第 9 章　時系列モデル

1　時系列モデルとは何か

確率過程：時点 t において実現する確率変数を y_t として，各時点ごとにこのような確率変数を考えたときの系列 $\{y_t\}$ のこと．時点が整数値の場合を**離散的確率過程**，実数値の場合を**連続的確率過程**という．

モデルの必要性：時系列データは確率過程からの標本の実現値であるが，各時点からは 1 個のデータしか得られない．ところが，確率過程においては期待値や分散は時点ごとに異なりうるから，追加的な情報なり制約がなければ，各時点で 1 個のデータから各時点の平均や分散を推定することは無意味である．したがって，統計的推測の観点からは，確率過程はあまりにも茫漠としすぎており，意味のある統計的推測を行うためには，確率過程を制約した確率モデルを想定する必要がある．

定常性：(i)　$E(y_t)=\mu$　（期待値は時点にかかわらず一定）

(ii)　$Cov(y_t, y_{t-h})=\gamma_h=\gamma_{-h}$　（分散は一定，共分散は時間差のみに依存）

γ_h ：時差 h の**自己共分散**

$\rho_h=\gamma_h/\gamma_0$ ：時差 h の**自己相関**（無名数）， $\rho_h=\rho_{-h}$， $\rho_0=1$

h の関数とみなしたものを**コレログラム**という．

時系列モデル：定常性を仮定することにより未知のパラメータは減少する．実際，平均と分散のパラメータはそれぞれ 1 個となる．しかし，自己共分散のパラメータは依然として多数である．**時系列モデル**は，確率過程の標本が内包するこのような難点を克服して統計的推測を可能にするためのモデルである．

時系列モデルによる分析：時系列モデルの例として，私たちはすでに AR モデルや ARMA モデルなどに出会っている．経済学に限らず，時系列モデルによる分析が広く使われるようになったのは，1970 年に出版されたボックス＝ジェンキンスの著書 Time Series Analysis: Forecasting and Control による所が大きい．1930 年代に起源をもつ時系列分析の方法は，その数学的な難しさもあって，しばらくは一部の専門家の研究対象にとどまっていた．しかし，ボックス＝ジェンキンスは，時系列モデルの特定化（identification），推定（estimation），診断（diagnostic checking），予測（forecasting）という一連の段階的手続きとして時系列分析を定式化した．このようなモデル・ビルディングの発想は，コンピュータの進展という追い風を受けて，すぐにパッケージ化され，多くの利用者を生み出すことになった．

2　さまざまな時系列モデル

AR(p) モデル

確率過程 $\{y_t\}$ の確率モデルが，各時点 t に対して，

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t \quad \Leftrightarrow \quad \phi(L) y_t = \phi_0 + \varepsilon_t \tag{1}$$

と表されるとする．ここで，ε_t は $i.i.d.(0,\sigma^2)$ の誤差項である．また，

$$\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p \tag{2}$$

は，ラグ・オペレータ L のラグ多項式である．

> **(1) の AR(p) モデルが定常となる条件**
> $\phi(x) = 1 - \phi_1 x - \cdots - \phi_p x^p = 0$ の解の絶対値がすべて 1 より大きいこと．

上の方程式 $\phi(x)=0$ を**特性方程式**という．特性方程式が定常条件をみたすならば，平均も分散も一定となる．したがって，$E(y_t)=\mu$ として，(1) の両辺の期待値をとることにより，

$$E(y_t) = \mu = \phi_0 + \phi_1\mu + \cdots + \phi_p\mu \quad \Leftrightarrow \quad E(y_t) = \mu = \phi_0 \big/ \left(1 - \phi_1 - \cdots - \phi_p\right)$$

を得る．このことから，定常な AR(p) モデルは次のようにも表現されることがわかる．

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \cdots + \phi_p(y_{t-p} - \mu) + \varepsilon_t \quad \Leftrightarrow \quad \phi(L)(y_t - \mu) = \varepsilon_t \tag{3}$$

> **定常な AR(1) モデル**　$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t$，　$\{\varepsilon_t\} \sim i.i.d.(0,\sigma^2)$
> (a) 定常性　$-1 < \phi_1 < 1$
> (b) 平均　$\dfrac{\phi_0}{1-\phi_1}$　分散　$\dfrac{\sigma^2}{1-\phi_1^2}$
> (c) 自己相関係数　$\rho_h = \phi_1^{|h|}$
> **定常な AR(2) モデル**　$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$，　$\{\varepsilon_t\} \sim i.i.d.(0,\sigma^2)$
> (a) 定常性　$\phi_2 + \phi_1 < 1$，$\phi_2 - \phi_1 < 1$，$-1 < \phi_2 < 1$
> (b) 平均　$\dfrac{\phi_0}{1-\phi_1-\phi_2}$　分散　$\dfrac{(1-\phi_2)\sigma^2}{(1+\phi_2)\{(1-\phi_2)^2 - \phi_1^2\}}$
> (c) 自己相関係数　$\rho_1 = \dfrac{\phi_1}{1-\phi_2}$，　$\rho_h = \phi_1\rho_{h-1} + \phi_2\rho_{h-2}$　（$h \geq 2$）

