ARモデルを用いた株価の予測

2017826 高橋竜眞
片岡ゼミ
令和二年度提出

要約

株価はダウ理論におけるトレンドを確認することで、ある程度の予測をすることができる。これをモデルを使った予測を用いることで、トレンドの把握をより簡単にしようとするのが今回の目的である。使用するモデルは、ARモデル、VARモデルである。どちらも自身の過去のデータで回帰する回帰モデルである。2章ではそれぞれのモデルについての説明、3章では実際にモデルにデータを入れてモデルがどのような動きをするか見る。また、データの表現の仕方を変えることによっての予測精度が変化するのかを観察した。詳しい話は3章になるが、モデルは今回使用するデータでは、1期後を有意に予測をすることができることがわかった。これを利用すると前営業日に次の営業日の終値が主要高値を超えるかどうかの判断が可能になる。統計モデルと、従来のテクニカル分析を合わせることが有用であるかもしれない。

今後の展望として、統計モデルの不確かさについて、どのようにカバーするのかといった点である。例えば、統計モデルの区間推定の下側の値が主要高値を超えるとき、どれほどの収益があるかといったことを調べる。これは区間推定の幅とトレンドのチェック方法を組み合わせて実際に資産の増減を記録するといった実験をすることで確認することができるだろう。統計モデルの不確かさをテクニカル分析でカバーできれば、モデルのある程度の予測誤差は許容できるかもしれない。

高橋竜眞　2017826　R2 12.27

目次

1.　はじめに

2020/02/20を境に日経平均株価が急激な下落を見せた。コロナショックである。2万3000円付近を推移していた株価は一時的に1万8000円を割り込んだ株価は、日銀の金融政策により3か月ほどで元の水準まで戻った。しかし株価に対して実態経済がコロナショック以前の水準に回復しているかと言われればそうではない。事実2020年6月調べの[1]日銀短観によれば、大、中堅、中小企業のほとんどの産業は先行きに不安を覚えている。[2]また図1の通り有効求人倍率は右肩下がりである。これらの指標は、現在の就職先がなくなる可能性や、再就職先の減少という問題を示唆している。また金融庁が老後資金は2000万円必要だと試算したことは記憶に新しい。以上のことから、今までの時代よりも、より個人がどのように生きるかが重要になってきていると考えている。

そこで私は資産形成に興味を持っていたので、在学中に実際に株の取引を行った。テクニカル分析を通して、株価の予測の難しさを体感した。例えばダウ理論という投資の基本の考え方がある。[3]新井邦宏[2003, 33p]によると、

“上昇トレンド=次の主要な高値は前の主要な高値より高く、次の主要な安値は前の主要な安値より高い
下落トレンド=次の主要な安値は前の主要な安値より安く、次の主要な高値は前の主要な高値より安い”

とある。ダウ理論とは相場におけるトレンドの定義と同義である。図2に上記ダウ理論を表した。テクニカル分析とは、トレンドが上昇トレンドであるか、または下落トレンドであるのか、そしてそれらが継続するのか反転するのかの確認作業である。そこで機械学習によって。1期先、2期先の予測が可能となれば、トレンドの把握もより容易になるのではないかと考えた。時系列モデルを使い株価の予測がどこまで可能かについて調べた。2章で、時系列モデルのARモデルとVARモデルについての説明、3章は実際に株価データをモデルに入れてシミュレーションをした。

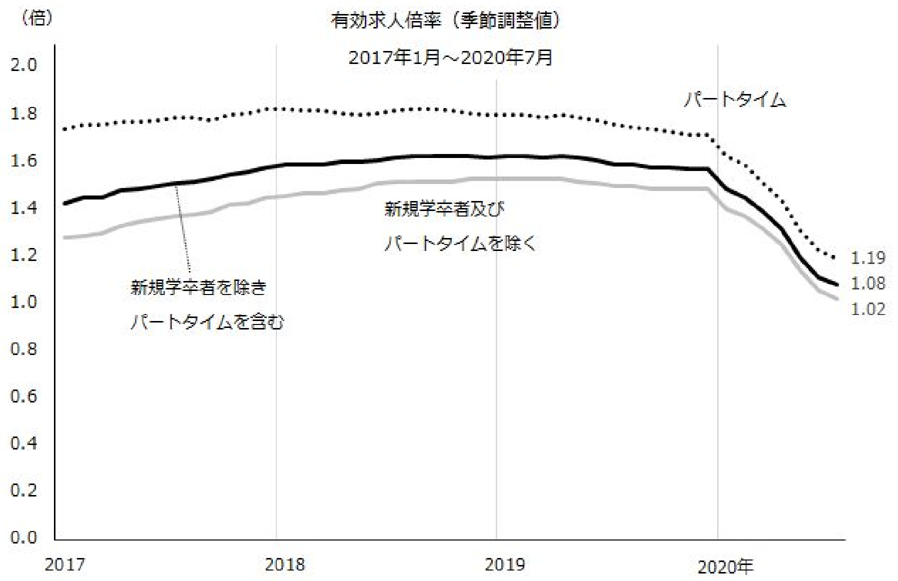


図1 独立行政法人　労働政策研究所・研修機構 国内統計：有効求人倍率　新型コロナウイルス感染症関連情報：新型コロナが雇用・就業・失業に与える影響

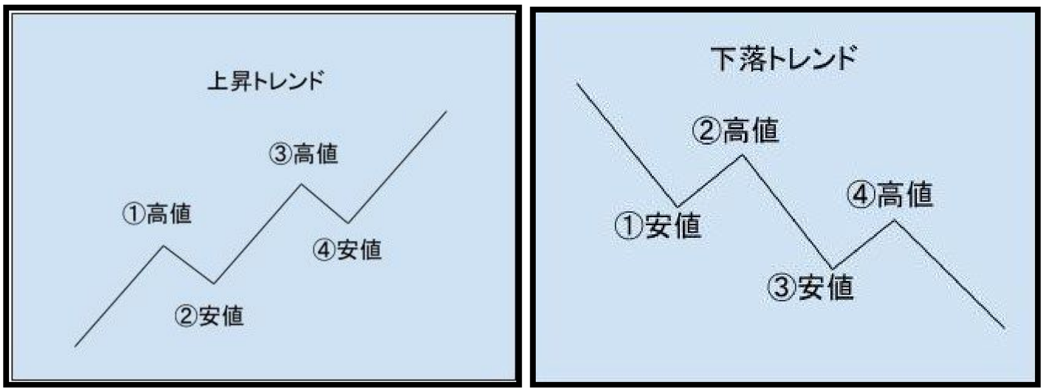


図2　上昇トレンドと下落トレンド　新井邦宏[2003]を参考に自作

2.　時系列モデル

時系列データとは時間の経過とともに観測されるデータであり、観測される順序に重要な意味がある。また、時系列モデルを扱うにあたり、時系列データが定常であるどうかも重要なポイントである。今回扱うモデルに入る前に定常性について少しまとめておく。

定常性には、弱定常性と強定常性が存在している。[4]弱定常性は、期待値が一定かつ自己共分散が時点によらず時間差によってのみ決まる過程である。これに対して、強定常性は同時分布が同一になることを求める。しかし、経済・ファイナンスの分野において、単に定常性というと、弱定常性を指すことが多い。なぜなら強定常性は非常に強い仮定であり、経済・ファイナンスの議論においてそこまで強い仮定は必要としていないからである。従って特に断りがなく定常性というときは、弱定常性を指すものとする。

さて、今回扱うモデルは自己回帰モデルとベクトル自己回帰モデルである。モデルについての詳しいまとめは次節以降に行う。簡単な説明を加えておくと、自己回帰モデルとは、過去の値で回帰分析をしたもので、ベクトル自己回帰モデルはそれを多変量に拡張したものである。

2.1.　自己回帰モデル

まずはじめにモデルの定義から始める。自己回帰(AR)モデル（autoregressive process）は、自身の過去の値に回帰された形で表現されるモデルである。p次AR(p)モデルは以下のように定義される。

次数pのARモデルAR(p)の時刻tにおける、状態y_t は

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \qquad \varepsilon_t \sim W.N(\sigma^2)$$

と計算される。但し、c, ϕ_iはモデルのパラメータで、ε_tは期待値0分散σ^2のホワイトノイズである。

AR(1)は以下のように初期値を設定すると逐次計算で求めることができる。

$y_1 = c + \phi_1 y_0 + \varepsilon_1$
$y_2 = c + \phi_1 y_1 + \varepsilon_2$
$y_3 = c + \phi_1 y_2 + \varepsilon_3$ ……

y_0は初期値といってyの確率分布が定まっているならば、その分布からの値を与えてもよい。この場合y_0は確率変数である。基本的に$y_0=0$として計算している文献が多いため、$y_0=0$でもよい。ただ初期値の影響は時間とともになくなっていくため大きな問題にはならない。

次に自己相関ついてである。時系列データを分析するにあたって、自己相関構造を把握することは重要である。AR(1)の自己相関は次のように表される。

$$\rho_k = \phi \rho_{k-1}$$

k次自己相関係数は差分方程式で表され、この方程式をユールウォーカー方程式といいAR(1)の場合は簡単に解ける。$\rho_0=1$として、

$$\rho_k = \phi \rho_{k-1} = \phi(\phi \rho_{k-2}) = \phi(\phi(\phi \rho_{k-3})) = \cdots = \phi(\phi(\phi(\ldots(\phi \rho_1)))) = \phi^k$$

AR(1)の定常条件は、$|\phi|<1$であるので、自己相関係数が指数的に減衰することがわかる。AR(p)は常に定常であるわけではないが、定常であると仮定すると、AR(p)の自己相関係数も同様に指数的に減衰する。

最適予測について触れる。平均二乗誤差(MSE; mean squared error)の点で最適予測は条件付き期待値で与えられることが知られている。ε が正規ホワイトノイズという仮定の下では以下の二つの結果を用いればAR過程の条件付き期待値を求めることができる。

$$E(y_\tau|\Omega_t) = y_\tau \quad (\tau \leqq t) \qquad (1.1)$$
$$E(\varepsilon_{t+k}|\Omega_t) = 0 \quad (k > 0) \qquad (1.2)$$

ここで、$\Omega_t=(y_t,y_{t-1},,,y_3,y_2,y_1)$という時点tにおいて使用可能な観測値からなる情報集合を表す。(1.2)は過去のy、つまり情報集合に含まれるyについては条件付き期待値実際の値をそのまま使うことを示すものである。(1.2)は将来の ε の期待値が0であることを表す。これは、ε が独立系列であれば成立するものである。但し一般的にホワイトノイズはiid系列ではないので、ホワイトノイズだけを仮定すると(1.2)は成り立たないことに注意しよう。では実際に以下の、

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t \quad \varepsilon_t \sim \text{iid } N(\sigma^2)$$

AR(1)を用いて計算してみる。

まず1期先予測である。y_{t+1}を過去のyと将来の ε で表すと、

$$y_{t+1} = c + \phi_1 y_t + \varepsilon_{t+1}$$

となる。(1.2),(1.3)を用いて、1期先最適予測は、

$$\hat{y}_{t+1|t} = c + \phi_1 y_t$$

となり、この時与えられる予測誤差とそのMSEは、

$$\hat{e}_{t+1|t} = y_{t+1} - \hat{y}_{t+1|t} = \varepsilon_{t+1}$$

$$MSE(\hat{y}_{t+1|t}) = E(\varepsilon^2_{t+1}) = \sigma^2$$

となることがわかる。同様に2期先予測を考える。y_{t+2}を過去のyと将来の ε で表すと、

$$\begin{aligned} y_{t+2} &= c + \phi_1 y_{t+1} + \varepsilon_{t+2} \qquad (1.3) \\ &= c + \phi_1(c + \phi_1 y_t + \varepsilon_{t+1}) + \varepsilon_{t+2} \\ &= (1 + \phi_1)c + \phi^2_1 y_t + \phi_1 \varepsilon_{t+1} + \varepsilon_{t+2} \end{aligned}$$

