は小さくなる，つまり，下平(2004)が述べているように簡単なモデルを選択しない誤りがより小さくなるよう調整されている(後述の表 3 参照)．自己回帰モデルにおいて AIC によるモデル選択は FPE によるモデル選択と漸近的に等しく，FPE は 1 期先の PMSE を不偏推定して導かれたモデル選択基準であることを思い出すと，この結果は面白い．

対して我々の提案するモデル選択法は，1 期先予測における ARE の意味で予測量の良さを決める相対的な関係式から導出している．(4.12)–(4.15)から分かるように，棄却域の計算は AIC と比べて複雑で比較も面倒であるが，有意水準をコントロールできるという良さを持つ．

ところで，AIC によるモデル選択は，(4.2)の $c_2'\boldsymbol{D}_{22}c_2$ を漸近的に不偏推定することからも導かれる．モデル p_1 とモデル j $(j=p,p+1,\ldots,q)$を ARE の観点で比較することを考える．このとき，定理 3.2 から $\xi_n(p_1,j)\stackrel{\mathrm{a}}{\sim}\chi^2(j-p_1,c_2'\boldsymbol{D}_{22}c_2)$ である．よって，$\xi_n(p_1,j)-(j-p_1)$ は，漸近分布の平均をとるという意味で $c_2'\boldsymbol{D}_{22}c_2$ の漸近的な不偏推定量となっている．これから，(4.2)のモデル選択の決定方式を $\xi_n(p_1,j)-(j-p_1)\le j-p_1$ であればモデル p_1 を選択，$\xi_n(p_1,j)-(j-p_1)\ge j-p_1$ であればモデル j を選択，とすることが考えられるが，これは(4.16)より AIC によるモデル選択であることが分かる．

4.5 棄却域と有意水準の数値例

以下の表 1–表 3 は，統計ソフトウェアS-PLUS を用いて，検定(4.5)で $q-p_1=1,2,3$ の場合の棄却域と有意水準を求めたものである．それぞれ，(4.12)と(4.13)より棄却域を求める方法をモデル選択法 1 とし表 1 にまとめ，(4.14)と(4.15)より棄却域を求める方法をモデル選択法 2 とし表 2 にまとめ，AIC によりモデル選択を行う方法を表 3 にまとめている．

表 1 は，5.3 節の方法で個々の検定の棄却域を求め，さらに検定全体の有意水準を 5.4 節の計算方法を利用して求めている．表 3 における検定の有意水準も 5.4 節の計算方法を利用して求めている．表 2 は，さらに Hosoya (1986)による Newton–Raphson 法も利用しており，初期値には Hosoya (1986)の Table 1 の結果を用いている．表に用いた計算の全ての収束条件は，高々 10^{-5} 未満とした．

表 1 は，$\beta=0.05,0.10,\ldots,0.50$ を所与として，5.3 節の方法で $F(d_i;i,i)=1-\beta$，$i=1,2,3$ を満たす棄却域 d_i を求めている．また，5.4 節の方法で，各 q での α (検定全体の有意水準)を $C_{p_1+1}^{p_1+2}(d_1,d_2)=\alpha$，$C_{p_1+1}^{p_1+3}(d_1,d_2,d_3)=\alpha$ より計算している．また，α_0 と β_{j0} は，それぞれ，d_j を所与として，

$$(4.17)\qquad \alpha_0=1-\Pr\left[\bigcap_{j=p_1+1}^{q}\left\{\sum_{i=p_1+1}^{j}Z_i^2<d_j\right\}\right],\qquad \beta_{j0}=\Pr\left\{\chi^2(j,0)>d_j\right\}\quad (j=1,2,3)$$

より計算される．これは，系 3.2 より導かれる $p=p_1$ のときの有意水準で，検定全体の有意水準 α_0 は，Hosoya (1986, Section 3)より計算している．これらを表示したのは，通常の GLR 検定での有意水準との対応を示すためと，overfitting のとき，つまり c_2 と c_3 が 0 であれば，系 2.1 より，ARE の意味で $y_n(h,\widetilde{\theta})$ よりも $y_n(h,\overline{\theta})$ の方がよく，棄却する確率が低く抑えられる

ほうが望ましいためである.

表2は, $\alpha=0.05, 0.10, \ldots, 0.50$ を所与として, (4.14)と(4.15)より, $C_{p_1+1}^{p_1+2}(d_1,d_2)=1-\alpha$, と $F(d_i,i,i)=1-\beta$ $(i=1,2)$を満たす d_1,d_2 と α を求めている. そして $q=p_1+3$ のときも同様に, $C_{p_1+1}^{p_1+3}(d_1,d_2,d_3)=1-\alpha$, と $F(d_i,i,i)=1-\beta$ $(i=1,2,3)$を満たす d_1,d_2,d_3 と α を求めている. また, α_0 と β_{j0} は, それぞれ求めた d_j を(4.17)に代入して計算している.

表1や表2では d_j を求めたのに対し, 表3の AIC によるモデル選択法では, $d_j=2j$ と

表1. モデル選択法1の棄却域と有意水準.

β	0.05	0.10	0.20	0.30	0.35	0.40	0.50
d_1	7.0021	5.2188	3.4205	2.3728	1.9800	1.6449	1.1036
d_2	10.8381	8.6754	6.4013	4.9947	4.4392	3.9464	3.0936
d_3	14.1887	11.7374	9.1080	7.4405	6.7694	6.1660	5.0985
β_{10}	0.0081	0.0223	0.0644	0.1235	0.1594	0.1997	0.2935
β_{20}	0.0044	0.0131	0.0407	0.0823	0.1087	0.1390	0.2129
β_{30}	0.0027	0.0083	0.0279	0.0591	0.0796	0.1038	0.1647
$q=p_1+2$							
α	0.0760	0.1468	0.2803	0.4052	0.4647	0.5223	0.6316
α_0	0.0105	0.0289	0.0828	0.1571	0.2015	0.2507	0.3625
$q=p_1+3$							
α	0.0938	0.1774	0.3289	0.4647	0.5273	0.5867	0.6957
α_0	0.0115	0.0318	0.0909	0.1720	0.2202	0.2731	0.3923

表2. モデル選択法2の棄却域と有意水準.

α	0.05	0.10	0.20	0.30	0.40	0.50
$q=p_1+2$						
β	0.0323	0.0667	0.1390	0.2153	0.2957	0.3804
d_1	8.1136	6.2640	4.3660	3.2288	2.4098	1.7700
d_2	12.1575	9.9510	7.6126	6.1500	5.0459	4.1328
α_0	0.0057	0.0159	0.0473	0.0929	0.1535	0.2309
β_{10}	0.0044	0.0123	0.0367	0.0724	0.1206	0.1834
β_{20}	0.0023	0.0069	0.0222	0.0462	0.0802	0.1266
$q=p_1+3$						
β	0.0255	0.0536	0.1142	0.1800	0.2511	0.3279
d_1	8.7124	6.8266	4.8763	3.6943	2.8315	2.1459
d_2	12.8615	10.6281	8.2514	6.7565	5.6214	4.6762
d_3	16.4525	13.9523	11.2518	9.5232	8.1885	7.0568
α_0	0.0044	0.0127	0.0387	0.0773	0.1297	0.1982
β_{10}	0.0032	0.0090	0.0272	0.0546	0.0924	0.1430
β_{20}	0.0016	0.0049	0.0162	0.0341	0.0602	0.0965
β_{30}	0.0009	0.0030	0.0104	0.0231	0.0423	0.0701

表 3. $AIC(k)$ によるモデル選択法の有意水準.

j	1	2	3
β_j	0.3472	0.3943	0.4146
β_{j0}	0.1573	0.1353	0.1116
α_j	0.3472	0.4841	0.5608
α_{j0}	0.1573	0.2126	0.2398

表 4. $AIC_d(k)$ によるモデル選択法の棄却域と有意水準.

α	0.05	0.10	0.20	0.30	0.40	0.50
$q=p_1+2$						
d	7.3263	5.7080	4.1012	3.1530	2.4667	1.9197
β_1	0.0440	0.0827	0.1539	0.2217	0.2892	0.3584
β_2	0.0139	0.0414	0.1159	0.2057	0.3052	0.4116
α_0	0.0071	0.0184	0.0502	0.0943	0.1520	0.2255
β_{10}	0.0068	0.0169	0.0429	0.0758	0.1163	0.1659
β_{20}	0.0007	0.0033	0.0166	0.0427	0.0849	0.1466
$q=p_1+3$						
d	7.3805	5.8224	4.3095	3.4284	2.7901	2.2760
β_1	0.0431	0.0792	0.1420	0.1994	0.2551	0.3116
β_2	0.0133	0.0383	0.1018	0.1747	0.2540	0.3393
β_3	0.0042	0.0187	0.0718	0.1477	0.2402	0.3454
α_0	0.0069	0.0173	0.0453	0.0827	0.1313	0.1936
β_{10}	0.0066	0.0158	0.0379	0.0641	0.0948	0.1314
β_{20}	0.0006	0.0030	0.0134	0.0324	0.0614	0.1027
β_{30}	0.0001	0.0006	0.0048	0.0163	0.0389	0.0776

決められている．これから，$q-p_1=j=1,2,3$ としたときの検定(4.5)において，各検定の有意水準を $\beta_j=\Pr\{\sum_{i=1}^{j}(Z_i+1)^2>2j\}$，検定全体の有意水準を $C_{p_1+1}^{p_1+2}(2,4)=\alpha_2$ $(j=2)$，$C_{p_1+1}^{p_1+3}(2,4,6)=\alpha_3$ $(j=3)$ より計算している．そして，系 3.2 より導かれる $p=p_1$ のときの有意水準を，(4.17) と同様に，$\alpha_{j0}=\Pr[\bigcup_{i=1}^{j}\{\sum_{l=1}^{i}Z_l^2>2i\}]$，$\beta_{j0}=\Pr(\sum_{i=1}^{j}Z_i^2>2j)$ より計算している．

表 1，2，3 は総じて $\beta_{i0}>\beta_{j0}$ for $i<j$ となっている．これは ARE の意味で望ましくより簡単な真のモデルを選択しやすくなるよう有意水準を調整する作用があることを示す．

表 3 は，β_j と α_j に着目すると，ARE の意味で帰無仮説と対立仮説のモデルが同程度のとき，モデルの候補が多くなるにつれ，より複雑なモデルを選択しがちになることを示している．AIC を改造し，検定問題(4.5)の有意水準 β_j と α_j をコントロールするには，例えば次のような方法が考えられる．(4.6)の $d(p_1,j)$ をある正の d により $d(p_1,j)=d(j-p_1)=dj-dp_1$，$j=p_1+1,p_1+2,\ldots,q$ とおく．そして(4.16)の式展開と同じアイデアで，母数を k 個推定したときのモデル選択基準

$$AIC_d(k)=-2l(\widehat{\boldsymbol{\theta}}(k),\widehat{\sigma}^2(k))+dk, \qquad k=p_1,p_1+1,\ldots,q \tag{4.18}$$

が最小となるような k を ARE の意味でよいモデルとして選択する．ただし，d は検定全体の有意水準 α を所与として，

$$C^q_{p_1+1}(d, 2d, \ldots, (q-p_1)d) = 1-\alpha \tag{4.19}$$

を満たし，個々の検定の有意水準を $\beta_i = 1-F(di; i, i)$，$j-p_1 = i = 1, 2, \ldots, q-p_1$ とする方法である．表 4 は，表 2 と同様に $\alpha = 0.05, 0.10, \ldots, 0.50$ を所与として，(4.19) と Newton–Raphson 法（初期値は $\alpha = 0.5, 0.4$ のとき $d = 2$，$\alpha = 0.3, 0.2, \ldots, 0.05$ のときはそれぞれ $\alpha = 0.4, 0.3, \ldots, 0.1$ のときの d の値）から d を求め，d を代入して β_i を求めている．また，α_0 と β_{j0} は，それぞれ求めた $d_j = jd$ $(j = 1, 2, 3)$ を (4.17) に代入して計算している．表 3 より $d = 2$ のとき β_i は単調増加していた．一方，表 4 をみると，$d \geq 2.8$ で β_i は単調減少しているようである．しかし $d = 2.0$ と 2.8 の間に増減の変化する d が存在するわけではない．具体的に $\{\beta_i, i = 1, 2, \ldots, 8\}$ は，$d = 2.3$ のとき $\{0.3087,\ 0.3349,\ 0.3398,\ 0.3391,\ 0.3361,\ 0.3322,\ 0.3279,\ 0.3234\}$，$d = 2.5$ のとき $\{0.2855,\ 0.2996,\ 0.2959,\ 0.2882,\ 0.2795,\ 0.2706,\ 0.2619,\ 0.2535\}$ と不定である．d の決定方法には，他に個々の検定の有意水準 β_i の最大を所与として求める方法も考えられる．

