

令和2年度 中級計量経済学
講義ノート6 定常時系列解析

このノートでは簡単な一変量の時系列モデルとその分析方法を紹介する。まず、時間を通じた相関を表す系列相関とそれを調べる統計的手法を紹介し、時系列の代表的なモデルである自己回帰移動平均モデルとその統計的分析法を解説する。また、自己回帰モデルを用いた予測法にも触れる。

6.1 時系列データ

時系列データとは、各年の我が国のGDPデータのように、時間を追って観測されるデータである。経済において主たるものはGDPを始め、金利や貿易その他のマクロデータである。また、近年はデータの収集、記録、保存の技術進歩に伴って、株価や為替レートの高頻度データの蓄積が進んでおり、これらも時系列データである。

$\{y_t\}_{t=1}^T$ を時系列データとする。これは、 $t = 1$ から $t = T$ 時点までの y_t の値を表す。時系列データでは、 y_t と y_{t-j} は相関しているかもしれない。そのような異時点間の相関を系列相関あるいは自己相関という。時系列分析は、その動学的性質（時間を通じた変動に関する性質）を分析の対象とする。

y_{t-j} を y_t の j 次のラグといい、それらの共分散 $cov(y_t, y_{t-j})$ を自己共分散という。

定常性 同時分布が時間を通じて変わらないことを（強）定常性という。

例えば、 y_t と y_{t-j} の同時分布が t に依存しないということである（通常、 j には依存する）。特に、定常な時系列においては、

- $E(y_t) = \mu$ (期待値が t に依存しない)。
- $var(y_t) = \gamma_0$ (分散が t に依存しない)。
- $cov(y_t, y_{t-j}) = \gamma_j$ (共分散が t に依存しない)。なお、 γ_j を j 次の自己共分散という。

が成り立ち、このように平均と自己共分散が j のみに依存し t に依存していない場合、弱定常または共分散定常であるという。定常なら弱定常であるが、逆は必ずしも成り立たない。また、平均が0で1次以上の自己共分散がすべて0の弱定常な時系列をホワイトノイズという。

定常性が満たされない例で、経済学上重要なものは、

- トレンドがある場合（期待値が時間と共に変化する）
- 確率的トレンドがある場合（単位根など）
- 構造変化がある場合（ある時点で期待値や分散、パラメータの値が変化する）

などである。定常性がない時系列を非定常時系列という。非定常時系列の分析は、この講義では取り扱わないが、経済時系列の解析においては非常に重要なトピックである。

6.2 自己共分散と自己相関係数

$\{y_t\}$ を共分散定常な時系列であるとき、自己共分散と自己相関係数は変数の系列相関を表現する基本的なパラメーターである。

$\gamma_j = cov(y_t, y_{t-j})$ を j 次の自己共分散として、

j 次の自己相関係数は、

$$\rho_j = \frac{\text{Cov}(y_t, y_{t-j})}{\sqrt{\text{var}(y_t)}\sqrt{\text{var}(y_{t-j})}} = \frac{\gamma_j}{\gamma_0} \quad (1)$$

である。

自己相関を調べることは時系列解析の第一歩である。横軸に j をとり、縦軸に ρ_j をとるグラフをコレログラムという。

推定 自己共分散は、標本自己共分散

$$\hat{\gamma}_j = \frac{1}{T} \sum_{t=j+1}^T (y_t - \bar{y})(y_{t-j} - \bar{y}) \quad (2)$$

によって推定できる。ここで、 $\bar{y} = \sum_{t=1}^T y_t / T$ である。

また自己相関係数は、標本自己相関係数

$$\hat{\rho}_j = \frac{\hat{\gamma}_j}{\hat{\gamma}_0} \quad (3)$$

によって推定できる。

これらの推定量は、ある条件のもとで、一致性と漸近正規性を持つ。しかし、その証明、漸近分散の式はかなり複雑なので、ここでは取り扱わない。

一方で、もし、 y_t が i.i.d. であるなら、 $\rho_0 = 1$ 、他の自己相関係数はすべて 0 であり、また

$$\sqrt{T}\hat{\rho}_j \rightarrow_d N(0, 1) \quad (4)$$

となる。推定された自己相関係数でコレログラムを描くとき、合わせて $\pm 1/\sqrt{T}$ や $\pm 1.96/\sqrt{T}$ の横線を書くことがよくある。それは、上の結果を用いて帰無仮説 $\rho_j = 0$ を検定する際の棄却域を示している。すなわち、 $\pm 1.96/\sqrt{T}$ の線を超えているなら、系列相関があると結論付ける。