となる。2期先最適予測とMSEは、

$$\hat{y}_{t+2|t} = (1 + \phi_1)c + \phi^2_1 y_t \qquad (1.4)$$

$$MSE(\hat{y}_{t+2|t}) = E((\phi_1 \varepsilon_{t+1} + \varepsilon_{t+2})^2) = (1 + \phi^2_1)\sigma^2$$

となる。ここでϕ^2_1は非負であるため、MSE($\hat{y}_{t+1|t}$) < MSE($\hat{y}_{t+2|t}$)が成り立つ。一般にh期先予測を考えると、y_{t+h}は上と同様に、

$$\begin{aligned}
y_{t+h} &= c + \phi_1 y_{t+h-1} + \varepsilon_{t+h} \\
&= (1 + \phi_1 + \phi^2_1 + \cdots + \phi^{h-1}_1)c + \phi^h_1 y_t \\
&\quad + \phi_1^{h-1}\varepsilon_{t+1} + \phi_1^{h-2}\varepsilon_{t+2} + \cdots + \phi_1 \varepsilon_{t+h-1} + \varepsilon_{t+h} \\
&= c\sum_{k=1}^{h} \phi_1^{k-1} + \phi^h_1 y_t + \sum_{k=0}^{h} \phi_1^k \varepsilon_{t+h-k}
\end{aligned}$$

と書き直すことができるので、h期先最適予測とMSEはそれぞれ、

$$\begin{aligned}
\hat{y}_{t+h|t} &= c\sum_{k=1}^{h} \phi_1^{k-1} + \phi^h_1 y_t \\
&= \frac{(1-\varphi 1)c}{1-\varphi 1} + \phi^h_1 y_t \qquad (1.5)
\end{aligned}$$

$$\begin{aligned}
MSE(\hat{y}_{t+h|t}) &= E((\sum_{k=0}^{h-1} \phi_1^k \varepsilon_{t+h-k})^2) = \sigma^2 \sum_{k=0}^{h-1} \phi_1^{2k} \\
&= (1 - \phi_1^{2h})\sigma^2(1 - \phi^2_1)^{-1} \qquad (1.6)
\end{aligned}$$

となる。補足であるが、(1.3)のy_{t+1}を$\hat{y}_{t+1|t}$ に入れ替えて計算してみると、

$$\begin{aligned}
y_{t+2} &= c + \phi_1 \hat{y}_{t+1|t} + \varepsilon_{t+2} \\
&= c + \phi_1(c + \phi_1 y_t) + \varepsilon_{t+2} \\
&= (1 + \phi_1)c + \phi^2_1 y_t + \varepsilon_{t+2}
\end{aligned}$$

これの2期先最適予測は、

$$\hat{y}_{t+2|t} = (1 + \phi_1)c + \phi^2_1 y_t$$

となり、(1.4)と一致する。これは予測に必要な将来のyは逐次予測値で置き換えて計算していいことを示す。しかし、この場合MSEは予測が過去のyのみを用いて表していないため計算が困難になる。だが、逐次予測を用いればARモデルの次数が大きくなった時にも計算が楽になる。

さて、AR(1)の最適予測の性質について話を戻そう。まず、最適予測は予測期間にかかわらずy_tにのみ依存するといった性質がある。これはAR(1)が過去1期までしか含まないことからも容易に理解できる。次に予測期間が長くなっていく時の最適予測についてみていく。過程の定常条件は$|\phi|<1$であるので、予測期間が長くなっていくにつれてy_tの影響が指数的に減衰((1.6) 第二項)していくことがわかる。その結果、最適予測は過程の期待値へと収束していく。これを平均回帰的(mean reverting)といわれる性質である。最後にMSEについて考える。MSEは予測期間が長くなるにつれて単調に増加する。これは予測期間が長くなるほど予測が難しくなっていくことを示している。$\phi_1^{2h}\rightarrow 0$より、MSEは過程の分散に近づいていくことがわかる。すなわち、MSEの上限は過程の分散である。

下記の図はAR(1)を用いて日経平均株価を予測したものを重ねてプロットしたものである。ピンク色の垂線は予測時点0から5つごとにひいてある。予測期間が長くなるにつれて予測が放射状に広がっていることがわかる。これはAR(1)が過去に一期分しか情報を持たないことに起因する。拡大した図をみるとわかるのだが、5期間離れただけで予測がかなりずれるのである。

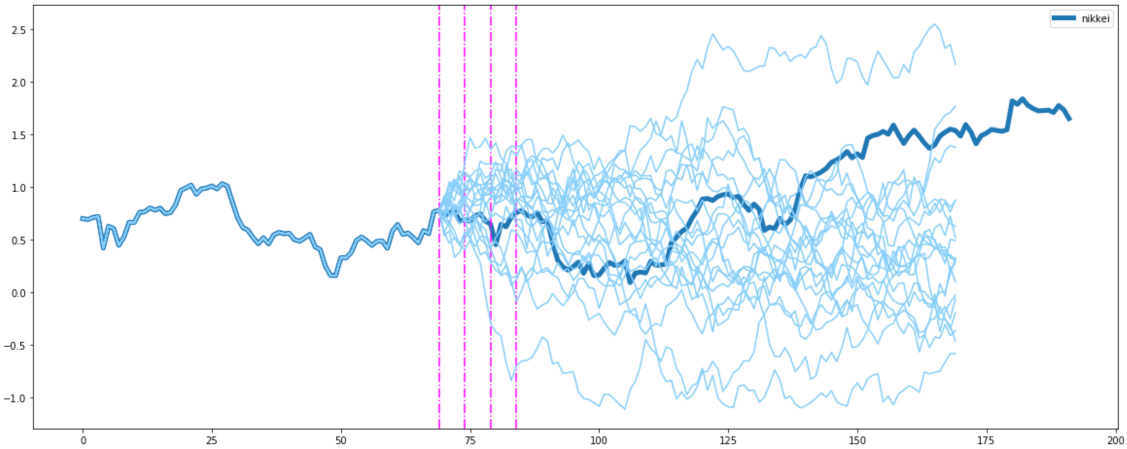


図3　日経平均株価　AR(1)　予測

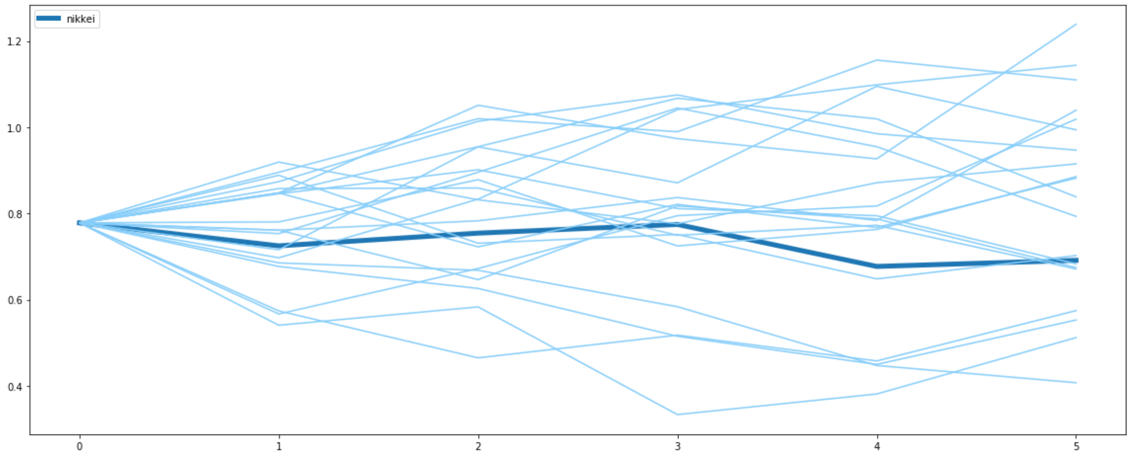


図4　図の予測の開始時点付近を拡大したもの

2.2.　ベクトル自己回帰モデル

ベクトル自己回帰モデル(VARモデル)は前章で扱ったARモデルを多変数に拡張したものである。VARモデルは、変数間の関係の分析に対して、グレンジャー因果性分析、インパルス応答関数、分散分解といったツールを提供するが、今回それらは扱わず、単にARモデルの拡張として見たものがどれほど予測精度の向上につながるかを考える。VARモデルの定義に入る前に期待値、自己共分散など、一変数の時に使用したものを多変数に拡張しておく。[4]

$n \times 1$の列ベクトル$\mathbf{y}$は次のようになる。

$$\mathbf{y}_t = (y_{1t,} y_{2t ,,,} y_{nt})'$$

ベクトルyの期待値は、各成分の期待値を集めベクトルの表示にしたものである。

$$E(\mathbf{y}_t) = (E(y_{1t}), E(y_{2t}) ,,, E(y_{nt}))'$$

自己共分散を多変量に拡張すると、行列になる。k次自己共分散行列は、(i, j)成分がy_{it}と $y_{j,t-k}$の共分散に等しい、n×n行列である。

$$Cov(y_t, y_{t-k}) = \begin{bmatrix} Cov_{(y_{1t}, y_{1,t-k})} & Cov_{(y_{1t}, y_{2,t-k})} & \cdots & Cov_{(y_{1t}, y_{n,t-k})} \\ Cov_{(y_{2t}, y_{1,t-k})} & Cov_{(y_{2t}, y_{2,t-k})} & \cdots & Cov_{(y_{2t}, y_{n,t-k})} \\ \vdots & \vdots & \ddots & \vdots \\ Cov_{(y_{nt}, y_{1,t-k})} & Cov_{(y_{nt}, y_{2,t-k})} & \cdots & Cov_{(y_{nt}, y_{n,t-k})} \end{bmatrix}$$

最後にホワイトノイズもここで多変数に拡張しておく。ε はベクトルホワイトノイズと言われる。自己相関を持たないことは、定義式から明らかではある。ここで、Σ はn×nの正定値行列であり、対角行列ではないことに注意しよう。異時点でノイズは相関を持たないが、同時点で各成分は相関を持ってよいのである。

$$\mathrm{E}(\epsilon_\mathrm{t}) = \mathbf{0}$$
$$\mathrm{E}(\epsilon_\mathrm{t}, \epsilon_\mathrm{t-k}^\mathrm{T}) = \begin{cases} \mathbf{\Sigma}, & k = 0 \\ \mathbf{0}, & k \neq 0 \end{cases}$$

次はVARモデルの定義をしていく。

---

n変数VAR(p)の時刻tにおける、状態$\mathbf{y}_t$ は

$$y_t = c + \mathbf{\Phi_1} y_{t-1} + \mathbf{\Phi_2} y_{t-2} + \cdots + \mathbf{\Phi}_p y_{t-p} + \epsilon_t, \quad \epsilon_t \sim \mathbf{W.N(\Sigma)}$$