AR モデルは単純な構造をしており，定常な AR モデルに関する推測は，本質的に重回帰モデルの場合と異なるところがないことが知られている．このような理由から，AR モデルは，時系列データを描写するモデルとして最も頻繁に使われている．

MA(q) モデル

現在までの異なる時点の誤差の一次結合で表現されるモデル

$$y_t = \theta_0 + \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q} \quad \Leftrightarrow \quad y_t = \theta_0 + \theta(L)\varepsilon_t \tag{5}$$

を考える．ここで，ε_t は $i.i.d.(0,\sigma^2)$ の誤差項であり，

$$\theta(L) = 1 - \theta_1 L - \cdots - \theta_q L^q \tag{6}$$

は，次数 q のラグ多項式である．AR モデルと異なり，MA モデルは常に定常である．

MA モデルの定常性に関しては，モデルのパラメータに何ら制約を必要としないが，モデルの**識別可能性**の観点からは必要である．ここで，定常な確率モデルの識別可能性は，平均および自己共分散が与えられたとき，パラメータが一意的に決められるかどうか，という性質である．

〔**例題** 9.3〕 $\{y_t\}$ が MA(1) モデルに従い，$E(y_t)$，γ_0，γ_1 が与えられたとき，MA(1) モデルの 3 個のパラメータの組（θ_0，θ_1，σ^2）が一意的に決められるかどうかを考察せよ．

（解）今，同一の平均と共分散をもたらすような別のパラメータの組（δ_0，δ_1，ω^2）があるものとすると，次の方程式が成り立つ．

$$E(y_t) = \theta_0 = \delta_0$$
$$\gamma_0 = \sigma^2(1+\theta_1^2) = \omega^2(1+\delta_1^2)$$
$$\gamma_1 = -\theta_1\sigma^2 = -\delta_1\omega^2$$

ここで，第 1 の方程式から，θ_0 は一意的であることがわかる．しかし，第 2，第 3 の方程式は，$\delta_1 = 1/\theta_1$，$\omega^2 = \theta_1^2\sigma^2$ としてもみたされる．すなわち，パラメータの組（θ_0，$1/\theta_1$，$\theta_1^2\sigma^2$）は，（θ_0，θ_1，σ^2）とまったく同一の平均と自己共分散をもたらすことになる．したがって，制約をおかない限り，識別不可能である．

MA(1) モデルを識別可能にする 1 つの方法は，θ_1 の値を絶対値が 1 以下に制約することである．一般の MA(q) モデルの場合には，特性方程式

$$\theta(x) = 1 - \theta_1 x - \cdots - \theta_q x^q = 0$$

の解の絶対値を 1 以上にすることにより識別可能となることが知られている．MA モデルを扱うときは，常にこのような識別可能性を前提とする．

ところで，定常な AR モデルは，無限の次数をもつ MA モデルで表現することが常に可能である．例えば，AR(1) モデルについては，次のようになる．

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t \quad \Leftrightarrow \quad y_t = \frac{\phi_0 + \varepsilon_t}{1-\phi_1 L} = \frac{\phi_0}{1-\phi_1} + \sum_{j=0}^{\infty}\phi_1^j\varepsilon_{t-j}$$

では，MA モデルについても，無限の次数をもつ AR モデルで表すことが常に可能であろうか．答えはノーである．そのためには，AR モデルの定常性と同様に，特性方程式の解の絶対値がすべて 1 より大きいという条件が必要になる．これを**反転可能性**の条件という．例えば，反転可能な MA(1) モデルでは，次のように表現される．

$$y_t = \theta_0 + \varepsilon_t - \theta_1\varepsilon_{t-1} \quad \Leftrightarrow \quad \varepsilon_t = \frac{-\theta_0 + y_t}{1-\theta_1 L} = -\frac{\theta_0}{1-\theta_1} + \sum_{j=0}^{\infty}\theta_1^j y_{t-j}$$