系列相関の検定 上の方法によって、各 j に対して $\rho_j = 0$ を検定することが可能である。一方、最初のいくつかの相関係数をまとめて系列相関の有無を検定することもできる。標準的に使われるのは以下の二つの検定統計量である。

$$\text{Box-Pierce 検定統計量: } Q = T \sum_{k=1}^p \hat{\rho}_k^2, \quad (5)$$

$$\text{Ljung-Box 検定統計量: } Q = T(T+2) \sum_{k=1}^p \frac{\hat{\rho}_k^2}{T-k}, \quad (6)$$

なお、 p は研究者が適当に選ぶ。これらの二つの検定統計量は、もし y_t が i.i.d. (従ってすべての自己相関係数が 0) であるという帰無仮説のもとで、 χ_p^2 の分布を持つ。他方、もし p 次までの自己相関係数のうちどれかが 0 でないなら、 $T \rightarrow \infty$ と共に検定統計量 Q の値が大きくなって発散するという性質を使って検定するものである。

これらの検定は、いくつかの自己相関をまとめて検定するという意味で、ふるしき検定やかばん検定と呼ばれる。

6.3 自己回帰移動平均モデル

時系列構造を表現するために最もよく使われるモデルは、自己回帰移動平均モデル (Autoregressive and moving average model, ARMA model) である。まず自己回帰の部分について述べ、次に移動平均を説明し、最後にそれらを合わせた ARMA モデルを紹介する。

自己回帰モデル (Autoregressive model) 現在の値が過去の値とホワイトノイズから決まるモデルを自己回帰モデルという。その最も単純なものが下の 1 次の自己回帰モデル (AR(1) モデル) で、過去一期分のみに依存している場合である。

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + u_t \quad (7)$$

ここで、 α_0 と α_1 はパラメーターであり、 $u_t \sim i.i.d.(0, \sigma^2)$ は誤差項で、時系列分析ではノイズ、イノベーション等とも呼ばれる。

これは、分散均一を仮定した回帰モデルと見ることができる。分散均一は強い仮定かもしれないが、理論分析を簡単にする。分散不均一を表現するものとして ARCH (Autoregressive Conditional Heteroscedasticity) や GARCH (generalized ARCH) 等のモデルが提案されており、マクロ時系列分析や金融時系列分析において広く使われている。これらは少し難しいためこの講義では取り扱わない。

AR(1) のラグ次数を一般化した

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + u_t \quad (8)$$

を AR(p) モデルという。

定常性の条件 AR モデルで表現できる時系列が弱定常性を持つための条件は、

$$1 - \alpha_1 x - \alpha_2 x^2 - \dots - \alpha_p x^p = 0 \quad (9)$$

という方程式の解の絶対値がすべて 1 より大きいことである。

AR(1) の場合、定常性の条件は、

$$|\alpha_1| < 1 \quad (10)$$

である。

AR(2) の場合は、 $\alpha_1 + \alpha_2 < 1$, $\alpha_2 - \alpha_1 < 1$, $-1 < \alpha_2 < 1$ となる。一般の AR(p) モデルの場合に定常性の条件を求めることも可能だが、その条件は複雑な式になる。

自己回帰モデルの平均と分散 定常な AR(1) モデルの平均と分散は次のようにして求められる。 $E(y_t) = \mu$ とする。

AR(1) の式の両辺の期待値をとると、

$$E(y_t) = \alpha_0 + \alpha_1 E(y_{t-1}) + E(u_t) \quad (11)$$

となるが、 $E(y_t) = E(y_{t-1}) = \mu$ かつ $E(u_t) = 0$ なので

$$\mu = \alpha_0 + \alpha_1 \mu \quad (12)$$

となり、期待値

$$\mu = \frac{\alpha_0}{1 - \alpha_1} \quad (13)$$

が得られる。

AR(1) の式から (12) を引くと

$$y_t - \mu = \alpha_1(y_{t-1} - \mu) + u_t \quad (14)$$

となる。この両辺に u_t をかけて期待値をとると、

$$E\{(y_t - \mu)u_t\} = \alpha_1 E\{(y_{t-1} - \mu)u_t\} + E(u_t^2) = \sigma^2 \quad (15)$$

