

株価の日次・月次推移と学習期間調整

吉田 健一^{1,a)}

受付日 2020年6月28日, 採録日 2020年12月1日

概要: 株価予測は学問的にも実務的にも重要な研究テーマであり, 伝統的なファイナンスの観点から古くから研究されてきた。また近年では深層学習などデータマイニングの手法を使った研究もさかんである。本報では, 日次の終値の変化と標準偏差のみを入力に用いた Gradient Boosting Decision Tree 法が代表的な株価インデックスである TOPIX や日経 225 先物の翌日の値を予測可能であり, 代表的な取引手法であるインデックス投資と比較して超過収益が得られることを示す。この結果は月次データに関して報告した代表的な株価インデックスが持つ特徴と同じ特徴 (株価の分析における混合分布分析と学習期間調整の重要性) が日次データにも存在することを示している。本報では, 学習期間調整の重要性は日次の方が顕著であり, また日経先物の分析結果からは限月という先物特有の要因が有効な学習期間に影響している可能性も指摘する。

キーワード: 株価予測, 先物取引, Gradient Boosting Decision Tree

Training Period of Daily and Monthly Stock Price Prediction

KENICHI YOSHIDA^{1,a)}

Received: June 28, 2020, Accepted: December 1, 2020

Abstract: This paper reports a fact that the Gradient Boosting Decision Tree (GBDT), which only uses the closing prices and standard deviation, can predict daily prices of TOPIX and Nikkei 225 futures. It shows that excess returns can be obtained compared to the standard index investment. The method used for the prediction is a straightforward application of a method that has been shown to be effective in the analysis of monthly stock prices. It is a combination of the GBDT that can analyze the mixture distribution and the adjustment procedure for the learning period. These results show that the same characteristics of the monthly stock prices also exist in the daily data. In other words, the importance of mixture distribution analysis and learning period adjustment in the analysis of stock prices are shown for different time scales such as daily and monthly. From the analysis results of Nikkei 225 futures, it is also pointed out that the futures-specific factor called contract month may affect the effective learning period.

Keywords: stock price prediction, futures trading, Gradient Boosting Decision Tree

1. はじめに

株価予測は学問的にも実務的にも重要な研究テーマであり, 伝統的なファイナンスの観点から古くから研究されてきた。また近年では深層学習などデータマイニングの手法を使った研究もさかんである。

本報では, 日次の終値の変化と標準偏差のみを入力に用

いた Gradient Boosting Decision Tree (GBDT) 法 [11] を代表的な株価インデックスである TOPIX や日経 225 先物の翌日の値を予測に適用し, 代表的な取引手法であるインデックス投資と比較した結果を示す。予測に用いた方法は著者達が月次データの分析に有効性を示した手法 [16] を素直に日次に適用したものであり, 混合分布が分析可能な GBDT 法に学習期間の調整手順を組み合わせたものである。

実験結果は月次データに関して報告した代表的な株価インデックスが持つ特徴と同じ特徴が日次データにも存在す

¹ 筑波大学大学院ビジネス科学研究科
University of Tsukuba, Bunkyo, Tokyo 112-0012, Japan
^{a)} yoshida.kenichi.ka@u.tsukuba.ac.jp

ることを示した。すなわち、株価の分析における混合分布分析と学習期間調整の重要性が日次と月次という異なる時間の尺度において重要であることを示している。また学習期間調整の重要性は日次の方が顕著であり、日経先物の分析結果からは限月という先物特有の要因が有効な学習期間に影響している可能性が示唆された。

以下2章で関連研究を概観した後、本研究が持つファイナンスや人工知能研究上の価値を議論する。3章で分析手法の説明を行い、4章でTOPIXと日経225先物を対象とした分析結果を示す。最後に5章で結論をまとめる。

2. 関連研究

株価予測は学問的にも実務的にも重要な研究テーマであり、伝統的なファイナンスの観点から古くから研究されてきた。代表的な方法論としては、複数企業群の時価総額などの属性に基いたクロスセクション分析 [5], [13] や、ある銘柄の過去の株価推移を分析する時系列分析 [8] などが上げられる。