5. 非心 χ^2 分布に関するいくつかの結果とプログラム方法

この節では，論文の主要結果で使われた非心 χ^2 分布に関する結果と統計ソフトウェア S-PLUS による簡単なプログラム方法を紹介する．

5.1 定義と記号

非心 χ^2 分布のいくつかの定義と記号をおく．記号や既知の結果の多くは，Johnson et al. (1995) の Chapter 29 を参照している．

$\chi^2(\nu, \lambda)$ は，自由度 ν，非心度 λ の非心 χ^2 分布に従う確率変数で，$\chi^2(\nu, \lambda) = \sum_{i=1}^{\nu}(Z_i + \mu_i)^2$ と非心度 $\lambda = \sum_{i=1}^{\nu}\mu_i^2$ で定義される．ここで $\{Z_i\} \sim iidN(0, 1)$ で $\{\mu_i\}$ は実数列である．

$\chi^2(\nu, \lambda)$ の確率密度関数 $p(x; \nu, \lambda)$ は，次で表される．

$$p(x; \nu, \lambda) = \sum_{j=0}^{\infty}\left\{\frac{(\lambda/2)^j}{j!}e^{-\lambda/2}\right\}p(x; \nu+2j, 0) \qquad (x > 0), \tag{5.1}$$

$p(x; \nu, \lambda) = 0$ $(x \leq 0)$ である．ここで $p(x; \nu, 0)$ は，自由度 ν の χ^2 分布の確率密度関数で，$p(x; \nu, 0) = \{2^{\nu/2}\Gamma(\nu/2)\}^{-1}x^{\nu/2-1}\exp(-x/2)$ $(x > 0)$ で表される．つまり $x > 0$ のときは，平均 $\lambda/2$ のポアソン分布と自由度 $\nu+2j$ の χ^2 分布の確率密度の積の和により表される．

$\chi^2(\nu, \lambda)$ の分布関数，$\Pr(\chi^2(\nu, \lambda) \leq x) = F(x; \nu, \lambda)$ は，

$$\begin{aligned} F(x; \nu, \lambda) &= e^{-\lambda/2}\sum_{j=0}^{\infty}\frac{(\lambda/2)^j}{2^{\nu/2+j}j!\Gamma(\nu/2+j)}\int_0^x y^{\nu/2+j-1}e^{-y/2}dy \\ &= \sum_{j=0}^{\infty}\left\{\frac{(\lambda/2)^j}{j!}e^{-\lambda/2}\right\}F(x; \nu+2j, 0) \qquad (x > 0), \end{aligned} \tag{5.2}$$

$F(x; \nu, \lambda) = 0$ $(x \leq 0)$ である．ここで $F(x; \nu, 0)$ は，自由度 ν の χ^2 分布の分布関数である．また，例えば Ruben (1974) の 1 節により，$F(x; \nu, \lambda)$ は ν, λ の単調減少関数である．

5.2 漸化式と不等式

確率密度関数について，次が成立する．

$$xp(x; \nu, \lambda) = \nu p(x; \nu+2, \lambda) + \lambda p(x; \nu+4, \lambda) \qquad (\nu \geq 1, \lambda \geq 0), \tag{5.3}$$

$$p(x;\nu,\lambda)=\frac{1}{2}\{F(x;\nu-2,\lambda)-F(x;\nu,\lambda)\} \qquad (\nu\geq 3,\lambda\geq 0). \tag{5.4}$$

上の 2 式は，例えば Ruben（1974）の 2 節を参照．これら 2 式を用いて，$x>0$，$\nu\geq 1$，$\lambda\geq 0$ で

$$\begin{aligned} p(x;\nu,\lambda) &= \frac{\nu}{x}p(x;\nu+2,\lambda)+\frac{\lambda}{x}p(x;\nu+4,\lambda) \\ &= \frac{\nu}{2x}\{F(x;\nu,\lambda)-F(x;\nu+2,\lambda)\}+\frac{\lambda}{2x}\{F(x;\nu+2,\lambda)-F(x;\nu+4,\lambda)\} \end{aligned}$$

と展開される．さらに最右辺に三角不等式を施し，分布関数が $[0,1]$ の値をとることを思い出すと，次の不等式を得る．

$$p(x;\nu,\lambda)\leq\frac{\nu+\lambda}{2x} \qquad (x>0,\nu\geq 1,\lambda\geq 0). \tag{5.5}$$

この不等式は，$\mathrm{E}[\chi^2(\nu,\lambda)]=\nu+\lambda$ であることから，$2xp(x;\nu,\lambda)\leq\int_0^\infty yp(y;\nu,\lambda)dy$ なる被積分関数とその積分による不等式ともとれる．また（5.4）より $\nu\geq 3$，$\lambda\geq 0$ のとき，$p(x;\nu,\lambda)\leq 1/2$ なる上限が与えられる．

確率密度関数 $p(x;\nu,\lambda)$ を計算する関数は`S-PLUS` にはないが，$F(x;\nu,\lambda)$ を計算する`S-PLUS` 関数`pchisq`，（5.3）と（5.4）を利用して求められる．また，$p(x;1,\lambda)$ については，Johnson et al.（1995）の p. 441 より，標準正規分布の密度関数を利用しても求められる．

5.3 メディアンとパーセント点

メディアン及びパーセント点は，明示的に求めることができず,（擬似）Newton 法等を用いて求めるのが一般的であろう．

ここでは，初期値に依存せず，反復回数の少ない収束のために，最適化を行う関数について吟味する．また，メディアンとパーセント点のプログラム方法を紹介する．

連続型確率変数 X の確率密度関数を $f(x)$，分布関数を $F(x)$ として，

$$\frac{\partial}{\partial x}G(x)=g(x)=C\{F(x)-\alpha\}, \qquad G(x)=\int_{x_0}^x g(y)dy$$

とする．ここで C はある正の定数，x_0 と α はそれぞれ $x_0<x$ と $\alpha\in(0,1)$ を満たす定数とする．このとき，$g(x)$ は x の単調増加関数で，$x\in\mathbb{R}$ 上の凸関数である．以降，異なる $G(x)$ の表現を求める．

$$\begin{aligned} G(x)/C &= \int_{x_0}^x\{F(y)-\alpha\}dy=[y\{F(y)-\alpha\}]_{x_0}^x-\int_{x_0}^x y\frac{\partial}{\partial y}\{F(y)-\alpha\}dy \\ &= x\{F(x)-\alpha\}-x_0\{F(x_0)-\alpha\}-\int_{x_0}^x yf(y)dy. \end{aligned}$$

よって $G(x)$ の一般式として，

$$G(x)=Cx\{F(x)-\alpha\}-C\int_{-\infty}^x yf(y)dy+C_1 \tag{5.6}$$

を得る．ここで C_1 はある定数．さらに，

$$C=1/\alpha, \qquad C_1=\mathrm{E}(X)=\int_{-\infty}^\infty yf(y)dy$$

とおくと，$G(x)$ は次式のように展開される．

$$G(x)=x\{CF(x)-1\}-(C-1)\int_{-\infty}^{x}yf(y)dy+\int_{x}^{\infty}yf(y)dy$$
$$=(C-1)\int_{-\infty}^{x}(x-y)f(y)dy+\int_{x}^{\infty}(y-x)f(y)dy.$$

ここで $CF(x)-1=(C-1)F(x)+F(x)-1=(C-1)F(x)-\int_x^\infty f(y)dy$ を使った．さらに積分範囲の交換 $\int_{-\infty}^{x}(x-y)dF(y)=\int_{-\infty}^{x}\int_y^x dtdF(y)=\int_{-\infty}^{x}\int_{-\infty}^{t}dF(y)dt=\int_{-\infty}^{x}F(y)dy$ を行い，次式を得る．

$$(5.7)\qquad G(x)=C\left\{(1-\alpha)\int_{-\infty}^{x}|y-x|f(y)dy+\alpha\int_{x}^{\infty}|y-x|f(y)dy\right\}$$
$$=C\left\{(1-\alpha)\int_{-\infty}^{x}F(y)dy+\alpha\int_{x}^{\infty}\{1-F(y)\}dy\right\}.$$

$G(x)/C$ は $|X-x|$ の α と $1-\alpha$ による重み付き平均であり，$F(x)=\alpha$ を満たす x で最小化される．特に $\alpha=1/2$ の時には，よく知られた $G(x)=\mathrm{E}|X-x|$ となり，x が X のメディアンのときに最小化される．

さて，X が $\chi^2(\nu,\lambda)$ のとき，$\mathrm{E}[\chi^2(\nu,\lambda)]=\nu+\lambda$ と(5.3)を用いて，(5.7)は次のように書き表せる．

$$(5.8)\qquad G(x)/C=x\{F(x;\nu,\lambda)-\alpha\}-\nu\{F(x;\nu+2,\lambda)-\alpha\}-\lambda\{F(x;\nu+4,\lambda)-\alpha\}.$$

この結果は Causey (1986)の拡張ともいえる．

S-PLUS における $\chi^2(\nu,\lambda)$ のパーセント点及びメディアンの計算は，例えば次のように行う．

1) 初期値：Johnson et al. (1995)の(29.61d)式を用いる．またメディアンの場合は，Sen (1989)の mean-median-mode 不等式より，平均 $\nu+\lambda$ を用いてもよい．
2) Newton 法：S-PLUS 関数`nlminb`と`pchisq`を用いて，目的関数を(5.8)，1 階微分を`pchisq`，2 階微分を 5.3 節の $p(x;\nu,\lambda)$ を求める方法により求め，`nlminb` により最適化させる．ただし，収束判定条件を $|F(x^*;\nu,\lambda)-\alpha|<10^{-6}$ とし，満たす場合は x^* を出力して終了．収束していない場合は 3)へ．
3) 擬似 Newton 法：2)の`nlminb`の 2 階微分を指定せずに行う．収束判定条件を $|F(x^{**};\nu,\lambda)-\alpha|<10^{-6}$ とし，満たす場合は x^{**} を出力して終了．収束していない場合は出力せず警告を出して終了．

自由度 1 で α が極めて小さい場合，5.2 節の $p(x;\nu,\lambda)$ を求める方法は，x で割られていることが影響しているせいか，Newton 法の収束がうまくいかないことがある．そこで 3)を追加している．

5.4 非心 χ^2 分布の確率密度関数の重積分

ここでは，Hosoya (1986)の計算法を利用し，(4.11)の計算とそのS-PLUS によるプログラム法の例を示す．

まず，(4.11)の計算方法の例を挙げる．簡単化して，$S_j=\sum_{i=1}^{j}(Z_i+1)^2$，$p_1=0$，$k=1$ とおいて，$C_1^q(d_1,d_2,\ldots,d_q)=\Pr(S_1\le d_1,S_2\le d_2,\ldots,S_q\le d_q)$ $(q=2,3,4)$を求める．計算方法は Hosoya (1986, Sections 2 and 3)と同様に Markov 性を利用して積分する回数を節約する．具体的には

$$
\begin{aligned}
C_1^2(d_1,d_2) &= \Pr(S_1 \le d_1, S_2 \le d_2) = \int_0^{d_1} p(x_1;1,1)\int_0^{d_2-x_1} p(x_2;1,1)dx_2dx_1 \\
&= \int_0^{d_1} p(x_1;1,1)F(d_2-x_1;1,1)dx_1, \\
C_1^3(d_1,d_2,d_3) &= \int_0^{d_2} \Pr(S_i \le d_i, i=1,3|S_2=x)p(x;2,2)dx \\
&= \int_0^{d_2} \Pr(S_1 \le d_1|S_2=x)\Pr(S_3 \le d_3|S_2=x)p(x;2,2)dx \\
&= \int_0^{d_2} B_1(x,d_1)F(d_3-x;1,1)p(x;2,2)dx, \\
C_1^4(d_1,d_2,d_3,d_4) &= \int_0^{d_3} \Pr(S_i \le d_i, i=1,2,4|S_3=x)p(x;3,3)dx \\
&= \int_0^{d_3} \Pr(S_i \le d_i, i=1,2|S_3=x)\Pr(S_4 \le d_4|S_3=x)p(x;3,3)dx \\
&= \int_0^{d_3} B_2(x,d_1,d_2)F(d_4-x;1,1)p(x;3,3)dx.
\end{aligned}
\tag{5.9}
$$