が得られる。

次に (14) 式の両辺に $(y_t - \mu)$ をかけて期待値をとると、

$$E(y_t - \mu)^2 = \alpha_1 E\{(y_{t-1} - \mu)(y_t - \mu)\} + E\{u_t(y_t - \mu)\} \quad (16)$$

となり、(15) 式の結果と γ_j の定義から

$$\gamma_0 = \alpha_1 \gamma_1 + \sigma^2 \quad (17)$$

が得られる。さらに、(14) 式に $(y_{t-1} - \mu)$ をかけて期待値をとると、

$$E\{(y_t - \mu)(y_{t-1} - \mu)\} = \alpha_1 E(y_{t-1} - \mu)^2 + E\{u_t(y_{t-1} - \mu)\} \quad (18)$$

となる。 t 期の誤差 u_t は 1 期前の y_{t-1} と相関を持たない（新しい誤差項はそれ以前の y の値とは関係ない）と仮定することは自然なので、右辺の第二項は 0 である。

$$\gamma_1 = \alpha_1 \gamma_0 \quad (19)$$

となる。(17), (19) の結果より、

$$\gamma_0 = \frac{\sigma^2}{1 - \alpha_1^2}, \quad (20)$$

$$\gamma_1 = \frac{\alpha_1 \sigma^2}{1 - \alpha_1^2} \quad (21)$$

となる。同様に、(14) 式の両辺に $(y_{t-j} - \mu)$ をかけて期待値をとり、逐次代入すると

$$\gamma_j = \alpha_1^j \gamma_0 = \frac{\alpha_1^j \sigma^2}{1 - \alpha_1^2} \quad (22)$$

であることを示すことができる。

AR(p) モデルの場合も同じように平均と自己共分散を計算することができる。

また AR(1) モデルの自己相関係数は、

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \alpha_1^j \quad (23)$$

である。

なお、上で述べたように y_t が定常のとき $|\alpha_1| < 1$ なので、仮に $t-1$ 期に y_{t-1} が μ からかなり離れた値をとっても、次の期の y_t は平均的に μ に近づく方向に力が働くことが (14) 式からわかるだろう。そのため、 y_t の振る舞いは安定的である（とんでもなく離れた値に飛んでいってしまっても戻ってこないというような振る舞いはしない）。もし $|\alpha_1| = 1$ や $|\alpha_1| > 1$ であれば、そのように期待値の方向に戻ろうとする力が働かず、振る舞いが安定的ではないことがわかるだろう。 $|\alpha_1| = 1$ のとき、 y_t は単位根 (unit root) をもつといい、 $|\alpha_1| > 1$ のとき y_t は爆発的 (explosive) であるという。

シミュレーションによって発生させた AR(1) のデータでそれらの例を見ておこう。本ノート末尾には、図 1～4 の 4 枚の図が描かれている。横軸は時間、縦軸は $y_t, t = 1, \dots, T$ の値である。 y_t を作るために、まず $N(0,1)$ から 200 個の乱数を発生させてそれを $u_t, t = 1, \dots, 200$ とした。それを用いて、定数項はすべて 0 だが、係数パラメータ α_1 が $-0.8, 0.5, 1.0, 1.1$ という 4 種類の AR(1) のデータを発生させた。それを図示したのが、図 1～4 である。図 1～3 では $T = 200$ 、図 4 では $T = 200$ まで描くと値が発散してしまうため、 $T = 50$ で止めてある。図 1 では、かなり上下動が激しいことがわかるだろう。それは、係数の値がマイナスなので、每期データの符号が入れ替わりやすいからである。他方、図 2 では図 1 に比べて穏やかな動きである。またいずれも期待値である 0 からあまり大きく離れていかず、安定した挙動である。他方、図 3 (係数が 1) では一旦大きく 0 から離れてしまっている。その後、また 0 方向に戻っているように見えるが、更に続けていけば、また大きく 0からはずれていく。その外れ方も、プラス方向かマイナス方向か偶然に左右される。図 4 は、ほぼ単調にマイナス方向に発散している。プラスでなくマイナス方向に発散しているのは偶然である。発生させた乱数の値が初めのころにマイナスの値が多く、それに影響されたためであり、もし初めの頃の乱数にプラスの値の影響が大きければ、プラス無限大に発散していく。どちらの方向に動くかは偶然であるが、一旦方向が定まると急激に発散する。