MA(q) モデルにおける識別可能性と反転可能性は，次のようにまとめることができる．

(5) の MA(q) **モデル の識別可能性と反転可能性**

識別可能性：特性方程式 $\theta(x)=1-\theta_1 x-\cdots-\theta_q x^q=0$ の解の絶対値がすべて 1 以上

反転可能性：特性方程式 $\theta(x)=1-\theta_1 x-\cdots-\theta_q x^q=0$ の解の絶対値がすべて 1 より大

識別可能性と反転可能性の違いは，特性方程式が絶対値 1 の解をもつかどうかということである．MA モデルを考える場合には識別可能性を常に前提とするので，反転不可能な MA モデルとは，特性方程式が絶対値 1 の解を含む場合である．例えば，

$$y_t=\theta_0+\varepsilon_t-\varepsilon_{t-1}$$

は反転不可能な MA(1) モデルである．このようなモデルは，y_t が原系列の過剰な階差変換により作られた場合にしばしば生じる．

ARMA モデル

AR モデルと MA モデルを混合させたモデル

$$\phi(L)(y_t-\mu)=\theta(L)\varepsilon_t \tag{7}$$

を考えよう．ここで，$\phi(L)$ は (2) で定義された次数 p のラグ多項式，$\theta(L)$ は (6) で定義された次数 q のラグ多項式である．

(7) の ARMA(p,q) モデルの定常性の条件は，AR モデルと同様に，特性方程式 $\phi(x)=0$ の解の絶対値がすべて 1 より大きくなることである．他方，識別可能性の条件は，MA モデルの場合の条件，すなわち，$\theta(x)=0$ の解の絶対値がすべて 1 以上，に加えて，2 つの特性方程式が共通解をもたない，という条件が必要である．なお，反転可能性の条件は，MA モデルと同じで，$\theta(x)=0$ の解の絶対値がすべて 1 より大きくなることである．

定常な ARMA(p,q) モデルの自己共分散の表現は複雑である．しかし，時差が q を越えると，AR(p) モデルの自己共分散と同じ構造をもつことがわかる．

ARIMA モデル

経済データにおいては，原系列そのものが非定常となっているものが多い．しかし，何回か階差を施したあとの系列は定常になる場合がある．そのような非定常性をもつ時系列に対しては，次のような定義が与えられる．

$I(d)$ **過程の定義**

非定常な時系列 $\{y_t\}$ に対して，d 回の階差変換

$$(1-L)^d y_t=(1-L)^{d-1}\{(1-l)y_t\}$$

がはじめて定常になるとき，$\{y_t\}$ は $I(d)$ 過程に従うという．

ここで，$I(d)$ の I は Integrated（集積的）の頭文字である．$\{y_t\}$ ～ $I(1)$ ならば，

$$(1-L)y_t = u_t \qquad \Leftrightarrow \qquad y_t = y_{t-1} + u_t = y_0 + u_1 + \cdots + u_t$$

の関係がある．すなわち，もとの系列 $\{y_t\}$ は定常な系列 $\{u_t\}$ を集積したものとなっている．このとき，$\{y_t\}$ は $\{u_t\}$ の次数 1 の**和分**である，という．特に，$\{u_t\}$ が $i.i.d.(0,\sigma^2)$ に従うならば，$\{y_t\}$ は y_0 から出発する**ランダム・ウォーク**と呼ばれる．ランダム・ウォークは，AR(1) モデルの特殊形でもある．

いずれにしろ，$I(d)$ 過程に従う系列の d 回の階差変換後の系列は定常となるから，変換後の系列に定常な ARMA モデルをあてはめることができる．すなわち，

$$\phi(L)\left\{(1-L)^d y_t - \mu\right\} = \theta(L)\varepsilon_t \tag{8}$$

のようなモデルを考えることができる．ここで，$\phi(L)$ は (2) で定義された次数 p のラグ多項式，$\theta(L)$ は (6) で定義された次数 q のラグ多項式である．このとき，(8) の $\{y_t\}$ は次数（p, d, q）の **ARIMAモデル**に従うといい，$\{y_t\}$ ～ ARIMA(p, d, q) とかく．ARIMA の I は Integrated の略である．特に，ランダム・ウォークは ARIMA(0,1,0) である．

3　特定化，推定および診断

特定化

定常かつ識別可能な ARMA(p, q) モデルを推定するためには，まず，モデルの次数 p，q を特定化しなければならない．その場合に有用な情報は，データから計算される自己相関である．今，$\{y_t\}$ に関するサイズ n の標本があるとして，

$$\hat{\rho}_h = \frac{\sum_{t=h+1}^{n}(y_t-\bar{y})(y_{t-h}-\bar{y})}{\sum_{t=1}^{n}(y_t-\bar{y})^2} = \frac{\hat{\gamma}_h}{\hat{\gamma}_0} \tag{9}$$

を時差 h の**標本自己相関**といい，h の関数とみて**標本コレログラム**という．ここで，

$$\hat{\gamma}_h = \frac{1}{n}\sum_{t=h+1}^{n}(y_t-\bar{y})(y_{t-h}-\bar{y}) \tag{10}$$