ここで，

$$
B_1(x,d_1) = \Pr(S_1 \le d_1|S_2=x) = \begin{cases} 1 & \text{if } x < d_1, \\ \displaystyle\int_0^{d_1} \frac{p(y;1,1)p(x-y;1,1)}{p(x;2,2)}dy & \text{if } x \ge d_1, \end{cases}
$$

$$
\begin{aligned}
B_2(x,d_1,d_2) &= \Pr(S_1 \le d_1, S_2 \le d_2|S_3=x) \\
&= \begin{cases} 1 & \text{if } x \le d_1, \\ \displaystyle\int_0^{d_1} \frac{p(y;1,1)p(x-y;2,2)}{p(x;3,3)}dy & \text{if } d_1 < x \le d_2, \\ \displaystyle\int_0^{d_1} p(y;1,1)\int_0^{d_2-y} \frac{p(z;1,1)p(x-y-z;1,1)}{p(x;3,3)}dzdy & \text{if } d_2 < x. \end{cases}
\end{aligned}
$$

数値計算では，$B_1(x,d_1)$ の積分，$B_2(x,d_1,d_2)$ の $d_1 < x \le d_2$ の積分，そして $d_2 < x$ の z の積分は，2 次形式の表現に簡単化できる．以下例として，$\int_0^a p(x;1,1)p(x-y;1,1)dy$ の計算の近似を紹介する．

(5.1) より $p(x;\nu,\lambda) = \sum_{j=0}^{\infty} P_o(j,\lambda/2)p(x;\nu+2j,0)$ と書ける．ここで $P_o(j,\lambda/2)$ は平均 $\lambda/2$ のポアソン分布の確率分布である．また，$g_\nu = \{\Gamma(\nu/2)2^{\nu/2}\}^{-1}$ とおくと，$x,\nu > 0$ で $p(x;\nu,0) = g_\nu e^{-x/2}x^{\nu/2-1}$ と書けるので，次のように展開できる．

$$
\begin{aligned}
&\int_0^a p(x;1,1)p(x-y;1,1)dy \\
&= \sum_{j,k=0}^{\infty} P_o\left(j,\frac{1}{2}\right)P_o\left(k,\frac{1}{2}\right)g_{1+2j}g_{1+2k}e^{-x/2}\int_0^a y^{\frac{1+2j}{2}-1}(x-y)^{\frac{1+2k}{2}-1}dy \\
&= xe^{x/2}\sum_{j,k=0}^{\infty} b_j(x;1,1)B_{a/x}\left(\frac{1+2j}{2},\frac{1+2k}{2}\right)b_j(x;1,1)
\end{aligned}
\tag{5.10}
$$

ここで $b_j(x;\nu,\lambda) = P_o(j,\lambda/2)p(x;\nu+2j,0)$，$B_x(p,q)$ は不完全ベータ関数で $B_x(p,q) = \int_0^x t^{p-1}(1-$

$t)^{q-1}dt$ $(0<x<1)$で定義される．よって，2 次形式

$$xe^{x/2}\sum_{j,k=0}^{n} b_j(x;1,1)B_{a/x}\left(\frac{1+2j}{2},\frac{1+2k}{2}\right)b_k(x;1,1) \tag{5.11}$$

は，(5.10)の近似となる．

(5.11)の適当な $n>1$ を決めるために近似による誤差を評価する．$B_x(p,q)<\int_0^1 t^{p-1}(1-t)^{q-1}dt=2\int_0^{\pi/2}\cos^{2p-1}\theta\sin^{2q-1}\theta d\theta\le 2\int_0^{\pi/2}d\theta=\pi$ が，$p,q\ge 1/2$ で成立し，$\sum_{j,k=0}^{\infty}a_{j,k}-\sum_{j,k=0}^{n}a_{j,k}=\sum_{j=0}^{n}\sum_{k=n+1}^{\infty}a_{j,k}+\sum_{j=n+1}^{\infty}\sum_{k=0}^{\infty}a_{j,k}$ であることから，(5.10)と(5.11)の差は次のように評価できる．

$$\begin{aligned}
&xe^{x/2}\left\{\sum_{j=0}^{n}b_j(x;1,1)\sum_{k=n+1}^{\infty}b_k(x;1,1)B_{a/x}\left(\frac{1+2j}{2},\frac{1+2k}{2}\right)\right.\\
&\qquad\left.+\sum_{j=n+1}^{\infty}b_j(x;1,1)\sum_{k=0}^{\infty}b_k(x;1,1)B_{a/x}\left(\frac{1+2j}{2},\frac{1+2k}{2}\right)\right\}\\
&\le \pi xe^{x/2}\left\{\sum_{j=0}^{n}b_j(x;1,1)\sum_{k=n+1}^{\infty}b_k(x;1,1)+\sum_{j=n+1}^{\infty}b_j(x;1,1)\sum_{k=0}^{\infty}b_k(x;1,1)\right\}\\
&\le 2\pi xe^{x/2}p(x;1,1)\sum_{j=n+1}^{\infty}P_o\left(j,\frac{1}{2}\right)p(x;1+2j,0)\le \pi e^{x/2}\Pr\{P_o(1/2)>n\}.
\end{aligned}$$

ここで，最初と 2 番目の不等式は $b_j(x;\nu,\lambda)$ が非負であることと $p(x;\nu,\lambda)=\sum_{j=0}^{\infty}b_j(x;\nu,\lambda)$ であることを使った．また最後の不等式は $j\ge 1$ で，(5.4)より $p(x;1+2j,0)\le 1/2$ であることと，(5.5)より $p(x;1,1)\le 1/x$ を用いている．$P_o(1/2)$ は平均 $1/2$ のポアソン分布の確率変数である．

ゆえに，十分小さい誤差の上限 ϵ と $x>0$ を所与とすると，(5.11)の $n=n_\epsilon$ は，ポアソン分布の上側パーセント点 $\Pr\{P_o(1/2)>n_\epsilon\}=(\pi e^{x/2})^{-1}\epsilon$ により決めればよい．なお，Johnson et al. (1995)の 29.7 節にあるように，ポアソン分布の上側パーセント点の計算は，$F(x;\nu,\lambda)$ の計算で使われることも多くそれらの手法が適用できる．

しかし，Hosoya (1986)の Section 3 の例とは異なり，$B_2(x,d_1,d_2)$ の $d_2<x$ のケースは積分が残るため，$C_1^q(d_1,d_2,\ldots,d_q)$, $q\ge 6$ では，3 重積分(またはそれ以上の重積分)が出現し計算が複雑になる．

上記の方針によるS-PLUS の $C_1^q(d_1,d_2,\ldots,d_q)$ の計算は，5.2 節で述べた非心 χ^2 分布の分布関数を求めるpchisq，積分関数integrate，ベータ分布の分布関数を求めるpbeta，そしてポアソン分布のパーセント点を計算するqpois 等を用いて計算できる．

6.　最後に

6.1　論文のまとめ

論文は，入れ子型の複数の時系列モデルに限定しなおかつ真のモデルを含んでいるという条件で，PMSE を基準にモデル選択を誤る過誤を考慮したモデル選択法を漸近論より導いた．論文の設定として真の母数を標本の大きさ n に依存させたが，竹内(2004)が主張するように n が大きくなるにつれ有意水準が小さくなるような設定も考えられる．標本数が増えるとモデルを区別する情報がそれだけ増えるのだから，より小さい危険率でモデルを選択することができるのが当然というのがその理由である．

本来の研究目的は，PMSE の観点から見てよりよいモデル選択法は何か？であったが，それについての明確な解答を導いたわけではない．というのも真の母数の次数が未知で，仮に既知であっても検定問題でそうであるように全ての過誤を低くすることは不可能だからである．し

かし，けちの原理や非線形最適化による解の不安定性を考えると，簡単なモデルの方が同等または望ましいのにそれを棄て，複雑なモデルを採用する過誤をコントロールするモデル選択法は必要であろう．

論文の結果より様々な拡張が考えられる．具体的には，想定したモデルの最大次数 q が真の母数の次数 p より小さい場合が挙げられる．これは仮定を弱めて $c_3 \neq \mathbf{0}$ で考える必要があるので，そのときの PMSE の良さを導く必要がある．他にも，非線形を含む他のモデルの適用，非線形制約への拡張，h 期先予測について最良なモデル選択法の探索等は，将来の研究として残っている．

謝　辞

本研究は，2004 年統計関連学会連合大会で筑波大学ビジネス科学研究科の椿広計教授からいただいたコメントより発想を得ました．また，一橋大学大学院経済学研究科の山本拓教授と田中勝人教授からも論文の草稿段階よりコメントを頂きました．匿名のレフェリーお二人からは，重要かつ適切なコメントをいただきました．4.4 節の AIC の解釈の一部は，匿名のレフェリーのコメントより導きました．ここに記して心より感謝いたします．

なお，本研究の一部は，文部科学省科学研究費補助金（平成 16 年度と平成 17 年度）の援助を受けました．

A.　補　遺

A.1　2 節の証明

(2.4) の証明． 2.3 節の記号と仮定の下で，法則収束と同じオーダーでモーメント収束が成立することを示す．ある $n_0 > 0$ より大の全ての n で行列 $\boldsymbol{I}(\boldsymbol{\theta}^0)$ は正値定符号で $\widehat{\boldsymbol{Q}}^{(2)}(\boldsymbol{\theta}^*) = \widehat{\boldsymbol{I}}(\boldsymbol{\theta}^0) + \widehat{\boldsymbol{R}}(\boldsymbol{\theta}^*)$ と $\widehat{\boldsymbol{I}}(\boldsymbol{\theta}^0)$ の最小の固有値は正となる．この結果と (2.3) より $\boldsymbol{I}(\boldsymbol{\theta}^0) = \sum_{j=1}^{\infty} \boldsymbol{d}_j(\boldsymbol{\theta}^0)\boldsymbol{d}_j(\boldsymbol{\theta}^0)'$ を用いれば，Katayama (2006, Corollary 2.1) の証明と同様にして導ける．

他の結果も仮定 (2.3) と Katayama (2006, Corollary 2.1) より導かれる．□

補助定理 A.1． 数列 $\{a_i\}$ が二乗総和可能 ($\sum_{i=1}^{\infty} a_i^2 < \infty$) ならば，$\sum_{i=1}^{n} |a_i| = o(\sqrt{n})$ as $n \to \infty$．

証明． Robinson (1991) の Theorem 1 の証明を参照．□

補助定理 A.2． 次の確率変数

$$x_{i,n} = \sum_{j=0}^{n-1} \alpha_{i,j} v_{n-j}, \qquad z_{i,n} = \sum_{j=0}^{n-2} \left(\sum_{k=1}^{n-j-1} \beta_{i,k} v_{n-j-k} \right) v_{n-j}, \qquad i = 1, 2,$$