ここで、cはn×1の列ベクトルで、$\mathbf{\Phi}_i$はn×nの係数行列である。

ARモデルと同様にVARモデルも常にモデルが定常であるわけではない。ここで定常条件について少し触れておく。VAR(p)の定常条件は、

$$\left| I_n - \Phi_1 z - \Phi_2 z^2 - \cdots - \Phi_p z^p \right| = 0$$

で与えらえれる、AR特性方程式の全ての解の絶対値が1より大きいことである。但し、Iは単位行列である。ここで、VAR(1)について、定常条件を考えてみる。AR特性方程式は次のようになる。

$$\left|I_n - \Phi_1 z\right| = 0$$

$$\left|z^{-1} I_n - \Phi_1\right| = 0$$

これは、係数行列の固有値を求めていることに等しい。|z|>1となるためには、係数行列の固有値の全てが1より大きくなくてはいけないのである。また、単位行列を1とすることでARモデルの定常条件になる。AR(1)を考えると、係数行列はスカラーになるので、|z|>1を満たすのは|Φ|<1ということになる。
VARモデルでも最適予測についても簡単に触れておこう。考え方はARモデルの時と同じである。εを多変量正規ホワイトノイズと仮定すると、式(1.1),(1.2)は次のように表せる。

$$E\left(y_\tau | \Omega_t\right) = y_\tau \qquad (\tau \leq t) \tag{1.7}$$

$$E\left(\epsilon_{t+k} | \Omega_t\right) = 0 \qquad (k > 0) \tag{1.8}$$

ここで、y, ε はそれぞれ、先ほど定義した列ベクトルであることに注意しよう。ARモデルの時と同様にVAR(1)を用いて計算してみる。

$$y_t = c + \Phi_1 y_{t-1} + \epsilon_t \qquad \epsilon_t \sim \mathrm{W.N.}(\Sigma)$$

1期先予測を過去のyと将来のノイズで表すと、

$$y_{t+1} = c + \Phi_1 y_t + \epsilon_{t+1}$$

となる。(1.7),(1.8)を用いて1期先最適予測は、

$$\hat{y}_{t+1|t} = c + \Phi_1 y_t$$

となり、この時与えられる予測誤差とそのMSEは、

$$\hat{\epsilon}_{t+1|t} = y_{t+1} - \hat{y}_{t+1|t} = \epsilon_{t+1}$$

$$MSE\left(\hat{y}_{t+1|t}\right) = E\left(\epsilon_{t+1}\epsilon_{t+1}^T\right) = \Sigma$$

となることがわかる。
※MSEを対角成分としていいか？区間推定に対角成分を使うとしか、論文には書いてない
εは同時点では各変数との相関を許すことに注意する。同様に2期先予測を考える。2期先予測を過去のyと将来のεで表すと、

$$\begin{aligned} y_{t+2} &= c + \Phi_1 y_{t+1} + \epsilon_{t+2} \\ &= c + \Phi_1\left(c + \Phi_1 y_t + \epsilon_{t+1}\right) + \epsilon_{t+2} \\ &= (\mathbf{1} + \Phi_1)c + \Phi_1^2 y_t + \Phi_1 \epsilon_{t+1} + \epsilon_{t+2} \end{aligned}$$

となる。2期先最適予測とMSEは、

$$
\begin{aligned}
\hat{y}_{t+2|t} &= (\mathbf{1} + \Phi_1)\, c + \Phi_1^2 y_t \\
MSE\left(\hat{y}_{t+2|t}\right) &= E\left((\Phi_1 \epsilon_{t+1} + \epsilon_{t+2})(\Phi_1 \epsilon_{t+1} + \epsilon_{t+2})^T\right) \\
&= \Sigma + \Phi_1 \Sigma \Phi_1^T
\end{aligned}
$$

となる。一般にh期先最適予測を考えると、

$$
\begin{aligned}
y_{t+h} &= c + \Phi_1 y_{t+h-1} + \epsilon_{t+h} \\
&= (\mathbf{1} + \Phi_1 + \Phi_1^2 + \cdots + \Phi_1^{h-1})c + \Phi_1^h y_t \\
&+ \Phi_1^{h-1} \epsilon_{t+1} + \Phi_1^{h-2} \epsilon_{t+2} + \cdots \\
&\cdots + \Phi_1 \epsilon_{t+h-1} + \epsilon_{t+h} \\
&= c \sum_{k=1}^{h} \Phi_1^{k-1} + \Phi_1^h y_t + \sum_{k=0}^{h} \Phi_1^k \epsilon_{t+h-k}
\end{aligned}
$$

$$
\begin{aligned}
\hat{y}_{t+h|t} &= c \sum_{k=1}^{h} \Phi_1^{k-1} + \Phi_1 y_t \\
&= (\mathbf{1} - \Phi_1) c (\mathbf{1} - \Phi_1)^{-1} + \Phi_1^h y_t
\end{aligned}
$$

$$
\begin{aligned}
MSE(\hat{y}) &= E\Big(\big(\sum_{k=0}^{h-1} \Phi_1^k \epsilon_{t+h-k}\big)^2\Big) \\
&= \Sigma + \Phi_1 \Sigma \Phi_1^T + \Phi_1^2 \Sigma (\Phi_1^2)^T + \cdots + \Phi_1^{h-1} \Sigma (\Phi_1^{h-1})^T \\
&= \sum_{k=0}^{h-1} \Phi_1^k \Sigma (\Phi_1^k)^T
\end{aligned}
$$

となる。多変数の場合も将来のMSEを計算するのは困難になるが、1変数の時と同様に、逐次予測値で置き換えて計算することができる。

3.　シュミレーション

実際にモデルにデータを入れて予測の精度を見ていく。データは日経平均組み入れ銘柄の中から10種類、日経平均株価自体をいれた計11種類を使用する。後述するが、どのデータもADF検定の帰無仮説を棄却できず、単位根過程ではないという仮説を指示するデータが今のところない、という結果になっている。今回使用する両モデルは定常性を過程においているため、データをそのまま使うと何らかの不具合が生じることがわかっている。そのため、シュミレーションの内容としては下記のような加工したデータと原データの予測精度の違いについて見ていく。

原データ = { x0 ,x1, x2, x3 }

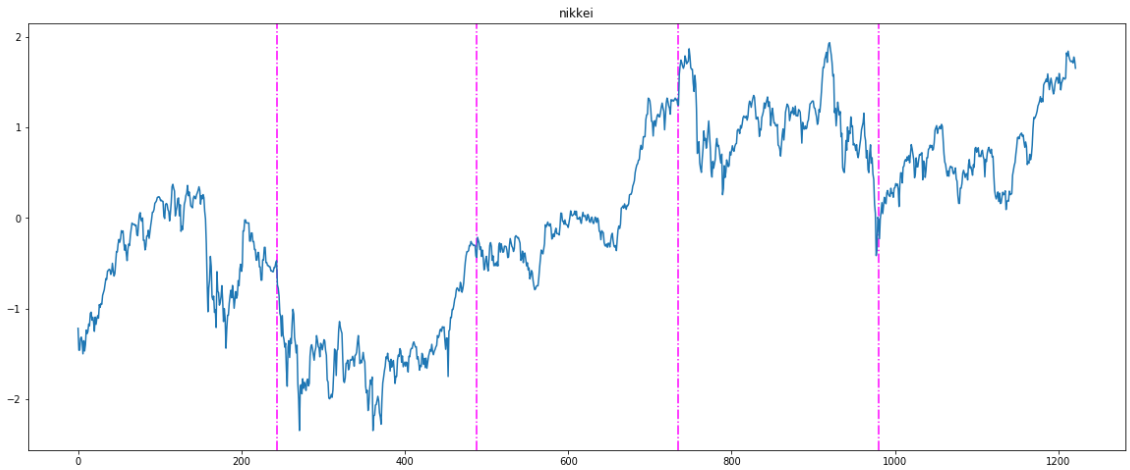
差分系列	復元
y1 = x1 − x0	x1 = y1 + x0
y2 = x2 − x1	x2 = y2 + x1
y3 = x3 − x2	x3 = y3 + x2
y4 → prediction	x4 = y4 + x3 (new)

1階差とることによって差分データを作り、その差分データで予測したものを原データに復元することで疑似的に予測を行う。これによりモデルに適用するデータは定常性を持つので、モデルの仮定を満たすことになる。

3.1.　データ

使用するデータは、日経平均株価、テルモ、花王、コマツ、トヨタ自動車、資生堂、マツダ、日立製作所、サイバーエージェント、キャノン、楽天の11種類。日経平均株価はヤフーファイナンスから、それ以外のデータは株式投資メモ株価DBからダウンロードした。データの期間は2015/01/05～2019/12/30のに日次データを用いる。なお、コロナショックの影響を除くため直近のデータは入れてない。日次データを用いる理由は、突然の災害などの不確定要素の排除するためである。トレンドに反した突然の下落や高騰はモデルで表現することは厳しい。以上のことから、今回のシュミレーションには日次データを採用した。