上記の伝統的な方法論による株価の予測可能性に関連して、長らく効率的市場仮説 [2] により過去データに基づく株価の予測はできないとされてきた [4] が、近年データマイニングの手法を使った研究などの成果により、予測可能性を示す研究 [12], [17], [18], [19], [20] も報告されつつあり、各種の提案手法を比較している報告 [3] もある。

本研究は2001年以降の代表的な株価インデックスの月次データに関して報告した規則性 [16] が日次データにもあるかを確認することで、株価予測における混合分布の取扱いと学習期間の重要性が一般的である可能性を指摘する。

近年の深層学習研究の発展 [9] により深層学習を応用して株価予測を試みる研究はさかんであるが、現状、株価の挙動の背景にあるモデルが、ある期間内では比較的安定しているが、時間が経過すると変化するという仮定において、その変化に対応したモデル化を行おうとした研究は少ない。学習期間を固定長に制限したり [14] 時期によって手法の予測性能が変わることを報告した研究 [7] はあるが、学習期間を調整し、時期による現象の変化に対応した研究は文献 [16] 以外に見当たらない。その意味で本研究の内容は今後の株価予測の研究に重要な知見を与えると考える。

また、深層学習の研究においては、学習したモデルで予測した場合の株価が上昇/下降するかの正解率や Mean Squared Error を用いて評価する研究が多いが、金融工学の研究では単純な正解率や Mean Squared Error を用いた評価の問題点が古くから指摘され、問題の少ない評価手法としてシャープ・レシオ [15] が用いられている。また、文献 [6] や文献 [10] ではこの問題に対処するため超過収益を評価している。本研究でもこれらにならぬ超過収益とシャープ・レシオを利用して学習モデルを評価する。

3. 学習期間を調整した株価分析

著者達は1954年から2016年までのTOPIXの値動きを分析し、1986年以降の月次の値動きがGBDT法で予測可能であることを示すと同時に、同じ方法で代表的な株価インデックスである2001年以降のCAC, DAX, NKY, SPX, およびUKXの値動きも予測可能であることを報告した [16]。この研究は株価の分析における混合分布分析と学習期間調整の重要性を議論しているが、査読されていないワーキングノートであり、結果も月次データに限定されている。本研究では、そこで提案された方法を日次データに適用し、一般性を確認するとともに、日経先物の分析結果から限月という先物特有の要因が有効な学習期間に影響している可能性を指摘し、株価データを扱ううえでの「時間」の扱いを議論する。

3.1 データ表現

次章の実験ではTOPIXの値を1秒ごとに記録した秒速データを各分の最終の値にまとめ、1分ごとのデータに変換したうえで、以下のような入力データ v_t を作成した。

$p_{t,m}$ = 日付 t における1分間隔の株価。 m は時刻情報

p_t = 株価の日付 t の終値

$$d_t = (p_t - p_{t-1}) / p_{t-1}$$

$$d_{2,t} = (d_t + 1) * (d_{t-1} + 1)$$

$$d_{3,t} = (d_t + 1) * (d_{t-1} + 1) * (d_{t-2} + 1)$$

$$s_t = \sqrt{\frac{\sum_{m=1}^{\text{取引終了時刻}} (p_{t,m} / p_{t-1} - 1)^2}{\text{データ数}}}$$

$$v_t = (d_t, s_t, d_{t-1}, d_{2,t}, s_{t-1}, d_{t-2}, d_{3,t}, s_{t-2})$$

すなわち、入力データ v_t の基本は終値の変化 d_t と、日中の株価変動の標準偏差 s_t である。

このデータ表現は文献 [16] で用いたデータ表現において、 d_t は月末の株価の変化、 s_t は月内の日ごとの終値を前月最後の終値で正規化した値の標準偏差を、月を日に日を分に置き換えた、すなわち時間の尺度を置き換えたものである。また4章では分を秒や2分間隔などに置き換えた実験も行い、時間の尺度が予測に及ぼす影響も調査した。