が，$\{v_j\} \sim iid(0, \sigma^2)$ で，$\mathrm{E}[v_j]^4 < \infty$ とする．また，$\sum_{j=0}^{\infty} \alpha_{i,j}^2$ と $\sum_{j=0}^{\infty} \beta_{i,j}^2$ は各 $i = 1, 2$ で有界とする．このとき $n \to \infty$ で次が成立する．

$$\mathrm{E}[x_{1,n} x_{2,n} z_{1,n} z_{2,n}]/n = \mathrm{E}[x_{1,n} x_{2,n}]\,\mathrm{E}[z_{1,n} z_{2,n}]/n + o(1), \tag{A.1a}$$

$$\mathrm{E}[x_{1,n} x_{2,n} z_{1,n}] = o(\sqrt{n}). \tag{A.1b}$$

証明. (A.1b)のみ証明する.(A.1a)についても証明の方針は同じである.与式を展開すると,

$$\mathrm{E}[x_{1,n}x_{2,n}z_{1,n}] = \sum_{j=0}^{n-1}\sum_{k=0}^{n-1}\sum_{l=0}^{n-2}\sum_{m=1}^{n-l-1}\alpha_{1,j}\alpha_{2,k}\beta_{1,m}\,\mathrm{E}[v_{n-j}v_{n-k}v_{n-l-m}v_{n-l}].$$

ところで独立性の仮定から,$j=l+m\neq k=l$ と $j=l\neq k=l+m$ のとき,$\mathrm{E}[v_{n-j}v_{n-k}v_{n-l-m}v_{n-l}] =\sigma^4$ で,他のケースは 0 である.よって Cauchy–Schwarz の不等式と補助定理 A.1 を用いて,

$$\begin{aligned}\mathrm{E}[x_{1,n}x_{2,n}z_{1,n}] &= \sigma^4\sum_{l=0}^{n-2}\alpha_{2,l}\sum_{m=1}^{n-l-1}\alpha_{1,l+m}\beta_{1,m} + \sigma^4\sum_{l=0}^{n-2}\alpha_{1,l}\sum_{m=1}^{n-l-1}\alpha_{2,l+m}\beta_{1,m}\\ &= O\left(\sum_{l=0}^{n-2}\alpha_{1,l}\right) + O\left(\sum_{l=0}^{n-2}\alpha_{2,l}\right) = o(\sqrt{n}).\end{aligned}$$ □

定理 2.1 の (2.7a) の証明. Katayama (2006), Theorem 3.1 の証明方針と同様なので概略のみ示す.(2.6)より,

$$\begin{aligned}\mathrm{E}[y_{n+h}-y_n(h,\widehat{\boldsymbol{\theta}})]^2 &= \mathrm{E}[y_{n+h}-y_n(h,\boldsymbol{\theta}^0)+y_n(h,\boldsymbol{\theta}^0)-y_n(h,\widehat{\boldsymbol{\theta}})]^2\\ &= \mathrm{E}\left[\sum_{j=0}^{h-1}\psi_j(\boldsymbol{\theta}^0)\varepsilon_{n+h-j}\right]^2 + \mathrm{E}[y_n(h,\boldsymbol{\theta}^0)-y_n(h,\widehat{\boldsymbol{\theta}})]^2\\ &= \sigma_0^2(h) + \mathrm{E}\left[\sum_{j=1}^{n}\left\{c_j(h,\widehat{\boldsymbol{\theta}})-c_j(h,\boldsymbol{\theta}^0)\right\}y_{n+1-j}\right]^2.\end{aligned} \tag{A.2}$$

よって(A.2)右辺の第 2 項を評価すれば十分.さて,Taylor 展開より右辺の第 2 項は,

$$\begin{aligned}&\sum_{j=1}^{n}\left\{c_j(h,\widehat{\boldsymbol{\theta}})-c_j(h,\boldsymbol{\theta}^0)\right\}y_{n+1-j} = \sum_{j=1}^{n}\frac{\partial c_j(h,\boldsymbol{\theta}^0)}{\partial\boldsymbol{\theta}'}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)y_{n+1-j}+R_{1,n},\\ &R_{1,n} = \sum_{j=1}^{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)'\frac{\partial^2 c_j(h,\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)y_{n+1-j}.\end{aligned} \tag{A.3}$$

ここで $\boldsymbol{\theta}^*$ は $\|\boldsymbol{\theta}^*-\boldsymbol{\theta}^0\|\le\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0\|$.さて,$\pi_j(\boldsymbol{\theta})$ は Θ 上一様に二乗総和可能なので,Weierstrass の 2 重級数定理と 2.2 節の $c_j(h,\boldsymbol{\theta})$ の定義より,$c_j^{(2)}(h,\boldsymbol{\theta}^*)$ の各要素は二乗総和可能である.また,Cauchy–Schwarz の不等式,補助定理 A.1,そして(2.4)より,ある $\sum_{j=1}^n c_j=o(\sqrt{n})$ なる正の実数列 $\{c_j\}$ が存在して,

$$\mathrm{E}[R_{1,n}]^2 \le \sqrt{\mathrm{E}\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0\|^8\left[\sum_{j=1}^{n}c_j\{\mathrm{E}(y_{n+1-j})^4\}^{1/4}\right]^4} = o(n^{-1}).$$

よって,(A.3)は,$\sum_{j=1}^n\{\partial c_j(h,\boldsymbol{\theta}^0)/\partial\boldsymbol{\theta}'\}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0)y_{n+1-j}$ の 2 次モーメントを評価すれば十分.証明の残りは,(2.4)と(A.1a)を用いて Katayama (2006), Theorem 3.1 の証明と同様の方針で導かれる.□

定理(2.1)の(2.7c)の証明. Katayama (2006), Theorem 3.2 の証明方針と,(2.7b),(2.7c)との証明方針は同様なので,(2.7c)の証明の概略のみ示す.

$\widetilde{\boldsymbol{\delta}}=(\widetilde{\boldsymbol{\delta}}_1',\widetilde{\boldsymbol{\delta}}_2')'$,対応させて $\boldsymbol{\delta}=(\boldsymbol{\delta}_1',\boldsymbol{\delta}_2')'$,$\boldsymbol{\delta}^0=((\boldsymbol{\delta}_1^0)',(\boldsymbol{\delta}_2^0)')'$ とおく.また,$Q_n(\boldsymbol{\theta})=Q_n(\boldsymbol{\delta},\boldsymbol{\delta}_3)=n^{-1}\sum_{t=1}^n e_t(\boldsymbol{\theta})^2$ そして $u_t(\boldsymbol{\theta})=u_t(\boldsymbol{\delta},\boldsymbol{\delta}_3)$ とおく.このとき,$\widetilde{\boldsymbol{\delta}}$ は $Q_n(\boldsymbol{\delta},\mathbf{0})$ を最小化して得られる.

まず強一致性について示す．Gourieroux and Monfort（1989）の Lemma 24.1 の証明法を利用する．

$$\begin{aligned}
&\sup_{\boldsymbol\delta}\left|Q_n(\boldsymbol\delta,\mathbf 0)-\mathrm E[u_t(\boldsymbol\delta,\boldsymbol\delta_3^0)]^2\right|\\
&\quad\le\sup_{\boldsymbol\delta}\left|Q_n(\boldsymbol\delta,\mathbf 0)-\mathrm E[u_t(\boldsymbol\delta,\mathbf 0)]^2\right|+\sup_{\boldsymbol\delta}\left|\mathrm E[u_t(\boldsymbol\delta,\mathbf 0)]^2-\mathrm E[u_t(\boldsymbol\delta,\boldsymbol\delta_3^0)]^2\right|\\
&\quad\le\sup_{\boldsymbol\delta,\boldsymbol\delta_3}\left|Q_n(\boldsymbol\delta,\boldsymbol\delta_3)-\mathrm E[u_t(\boldsymbol\delta,\boldsymbol\delta_3)]^2\right|+\sup_{\boldsymbol\delta}\left|\mathrm E[u_t(\boldsymbol\delta,\mathbf 0)]^2-\mathrm E[u_t(\boldsymbol\delta,\boldsymbol\delta_3^0)]^2\right|\xrightarrow{a.c.}\mathbf 0
\end{aligned}$$

ここで最右辺の第 1 項は（2.3）の最初の式，第 2 項は $\mathrm E[u_t(\boldsymbol\theta)]^2$ の一様連続性を使う：任意の $\epsilon>0$ で δ に独立なある $n_0>0$ が存在して，全ての $n>n_0$ で $\|(\boldsymbol\delta',\mathbf 0')'-(\boldsymbol\delta',(\boldsymbol\delta_3^0)')'\|<\epsilon$ かつ $|\mathrm E[u_t(\boldsymbol\delta,\mathbf 0)]^2-\mathrm E[u_t(\boldsymbol\delta,\boldsymbol\delta_3^0)]^2|<\epsilon$ となる．故に Gallant and White（1988）の Theorem 3.3 より $\widetilde\theta$ の強一致性が証明される．また，（2.3）の最初の式と $\mathrm E[u_t(\boldsymbol\theta)]^2$ の連続性から $|\widetilde\sigma^2-\sigma_0^2|\le|Q_n(\widetilde{\boldsymbol\theta})-\mathrm E[u_t(\widetilde{\boldsymbol\theta})]^2|+|\mathrm E[u_t(\widetilde{\boldsymbol\theta})]^2-\sigma_0^2|\le\sup_{\boldsymbol\theta}|Q_n(\boldsymbol\theta)-\mathrm E[u_t(\boldsymbol\theta)]^2|+|\mathrm E[u_t(\widetilde{\boldsymbol\theta})]^2-\sigma_0^2|\xrightarrow{a.c.}0$ なので，$\widetilde\sigma^2$ の強一致性が成立する．

次に $\widetilde\delta$ の漸近分布を導く．Taylor 展開と（2.3）より，

$$\text{(A.4)}\qquad\begin{aligned}
&\frac1n\frac{\partial Q(\widetilde{\boldsymbol\theta})}{\partial\boldsymbol\delta}=\mathbf 0=\frac1n\frac{\partial Q(\boldsymbol\theta^0)}{\partial\boldsymbol\delta}+\frac1n\frac{\partial^2Q(\boldsymbol\theta^*)}{\partial\boldsymbol\delta\partial\boldsymbol\delta'}(\widetilde{\boldsymbol\delta}-\boldsymbol\delta^0)+\frac1n\frac{\partial^2Q(\boldsymbol\theta^*)}{\partial\boldsymbol\delta\partial\boldsymbol\delta_3'}(\mathbf 0-\boldsymbol\delta_3^0),\\
&\sqrt n(\widetilde{\boldsymbol\delta}-\boldsymbol\delta^0)\overset{\mathrm a}{=}-\boldsymbol I_{12,12}^{-1}\left\{\frac1{\sqrt n}\frac{\partial Q(\boldsymbol\theta^0)}{\partial\boldsymbol\delta}-\frac1n\frac{\partial^2Q(\boldsymbol\theta^0)}{\partial\boldsymbol\delta\partial\boldsymbol\delta_3'}\boldsymbol c_3\right\}\overset{\mathrm a}{\sim}N\big(\boldsymbol I_{12,12}^{-1}\boldsymbol I_{12,3}\boldsymbol c_3,\ \boldsymbol I_{12,12}^{-1}\big).
\end{aligned}$$

ここで θ^* は $\|\theta^*-\theta^0\|\le\|\widetilde\theta-\theta^0\|$．また，ある $n_0>0$ より大の全ての n で行列 $\boldsymbol I(\theta^0)=\sum_{j=1}^\infty\boldsymbol d_j(\theta^0)\boldsymbol d_j(\theta^0)'$ が正値定符号であることと，（2.3）を用いると，$\mathrm E\|\widetilde\delta-\delta^0\|^r=O(n^{-r/2})$ が $r\ge2$ で成立することが，Katayama（2006）の（A.14）式の証明方針と同様にして導ける．

さて，証明の残りは，（2.7a）の議論と Katayama（2006），Theorem 3.2 の証明から，

$$\text{(A.5)}\quad\begin{aligned}
&\mathrm E\left[\sum_{j=1}^n\frac{\partial c_j(h,\boldsymbol\theta^0)}{\partial\boldsymbol\theta'}(\widetilde{\boldsymbol\theta}-\boldsymbol\theta^0)y_{n+1-j}\right]^2=\mathrm E\left[\sum_{j=0}^{n-1}\boldsymbol\varphi_j(h,\boldsymbol\theta^0)'(\widetilde{\boldsymbol\theta}-\boldsymbol\theta^0)\varepsilon_{n-j}\right]^2\\
&=\sum_{j=0}^{n-1}\sum_{k=0}^{n-1}\boldsymbol\varphi_j(h,\boldsymbol\theta^0)'\,\mathrm E\left[\begin{pmatrix}(\widetilde{\boldsymbol\delta}-\boldsymbol\delta^0)(\widetilde{\boldsymbol\delta}-\boldsymbol\delta^0)' & -(\widetilde{\boldsymbol\delta}-\boldsymbol\delta^0)\boldsymbol c_3'/\sqrt n\\ -\boldsymbol c_3(\widetilde{\boldsymbol\delta}-\boldsymbol\delta^0)'/\sqrt n & \boldsymbol c_3\boldsymbol c_3'/n\end{pmatrix}\varepsilon_{n-j}\varepsilon_{n-k}\right]\boldsymbol\varphi_k(h,\boldsymbol\theta^0)
\end{aligned}$$