移動平均モデル (Moving average model) 以下のように、現在の値が現在と過去 q 期のイノベーションに依存して決まるモデルを移動平均モデル (MA(q) モデル) という。

$$y_t = \theta_0 + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} \quad (24)$$

ここで、 $\epsilon_t \sim i.i.d.(0, \sigma^2)$ とする。

移動平均モデルが常に弱定常であることは、次のように簡単に示すことができる。

MA(q) モデルに従う y_t の期待値は、

$$\mu = E(y_t) = E(\theta_0 + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q}) = \theta_0 \quad (25)$$

であり、分散は、

$$\gamma_0 = E(y_t - \theta_0)^2 = E(\epsilon_t^2) + \theta_1^2 E(\epsilon_{t-1}^2) + \dots + \theta_q^2 E(\epsilon_{t-q}^2) = (1 + \theta_1^2 + \dots + \theta_q^2) \sigma^2 \quad (26)$$

である。また 1 次の自己共分散は、

$$\gamma_1 = E(y_t y_{t-1}) - \mu^2 = (-\theta_1 + \theta_1 \theta_2 + \dots + \theta_{q-1} \theta_q) \sigma^2 \quad (27)$$

のようになる。なお、 $(q+1)$ 次以上の自己共分散は 0 である。従っていずれも t に依存しない。

MA モデルで注意しないといけないのは、識別性の問題である。AR モデルの場合は、モデルのパラメータ (係数とイノベーションの分散) が与えられれば自己共分散の系列が決まり、逆に自己共分散の系列を決めるとモデルのパラメータが一意に決まる。この意味で、AR モデルのパラ

メータは識別性をもつという。それに対して、MA モデルの場合はそうならない。例えば、MA(1) モデルについて計算すると、自己共分散は、

$$\gamma_0 = (1 + \theta_1^2)\sigma^2, \quad \gamma_1 = -\theta_1\sigma^2 \quad (28)$$

それ以上の自己共分散は 0 となる。逆に、 γ_0, γ_1 を定めたときに、ある θ_1^* と σ^{*2} の組み合わせが (28) 式を満たすとき、 $1/\theta_1^*$ と $\theta_1^{*2}\sigma^{*2}$ の組み合わせも同じく (28) 式を満たすことが分かる。(つまり (28) を θ_1 と σ^2 について解くと解が二つある) それゆえ、データから推定できる自己共分散が得られても、MA モデルのパラメータの値を一意に決めることができない。しかし、 $|\theta_1| \leq 1$ とすると、唯一に決めることができる。通常は、この識別性の条件を仮定する。一般に、

$$1 - \theta_1x - \theta_2x^2 - \dots - \theta_qx^q = 0 \quad (29)$$

の方程式の解の絶対値がすべて 1 以上 (1 を含む) であれば、識別可能であることが知られている。

AR モデルと MA モデルの関係 定常な AR モデルは、MA(∞) モデルで表現することができる。例えば、AR(1) モデルは、

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + u_t \quad (30)$$

$$= \alpha_0 + \alpha_1(\alpha_0 + \alpha_1 y_{t-2} + u_{t-1}) + u_t \quad (31)$$

$$= \alpha_0 + \alpha_1\alpha_0 + \alpha_1^2(\alpha_0 + \alpha_1 y_{t-3} + u_{t-2}) + u_t + \alpha_1 u_{t-1} \quad (32)$$

$$\dots \quad (33)$$

$$= \frac{\alpha_0}{1 - \alpha_1} + \sum_{j=0}^{\infty} \alpha_1^j u_{t-j} \quad (34)$$