ではまずそれぞれの株価データをプロットしていく。データは全て標準化をしたものを使用し、グラフは年度ごとに垂線を引いてある。



高橋竜眞　2017826　R2 12.27

図5　日経平均株価　標準化データ

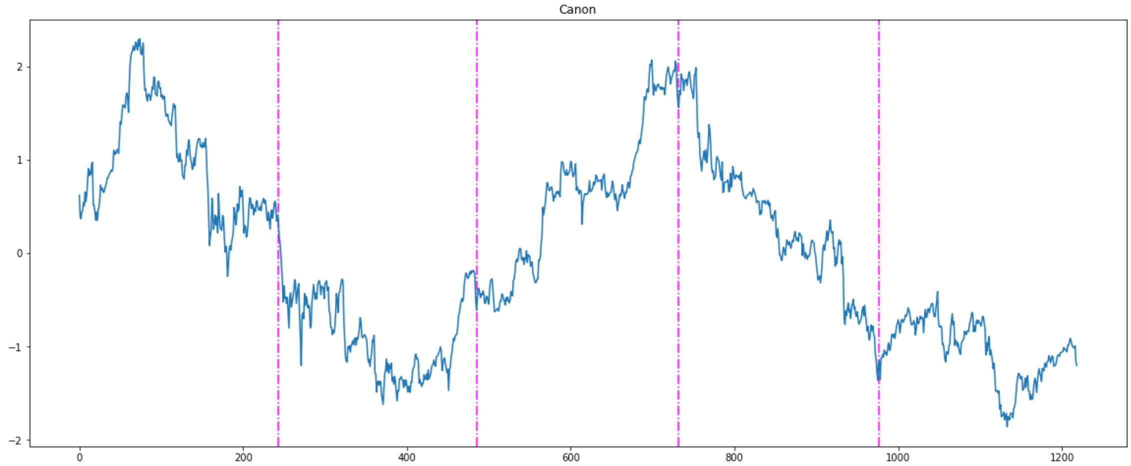


図6　キャノン　標準化データ

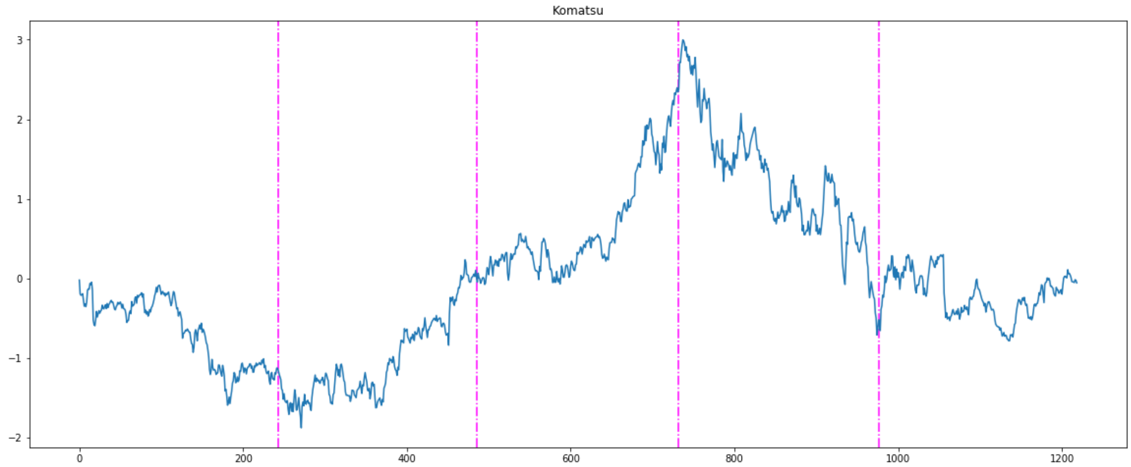


図7　コマツ　標準化データ

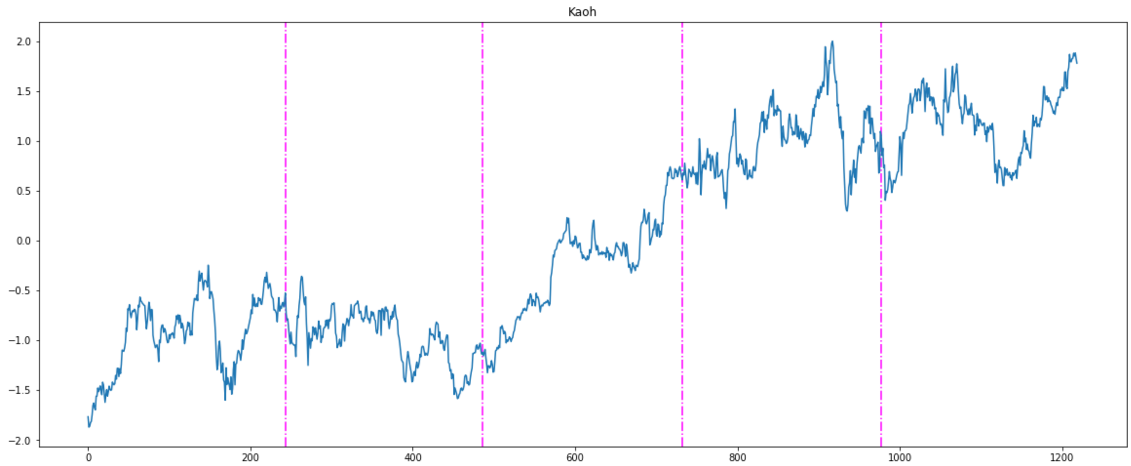


図8　花王　標準化データ

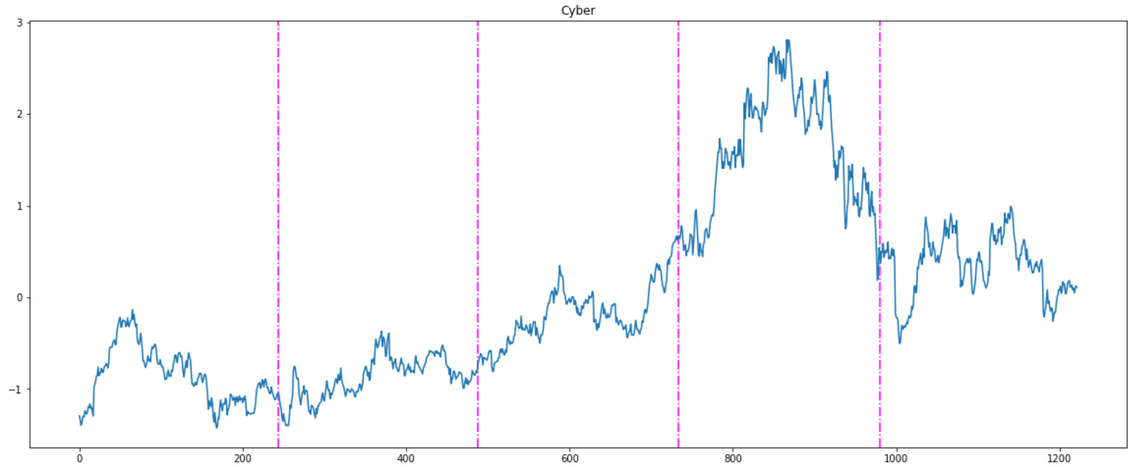


図9　サイバーエージェント　標準化

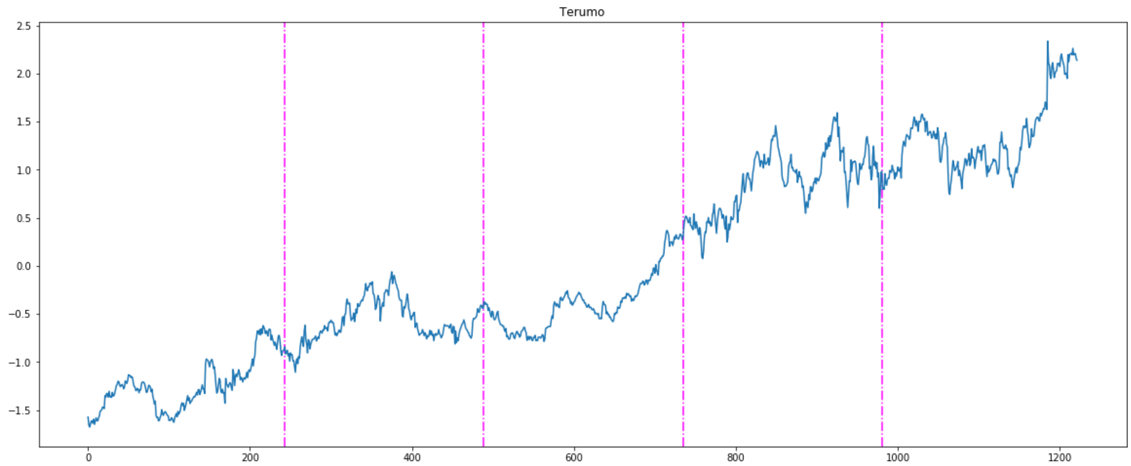


図10　テルモ　標準化データ

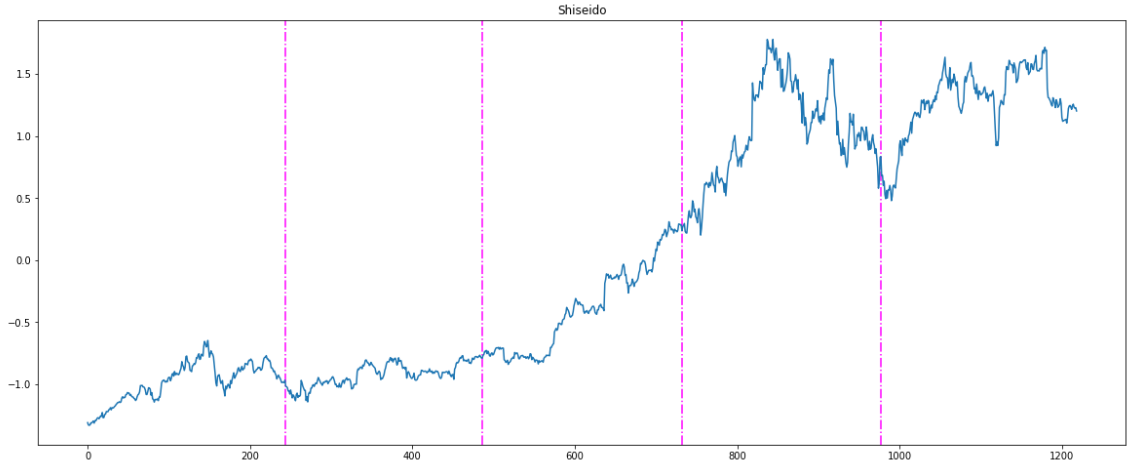


図11　資生堂　標準化データ

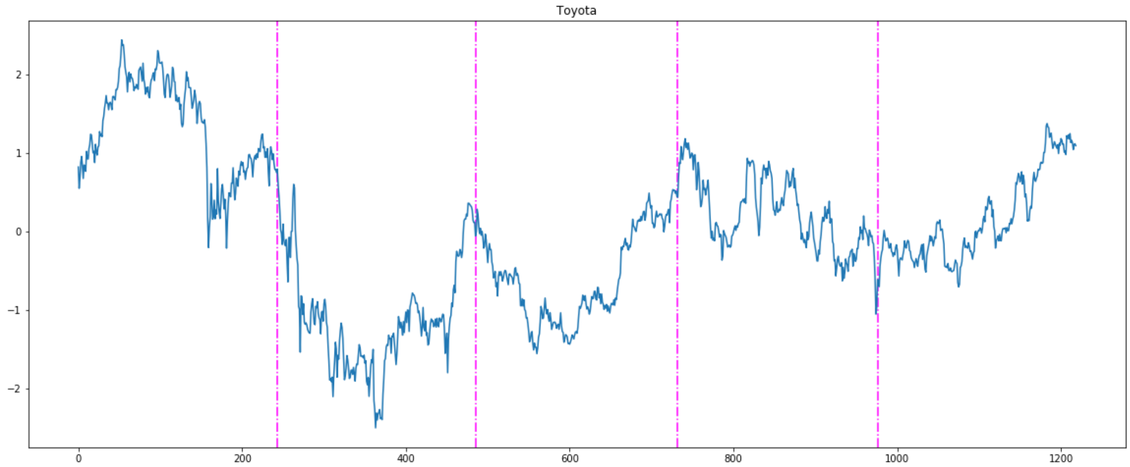


図12 トヨタ自動車 標準化データ

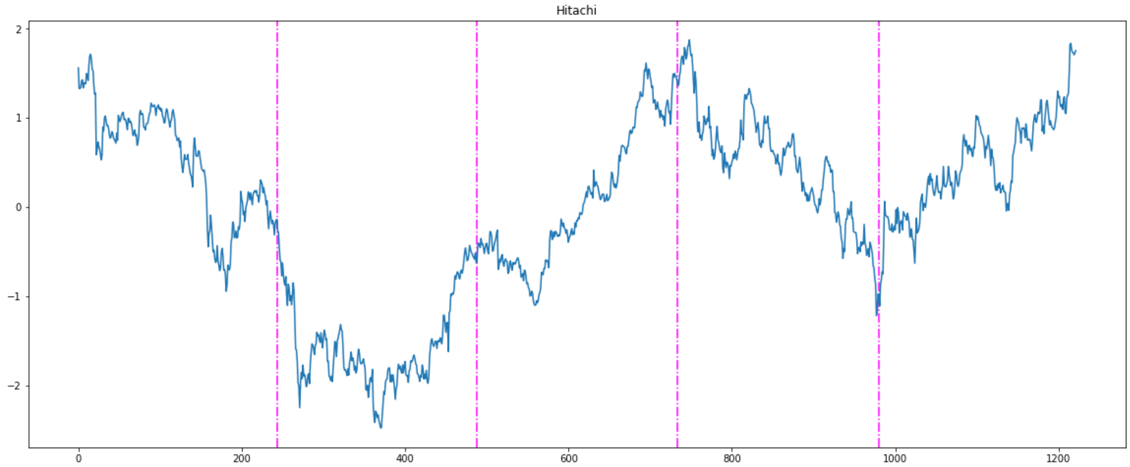


図13 日立製作所 標準化データ

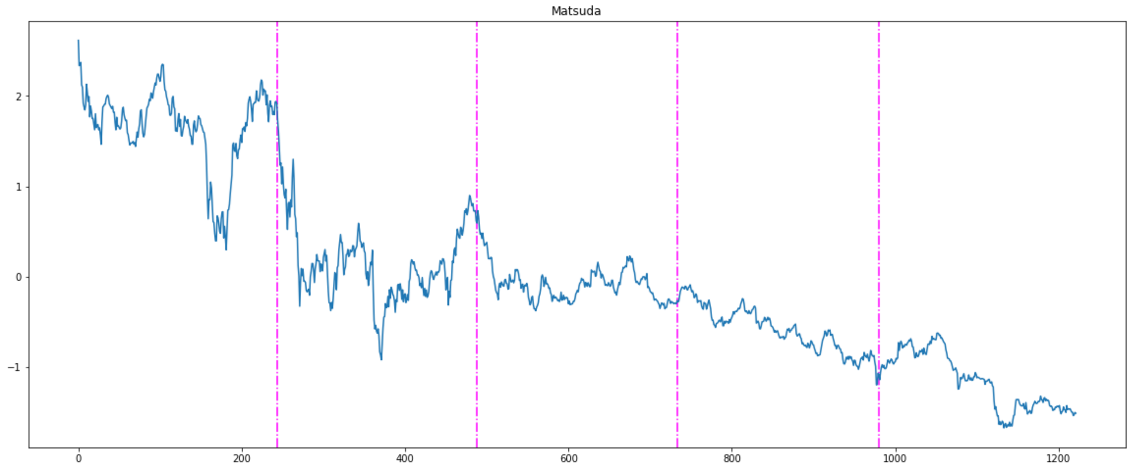


図14 マツダ 標準化データ

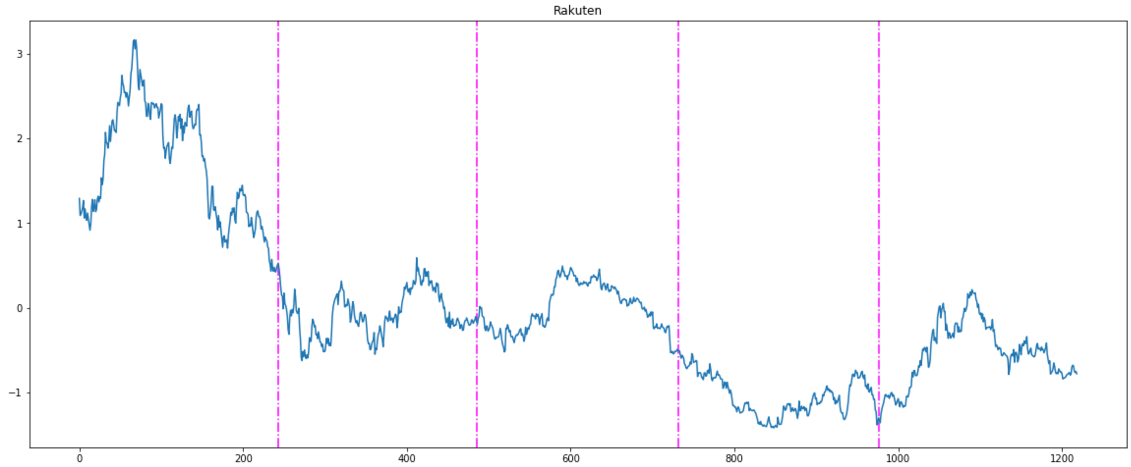


図15　楽天　**標準化データ**

ここで各データが定常性を持つかどうかADF検定を用いて確認をする。GDPが長期で見れば上昇しているように、経済は基本的に成長し続けいている。そのため、株価データも何らかのトレンドを含む可能性が高いことが予想される。そのため、検定に用いるモデルは、トレンド項を含むモデルを使用した。

表1 ADF検定　有意水準95%におけるP値

銘柄/p値	一次トレンド	二次トレンド	一次トレンド　差分	二次トレンド　差分
Toyota	0.537429	0.305928	0.00E+00	0.00E+00
Shiseido	0.470285	0.801919	0.00E+00	0.00E+00
Matsuda	0.02325	0.04921	2.85E−14	7.33E−14