を評価すれば十分．ところで，Taylor 展開を用いて，

$$\begin{aligned}
\widetilde{\boldsymbol\delta}-\boldsymbol\delta^0&=-\boldsymbol I_{12,12}^{-1}\left\{\frac1n\frac{\partial Q(\boldsymbol\theta^0)}{\partial\boldsymbol\delta}-\frac1{\sqrt n}\boldsymbol I_{12,3}\boldsymbol c_3\right\}+\boldsymbol R_{2,n},\\
\boldsymbol R_{2,n}&=-\boldsymbol I_{12,12}^{-1}\left[\left\{\frac1n\frac{\partial^2Q(\boldsymbol\theta^*)}{\partial\boldsymbol\delta\partial\boldsymbol\delta'}-\boldsymbol I_{12,12}\right\}(\widetilde{\boldsymbol\delta}-\boldsymbol\delta^0)+\left\{\boldsymbol I_{12,3}-\frac1n\frac{\partial^2Q(\boldsymbol\theta^*)}{\partial\boldsymbol\delta\partial\boldsymbol\delta_3'}\right\}\frac{\boldsymbol c_3}{\sqrt n}\right],
\end{aligned}$$

ここで $\sqrt n\boldsymbol R_{2,n}\xrightarrow{a.c.}\mathbf 0$ かつ $\mathrm E\|\boldsymbol R_{2,n}\|^r=o(n^{-r/2})$ である．よって $-\boldsymbol I_{12,12}^{-1}\{n^{-1}\partial Q(\boldsymbol\theta^0)/\partial\boldsymbol\delta-n^{-1/2}\,\boldsymbol I_{12,3}\boldsymbol c_3\}$ を（A.5）の $\widetilde\delta-\delta^0$ に代入し，補助定理 A.2 を用いれば結果を得る．□

系 2.1 の証明．（2.9a）-（2.9c）と（2.10a）は，定理 2.1 より $\varphi_j(1,\theta^0)=\boldsymbol d_{j+1}(\theta^0)$ であることと（2.3）から，簡単な行列演算で導ける．そこで（2.10b）を示す．定理 2.1 より，$\overline\Gamma(\theta^0)-\widetilde\Gamma(\theta^0)$ は次の行列となる．

$$\text{(A.6)}\qquad\begin{pmatrix}
\boldsymbol I_{11}^{-1}(\boldsymbol I_{12}\boldsymbol F\boldsymbol I_{12}'+\boldsymbol I_{13}\boldsymbol G'\boldsymbol I_{12}'+\boldsymbol I_{12}\boldsymbol G\boldsymbol I_{13}')\boldsymbol I_{11}^{-1} & -\boldsymbol I_{11}^{-1}(\boldsymbol I_{12}\boldsymbol F+\boldsymbol I_{13}\boldsymbol G') & -\boldsymbol I_{11}^{-1}\boldsymbol I_{12}\boldsymbol G\\
-(\boldsymbol F'\boldsymbol I_{12}'+\boldsymbol G\boldsymbol I_{13}')\boldsymbol I_{11}^{-1} & \boldsymbol F & \boldsymbol G\\
-\boldsymbol G'\boldsymbol I_{12}'\boldsymbol I_{11}^{-1} & \boldsymbol G' & \mathbf 0
\end{pmatrix}.$$

ここで $\boldsymbol{F}=\boldsymbol{F}_{22}=\boldsymbol{c}_2\boldsymbol{c}_2'-\boldsymbol{D}_{22}^{-1}-\boldsymbol{D}_{22}^{-1}\boldsymbol{E}_{23}\boldsymbol{c}_3\boldsymbol{c}_3'\boldsymbol{E}_{23}'\boldsymbol{D}_{22}^{-1}$，$\boldsymbol{G}=\boldsymbol{G}_{23}=(\boldsymbol{c}_2+\boldsymbol{D}_{22}^{-1}\boldsymbol{E}_{23}\boldsymbol{c}_3)\boldsymbol{c}_3'$．さて $\boldsymbol{c}_3=\boldsymbol{0}$ とする．このとき $\boldsymbol{F}=\boldsymbol{c}_2\boldsymbol{c}_2'-\boldsymbol{D}_{22}^{-1}$，$\boldsymbol{G}=\boldsymbol{0}$ となり，

$$
\text{(A.7)}\qquad \overline{\boldsymbol{\Gamma}}(\boldsymbol{\theta}^0)-\widetilde{\boldsymbol{\Gamma}}(\boldsymbol{\theta}^0)=\begin{pmatrix} \boldsymbol{I}_{11}^{-1}\boldsymbol{I}_{12}(\boldsymbol{c}_2\boldsymbol{c}_2'-\boldsymbol{D}_{22}^{-1})\boldsymbol{I}_{12}'\boldsymbol{I}_{11}^{-1} & -\boldsymbol{I}_{11}^{-1}\boldsymbol{I}_{12}(\boldsymbol{c}_2\boldsymbol{c}_2'-\boldsymbol{D}_{22}^{-1}) & \boldsymbol{0} \\ -(\boldsymbol{c}_2\boldsymbol{c}_2'-\boldsymbol{D}_{22}^{-1})\boldsymbol{I}_{12}'\boldsymbol{I}_{11}^{-1} & \boldsymbol{c}_2\boldsymbol{c}_2'-\boldsymbol{D}_{22}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}
$$
$$
=\big(\boldsymbol{I}_{12}'\boldsymbol{I}_{11}^{-1}\quad -\mathbb{I}\quad \boldsymbol{0}\big)'\big(\boldsymbol{c}_2\boldsymbol{c}_2'-\boldsymbol{D}_{22}^{-1}\big)\big(\boldsymbol{I}_{12}'\boldsymbol{I}_{11}^{-1}\quad -\mathbb{I}\quad \boldsymbol{0}\big)
$$

ここで $\mathbb{I}$ は $p_2\times p_2$ の単位行列である．よって，

$$
\begin{aligned}
\mathrm{ARE}\big[\,y_n(h,\overline{\boldsymbol{\theta}}),\ y_n(h,\widetilde{\boldsymbol{\theta}})\,\big] &\overset{\mathrm{a}}{=}\sigma_0^2\sum_{j=0}^{\infty}\boldsymbol{\varphi}_j(h,\boldsymbol{\theta}^0)'\Big(\overline{\boldsymbol{\Gamma}}(\boldsymbol{\theta}^0)-\widetilde{\boldsymbol{\Gamma}}(\boldsymbol{\theta}^0)\Big)\boldsymbol{\varphi}_j(h,\boldsymbol{\theta}^0)\\
&=\sigma_0^2\sum_{j=0}^{\infty}\boldsymbol{\kappa}_j(h,\boldsymbol{\theta}^0)'(\boldsymbol{c}_2\boldsymbol{c}_2'-\boldsymbol{D}_{22}^{-1})\boldsymbol{\kappa}_j(h,\boldsymbol{\theta}^0).
\end{aligned}
$$

(2.11)は，(2.10b)の証明の $p_2=1$ のケースとして導かれる．□

A.2 3 節の証明

定理 3.1 の証明． 定理 2.1 の(2.7c)の証明で用いた記号を使う．

まず ξ_n^W の漸近分布を導く．(A.4)と $\boldsymbol{\delta}_2^0=\boldsymbol{c}_2/\sqrt{n}$ より次のように展開できる．

$$
\text{(A.8)}\qquad
\begin{aligned}
&\sqrt{n}(\widetilde{\boldsymbol{\delta}}_2-\boldsymbol{\delta}_2^0)\overset{\mathrm{a}}{\sim}N\big(\boldsymbol{D}_{22}^{-1}\boldsymbol{E}_{23}\boldsymbol{c}_3,\ \boldsymbol{D}_{22}^{-1}\big),\\
&\sqrt{n}\widetilde{\boldsymbol{\delta}}_2\overset{\mathrm{a}}{\sim}N\big(\boldsymbol{c}_2+\boldsymbol{D}_{22}^{-1}\boldsymbol{E}_{23}\boldsymbol{c}_3,\ \boldsymbol{D}_{22}^{-1}\big),\\
&\sqrt{n}\boldsymbol{T}_{22}^{-1}\widetilde{\boldsymbol{\delta}}_2\overset{\mathrm{a}}{\sim}N\big(\boldsymbol{T}_{22}^{-1}(\boldsymbol{c}_2+\boldsymbol{D}_{22}^{-1}\boldsymbol{E}_{23}\boldsymbol{c}_3),\ \mathbb{I}\big).
\end{aligned}
$$

ここで，$\boldsymbol{T}_{22}$ は，$\boldsymbol{D}_{22}^{-1}=\boldsymbol{T}_{22}\boldsymbol{T}_{22}'$ となる非特異な下側三角行列．この結果から $\xi_n^W\overset{\mathrm{a}}{\sim}\chi^2(p_2,\lambda)$ が導かれる．

次に $\xi_n^{LR}\overset{\mathrm{a}}{=}\xi_n^S$ を示す．目的関数 $Q(\widetilde{\theta})$ と $Q(\overline{\theta})$ はそれぞれ θ^0 まわりの Taylor 展開により，

$$
\boldsymbol{0}=\frac{1}{\sqrt{n}}\frac{\partial Q(\widetilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\delta}}\overset{\mathrm{a}}{=}\frac{1}{\sqrt{n}}\frac{\partial Q(\boldsymbol{\theta}^0)}{\partial\boldsymbol{\delta}}+\frac{1}{n}\frac{\partial^2 Q(\boldsymbol{\theta}^0)}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}\sqrt{n}\begin{pmatrix}\widetilde{\boldsymbol{\delta}}_1-\boldsymbol{\delta}_1^0\\ \widetilde{\boldsymbol{\delta}}_2-\boldsymbol{\delta}_2^0\end{pmatrix}+\frac{1}{n}\frac{\partial^2 Q(\boldsymbol{\theta}^0)}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}_3'}\sqrt{n}(\boldsymbol{0}-\boldsymbol{\delta}_3^0),
$$
$$
\boldsymbol{0}\neq\frac{1}{\sqrt{n}}\frac{\partial Q(\overline{\boldsymbol{\theta}})}{\partial\boldsymbol{\delta}}\overset{\mathrm{a}}{=}\frac{1}{\sqrt{n}}\frac{\partial Q(\boldsymbol{\theta}^0)}{\partial\boldsymbol{\delta}}+\frac{1}{n}\frac{\partial^2 Q(\boldsymbol{\theta}^0)}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}\sqrt{n}\begin{pmatrix}\overline{\boldsymbol{\delta}}_1-\boldsymbol{\delta}_1^0\\ \boldsymbol{0}-\boldsymbol{\delta}_2^0\end{pmatrix}+\frac{1}{n}\frac{\partial^2 Q(\boldsymbol{\theta}^0)}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}_3'}\sqrt{n}(\boldsymbol{0}-\boldsymbol{\delta}_3^0).
$$

両辺の差をとり，$\partial Q(\overline{\boldsymbol{\theta}})/(\partial\boldsymbol{\delta}_1)=\boldsymbol{0}$ を用いて整理すると次式を得る．

$$
\text{(A.9)}\qquad \boldsymbol{I}_{12,12}^{-1}\frac{1}{\sqrt{n}}\frac{\partial Q(\overline{\boldsymbol{\theta}})}{\partial\boldsymbol{\delta}}=\boldsymbol{I}_{12,12}^{-1}\begin{pmatrix}\boldsymbol{0}\\ \dfrac{1}{\sqrt{n}}\dfrac{\partial Q(\overline{\boldsymbol{\theta}})}{\partial\boldsymbol{\delta}_2}\end{pmatrix}\overset{\mathrm{a}}{=}\sqrt{n}\begin{pmatrix}\overline{\boldsymbol{\delta}}_1-\widetilde{\boldsymbol{\delta}}_1\\ -\widetilde{\boldsymbol{\delta}}_2\end{pmatrix}.
$$