効率的市場仮説は、過去の株価のみからなる v_t のような情報から未来の株価の予測はできないとしている。この点で1986年以降の月次の値動きがGBDT法で予測可能であることを示した文献 [16] の結果は著者達にとっても予想外であった。そこで本研究では、月次データの予測可能性が日次データでも観測できないかを確認するため（すなわち、その一般性を見るため）に、基本的には v_t は文献 [16] の単純な置き換えを行い（すなわち、時間の尺度のみ変更し）、結果を分析した。

```

1: function MODEL(d, v)
2:   return Model for estimating  $d_t$  from  $v_{t-1}$ 
3: end function
4: function SELECT(d, v)
5:   for  $i$  in [1:length(d)-test] do
6:     for  $j$  in [1:test] do
7:        $m = \text{MODEL}(d[\text{data of length } i \text{ for position } j],$ 
8:                    $v[\text{data of length } i \text{ for position } j])$ 
9:        $e[i, j] = m(v[\text{for } j])$ 
10:    end for
11:   end for
12:   return  $i$  that gives max correlation
13:     between  $d[t\text{-test}+1:t]$  and  $e[i,1:test]$ 
14: end function
15: function ESTIMATE(d, v)
16:    $i = \text{SELECT}(d, v)$ 
17:    $m = \text{MODEL}(d[\text{last part of length } (i+1)],$ 
18:                $v[\text{last part of length } (i+1)])$ 
19:   return  $m(v_t)$  ▷ i.e., return estimation of  $d_{t+1}$ 
20: end function

```

図 1 株価予測のアルゴリズム

Fig. 1 Stock price prediction algorithm.

3.2 株価の予測手法

図 1 に株価予測のアルゴリズムを示す。このアルゴリズムは、始めに過去の日次データの推移から翌日の株価を予測したときに相関が最も高くなる学習期間 i を選択 (図 11 行目) した後、過去 $i+1$ 日間の株価推移から予測モデルを作成している (図 15 行目)。

具体的には、Estimate は Select を使って最適な学習期間 i を過去の d, v から選択 (図 14 行目) した後、選択した学習期間長 $i+1$ の過去データ d と v からモデル m を作成 (図 15 行目) し、現時点のデータ v_t から 1 時点未来の d_{t+1} を計算し、予測値とする (図 16 行目)。

また Model は過去の v_{t-1} から 1 つ先の時点の d_t を計算する回帰モデルを GBDT 法を使って作成を行っている (図 2 行目)。実装は sklearn を用い、パラメータはすべて default のものを使用した。

Select は学習期間の長さ i を変えながら (図 5 行目)、それぞれの i で作成したモデル m (図 7 行目) で過去の d の予測を行い $e[i, j]$ に記憶する (図 8 行目)。すなわち e は過去の時点で利用可能な学習期間 i による d の予測値である。その後、過去に計測した d の値と、過去の時点で学習期間 i のもと値を予測した結果 e の値の相関が最も大きくなる i を最適な学習期間 i として Estimate に渡す (図 11 行目)。

データ表現とアルゴリズムの背景にある基本的な仮説は「投資家が終値の変化や標準偏差で表現される株価の推移を見て判断した結果が将来の株価の推移に影響を及ぼしている」また「投資家の行動は変化しており、変化前の古過ぎるデータの使用は予測を阻害するので、モデル作成には予測した場合の相関が高くなる期間 (すなわち学習期間

```

procedure INVESTMENT STEP(d, v)
 $e_{t+1} = \text{ESTIMATE}(d, v)$ 
if  $e_{t+1} > 0$  then
  Buy
else
  Sell
end if
end procedure

```

図 2 取引アルゴリズム

Fig. 2 Investment algorithm.