Hitachi	0.484944	0.729997	0.00E+00	0.00E+00
Cyber	0.852679	0.925395	1.47E−21	1.33E−22
Canon	0.699574	0.865388	0.00E+00	0.00E+00
Rakuten	0.654224	0.388973	0.00E+00	0.00E+00
Nikkei	0.283052	0.468506	0.00E+00	0.00E+00
Terumo	0.064145	0.082562	7.83E−22	5.57E−23
Kaoh	0.070151	0.119239	0.00E+00	0.00E+00
Komatsu	0.76216	0.634568	0.00E+00	0.00E+00

表1の4, 5列目は、各データの階差をとった差分系列に対してADF検定を行った時のP値である。帰無仮説が棄却されると過程は単位根過程ではないという解釈になる。単位根過程(一次和分過程)とは、元の過程が非定常であり、かつその差分が定常であるような過程を指す。表1を見ると、マツダだけが帰無仮説を棄却しており、それ以外のデータは帰無仮説を棄却できていない。差分データは、全て帰無仮説が棄却できているのでマツダを除く全てのデータは

単位根過程であるといえる。このことから、モデルでの予測はデータに定常性を仮定するので、何らかの不具合が生じることが予想される。その一方でマツダは定常性が保障されているので予測はほかのデータよりもうまくいくかもしれない。

次に自己相関係数、偏自己相関係数について見ていく。偏自己相関係数は通常の自己相関係数とは異なる。k次自己相関係数は$y_{t,}y_{t-k}$の相関を計算したものであるのに対し、k次偏自己相関係数は、y_{t-k}から、$y_{t-1,}$ $y_{t-2,}$ $y_{t-3\ ,,,}$ y_{t-k+1}の影響を取り除いたものとy_tの相関を計算したものである。ところでAR(p)の場合、y_tは$y_{t-1,}$ $y_{t-2,}$ $y_{t-3\ ,,,}$ y_{t-p}の線形関数で表現される。ここでy_tから$y_{t-1,}$ $y_{t-2,}$ $y_{t-3\ ,,,}$ y_{t-p}のを取り除いたとき、残るのは自己相関を持たないホワイトノイズだけとなる。すなわち、p+1次以降の偏自己相関は0ということである。今回はこの事実を参考にモデルの次数を決めていく。