となる。

また追加的な条件を置けば、MA モデルも AR(∞) モデルで表現できる。その条件を、反転可能性の条件という。それは、

$$1 - \theta_1x - \theta_2x^2 - \dots - \theta_qx^q = 0 \quad (35)$$

の方程式の解の絶対値がすべて 1 より大きい (1 を含まない) ということである。識別可能性との違いは、解の絶対値が 1 である場合が許されるかどうかである。例えば、MA(1) のとき、 $|\theta_1| < 1$ のとき反転可能であり、

$$y_t = \frac{\theta_0}{1 - \theta_1} - \sum_{j=1}^{\infty} \theta_1^j y_{t-j} + \epsilon_t \quad (36)$$

となる。MA(1) で $|\theta_1| = 1$ のとき、識別可能ではあるが、反転可能ではない。

自己回帰移動平均モデル AR モデルと MA モデルを組み合わせたものを自己回帰移動平均モデル (ARMA(p, q) モデル) という。 p, q はそれぞれ、自己回帰部分と移動平均部分のラグの次数である。モデルは、

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} \quad (37)$$

とかける。 $\epsilon_t \sim i.i.d.(0, \sigma^2)$ と仮定する。

定常性の条件は、AR モデルと同じで、方程式

$$1 - \alpha_1 x - \alpha_2 x^2 - \dots - \alpha_p x^p = 0 \quad (38)$$

の解の絶対値がすべて 1 より大きいことである。また、反転可能性の条件は、MA モデルと同じで、

$$1 - \theta_1 x - \theta_2 x^2 - \dots - \theta_q x^q = 0 \quad (39)$$

の解の絶対値がすべて 1 より大きいことである。

6.4 自己回帰モデルの推定

AR モデルは、OLS で推定できる。

つまり、定数項を含む AR(p) モデルは、

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + u_t \quad (40)$$

であるが、この係数 $\alpha = (\alpha_0, \dots, \alpha_p)'$ の推定量は、 $x_t = (1, y_{t-1}, \dots, y_{t-p})'$ と表記すると、 α の OLS 推定量は

$$\hat{\alpha} = \left(\sum_{t=p+1}^T x_t x_t' \right)^{-1} \sum_{t=p+1}^T x_t y_t \quad (41)$$

である。時系列が定常で誤差項が $\epsilon_t \sim i.i.d.(0, \sigma^2)$ なら、 $\hat{\alpha}$ は一貫性と漸近正規性をもつ推定量である。ただし、通常の線形回帰の場合と異なり、不偏性は持たない。これらの証明はこの講義では取り扱わない。

なお MA モデルや ARMA モデルについては、通常イノベーションに正規分布を仮定した最尤法を使って推定する。複雑なのでここでは取り扱わないが、EViews 等の統計ソフトではプログラムが実装されており、簡単に推定できる。

ラグの選び方 AR モデルでは、どこまでの次数のラグをモデルに含めるかによって、推定結果や（以下で扱う）予測が変わってしまう。経済モデルからラグの長さが決まる場合もあるが、通常はラグの長さは先験的に明らかではない。

良く使用されるラグの選び方は、情報量基準によるものである。代表的な情報基準としては、赤池情報量基準 (AIC; Akaike Information Criterion) やベイズ情報量基準 (BIC; Bayes Information Criterion) 等がある。

AIC は、

$$AIC(p) = \log \left(\frac{\sum_{t=p+1}^T \hat{u}_t(p)^2}{T} \right) + (p+1) \frac{2}{T} \quad (42)$$

である。ここで、 $\hat{\alpha}_j, j = 1, \dots, p$ を AR(p) モデルをあてはめた時の係数の OLS 推定量、 $\hat{u}_t(p) = y_t - \hat{\alpha}_0 - \hat{\alpha}_1 y_{t-1} - \dots - \hat{\alpha}_p y_{t-p}$ はその残差である。AIC の第一項は残差二乗和を変形したもので、フィットの良さを測る指標で、小さいほどフィットが良い。しかし、線形回帰分析の章で説明したとおり、説明変数の数（ここでは定数項も含めて $p+1$ 個）を増やしていけば自動的にフィットは良くなる。ここでいえば、 p を大きくする、つまり長いラグを取った AR(p) を使うことで第一項は小さくなってしまふ。そこで、第二項を加えることによって、説明変数のを増やし過ぎに対す

る罰則（ペナルティ）を与えている（第二項は p を大きくすると大きくなる）。色々な p に対して推定を行って、各 p に対して AIC を計算し、それが最小化になる p が選ばれる。