この式を念頭において LR 統計量の表現を導く．$\widetilde{\theta}$ まわりで $Q(\overline{\theta})$ を Taylor 展開すると，

$$
Q(\overline{\boldsymbol{\theta}})\overset{\mathrm{a}}{=}Q(\widetilde{\boldsymbol{\theta}})+\frac{\partial Q(\widetilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\delta}}\begin{pmatrix}\overline{\boldsymbol{\delta}}_1-\widetilde{\boldsymbol{\delta}}_1\\ -\widetilde{\boldsymbol{\delta}}_2\end{pmatrix}+\frac{1}{2}\begin{pmatrix}\overline{\boldsymbol{\delta}}_1-\widetilde{\boldsymbol{\delta}}_1\\ -\widetilde{\boldsymbol{\delta}}_2\end{pmatrix}'\frac{\partial^2 Q(\widetilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}\begin{pmatrix}\overline{\boldsymbol{\delta}}_1-\widetilde{\boldsymbol{\delta}}_1\\ -\widetilde{\boldsymbol{\delta}}_2\end{pmatrix}.
$$

$\partial Q(\widetilde{\theta})/(\partial\delta)=\boldsymbol{0}$ と誤差項の分散の MLE は強一致性が成立することから，

$$
\xi_n^{LR}=-2\Big\{l(\overline{\boldsymbol{\theta}},\overline{\sigma}^2)-l(\widetilde{\boldsymbol{\theta}},\widetilde{\sigma}^2)\Big\}\overset{\mathrm{a}}{=}n\begin{pmatrix}\overline{\boldsymbol{\delta}}_1-\widetilde{\boldsymbol{\delta}}_1\\ -\widetilde{\boldsymbol{\delta}}_2\end{pmatrix}'\boldsymbol{I}_{12,12}\begin{pmatrix}\overline{\boldsymbol{\delta}}_1-\widetilde{\boldsymbol{\delta}}_1\\ -\widetilde{\boldsymbol{\delta}}_2\end{pmatrix}.
$$

ここで(A.9)を用いて，

$$\xi_n^{LR} \stackrel{\rm a}{=} \begin{pmatrix} \mathbf{0} \\ \dfrac{1}{\sqrt{n}}\dfrac{\partial l(\overline{\boldsymbol{\theta}},\sigma_0^2)}{\partial \boldsymbol{\delta}_2} \end{pmatrix}' \boldsymbol{I}_{12,12}^{-1} \begin{pmatrix} \mathbf{0} \\ \dfrac{1}{\sqrt{n}}\dfrac{\partial l(\overline{\boldsymbol{\theta}},\sigma_0^2)}{\partial \boldsymbol{\delta}_2} \end{pmatrix} = \frac{1}{n}\frac{\partial l(\overline{\boldsymbol{\theta}},\sigma_0^2)}{\partial \boldsymbol{\delta}_2'}\boldsymbol{D}_{22}^{-1}\frac{\partial l(\overline{\boldsymbol{\theta}},\sigma_0^2)}{\partial \boldsymbol{\delta}_2'} \stackrel{\rm a}{=} \xi_n^S.$$

続いて $\xi_n^S \stackrel{\rm a}{=} \xi_n^W$ を示す．再び(A.9)より $n\boldsymbol{D}_{22}\widetilde{\boldsymbol{\delta}}_2 \stackrel{\rm a}{=} \partial l(\overline{\boldsymbol{\theta}},\sigma_0^2)/(\partial \boldsymbol{\delta}_2)$ を用いて，

$$\xi_n^S \stackrel{\rm a}{=} \frac{1}{n}\frac{\partial l(\overline{\boldsymbol{\theta}},\sigma_0^2)}{\partial \boldsymbol{\delta}_2'}\boldsymbol{D}_{22}^{-1}\frac{\partial l(\overline{\boldsymbol{\theta}},\sigma_0^2)}{\partial \boldsymbol{\delta}_2'} \stackrel{\rm a}{=} n\widetilde{\boldsymbol{\delta}}_2'\boldsymbol{D}_{22}\widetilde{\boldsymbol{\delta}}_2 \stackrel{\rm a}{=} \xi_n^W.$$

最後に $\xi_n^{LR} \stackrel{\rm a}{\sim} \chi^2(p_2,\lambda)$ と $\xi_n^S \stackrel{\rm a}{\sim} \chi^2(p_2,\lambda)$ を示す．$\xi_n^W \stackrel{\rm a}{=} \xi_n^{LR} \stackrel{\rm a}{=} \xi_n^S$ と Slutsky の定理より，ξ_n^{LR} と ξ_n^S も ξ_n^W と漸近的に同じ分布，$\chi^2(p_2,\lambda)$ に従う．□

定理 3.2 の証明． $j=q$ のケースのみ示す．$j=p+1,p+2,\ldots,q-1$ のケースも同様である．定理 3.1 より，3 つの検定統計量は漸近的に等しいものとして扱えるので，$\xi_n(i,j)$ は，

$$\begin{aligned}\xi_n(p_1,q) &\stackrel{\rm a}{=} \xi_n^{LR}(p_1,q) = -2\Big\{l(\overline{\boldsymbol{\theta}},\overline{\sigma}^2) - l(\widehat{\boldsymbol{\theta}},\widehat{\sigma}^2)\Big\} \\ &= -2\Big\{l(\overline{\boldsymbol{\theta}},\overline{\sigma}^2) - l(\widetilde{\boldsymbol{\theta}},\widetilde{\sigma}^2) + l(\widetilde{\boldsymbol{\theta}},\widetilde{\sigma}^2) - l(\widehat{\boldsymbol{\theta}},\widehat{\sigma}^2)\Big\} \\ &= \xi_n^{LR}(p_1,p) + \xi_n^{LR}(p,q) \stackrel{\rm a}{=} \xi_n^W(p_1,p) + \xi_n^W(p,q).\end{aligned}$$

よって，非心 χ^2 分布の再生性と $\xi_n(p_1,q) \stackrel{\rm a}{=} \xi_n(p_1,p)+\xi_n(p,q)$ から，$\xi_n^W(p_1,p)$ と $\xi_n^W(p,q)$ が漸近的に独立で，それぞれ，$\xi_n^W(p_1,p) \stackrel{\rm a}{\sim} \chi^2(p_2,\boldsymbol{c}_2'\boldsymbol{D}_{22}\boldsymbol{c}_2)$ かつ $\xi_n^W(p,q) \stackrel{\rm a}{\sim} \chi^2(p_3,0)$ を示せば十分である．そのために，定理 2.1 の(2.7c)の証明で用いた記号を使うと，(A.8)の式展開より，次を示せばよい．

$$\text{(A.10)}\qquad \begin{pmatrix} \sqrt{n}(\widetilde{\boldsymbol{\delta}}-\boldsymbol{\delta}^0) \\ \sqrt{n}(\widehat{\boldsymbol{\delta}}_3-\boldsymbol{\delta}_3^0) \end{pmatrix} \stackrel{\rm a}{\sim} N\left(\mathbf{0},\ \begin{pmatrix} \boldsymbol{I}_{12,12}^{-1} & \mathbf{0} \\ \mathbf{0} & {\boldsymbol{D}_{33}^*}^{-1} \end{pmatrix}\right).$$

ここで，$\widehat{\boldsymbol{\delta}}_3$ は，$\boldsymbol{\theta}^0=((\boldsymbol{\delta}_1^0)',(\boldsymbol{\delta}_2^0)',(\boldsymbol{\delta}_3^0)')'$ に対応させて，$\widehat{\boldsymbol{\theta}}=(\widehat{\boldsymbol{\delta}}_1',\widehat{\boldsymbol{\delta}}_2',\widehat{\boldsymbol{\delta}}_3')'$ により定義される．また $\boldsymbol{D}_{33}^*=\boldsymbol{I}_{33}-\boldsymbol{I}_{12,3}'\boldsymbol{I}_{12,12}^{-1}\boldsymbol{I}_{12,3}$ で，${\boldsymbol{D}_{33}^*}^{-1}$ は，$\sqrt{n}(\widehat{\boldsymbol{\delta}}_3-\boldsymbol{\delta}_3^0)$ の漸近分散である．定理 2.1 の(2.7c)の証明と同様に，Taylor 展開から，

$$\sqrt{n}(\widetilde{\boldsymbol{\delta}}-\boldsymbol{\delta}^0) \stackrel{\rm a}{=} \boldsymbol{I}_{12,12}^{-1}\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta}^0,\sigma_0^2)}{\partial \boldsymbol{\delta}},$$

$$\begin{aligned}\begin{pmatrix} \sqrt{n}(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta}^0) \\ \sqrt{n}(\widehat{\boldsymbol{\delta}}_3-\boldsymbol{\delta}_3^0) \end{pmatrix} &\stackrel{\rm a}{=} \begin{pmatrix} \boldsymbol{I}_{12,12} & \boldsymbol{I}_{12,3} \\ \boldsymbol{I}_{12,3}' & \boldsymbol{I}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \dfrac{1}{\sqrt{n}}\dfrac{\partial l(\boldsymbol{\theta}^0,\sigma_0^2)}{\partial \boldsymbol{\delta}} \\ \dfrac{1}{\sqrt{n}}\dfrac{\partial l(\boldsymbol{\theta}^0,\sigma_0^2)}{\partial \boldsymbol{\delta}_3} \end{pmatrix} \\ &= \begin{pmatrix} {\boldsymbol{I}_{12,12}^*}^{-1} & -\boldsymbol{I}_{12,12}^{-1}\boldsymbol{I}_{12,3}{\boldsymbol{D}_{33}^*}^{-1} \\ -{\boldsymbol{D}_{33}^*}^{-1}\boldsymbol{I}_{12,3}'\boldsymbol{I}_{12,12}^{-1} & {\boldsymbol{D}_{33}^*}^{-1} \end{pmatrix} \begin{pmatrix} \dfrac{1}{\sqrt{n}}\dfrac{\partial l(\boldsymbol{\theta}^0,\sigma_0^2)}{\partial \boldsymbol{\delta}} \\ \dfrac{1}{\sqrt{n}}\dfrac{\partial l(\boldsymbol{\theta}^0,\sigma_0^2)}{\partial \boldsymbol{\delta}_3} \end{pmatrix}.\end{aligned}$$

ここで ${\boldsymbol{I}_{12,12}^*}^{-1}=\boldsymbol{I}_{12,12}^{-1}+\boldsymbol{I}_{12,12}^{-1}\boldsymbol{I}_{12,3}{\boldsymbol{D}_{33}^*}^{-1}\boldsymbol{I}_{12,3}'\boldsymbol{I}_{12,12}^{-1}$，$\widehat{\boldsymbol{\delta}}=(\widehat{\boldsymbol{\delta}}_1',\widehat{\boldsymbol{\delta}}_2')'$ とおいた．これらの式から，$(\sqrt{n}(\widetilde{\boldsymbol{\delta}}-\boldsymbol{\delta}^0)',\sqrt{n}(\widehat{\boldsymbol{\delta}}_3-\boldsymbol{\delta}_3^0)')'$ は，漸近的に次のように行列表現できる．

$$\begin{pmatrix} \sqrt{n}(\widetilde{\boldsymbol{\delta}}-\boldsymbol{\delta}^0) \\ \sqrt{n}(\widehat{\boldsymbol{\delta}}_3-\boldsymbol{\delta}_3^0) \end{pmatrix} \stackrel{\rm a}{=} \begin{pmatrix} \boldsymbol{I}_{12,12}^{-1} & \mathbf{0} \\ -{\boldsymbol{D}_{33}^*}^{-1}\boldsymbol{I}_{12,3}'\boldsymbol{I}_{12,12}^{-1} & {\boldsymbol{D}_{33}^*}^{-1} \end{pmatrix} \begin{pmatrix} \dfrac{1}{\sqrt{n}}\dfrac{\partial l(\boldsymbol{\theta}^0,\sigma_0^2)}{\partial \boldsymbol{\delta}} \\ \dfrac{1}{\sqrt{n}}\dfrac{\partial l(\boldsymbol{\theta}^0,\sigma_0^2)}{\partial \boldsymbol{\delta}_3} \end{pmatrix}$$

$$= \boldsymbol{A}\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta}^0,\sigma_0^2)}{\partial \boldsymbol{\theta}} \overset{\mathrm{a}}{\sim} N\big(\mathbf{0},\ \boldsymbol{A}\boldsymbol{I}(\boldsymbol{\theta}^0)\boldsymbol{A}'\big).$$