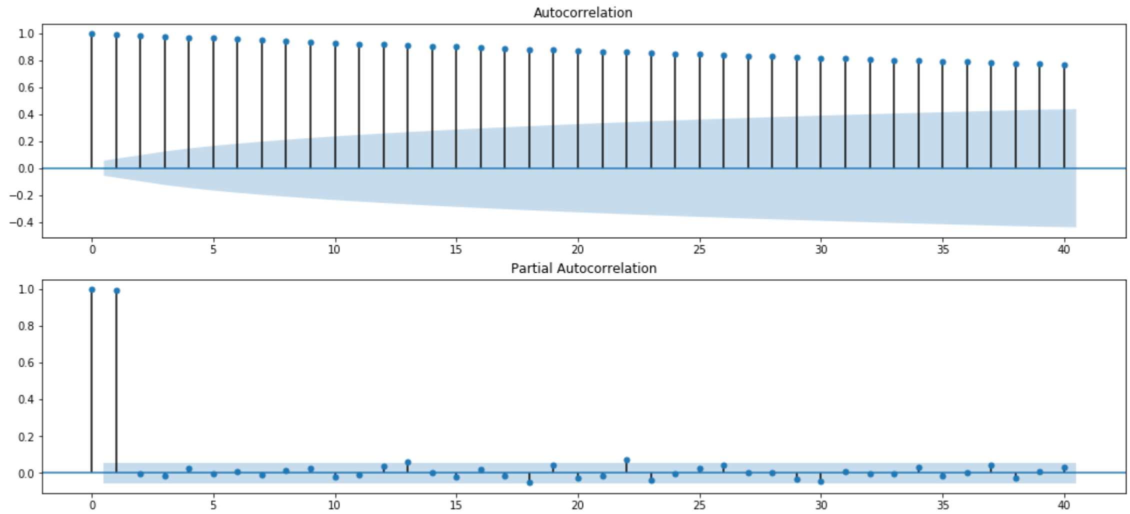


図16　コレログラム　日経平均株価

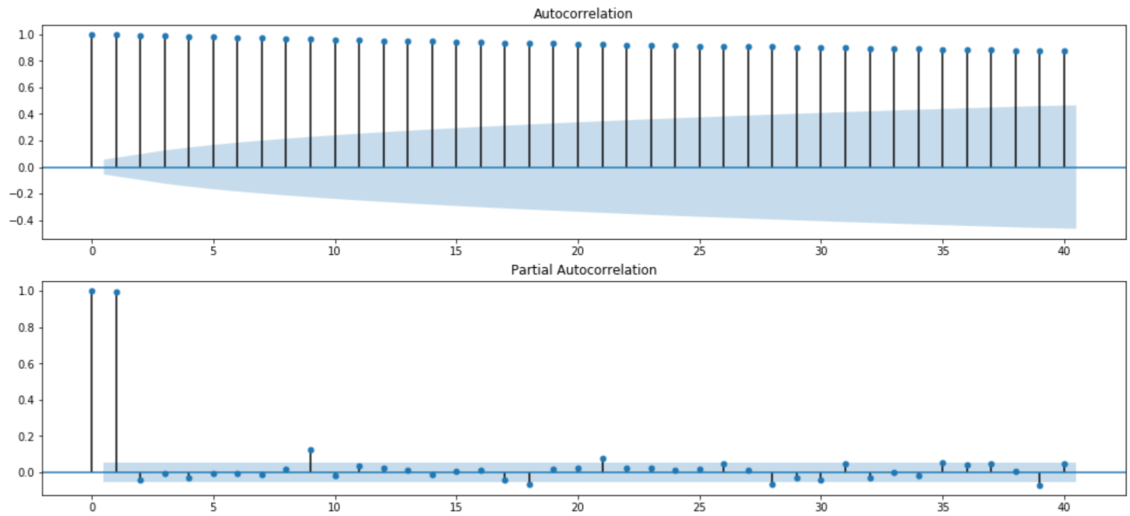


図17　コレログラム　コマツ

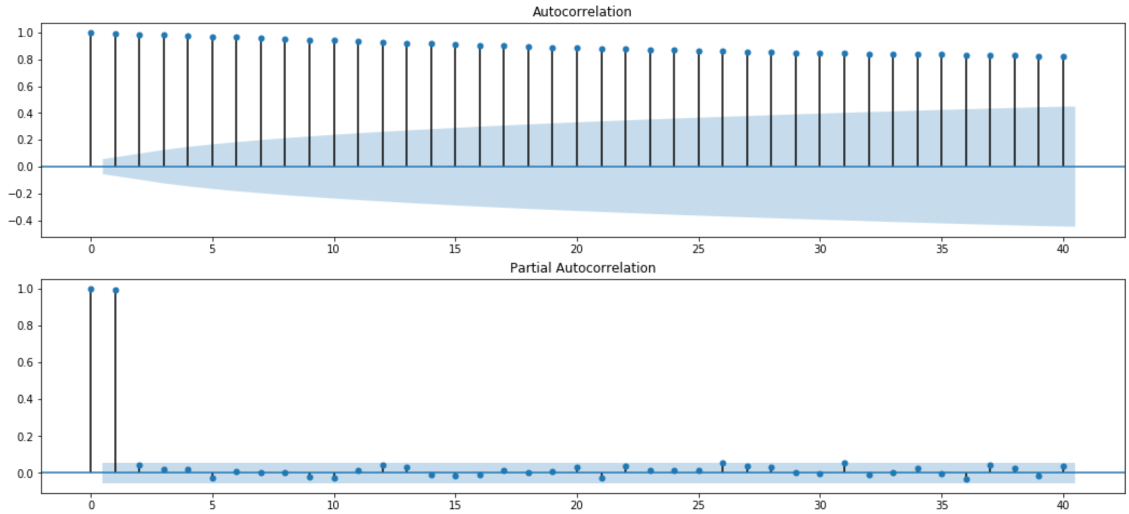


図18　コレログラム　花王

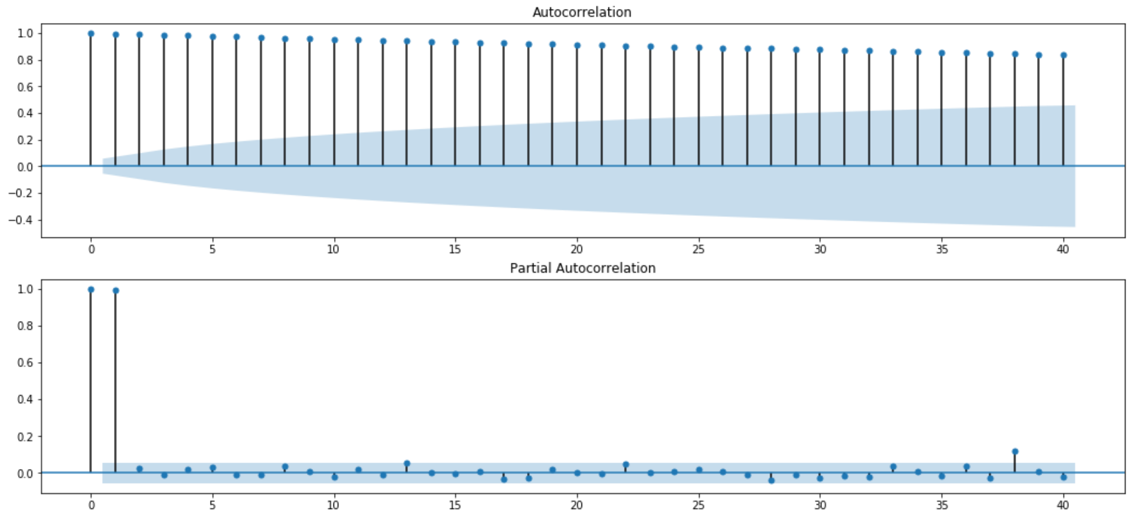


図19　コレログラム　テルモ

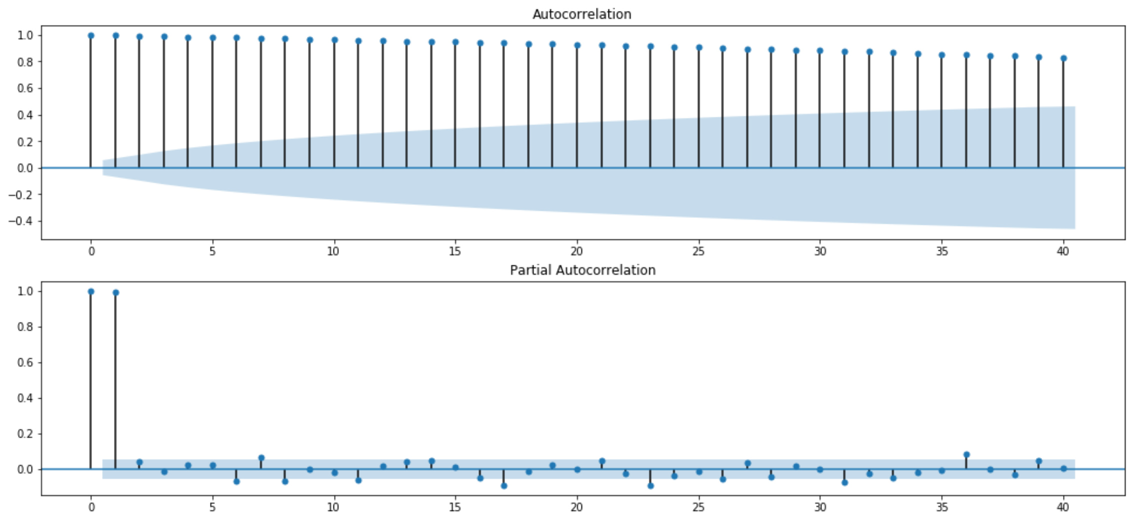


図20　コレログラム　楽天

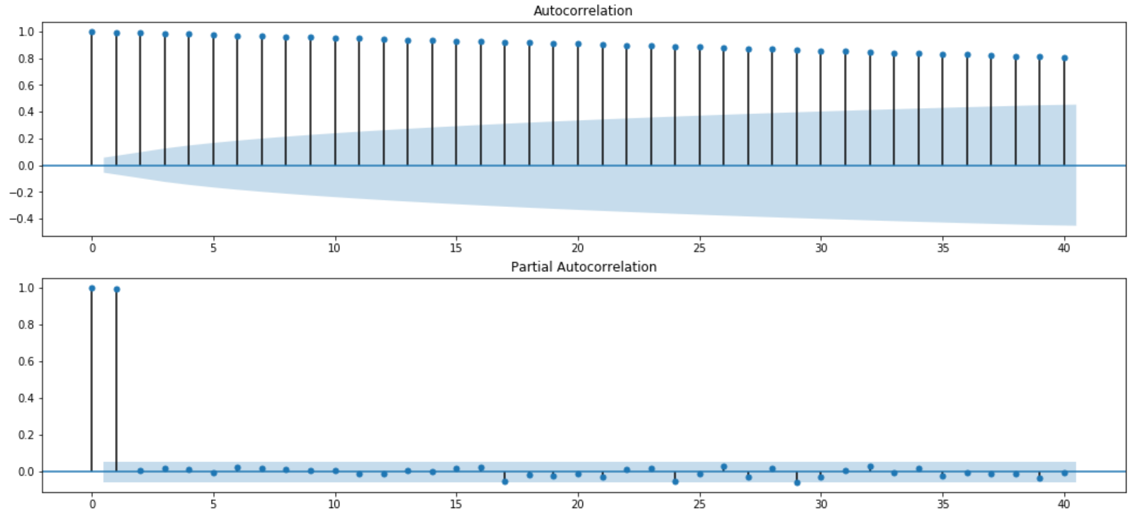


図21　コレログラム　キャノン

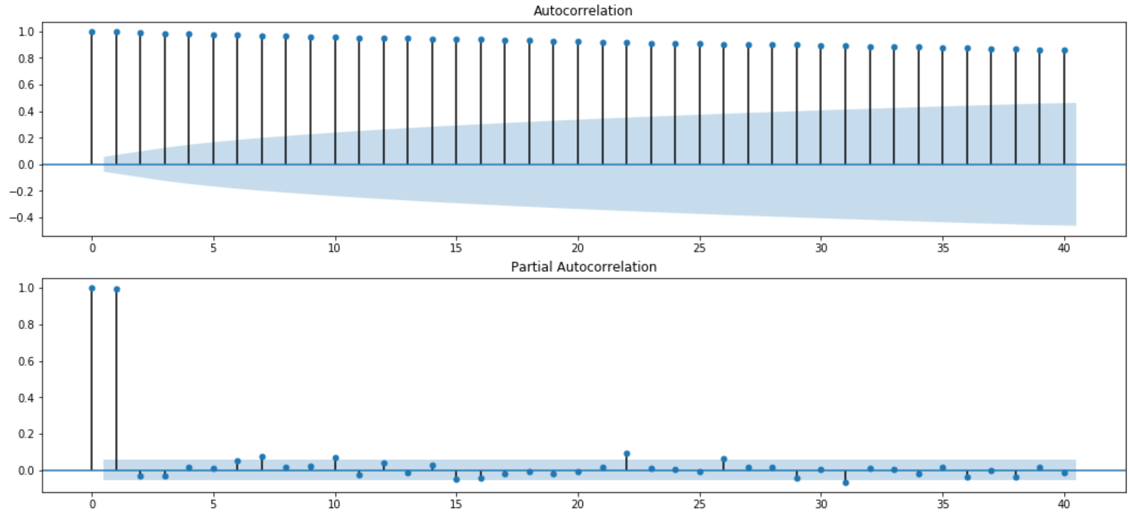


図22　コレログラム　サイバーエージェント

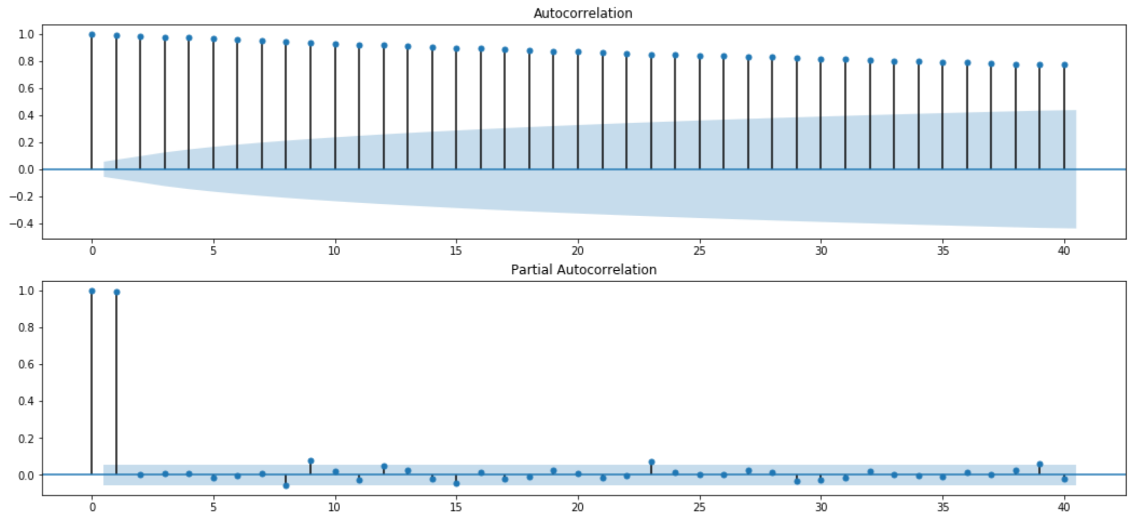


図23　コレログラム　日立

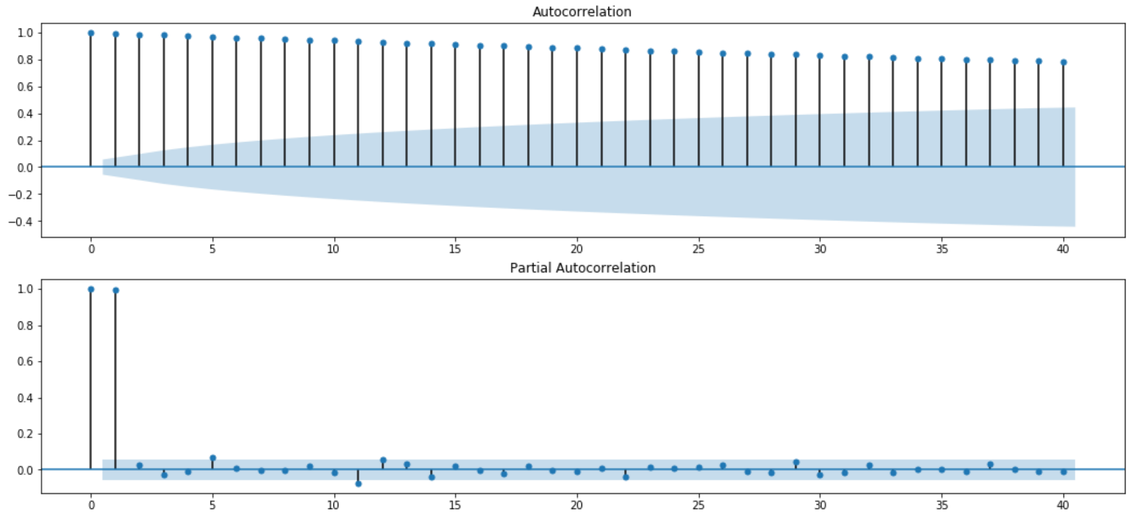


図24　コレログラム　マツダ

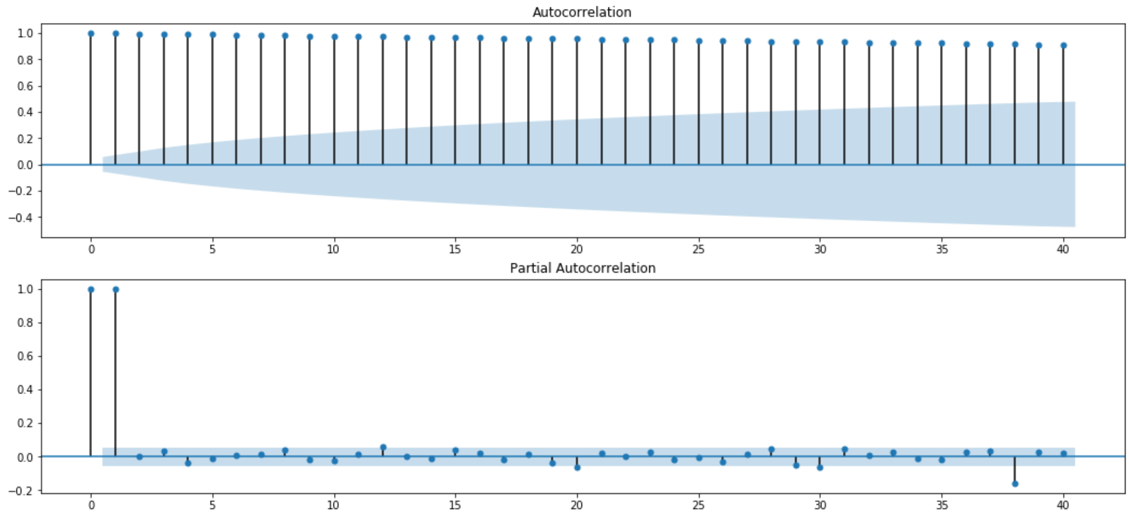


図25　コレログラム　資生堂

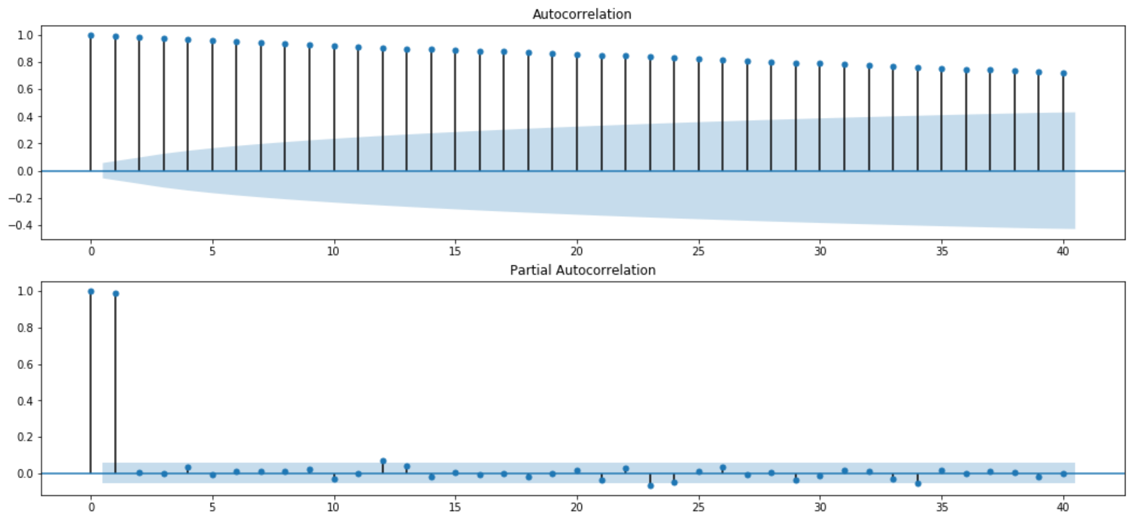


図26　コレログラム　トヨタ

ラグ40までの相関係数と偏相関係数のコレログラムをプロットした。青い部分は、有意水準5%で帰無仮説を自己相関係数、偏自己相関係数が0としたときの範囲である。どのデータも偏相関係数がラグ1の時有意であるので、モデルの次数はp=1が妥当だと考えられる。

3.2.　時系列モデルによる予測

二章で示したように予測は予測期間が長くなるにつれて不確かなものとなっていく。それらを考慮し、今回は1期先予測についてデータの予測精度を見ていく。
まずは1期先区間推定を用いて、ARモデルの次数を変化させ、予測精度を見る。

表2　ARモデル一期先95%区間推定

銘柄/次数(p)	p=1	p=2	p=10	p=1 差分	p=2 差分	p=10 差分
Toyota	(−0.5619409700248914, 0.056076852370124775)	(−0.554128125281661, 0.05371599202327826)	(−0.5677178144525494, 0.04596220706382881)	(−0.3196099419262914, 0.3219647632175023)	(−0.32274000927846314, 0.3267762036703574)	(−0.3303469298638323, 0.3110418348774404)
Shiseido	(−0.8136681826380562, −0.7203107408669073)	(−0.8132130514071549, −0.7224713797094215)	(−0.8160517295432057, −0.7214106335078556)	(−0.04545219005661449, 0.04673618296366222)	(−0.04741637738807719, 0.04855273710184661)	(−0.0472792086128848, 0.0476270322343373)
Matsuda	(0.1510825386941483, 0.5917805942748525)	(0.1595859142738645, 0.5851975656538637)	(0.15940011227451656, 0.5861767197369391)	(−0.21007623900137865, 0.20398220685448568)	(−0.21517151797064143, 0.2101202735003031)	(−0.20662818348788947, 0.1915368884569305)
Hitachi	(−0.6215665554449376, −0.2162857969366205)	(−0.6259193915881003, −0.2068949397490563)	(−0.6187718543826434, −0.20135203971387275)	(−0.20801645181881642, 0.20498667493009656)	(−0.20652514866545876, 0.20278664827365658)	(−0.1958221340794078, 0.2067635385141579)
Cyber	(−0.8077917768966348, −0.5762884169880719)	(−0.8090738253180679, −0.5767017747032567)	(−0.8126209883278553, −0.5713719337651838)	(−0.11590085191109786, 0.11756981530917246)	(−0.1144556981425947, 0.11664916128212642)	(−0.11088044812314372, 0.11806403043245557)
Canon	(−0.731898845435647, −0.2863265883716035)	(−0.722490403832334, −0.29662277524417624)	(−0.7291647292839866, −0.29778921141864034)	(−0.21767131623421554, 0.21265005098526885)	(−0.21898145392260296, 0.21416908299828355)	(−0.22031879918505456, 0.20900436573689138)
Rakuten	(−0.4676650343734283, −0.08112457429753653)	(−0.4591238257610095, −0.09013176110234819)	(−0.4688320469017423, −0.10280548815374299)	(−0.19278817102464801, 0.18656816032783713)	(−0.19976162804678702, 0.19325381530726646)	(−0.20301396316963713, 0.17253345708782988)