は，時差 h の**標本自己共分散**である．

ところで，相関係数に対して偏相関係数の概念があるように，自己相関に対しては**偏自己相関**が定義される．定常過程 $\{y_t\}$ に対する時差 h の偏自己相関とは，ひき続く $h+1$ 時点の変数 $y_t, y_{t-1}, \cdots, y_{t-h+1}, y_{t-h}$ において，$y_{t-1}, \cdots, y_{t-h+1}$ の影響を差し引いたあとの y_t と y_{t-h} の相関のことをいう．偏自己相関を計算するためには，$\{y_t\}$ の自己相関 $\rho_1, \cdots, \rho_h$ が与えられたとき，未知数 $\phi_{h1}, \cdots, \phi_{hh}$ に関する次の h 本の方程式を考える．

$$\begin{aligned}
&\phi_{h1} + \phi_{h2}\rho_1 + \cdots + \phi_{hh}\rho_{h-1} = \rho_1 \\
&\phi_{h1}\rho_1 + \phi_{h2} + \cdots + \phi_{hh}\rho_{h-2} = \rho_2 \\
&\quad \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot
\end{aligned} \tag{11}$$

$$\phi_{h1}\rho_{h-1}+\phi_{h2}\rho_{h-2}+\cdots+\phi_{hh}=\rho_h$$

このとき，定常過程 $\{y_t\}$ に対する時差 h の偏自己相関とは，h 本からなるこれらの方程式の解 $\phi_{h1},\cdots,\phi_{hh}$ のうち，最後の値 ϕ_{hh} となる．当然ながら，時差 1 の偏自己相関は自己相関と一致して，γ_1/γ_0 であることがわかる．

なお，**標本偏自己相関**は，(11) の方程式の中の自己相関を標本自己相関におきかえたときの解 $\hat{\phi}_{hh}$ である．$\hat{\phi}_{hh}$ は，y_t を $y_{t-1},\cdots,y_{t-h+1}$ に回帰したときの残差と，y_{t-h} を $y_{t-1},\cdots,y_{t-h+1}$ に回帰したときの残差との間の標本相関係数である．

〔**例題** 9.5〕 $\{y_t\}$ が定常な AR(p) モデル

$$y_t=\phi_0+\phi_1 y_{t-1}+\cdots+\phi_p y_{t-p}+\varepsilon_t$$

に従うならば，偏自己相関は時差が p の所で切断が生じることを示せ．

（解）上で定義された定常な AR(p) モデルは，

$$y_t=\phi_0+\phi_1 y_{t-1}+\cdots+\phi_p y_{t-p}+\phi_{p+1}y_{t-p-1}+\cdots+\phi_k y_{t-k}+\varepsilon_t$$

とも表される．ただし，$\phi_{p+1}=\cdots=\phi_k=0$ である．このとき，(4) と同様の k 本からなるユール＝ウォーカー方程式が得られる．特に，$h=p=k$ の場合には，(11) の解は $\phi_{h1}=\phi_1$，$\cdots,\phi_{hh}=\phi_p$ である．したがって，時差 p の偏自己相関は ϕ_p となる．しかし，$h=k>p$ ならば，$\phi_{hh}=0$ でなければならない．したがって，偏自己相関は時差が p の所で切断が生じることになる．

	自己相関	偏自己相関
定常な AR モデル:	切断なし	切断あり
MA モデル：	切断あり	切断なし
定常な ARMA モデル：	切断なし	切断なし

図 9-8(i)（164 ページ）は，**図 9-7(i)**（163 ページ）の対数値系列（原系列は日経平均株価）に対して 1 回の階差をとった系列の時系列プロットである．ここで，原系列を Y_t とすれば，対数変換後の 1 回の階差系列は，

$$\log Y_t-\log Y_{t-1}=\log\frac{Y_t}{Y_{t-1}}=\log\left[1+\frac{Y_t-Y_{t-1}}{Y_{t-1}}\right]$$
$$\cong\frac{Y_t-Y_{t-1}}{Y_{t-1}}$$

と表すことができる．したがって，これを 100 倍したものはパーセント収益率の意味合いをもっている．ただし，厳密には，収益率の方が若干大きめな値となる．

推定

まず，AR(p) モデル

$$y_t=\phi_0+\phi_1 y_{t-1}+\cdots+\phi_p y_{t-p}+\varepsilon_t \qquad (t=p+1,\cdots,n)$$