ここで，

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{I}_{12,12}^{-1} & \mathbf{0} \\ -\boldsymbol{D}_{33}^{*}{}^{-1}\boldsymbol{I}_{12,3}'\boldsymbol{I}_{12,12}^{-1} & \boldsymbol{D}_{33}^{*}{}^{-1} \end{pmatrix}$$

とおいた．これから，$\boldsymbol{A}\boldsymbol{I}(\theta^0)\boldsymbol{A}'$ を計算すると，(A.10)が得られる．□

定理 3.3 の証明のために次の補助定理を与える．そのために記号を準備する．δ_2^0 を $p_2 = p_\alpha + p_\beta + p_\gamma, p_\alpha \geq 1, p_\beta \geq 1, p_\gamma \geq 0$ となるように，$\delta_2^0 = \boldsymbol{c}_2/\sqrt{n} = ((\boldsymbol{\delta}_\alpha^0)', (\boldsymbol{\delta}_\beta^0)', (\boldsymbol{\delta}_\gamma^0)')' = (\boldsymbol{c}_\alpha', \boldsymbol{c}_\beta', \boldsymbol{c}_\gamma')'/\sqrt{n}$ と分解して表記する．対応させて，$\delta_2 = (\delta_\alpha', \delta_\beta', \delta_\gamma')'$ とおく．また $q - p_1 - p_\alpha$ の 0 制約による MLE を $\widehat{\boldsymbol{\theta}}(p_1 + p_\alpha) = (\widehat{\boldsymbol{\delta}}'_{1\alpha}, \widehat{\boldsymbol{\delta}}'_{\alpha\alpha}, 0, \ldots, 0)'$ とおく．ここで $\widehat{\delta}_{1\alpha}$ と $\widehat{\delta}_{\alpha\alpha}$ はそれぞれ p_1 次と p_α 次ベクトルである．$q - p_1 - p_\alpha - p_\beta$ の 0 制約による MLE を $\widehat{\boldsymbol{\theta}}(p_1 + p_\alpha + p_\beta) = (\widehat{\boldsymbol{\delta}}'_{1\beta}, \widehat{\boldsymbol{\delta}}'_{\alpha\beta}, \widehat{\boldsymbol{\delta}}'_{\beta\beta}, 0, \ldots, 0)'$ とおく．ここで $\widehat{\delta}_{1\beta}$，$\widehat{\delta}_{\alpha\beta}$ そして $\widehat{\delta}_{\beta\beta}$ はそれぞれ p_1 次，p_α 次，p_β 次ベクトルである．さらに，それぞれ同じ次数のベクトルでそろえて，$\widehat{\theta}_\alpha = (\widehat{\boldsymbol{\delta}}'_{1\alpha}, \widehat{\boldsymbol{\delta}}'_{\alpha\alpha})'$，$\boldsymbol{\theta}_\alpha^0 = ((\delta_1^0)', (\delta_\alpha^0)')'$，$\boldsymbol{\theta}_\alpha = (\delta_1', \delta_\alpha')'$ とおく．同様にして，$\widehat{\theta}_\beta = (\widehat{\boldsymbol{\delta}}'_{1\beta}, \widehat{\boldsymbol{\delta}}'_{\alpha\beta}, \widehat{\boldsymbol{\delta}}'_{\beta\beta})'$，$\boldsymbol{\theta}_\alpha^0 = ((\delta_1^0)', (\delta_\alpha^0)', (\delta_\beta^0)')'$ とおく．行列の記号を追加して，$\boldsymbol{I}_{12} = (\boldsymbol{I}_{1\alpha}\ \boldsymbol{I}_{1\beta}\ \boldsymbol{I}_{1\gamma})$，$\boldsymbol{I}_{1\alpha,\beta} = (\boldsymbol{I}_{1\beta}'\ \boldsymbol{I}_{\alpha\beta}')'$，$\boldsymbol{I}_{1\alpha,\gamma} = (\boldsymbol{I}_{1\gamma}'\ \boldsymbol{I}_{\alpha\gamma}')'$，

$$\boldsymbol{I}_{22} = \begin{pmatrix} \boldsymbol{I}_{\alpha\alpha} & \boldsymbol{I}_{\alpha\beta} & \boldsymbol{I}_{\alpha\gamma} \\ \boldsymbol{I}_{\alpha\beta}' & \boldsymbol{I}_{\beta\beta} & \boldsymbol{I}_{\beta\gamma} \\ \boldsymbol{I}_{\alpha\gamma}' & \boldsymbol{I}_{\beta\gamma}' & \boldsymbol{I}_{\gamma\gamma} \end{pmatrix},\quad \boldsymbol{I}_{1\alpha,1\alpha} = \begin{pmatrix} \boldsymbol{I}_{11} & \boldsymbol{I}_{1\alpha} \\ \boldsymbol{I}_{1\alpha}' & \boldsymbol{I}_{\alpha\alpha} \end{pmatrix},\quad \boldsymbol{I}_{1\alpha\beta,1\alpha\beta} = \begin{pmatrix} \boldsymbol{I}_{1\alpha,1\alpha} & \boldsymbol{I}_{1\alpha,\beta} \\ \boldsymbol{I}_{1\alpha,\beta}' & \boldsymbol{I}_{\beta\beta} \end{pmatrix}$$

とおく．ここで，$\boldsymbol{I}_{ij}$ $(i, j = 1, \alpha, \beta, \gamma)$は，それぞれ，$p_i \times p_j$ 行列である．

補助定理 A.3. 定理 3.2 を仮定し，上記の記号を用いて，$n \to \infty$ で次が成立する．

$$\xi_n(p_1, p_1 + p_\alpha + p_\beta) \overset{\mathrm{a}}{=} \xi_n(p_1, p_1 + p_\alpha) + \xi_n(p_1 + p_\alpha, p_1 + p_\alpha + p_\beta).$$

ここで，$\xi_n(p_1, p_1 + p_\alpha)$ と $\xi_n(p_1 + p_\alpha, p_1 + p_\alpha + p_\beta)$ は漸近的に独立で，

$$\xi_n(p_1, p_1 + p_\alpha) \overset{\mathrm{a}}{\sim} \chi^2\big(p_\alpha, (\boldsymbol{c}_\alpha' + \boldsymbol{c}_{\beta\gamma}'\boldsymbol{E}_{\alpha,\beta\gamma}'\boldsymbol{D}_{\alpha\alpha}^{-1})\boldsymbol{D}_{\alpha\alpha}(\boldsymbol{c}_\alpha + \boldsymbol{D}_{\alpha\alpha}^{-1}\boldsymbol{E}_{\alpha,\beta\gamma}\boldsymbol{c}_{\beta\gamma})\big),$$
$$\xi_n(p_1 + p_\alpha, p_1 + p_\alpha + p_\beta) \overset{\mathrm{a}}{\sim} \chi^2\big(p_\beta, (\boldsymbol{c}_\beta' + \boldsymbol{c}_\gamma'\boldsymbol{E}_{\beta\gamma}'\boldsymbol{D}_{\beta\beta}^{-1})\boldsymbol{D}_{\beta\beta}(\boldsymbol{c}_\beta + \boldsymbol{D}_{\beta\beta}^{-1}\boldsymbol{E}_{\beta\gamma}\boldsymbol{c}_\gamma)\big).$$

ここで，$\boldsymbol{c}_{\beta\gamma} = (\boldsymbol{c}_\beta', \boldsymbol{c}_\gamma')'$，$\boldsymbol{E}_{\alpha,\beta\gamma} = \{(\boldsymbol{I}_{\alpha\beta}\ \boldsymbol{I}_{\alpha\gamma}) - \boldsymbol{I}_{1\alpha}'\boldsymbol{I}_{11}^{-1}(\boldsymbol{I}_{1\beta}\ \boldsymbol{I}_{1\gamma})\}\boldsymbol{c}_{\beta\gamma}$，$\boldsymbol{E}_{\beta\gamma} = (\boldsymbol{I}_{\beta\gamma} - \boldsymbol{I}_{1\alpha,\beta}'\boldsymbol{I}_{1\alpha,1\alpha}^{-1}\boldsymbol{I}_{1\alpha,\gamma})$，$\boldsymbol{D}_{\alpha\alpha} = \boldsymbol{I}_{\alpha\alpha} - \boldsymbol{I}_{1\alpha}'\boldsymbol{I}_{11}^{-1}\boldsymbol{I}_{1\alpha}$，$\boldsymbol{D}_{\beta\beta} = \boldsymbol{I}_{\beta\beta} - \boldsymbol{I}_{1\alpha,\beta}'\boldsymbol{I}_{1\alpha,1\alpha}^{-1}\boldsymbol{I}_{1\alpha,\beta}$ である．

証明. 定理 3.2 の証明と方針は同様である．Taylor 展開により

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_\alpha^0) - \boldsymbol{I}_{1\alpha,1\alpha}^{-1}(\boldsymbol{I}_{1\alpha,\beta}\boldsymbol{c}_\beta + \boldsymbol{I}_{1\alpha,\gamma}\boldsymbol{c}_\gamma) \overset{\mathrm{a}}{=} \boldsymbol{I}_{1\alpha,1\alpha}^{-1}\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta}^0,\sigma_0^2)}{\partial \boldsymbol{\theta}_\alpha},$$

$$\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\delta}}_{1\beta} - \boldsymbol{\delta}_1^0 \\ \widehat{\boldsymbol{\delta}}_{\alpha\beta} - \boldsymbol{\delta}_\alpha^0 \\ \widehat{\boldsymbol{\delta}}_{\beta\beta} - \boldsymbol{\delta}_\beta^0 \end{pmatrix} - \boldsymbol{I}_{1\alpha\beta,1\alpha\beta}^{-1}\begin{pmatrix} \boldsymbol{I}_{1\gamma} \\ \boldsymbol{I}_{\alpha\gamma} \\ \boldsymbol{I}_{\beta\gamma} \end{pmatrix}\boldsymbol{c}_\gamma \overset{\mathrm{a}}{=} \boldsymbol{I}_{1\alpha\beta,1\alpha\beta}^{-1}\frac{1}{\sqrt{n}}\begin{pmatrix} \dfrac{\partial l(\boldsymbol{\theta}^0,\sigma_0^2)}{\partial \boldsymbol{\theta}_\alpha} \\ \dfrac{\partial l(\boldsymbol{\theta}^0,\sigma_0^2)}{\partial \boldsymbol{\delta}_\beta} \end{pmatrix}.$$