Nikkei	(−0.14564378777461934, 0.14284385260208393）	(−0.1380129635232044, 0.13735071241978025）	(−0.14797512662034012, 0.13826943529027952）	(−0.1641614529301938, 0.18549277426385097）	(−0.1158445615435563, 0.22575381031310918）	(−0.08363405914638475, 0.20259674235868577）
Terumo	(−0.10762184794732185, 0.09845457468179886）	(−0.10299203341063776, 0.09559025733447662）	(−0.10340169973070468, 0.10285889037749651）	(−0.10762184794732185, 0.09845457468179886）	(−0.10299203341063776, 0.09559025733447662）	(−0.10340169973070468, 0.10285889037749651）
Kaoh	(−1.4661349088620546, −1.1436350340783472）	(−1.4613186393527793, −1.138903598451525）	(−1.4612285810544046, −1.137422805051693）	(−0.15406684114405447, 0.16758744111823617）	(−0.1574017777276367, 0.17172775549936697）	(−0.16023450420727298, 0.17039087499347938）
Komatsu	(−0.04575927019051254, 0.2356978203655031）	(−0.04957031877690746, 0.23880656726080132）	(−0.057128433068695056, 0.24008635320991442）	(−0.15130473760542507, 0.15030916075411538）	(−0.1459516381297936, 0.13951155947131846）	(−0.15231471299607607, 0.13767465848343247）

表をみるとp=1,p=2,p=10の時の区間推定の幅がほぼ変わらないことがわかる。これは偏自己相関がラグ1の時有意であり、それ以前のデータは予測に対して何ら情報を与えないことと同じ結果を表す。差分をとったデータの場合もほぼ同じであることがわかる。

ここまでで、予測は一期先が最も精度が高く、またモデルの次数はp＝1が妥当であることがわかった。次に行うことは加工したデータで精度は上がるのかどうかを検証する。何度も言うようにモデルは定常性を背景に持つので、データが定常でなければ不具合が出る可能性がある。そこで、差分をとってデータを定常にしたあとに、そのデータを二章の最初で説明した方法により復元する、といった加工をデータに施す。そのデータともともとのデータを比べどちらが予測精度が良いかを検証する。差分を差分に戻す、差分を原データに戻す、差分の差分を原データに戻す、をそれぞれd2→d1、d1→d0、d2→d0、と表すことにする。実験回数は10,000回行う。実験1の結果については次のように考えることにする。

ARモデルにそのままデータを入れて予測した場合とデータを復元して予測した場合の結果に差がないと仮定すると、二項分布を用いて結果に有意な差異があるかどうかを調べることができる。N 回比較して結果の良い方をピックアップする方法は確率 0.5 のコインを投げて表の数を数える操作と等価である。ここでサンプル数がN＝10,000と十分に大きいので、二項分布を正規分布として近似して考えることができる。すなわち、N＝10,000、p=0.5、の二項分布は、期待値np、分散np(1−p)の正規分布に従う。ここで確率変数Xを、X〜Bin(10,000, 0.5)とし、確率変数Zを、$Z = (X - np)/(np(1-p))^{1/2}$, Z〜N(0,1)とすると、有意水準95%でのXの範囲は、$4902 < X < 5098$になる。この事実を用いて仮説を棄却できる値は色をつけた。

なお、結果の数値は復元データを用いたほうが、そうじゃない場合と比べて観測値との差が小さかった場合にカウントアップされる。ARモデル劣位に棄却できる値に関しては下線を引いておいた。

表3　AR(1)一期先予測　比較

銘柄/実験1	d2→d1	d1→d0	d2→d0
Toyota	**5401**	4962	**5374**
Shiseido	4226	5045	4300
Matsuda	4540	4996	4453
Hitachi	4151	4983	4229
Cyber	4314	**5192**	4415
Canon	4424	4990	4465
Rakuten	4121	5075	4163
Nikkei	4395	5012	4315
Terumo	4009	5015	4026
Kaoh	3829	5039	3883
Komatsu	4075	4727	2864

トヨタを除くとd2までとるとARモデルにそのままデータを入れたほうが有意に良くなる。d1までとるとサイバーエージェント、コマツ以外はほとんどデータによる予測精度に差が見えないことになる。二階差をとり復元することで精度が落ちる原因として、階差をとるごとに符号が反転する部分が強調されることが原因であることが考えられる。差分をとるごとに数値が大きくなることを考慮し、差分をとるごとに標準化を施し、実験1と同様の試行を考えてみる。但し、原系列の差分をとり、それを標準化したものは、原系列を標準化し、それの差分をとってからまた標準化したものと等しくなるので、原系列の差分をとるタイミングは特にこだわる必要がない。今回は初めからデータを標準化してあるものをd0として用いる。実験2の結果は以下の通りである。