を考えよう．誤差項 ε_t は $i.i.d.(0,\sigma^2)$ に従うものとする．未知のパラメータは，定

数項 ϕ_0，係数パラメータ $\phi_1,\cdots,\phi_p$，および誤差項の分散 σ^2 である．

ここで，パラメータ $\phi_0,\phi_1,\cdots,\phi_p$ の推定量としては LSE が考えられる．ただし，通常の回帰モデルの場合と異なり，p 期前までのラグを含むので，実質的なデータ数は p 個だけ減少する．LSE を $\hat{\phi}_0,\hat{\phi}_1,\cdots,\hat{\phi}_p$ とすると，具体的には，これらは次の正規方程式の解である．

$$\hat{\phi}_0\sum_{t=p+1}^{n}1+\hat{\phi}_1\sum_{t=p+1}^{n}y_{t-1}+\cdots+\hat{\phi}_p\sum_{t=p+1}^{n}y_{t-p}=\sum_{t=p+1}^{n}y_t$$

$$\hat{\phi}_0\sum_{t=p+1}^{n}y_{t-1}+\hat{\phi}_1\sum_{t=p+1}^{n}y_{t-1}^2+\cdots+\hat{\phi}_p\sum_{t=p+1}^{n}y_{t-1}y_{t-p}=\sum_{t=p+1}^{n}y_{t-1}y_t$$

$$\cdots\cdots$$

$$\hat{\phi}_0\sum_{t=p+1}^{n}y_{t-p}+\hat{\phi}_1\sum_{t=p+1}^{n}y_{t-p}y_{t-1}+\cdots+\hat{\phi}_p\sum_{t=p+1}^{n}y_{t-p}^2=\sum_{t=p+1}^{n}y_{t-p}y_t$$

定常な AR(p) モデルにおいては，誤差項 ε_t が 4 次のモーメントをもつならば，第 7 章で説明した「マルチンゲール差に対する中心極限定理」により，$\hat{\phi}_0,\hat{\phi}_1,\cdots,\hat{\phi}_p$ は漸近的に正規分布に従うことが示される．特に，AR(1) モデルの場合については，第 7 章で示したように，次の結果が成り立つ．

$$\sqrt{n}(\hat{\phi}_0-\phi_0)\ \to\ N\left(0,\sigma^2+(1+\phi_1)\phi_0^2/(1-\phi_1)\right),\qquad \sqrt{n}(\hat{\phi}_1-\phi_1)\ \to\ N\left(0,1-\phi_1^2\right)$$

パラメータ $\phi_0,\phi_1,\cdots,\phi_p$ の推定量としては，LSE のほかに，(4) のユール＝ウォーカー方程式の標本バージョンを使った**ユール＝ウォーカー推定量**を考えることができる．ユール＝ウォーカー推定量は，**モーメント法**と呼ばれる推定方式に基づく推定量である．モーメント法とは，母集団モーメントがみたす方程式において，母集団モーメントを標本モーメントにおきかえてパラメータを推定する方法である．例えば，定常な AR(1) モデルにおいては，モーメントに関する次の 2 つの方程式を考えることができる．

$$E(y_t)=\frac{\phi_0}{1-\phi_1},\qquad\qquad \gamma_1=\phi_1\gamma_0$$

ここで，$E(y_t)$ を標本平均 $\bar{y}$ でおきかえ，γ_0 と γ_1 を (10) で定義された標本自己共分散でおきかえることにより，ϕ_0 と ϕ_1 の推定量を得ることができる．

また，σ^2 は，LSE に基づく残差から，

$$\hat{\sigma}^2=\frac{1}{n-p-1}\sum_{t=p+1}^{n}\left(y_t-\hat{\phi}_0-\hat{\phi}_1y_{t-1}-\cdots-\hat{\phi}_py_{t-p}\right)^2$$

により推定することができる．もちろん，上で述べた LSE 以外の推定量に基づく残差を使った推定量を考えることもできる．これらは，すべて σ^2 の一致推定量となる．

AR モデルの推定に比べて，MA モデルや ARMA モデルの推定は面倒である．その理由は，最小 2 乗法を適用するにしても，最小化する関数がパラメータに関して複雑な非線形関数となるからである．例えば，反転可能な MA(1) モデル

$$y_t=\theta_0+\varepsilon_t-\theta_1\varepsilon_{t-1}\quad\Leftrightarrow\quad \varepsilon_t=-\frac{\theta_0}{1-\theta_1}+\frac{y_t}{1-\theta_1L}$$

のパラメータ θ_0 と θ_1 の推定を考えてみよう．この場合，θ_0 と θ_1 の LSE は，

$$\sum_{t=1}^{n}\varepsilon_t^2=\sum_{t=1}^{n}\left\{-\frac{\theta_0}{1-\theta_1}+\frac{y_t}{1-\theta_1 L}\right\}^2$$