と書けるので，$\widehat{\theta}_\alpha$ と $\widehat{\delta}_{\beta\beta}$ に対応する行の小行列を計算して整理すると，次式が導かれる．

$$\begin{pmatrix} \sqrt{n}(\widehat{\boldsymbol{\theta}}_\alpha - \boldsymbol{\theta}_\alpha^0) - \boldsymbol{I}_{1\alpha,1\alpha}^{-1}(\boldsymbol{I}_{1\alpha,\beta}\boldsymbol{c}_\beta + \boldsymbol{I}_{1\alpha,\gamma}\boldsymbol{c}_\gamma) \\ \sqrt{n}(\widehat{\boldsymbol{\delta}}_{\beta\beta} - \boldsymbol{\delta}_\beta^0) - \boldsymbol{D}_{\beta\beta}^{-1}\boldsymbol{E}_{\beta\gamma}\boldsymbol{c}_\gamma \end{pmatrix}$$

$$\stackrel{\mathrm{a}}{=} \begin{pmatrix} \boldsymbol{I}_{1\alpha,1\alpha}^{-1} & \boldsymbol{0} \\ \boldsymbol{D}_{\beta\beta}^{-1}\boldsymbol{I}_{1\alpha,\beta}'\boldsymbol{I}_{1\alpha,1\alpha}^{-1} & \boldsymbol{D}_{\beta\beta}^{-1} \end{pmatrix} \frac{1}{\sqrt{n}} \begin{pmatrix} \dfrac{\partial l(\boldsymbol{\theta}^0,\sigma_0^2)}{\partial \boldsymbol{\theta}_\alpha} \\ \dfrac{\partial l(\boldsymbol{\theta}^0,\sigma_0^2)}{\partial \boldsymbol{\delta}_\beta} \end{pmatrix} \stackrel{\mathrm{a}}{\sim} N\left(\boldsymbol{0}, \begin{pmatrix} \boldsymbol{I}_{1\alpha,1\alpha}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_{\beta\beta}^{-1} \end{pmatrix}\right).$$

さらに，$\widehat{\delta}_{\alpha\alpha}$ と $\widehat{\delta}_{\beta\beta}$ に対応する行の小行列を計算すると，

$$\sqrt{n}\begin{pmatrix} \widehat{\boldsymbol{\delta}}_{\alpha\alpha} \\ \widehat{\boldsymbol{\delta}}_{\beta\beta} \end{pmatrix} \stackrel{\mathrm{a}}{\sim} N\left(\begin{pmatrix} \boldsymbol{c}_\alpha + \boldsymbol{D}_{\alpha\alpha}^{-1}\boldsymbol{E}_{\alpha,\beta\gamma}\boldsymbol{c}_{\beta\gamma} \\ \boldsymbol{c}_\beta + \boldsymbol{D}_{\beta\beta}^{-1}\boldsymbol{E}_{\beta\gamma}\boldsymbol{c}_\gamma \end{pmatrix}, \begin{pmatrix} \boldsymbol{D}_{\alpha\alpha}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_{\beta\beta}^{-1} \end{pmatrix}\right)$$

を得る．この式は，$\xi_n^W(p_1,p_1+p_\alpha)$ と $\xi_n^W(p_1+p_\alpha,p_1+p_\alpha+p_\beta)$ が漸近的に独立に非心 χ^2 分布に従うことを示しており，定理 3.1 を用いて結論を導ける．□

定理 3.3 の証明． まず(3.9)を示す．系 3.1 より，$\xi_n(p_1,j)$，$j=p_1+1,p_1+2,\ldots,p$ の同時分布を求めれば十分である．まず $\xi_n(p_1,p)$ について．定理 3.1 より $\xi_n(p_1,p)\stackrel{\mathrm{a}}{=}\xi_n(p_1,p-1)+\xi_n(p-1,p)$ である．補助定理 A.3 の $(p_\alpha,p_\beta,p_\gamma)=(p_2-1,1,0)$ のケースを適用すると，$\xi_n(p_1,p-1)$ と $\xi_n(p-1,p)$ は漸近的に独立に非心 χ^2 分布に従う．特に $\xi_n(p-1,p)\stackrel{\mathrm{a}}{\sim}\chi^2(1,\boldsymbol{c}_\beta^2\boldsymbol{D}_{\beta\beta})$．次に $\xi_n(p_1,p-1)$ について，定理 3.1 より $\xi_n(p_1,p-1)\stackrel{\mathrm{a}}{=}\xi_n(p_1,p-2)+\xi_n(p-2,p-1)$ であり，補助定理 A.3 の $(p_\alpha,p_\beta,p_\gamma)=(p_2-2,1,1)$ のケースを適用する．これを繰り返して，最後に $\xi_n(p_1,p_1+2)$ について，$\xi_n(p_1,p_1+2)\stackrel{\mathrm{a}}{=}\xi_n(p_1,p_1+1)+\xi_n(p_1+1,p_1+2)$ について，補助定理 A.3 の $(p_\alpha,p_\beta,p_\gamma)=(1,1,p_2-2)$ のケースを適用すると，非心 χ^2 分布の再生性から(3.9)が得られる．$\mu(i)$ と(3.10)は，(3.9)の $\xi_n(p_1,p)$ に対応する行の結果，系 3.1，そして非心 χ^2 分布の再生性から得られる．□

A.3　4 節の証明

定理 4.1 の証明． (4.9)と(4.10)の最初の漸近的に成立する等式は，定理 3.3 により直ちに導かれる．また，(4.9)の不等式は分布関数 $F(x;\nu,\lambda)$ が λ の単調減少関数であることと，$(\mathcal{M}_{Gk}\bigcap\mathcal{A}_p)\subset\mathcal{M}_{Gk}$ より導かれる．よって(4.10)の不等式を帰納法を用いて示す．

$q=k+1$ $(k=q-1)$のとき，部分積分と，分布関数 $F(x;\nu,\lambda)$ が λ の単調減少関数であることを用いて，

$$\begin{aligned}
&\Pr\left[\bigcap_{j=k}^{k+1}\left\{\sum_{i=p_1+1}^{j}V(i)\le d_j\right\}\Big|H_0^{(Gk)}\right] \\
&=\int_0^{d_k}p\left(x_k;k-p_1,\sum_{i=p_1+1}^{k}\mu(i)^2\right)\int_0^{d_{k+1}-x_k}p\big(x_{k+1};1,\mu(k+1)^2\big)dx_{k+1}dx_k \\
&\ge\int_0^{d_k}p\left(x_k;k-p_1,\sum_{i=p_1+1}^{k}\mu(i)^2\right)F\big(d_{k+1}-x_k;1,1\big)dx_k \\
&\ge\big[F\big(x_k;k-p_1,k-p_1\big)F\big(d_{k+1}-x_k;1,1\big)\big]_0^{d_k} \\
&\quad-\int_0^{d_k}F\big(x_k;k-p_1,k-p_1\big)\frac{\partial}{\partial x_k}F\big(d_{k+1}-x_k;1,1\big)dx_k \\
&=\int_0^{d_k}p\big(x_k;k-p_1,k-p_1\big)F\big(d_{k+1}-x_k;1,1\big)dx_k=C_k^{k+1}(d_k,d_{k+1})
\end{aligned}$$

が $H_0^{(Gk)}$ 下で成立する．次に $q=k+2$（$k=q-2$）のとき，同様にして，

$$\begin{aligned}
&\Pr\left[\bigcap_{j=k}^{k+2}\left\{\sum_{i=p_1+1}^{j} V(i)\le d_j\right\}\middle| H_0^{(Gk)}\right]\\
&=\int_0^{d_k} p\left(x_k; k-p_1, \sum_{i=p_1+1}^{k}\mu(i)^2\right)\int_0^{d_{k+1}-x_k} p\big(x_{k+1};1,\mu(k+1)^2\big)\\
&\quad\times\int_0^{d_{k+2}-x_{k+1}-x_k} p\big(x_{k+2};1,\mu(k+2)^2\big)dx_{k+2}dx_{k+1}dx_k\\
&\ge\int_0^{d_k} p\left(x_k; k-p_1, \sum_{i=p_1+1}^{k}\mu(i)^2\right)C_{p_1+1}^{p_1+2}(d_{k+1}-x_k, d_{k+2}-x_k)dx_k\\
&\ge\big[F\big(x_k;k-p_1,k-p_1\big)C_{p_1+1}^{p_1+2}(d_{k+1}-x_k,d_{k+2}-x_k)\big]_0^{d_k}\\
&\quad-\int_0^{d_k} F\big(x_k;k-p_1,k-p_1\big)\frac{\partial}{\partial x_k}C_{p_1+1}^{p_1+2}(d_{k+1}-x_k,d_{k+2}-x_k)dx_k\\
&=\int_0^{d_k} p\big(x_k;k-p_1,k-p_1\big)C_{p_1+1}^{p_1+2}(d_{k+1}-x_k,d_{k+2}-x_k)dx_k=C_k^{k+2}(d_k,d_{k+1},d_{k+2})
\end{aligned}$$

が $H_0^{(Gk)}$ 下で成立する．ここで $C_k^q(d_k,d_{k+1},\dots,d_q)$ は，$d_k,d_{k+1},\dots,d_q$ の単調増加関数であることを使った．同じように $k=q-1,q-2,\dots,k+1$ まで成立するとし，一般の k についても，$k=q-2$ のケースと同様にして成立する．□

参 考 文 献

Akaike, H.（1969）. Fitting autoregressive models for prediction, *Annals of the Institute of Statistical Mathematics*, **21**, 243–247.

Akaike, H.（1973）. Information theory and an extention of the maximum likelihood principle, *2nd International Symposium on Information Theory*（eds. B. N. Petrov and F. Csaki）, 267–281, Akademiai Kiado, Budapest.

Box, G. E. P. and Jenkins, G. M.（1976）. *Time Series Analysis*: *Forecasting and Control*, Holden-Day, San Fransisco.

Causey, B. D.（1986）. Expected absolute departure of chi-square from its median, *Communications in Statistics—Simulation and Computation*, **15**, 181–183.

Fuller, W. A.（1996）. *Introduction to Statistical Time Series,* 2nd ed., John Wiley, New York.

Gallant, A. R. and White, H.（1988）. *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Basil Blackwell, New York.

Gourieroux, C. and Monfort, A.（1989）. *Statistique et Modeles Econometriques*, Volume II, Economica, Paris. (Translated by Vuong, Q.（1995）. *Statistics and Econometric Models*, Volume II, Cambridge University Press, New-York.)

Hosoya, Y.（1984）. Information criteria and tests for time-series models, *Time Series Analysis*: *Theory and Practice 5*（ed. O. D. Anderson）, 39–52, North-Holland, Amsterdam.

Hosoya, Y.（1986）. A simultaneous test in the presence of nested alternative hypotheses, *Journal of Applied Probability*, **23A**, 187–200.

Hosoya, Y.（1989）. Hierarchical statistical models and a generalized likelihood ratio test, *Journal of the Royal Statistical Society*, **B51**, 435–447.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Volume 2, Wiley, New York.

Katayama, N. (2006). Asymptotic prediction mean squared error for strongly dependent processes with estimated parameters, Discussion Paper Series, No. 10, Hitotsubashi University Research Unit for Statistical Analysis in Social Sciences, A 21st-century COE program, Institute of Economic Research, Hitotsubashi University, Tokyo.

Mallows, C. L. (1973). Some comments on C_p, *Technometrics*, **15**, 661–675.

Robinson, P. M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression, *Journal of Econometrics*, **47**, 67–84.

Ruben, H. (1974). Non-central chi-square and gamma revisited, *Communications in Statistics*, **3**, 607–633.

Sen, P. K. (1989). The mean-median-mode inequality and noncentral chi square distributions, *Sankhyā, Series A*, **51**, 106–114.

佐和隆光(1979). 『回帰分析』, 朝倉書店, 東京.

下平英寿(2004). 情報量基準によるモデル選択とその信頼性評価, 『統計科学のフロンティア 3』(甘利俊一, 竹内 啓, 竹村彰通, 伊庭幸人 編), 1–76, 岩波書店, 東京.

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, **153**, 12–18.

竹内 啓(2004). 分布の検定とモデルの選択, 『統計科学のフロンティア 3』(甘利俊一, 竹内 啓, 竹村彰通, 伊庭幸人 編), 199–227, 岩波書店, 東京.

On Model-selection Problems in Terms of Prediction Mean Squared Error

Naoya Katayama

Graduate School of Economics, Hitotsubashi University

This paper considers model-selection problems in terms of mean squared error of predictors on linear time series models with a nested structure. It investigates asymptotic relative efficiency of prediction mean squared error (PMSE) of two predictors using estimated parameters, and then proposes new model selection procedures using the generalized likelihood ratio (GLR) test, whose critical regions are decided by percentile points of the noncentral chi-squared variables with the same degrees of freedom and noncentrality parameter. It also explains Akaike's information criterion (AIC) in terms of our testing procedures.

Key words: Generalized likelihood ratio test, AIC, noncentral chi squared distribution, linear time series models, PMSE, multiple tests.