また BIC は

$$BIC(p) = \log \left(\frac{\sum_{t=p+1}^T \hat{u}_t(p)^2}{T} \right) + (p+1) \frac{\log T}{T} \quad (43)$$

である。AIC と同じように、BIC と最小化する p を選ぶ。AIC と BIC では第二項の T に関する部分のみが異なっている。

AIC と BIC のどちらを使用するかは状況や目的による。もし、真のモデルが AR(p) モデルの場合は、BIC で p を一貫性をもって推定できる。つまり、 $T \rightarrow \infty$ のとき BIC を使えば正しい p が選ばれることが知られている。他方、AIC は一般にそれよりも大きめの p を選びやすい。

ここでは、 T で割っているが、Ng and Perron (2000) によると、 $T - p_{\max}$ で割るのが適切であろうとのことである。 p_{\max} は、考慮する最大の p の値である。ただしこの場合、 p_{\max} を前もって決める必要がある。

AIC や BIC の厳密な定義は、使用する統計ソフトによって違うことがあるので、注意が必要である。

6.5 予測

次に、AR モデルを使用して、 h 期先の y の値を予測する方法を解説する。

まず、AR(1) から生成される T 期までのデータあって、そのパラメータの値がわかっているものとしよう。逐次代入によって、

$$y_{T+h} = \alpha_0 + \alpha_1 y_{T+h-1} + u_{T+h} \quad (44)$$

$$= \alpha_0 + \alpha_0 \alpha_1 + \alpha_1^2 y_{T+h-2} + u_{T+h} + \alpha_1 u_{T+h-1} \quad (45)$$

$$\dots \quad (46)$$

$$= \frac{(1 - \alpha_1^h) \alpha_0}{1 - \alpha_1} + \alpha_1^h y_T + \sum_{j=0}^{h-1} \alpha_1^j u_{T+h-j} \quad (47)$$

である。ここで、 $E_T(\cdot)$ を T 期までの情報での条件付き期待値とすると、 $E_T(u_{T+1}) = \dots = E_T(u_{T+h}) = 0$ なので

$$E_T(y_{T+h}) = \frac{(1 - \alpha_1^h) \alpha_0}{1 - \alpha_1} + \alpha_1^h y_T \quad (48)$$

となる。これを y_{T+h} の T 期までの情報に基づいた予測値として使うことができる。

この予測誤差の分散は、

$$\text{var}_T(y_{T+h}) = \sigma^2 \sum_{j=0}^{h-1} \alpha_1^{2j} = \sigma^2 \frac{1 - \alpha_1^{2h}}{1 - \alpha_1^2} \quad (49)$$

である。

なお、 $h \rightarrow \infty$ とすると、予測値は、

$$\frac{\alpha_0}{1 - \alpha_1} \quad (50)$$

に近づき、それは y_t の（条件のない）期待値に一致する。また予測誤差の分散は、

$$\sigma^2 \frac{1}{1 - \alpha_1^2} \quad (51)$$

に近づき、同じく y_t の（条件のない）分散に一致する。つまり、はるか先の y_t の値を予測する際には、それまでに観察された y_t の値は役に立たず、単に、 y_t の期待値を使用することと同じになる。

一般に AR(p) モデルの場合も同じように議論できる。

パラメーターの値がわかっていない場合には、パラメータを推定値で置き換え、

$$\frac{(1 - \hat{\alpha}_1^h)\hat{\alpha}_0}{1 - \hat{\alpha}_1} + \hat{\alpha}_1^h y_T \quad (52)$$

で予測を行うことが考えられる。その場合、予測誤差にはパラメータの推定誤差も考慮する必要がある。しかし、 T が十分に大きい場合には、一致性により、推定誤差の予測誤差への影響は軽微であることが分かっているので、予測誤差の分散を単に

$$\hat{\sigma}^2 \frac{1 - \hat{\alpha}_1^{2h}}{1 - \hat{\alpha}_1^2} \quad (53)$$

として推定することもよく行われる。ただし、 $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=2}^T (y_t - \hat{\alpha}_0 - \hat{\alpha}_1 y_{t-1})^2$ とする。

MA モデルや ARMA モデルの場合も同様の議論が成り立つが、予測には、 ϵ_t の値を推定する必要があり、若干面倒になる。