を最小にする値であり，その解を明示的に求めることは不可能である．ただし，統計計算用の多くのコンピュータ・パッケージでは，AR モデルだけでなく，MA や ARMA モデルの推定が可能であり，LSE だけでなく MLE も計算できるようになっている．そして，これらの推定量は漸近的に正規分布に従うことが知られている．例えば，上の MA(1) モデルにおいては，θ_0 と θ_1 の LSE $\hat{\theta}_0$ と $\hat{\theta}_1$ は，MLE と漸近的に同等であり，次の結果が成り立つ．

$$\sqrt{n}(\hat{\theta}_0-\theta_0)\ \to\ N\left(0,(1-\theta_1)^2\sigma^2\right),\qquad \sqrt{n}(\hat{\theta}_1-\theta_1)\ \to\ N\left(0,1-\theta_1^2\right)$$

時系列モデルを推定した場合には，回帰モデルの場合と同様に t 値を報告するのが普通である．パラメータが定常性や反転可能性の条件をみたす限り，t 値は漸近的に標準正規分布と関連付けて解釈することができる．

診断

推定したモデルがどの程度よいのかを調べる最も簡単な方法は，残差のふるまいを見ることである．例えば，AR(p) モデルを LSE で推定したときの残差は，

$$\hat{\varepsilon}_t=y_t-\hat{\phi}_0-\hat{\phi}_1 y_{t-1}-\cdots-\hat{\phi}_p y_{t-p}\qquad (p=t+1,\cdots,n)$$

で定義されるが，残差を時間軸に対してプロットすることにより，特定のパターンがないかどうかをチェックすることができる．パターンがあれば，AR(p) モデルは適切でなく，別のモデルを推定する必要がある．

モデルの診断に使われる統計量としては，**ボックス＝ピアス統計量**

$$Q=m\sum_{j=1}^{K}\hat{\rho}_j^2\left(\hat{\varepsilon}\right)$$

や，その修正バージョンである**リュング＝ボックス統計量**

$$\tilde{Q}=m(m+2)\sum_{j=1}^{K}\frac{\hat{\rho}_j^2\left(\hat{\varepsilon}\right)}{n-K}$$

がある．ここで，m は推定において実際に使われるデータ数，K は推定したモデルに含まれるパラメータの数より大きな整数，$\hat{\rho}_j\left(\hat{\varepsilon}\right)$ は残差から計算される時差 j の標本自己相関である

推定したモデルが ARMA(p,q) の場合，このモデルが適切であるならば，上の 2 つの統計量 Q と $\tilde{Q}$ は，ともに漸近的に自由度 $K-p-q$ の χ^2 分布に従う．モデルが不適切ならば，統計量の値は大きくなる．このことを使って，モデルの適切さに関する検定を行うことができる．

ところで，このような検定による方法とは別に，候補となるモデルの中で，どれがベストかを決めるモデル選択の方法もいくつか提案されている．その中で，ここでは **AIC**（Akaike's Information Criterion：**赤池情報量規準**）についてのみ説明する．今，ARMA(p,q) モデルをあてはめたとき，統計量

$$AIC(p,q) = -2\log \hat{L}(p,q) + 2(p+q)$$

を計算する．ここで，$\hat{L}(p,q)$ は ARMA(p , q) モデルの誤差項に正規分布を仮定して得られる尤度関数の最大値である．このとき，$AIC(p,q)$ を最小にするような p と q の組み合わせからなるモデルが最適である，とするのが AIC によるモデル選択である．

統計量 $AIC(p,q)$ の意味合いを考えると，モデルを複雑にする（＝モデルの次数を高める）にしたがって，第 1 項は減少し，第 2 項は増加する．第 2 項はモデルの複雑さに対するペナルティーであり，両者の和を最小にするモデルを最適なものと考える，というのが AIC の発想である．時系列モデルの分析においては，AIC やその修正バージョンが頻繁に使われている．

〔**例題 9.6**〕 **付録 8** に収録されている日経平均株価の日次データに対して，対数変換したあとに 1 回の階差をとって 100 倍した系列を考える．AIC を使って，この系列は，独立，AR(1)，AR(2)，ラグ 9 の係数 ϕ_9 以外は 0 の特殊な AR(9) モデルのいずれと考えることができるかを診断せよ．

（解）対数変換後に 1 回の階差をとって 100 倍した系列を $\{X_t\}$ とする．このとき，求めるモデルを (3) の形に表現して，$\mu = E(X_t)$, $\phi_1,\cdots,\phi_p$, および誤差項の分散 σ^2 を推定した結果は次の通りとなる．なお，係数推定値の下のカッコ内の値は t 値である．また，それぞれのモデルの AIC も求めてある．

独立：　$\hat{\mu} = 0.0053$,　　$\hat{\sigma}^2 = 1.489$,　AIC=402.2

　　(0.05)

AR(1)：　$\hat{\mu} = 0.0046$，$\hat{\phi}_1 = -0.112$,　　$\hat{\sigma}^2 = 1.482$,　AIC=402.7

　　(0.05)　(-1.24)