表4　AR(1)一期先予測　比較

銘柄/実験2	d2→d1	d1→d0	d2→d0
Toyota	**6880**	1394	1307
Shiseido	4836	177	170
Matsuda	**5438**	724	789
Hitachi	4981	682	635
Cyber	4965	478	543
Canon	**5624**	735	821
Rakuten	4860	613	623
Nikkei	**5486**	837	808
Terumo	4509	351	373
Kaoh	4383	537	422
Komatsu	4875	556	648

階差をとるごとに標準化を行うと、有意にARモデルにそのままデータを入れたほうが予測精度が良くなる。ただ実験2のほうがARモデル劣位で棄却している値が増えてるので、差分系列から差分系列を予測する場合は復元したほうが良いといえる。

ではVARモデルを使用するとどうなるか見てみよう。まずは区間推定である。次数を変えても予測精度はp=1の時とさほど変わらないように見える。この結果と3.1節の自己相関構造の結果を併せて考えると、VARモデルの次数に1を採用していいと考える。

表5　11変数VARモデル一期先95％区間推定

銘柄/次数(p)	p=1	p=2	p=10	p=1 差分	p=2 差分	p=10 差分
Toyota	−0.341209 −0.158124	−0.329359 −0.147455	−0.185793 −0.0222296	−0.0926289 0.0921022	−0.0864574 0.0984502	−0.0875251 0.0781401
Shiseido	−0.776968 −0.748913	−0.773995 −0.746604	−0.76025 −0.735212	−0.0118261 0.0164041	−0.00984578 0.0176446	−0.00611622 0.0189736
Matsuda	0.216849 0.343421	0.215094 0.339831	0.25456 0.36602	−0.0797893 0.0496819	−0.0794262 0.0477748	−0.0957996 0.0152236
Hitachi	−0.559576 −0.437913	−0.554604 −0.434301	−0.501602 −0.393392	−0.075645 0.046972	−0.0789824 0.0418472	−0.0481228 0.0590006
Cyber	−0.654331 −0.584345	−0.652927 −0.58311	−0.639592 −0.576439	−0.0299188 0.0400678	−0.0274604 0.0432132	−0.0256461 0.039763
Canon	−0.578655 −0.44963	−0.576287 −0.448269	−0.485999 −0.369415	−0.0802171 0.0504437	−0.0683304 0.0590703	−0.0569853 0.0592279
Rakuten	−0.31519 −0.207157	−0.315508 −0.204531	−0.251862 −0.154485	−0.0521392 0.0573367	−0.0567126 0.0543562	−0.0741441 0.024537
Nikkei	−0.610673 −0.47003	−0.59279 −0.455095	−0.548539 −0.42325	−0.0668498 0.0745814	−0.0601536 0.0781415	−0.0387837 0.0855817
Terumo	−0.580931 −0.518972	−0.568518 −0.507437	−0.582551 −0.526752	−0.0211842 0.0414947	−0.0184253 0.0426455	−0.00290564 0.0531638
Kaoh	−1.38461 −1.28565	−1.37383 −1.27805	−1.34289 −1.2565	−0.0459553 0.0539048	−0.0437002 0.0534189	−0.0109801 0.0749212
Komatsu	0.0667815 0.151317	0.0697473 0.153321	0.0828659 0.157583	−0.0434616 0.0416453	−0.0311484 0.0523994	−0.0405227 0.0332691

次はVARモデルに入れるデータを加工して予測精度を見る。ARモデル同様のシュミレーションを行った。以下の表がその結果である。

表6　11変数VARモデル　一期先予測比較

銘柄/実験1	d2→d1	d1→d0	d2→d0
Toyota	**9426**	4127	**9172**
Shiseido	3392	**5552**	3838
Matsuda	**7411**	5042	**7515**
Hitachi	**6558**	3817	**5436**
Cyber	**6991**	**5837**	**7636**
Canon	**6257**	4137	**5339**
Rakuten	2730	4802	2601
Nikkei	**7810**	3186	**6230**
Terumo	**5435**	2665	3394
Kaoh	1596	**5657**	2069
Komatsu	3147	4773	3047

表7　11変数VARモデル　一期先予測比較

銘柄/実験2	d2→d1	d1→d0	d2→d0
Toyota	**10000**	11	0
Shiseido	0	207	0
Matsuda	**10000**	**7407**	0
Hitachi	**10000**	252	0
Cyber	2775	2794	0
Canon	**10000**	1840	0
Rakuten	0	0	0
Nikkei	**10000**	1332	0
Terumo	0	487	0
Kaoh	0	74	0
Komatsu	626	4491	486

実験1について、ほとんどのデータでMatsuda(d1->d0)以外では有意差が有といった結果になった。d1→d0の列を見てみると、11個のデータのうち、6個のデータが予測精度が改善した。しかし、反対に予測精度が落ちているデータもある。実験2について、標準化のタイミングを変えた場合の結果は、表7で有意に仮説を棄却したいくつかが、d2→d1でほとんどの場合で予測精度を改善している。

3.3. 考察

実験1について、d2→d0の精度がいいデータは、少なくともd2→d1の精度もいいことが結果からわかる。このことから予測の精度の改善は差分データの二階復元によってもたらされる可能性があることがわかる。逆に予測精度が改善しなかったデータは、d1→d0で、多くの場合予測精度の改善が見られる。3章で説明したようにデータの復元は、差分系列の予測値と原系列の最新の値の和で行うことができる。つまり、差分系列の予測精度がそのままダイレクトに復元データの予測精度にかかってくるのだ。このことから、データは階差をとる回数によって予測精度が変化することがわかる。

実験2について、差分をとるごとに標準化をすると、差分系列への予測精度は改善が見られたデータがあった。差分をとるとデータの分散が小さくなるので、分散の影響が標準化によって改善したのかもしれない。

全体を通すと、今回のデータ群の中だと、ARモデルはそのままデータを予測に用いる、すなわち定常性の条件を無視して予測をしたほうが予測精度は良くなっている。反対にVARモデルは、定常性の条件を満たすデータを入力とすることで予測精度を改善するデータもある。VARモデルは複数の変数の影響を受けるので、定常性の条件がより強く受けることができたのではないかと考える。ARモデルはどうなのかといわれると難しいところである。ただ、モデルにおける定常性の前提は、長期にわたる予測の場合である。このこと踏まえると、直近の予測にどれほど影響があるのかといったところであろう。長期予測といった視点で見てみると、もしかしたらデータを復元したほうが予測精度はよくなるのかもしれない。

4.　おわりに

両実験を通してARモデル、VARモデルの両方で、予測精度が改善したデータもあれば、逆に精度が落ちたデータもある。ARモデルはデータを復元することで予測をすることよりも、そのまま予測するほうが精度がいいことがわかる。差分をとるごとに標準化を施しても、予測精度の改善はあまり見られなかった。反対にVARモデルではデータを加工することによって予測の改善が見られるデータもある。予測精度を改善するにはこれらのデータの違いについてもう少し調べることが必要である。例えば、予測改善したデータに相関があるのかなど。

卒論製作を通して、データの表現を変えるだけで、モデルの予測精度が変化するので、簡単なモデルでも予測精度を向上させることはできる可能性があることが今回わかった。加えて、統計モデルとは様々な前提の上に成り立っているものであることを再認識することができた。また、今回は触れなかったが、データの対数差分をとることによって、収益率を出すことができる。この収益率も定常性をもつのでこのデータを予測に用いたのち、実数値に予測収益率をかけることで、実質的な予測が可能となる。これらのデータの予測精度を比べ、どの場合が一番予測がいいのか検証するのも面白いかもしれない。

最後に、私の卒論製作は片岡先生のお力添えをいただいたおかげでここまで来ることができました。本当にありがとうございました。

5.　付録

5.1.　ADF検定

GDPや景気などは、長期で見れば上昇傾向にある。また株価や為替は将来の動きを予測することは非常に困難であり、平均回帰性を持つとは言い難い。このようにデータが定常性とは異なる動きをすることは現実問題として少なくない。このような非定常なデータについて少し補足説明をしておく。

[4]非定常過程の1つである単位根過程は次のように定義される。原系列をy_tとすると、その差分系列、$\Delta y_t = y_t - y_{t-1}$ が定常過程であるときその原系列を単位根過程という。データが単位根過程であるかどうかは、時系列モデルを記述するうえで非常に重要な問題である。そこで付録として、今回使用した単位根検定についてまとめていく。

今回使用した単位根検定はADF検定（拡張 Dickey-Fuller 検定）という、名前の通りDF検定を拡張したものである。であるならば、DF検定の説明から入ろう。

Dickey-Fuller （DF）検定は真の過程をAR(1)と仮定し、過程を単位根AR(1)過程であるという帰無仮説に対して検定するものである。DF検定含む多くの単位根検定と通常の仮説検定の大きな違いの一つに、帰無仮説のモデルが定数項を含むかどうかと、対立仮説のモデルが定数項とトレンド項を含むかどうか、によって検定統計量の漸近分布が異なるため、用いる棄却点が違うという点である。組み合わせによって、いくつかの場合に分けられるが、主に次の三つの場合が用いられる。

$$
\begin{array}{lll}
[\text{場合}1]\boldsymbol{H}_0 : y_t = y_t + u_t, & \boldsymbol{H}_1 : y_t = \rho y_t + u_t & |\rho| < 1 \\
[\text{場合}2]\boldsymbol{H}_0 : y_t = y_t + u_t, & \boldsymbol{H}_1 : y_t = \alpha + \rho y_t + u_t & |\rho| < 1 \\
[\text{場合}3]\boldsymbol{H}_0 : y_t = \alpha + y_t + u_t, & \boldsymbol{H}_1 : y_t = \alpha + \rho y_t + \delta t + u_t & |\rho| < 1
\end{array}
$$

H_0は帰無仮説を表し、H_1は対立仮説を表す。データがトレンドを持つ場合、単位根過程は定数項(ドリフト)を持つので場合3を採択する。トレンドを持たない場合、対立仮説の過程の期待値が0であるかどうかで判断する。期待値が0ならば場合1を、0でないならば場合2といった具合だ。データのトレンドや期待値の判断がつかないとき、場合3を使うこと安全である（場合1,2,3の順に一般的になっているから）。しかしながら、場合1,2が正しいときに、場合3を採択すると、過程が定常過程であるのに、帰無仮説を棄却できない可能性がある。

検定統計量についてだが、DF検定の検定統計量は、ブラウン運動の汎関数になることが知られている。今回の論文には検定統計量についての詳しい知識を必要としなかったので深くは調べなかった。検定統計量については、次の2つの事実を使用した。1つ目は、DF検定統計量は、DF分布という漸近分布に従うのだが、この漸近分布は帰無仮説、対立仮説のペアによって変化するということ。2つ目は、DF分布は右裾が広い分布の形をしていることに注意すること。通常のt検定を用いて、$\rho=1$を検定すると、帰無仮説を棄却しすぎてしまうのである。

さてようやく本題のADF検定である。ADF検定はDF検定の拡張であると始めに言った。DF検定では真の過程をAR（1）としたのに対して、ADF検定は真の過程をAR（p）であること許したのだ。ADF検定はほとんどDF検定と同じ考え方、手順で行うことができる。但し、推定するモデルを少し調整が必要であるが。

6.　参考文献

[1] 日銀短観　2020年6月調査https://www.boj.or.jp/statistics/tk/gaiyo/2016/tka2006.pdf
[2] 国内統計　有効求人倍率
https://www.jil.go.jp/kokunai/statistics/covid-19/c07.html
[3] 投資の王道　株式市場のテクニカル分析　新井邦宏著
[4] 経済・ファイナンスデータの計量時系列分析　沖本竜義著

片岡先生の資料
その他講義資料