AR(2)：　$\hat{\mu} = 0.0046$，$\hat{\phi}_1 = -0.113$，$\hat{\phi}_2 = -0.010$,　$\hat{\sigma}^2 = 1.494$,　AIC=404.7

　　(0.05)　(-1.24)　(-0.11)

特殊な AR(9)：

　$\hat{\mu} = 0.0033$，$\hat{\phi}_9 = 0.206$,　　$\hat{\sigma}^2 = 1.438$,　AIC=399.3

　　(0.02)　(2.26)

AIC の観点からは，ここで考えた 4 つのモデルのうち，ラグ 9 の係数以外は 0 の特殊な AR(9) モデルが最適とされる．ただし，独立な場合や AR(1) の場合との差はわずかである．なお，通常の AR(9) モデルを推定した場合の AIC は，407.8 となる．t 値の観点からは，AR(1) の係数も AR(2) の 2 つの係数も，ともに有意でない．しかし，特殊な
AR(9) の係数は有意である．

この例題で対象とする期間においては，日次の日経平均株価の収益率は，次数 9 の特殊な AR モデルに従う可能性がある，という結論を得た．なお，一般に，株価の系列に関する実証分析においては，収益率は独立，定常とみなすことができる場合が多い．ただし，独立性については再考の余地があり，この点については次の章で考察する．

4　予測

時点 n までの時系列データ $H_n=(y_1,\cdots,y_n)$ が与えられて，h 時点先の値 y_{n+h} を予測する最適な予測量を求めよう．最適性の基準として，予測の平均 2 乗誤差（MSE）を最小にすることを考える．すなわち，

$$MSE\left(y_{n+h}^*|H_n\right)=E\left[\left(y_{n+h}-y_{n+h}^*\right)^2|H_n\right] \tag{12}$$

を最小にする予測量 y_{n+h}^* を求めることを考える．

〔**例題** 9.7〕（12）で定義された予測の MSE を最小にする予測量は，条件付き期待値 $E(y_{n+h}|H_n)$ であることを示せ．

（解）（12）式は次のように変形することができる．

$$\begin{aligned}MSE\left(y_{n+h}^*|H_n\right)&=E\left[\left\{\left(y_{n+h}-E\left(y_{n+h}|H_n\right)\right)-\left(y_{n+h}^*-E\left(y_{n+h}|H_n\right)\right)\right\}^2|H_n\right]\\&=E\left[\left(y_{n+h}-E\left(y_{n+h}|H_n\right)\right)^2|H_n\right]+\left(y_{n+h}^*-E\left(y_{n+h}|H_n\right)\right)^2\\&\quad-2\left(y_{n+h}^*-E\left(y_{n+h}|H_n\right)\right)E\left[y_{n+h}-E\left(y_{n+h}|H_n\right)|H_n\right]\end{aligned}$$

ここで，最後の項の期待値は 0 となることがわかる．このことから，MSE は 2 つの非負の項の和となり，第 2 項を 0 とする予測量が MSE を最小とする．それは，条件付き期待値 $E(y_{n+h}|H_n)$ にほかならない．最適な予測量 $E(y_{n+h}|H_n)$ は，当然のことながら不偏な予測量である．

上の結果から，最適な予測量は条件付き期待値であることがわかったが，条件付き期待値を計算することは一般には困難である．しかし，時系列モデルの 1 つの利点は，その計算の簡便さにあるといえる．例として，定常な AR(1) モデル

$$y_t-\mu=\phi\left(y_{t-1}-\mu\right)+\varepsilon_t,\quad |\phi|<1,\quad \varepsilon_t \sim i.i.d.(0,\sigma^2)$$

に基づく予測を考えよう．このとき，

$$\begin{aligned}y_{n+h}-\mu&=\phi\left(y_{n+h-1}-\mu\right)+\varepsilon_{n+h}\\&=\phi^h\left(y_n-\mu\right)+\phi^{h-1}\varepsilon_{n+1}+\phi^{h-2}\varepsilon_{n+2}+\cdots+\phi\varepsilon_{n+h-1}+\varepsilon_{n+h}\end{aligned}$$

を得る．したがって，最適な予測量は，

$$E(y_{n+h}|H_n)=\mu+\phi^h\left(y_n-\mu\right) \tag{13}$$

となる．現実には，パラメータは未知であるから，これらを推定量でおきかえて，予測量として，

$$\hat{y}_{n+h}=\hat{\mu}+\hat{\phi}^h\left(y_n-\hat{\mu}\right)$$

を計算すればよい．

〔**例題** 9.8〕（13）で与えられた定常な AR(1) モデルの最適な予測量について，予測の間隔 h が大きくなるときのふるまいを調べよ．

（解）まず，最適な予測量の期待値は，

$$E\left(E\left(y_{n+h}|H_n\right)\right)=E\left(\mu+\phi^h\left(y_n-\mu\right)\right)=\mu$$

となり，モデルの期待値 $E(y_t)$ に一致する．また，その分散を求めると，

$$V\left(E\left(y_{n+h}|H_n\right)\right)=V\left(\mu+\phi^h\left(y_n-\mu\right)\right)=\frac{\phi^{2h}\sigma^2}{1-\phi^2}$$

となり，予測の間隔 h が大きくなるとともに 0 に収束する．したがって，最適な予測量は，h が大きくなるとともにモデルの期待値 $E(y_t)$ に確率収束する．他方，予測誤差は，

$$y_{n+h}-E\left(y_{n+h}|H_n\right)=\phi^{h-1}\varepsilon_{n+1}+\phi^{h-2}\varepsilon_{n+2}+\cdots+\phi\varepsilon_{n+h-1}+\varepsilon_{n+h}$$

と表されるから，その分散は，

$$V\left[y_{n+h}-E\left(y_{n+h}|H_n\right)\right]=\sigma^2\sum_{j=0}^{h-1}\phi^{2j}=\frac{1-\phi^{2h}}{1-\phi^2}\sigma^2$$

となる．当然ながら，予測誤差分散は h とともに単調に増加する．そして，最終的にはモデルの分散 $V(y_t)=\sigma^2/(1-\phi^2)$ に近づくことがわかる．

一般の ARMA や ARIMA モデルの場合の予測も，同様に考えることができる．ただし，MA 部分が入っているモデルでは，実際の計算においては，未知のパラメータを推定量でおきかえるだけでなく，時点 n までの誤差も残差でおきかえる必要がある．また，非定常な時系列モデルの予測では，予測の対象は階差変換後の定常な系列ではなく，原系列であることに注意すべきである．

〔**例題 9.9**〕 次の ARIMA(0,1,1) モデルにおける h 時点先の値 y_{n+h} の最適な予測量を求めよ．

$$(1-L)y_t=\mu+\varepsilon_t-\theta\varepsilon_{t-1},\quad y_0=0,\quad |\theta|<1,\quad \varepsilon_t \sim i.i.d.(0,\sigma^2)$$

（解）上の ARIMA(0,1,1) モデルは，時点 $n+h$ において次のように表現される．

$$y_{n+h}=y_{n+h-1}+\mu+\varepsilon_{n+h}-\theta\varepsilon_{n+h-1}$$

これより，最適な予測量は，

$$h=1\ :\quad E\left(y_{n+1}|H_n\right)=E\left(y_n+\mu+\varepsilon_{n+1}-\theta\varepsilon_n|H_n\right)=y_n+\mu-\theta\varepsilon_n$$

$$\begin{aligned}h>1\ :\quad E\left(y_{n+h}|H_n\right)&=E\left(y_{n+h-1}+\mu+\varepsilon_{n+h}-\theta\varepsilon_{n+h-1}|H_n\right)=E\left(y_{n+h-1}|H_n\right)+\mu\\&=E\left(y_{n+1}|H_n\right)+(h-1)\mu=y_n+h\mu-\theta\varepsilon_n\end{aligned}$$

となる．ところで，$v_t=\varepsilon_t-\theta\varepsilon_{t-1}$ とおくと，

$$\begin{aligned}y_{n+h}&=y_n+h\mu+v_{n+1}+v_{n+2}+\cdots+v_{n+h}\\&=(n+h)\mu+v_1+\cdots+v_{n+h}\end{aligned}$$

とも表すことができるので，

$$E\left(y_{n+h}\right)=E\left(E\left(y_{n+h}|H_n\right)\right)=(n+h)\mu$$

$$V\left(y_{n+h}\right)=V\left(v_1+\cdots+v_{n+h}\right)=\sigma^2\left(\theta^2+1+\left(1-\theta\right)^2(n+h-1)\right)$$

を得る．このことから，予測量は予測の間隔 h が大きくなると直線的に発散していくことがわかる．他方，予測誤差は，$h>1$ のとき，

$$
\begin{aligned}
y_{n+h} - E\left(y_{n+h}|H_n\right) &= \theta\varepsilon_n + v_{n+1} + \cdots + v_{n+h} \\
&= (1-\theta)\left(\varepsilon_{n+1} + \cdots + \varepsilon_{n+h-1}\right) + \varepsilon_{n+h}
\end{aligned}
$$

となることから，その分散は，

$$
V\left[y_{n+h} - E\left(y_{n+h}|H_n\right)\right] = \sigma^2\left(1 + (1-\theta)^2 (h-1)\right)
$$

である．したがって，定常な場合と異なり，予測誤差は h とともに発散していくことになる．このことは，非定常な系列では，あまり先の予測は無意味であることを示唆している．