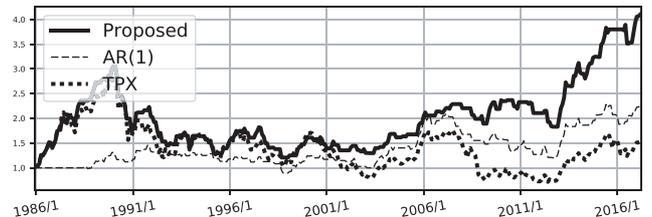


図 3 TOPIX 月次：収益比較

Fig. 3 Results on Monthly TOPIX.

$i+1$ のデータのみ利用する」というものである。

ここで次章の実験では相関の計算には 12 日間のデータを用いた (すなわち 6 行目の $test = 12$ とした)。本来は最適値を調整すべきであるが、12 日とした場合の結果でも日次データも月次 [16] と同じ手法で価格の予測ができるのを示すには十分と考えたので調整は行わなかった。GBDT 法のパラメータに default の値を用い最適化しなかったのも同じ考えによる。また過去 $i+1$ 日間の株価推移から最終的な予測モデルを作成したのは、最適値 i を選んだ過去の状況が翌日も継続するという考えによる。

上記アルゴリズムは過去のデータのみから学習モデルの作成と選択を行い、学習したモデルへの入力も過去のデータのみ用いるように設計されており、クロスバリデーションの不適切な適用により予測精度を過大に見積らないような設計方針のもとに学習期間の調整を行うアルゴリズムである。

3.3 月次データの結果

日次データの結果を報告する前に月次データの実験結果 [16] の要点を示す。実験では取り引きに手数料はかからないと仮定し、値上がりが予想されれば買い (すなわち手持ちの現金をすべて購入にあてる)、値下がりが予想されればすべて売るという想定 (図 2) で行った。

図 3 は 1986 年から 2016 年の TOPIX 月次データを使い図 1 に示したアルゴリズムの予測に従い取引したと仮定した場合の収益 (Proposed) を、TOPIX と同じ銘柄を同じ比率で組み入れたインデックス投資法 (TPX)、予測のアルゴリズムを図 1 の Estimate から Auto Regression に置き換え入力として v_t 内の d_t のみを用いたもの (AR(1)) と比較したものである。図 3 に示したとおり図 1 のアルゴリズムによる収益は Auto Regression やインデックス投

表 1 代表的な株価インデックスのシャープ・レシオ [16]

Table 1 Sharpe Ratio of Representative Indexes.

Index	Period	Proposed Method			Index	
		Accuracy	Sharp Ratio	Return*	Sharp Ratio	Return*
TPX	1986/1~	0.5525	0.1102	5.41%	0.0465	3.15%
CAC	2001/1~	0.5375	0.0639	2.99%	0.0062	0.39%
DAX	2001/1~	0.5234	0.0963	5.15%	0.0825	6.15%
NKY	2001/1~	0.5879	0.1960	8.44%	0.0587	3.93%
SPX	2001/1~	0.6012	0.1234	4.86%	0.0928	4.71%
UKX	2001/1~	0.5417	0.0909	3.30%	0.0398	1.88%

*：年間収益

DAXも提案手法のシャープ・レシオがインデックス投資法のそれを上回っており、16年間全体での収益は提案手法の方が大きい。

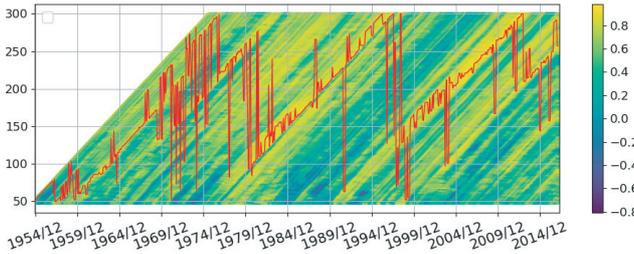


図 4 TOPIX 月次：学習期間

Fig. 4 Learning Period of Monthly TOPIX.

資による収益を上回っている。月次データは、 p_t を月末の終値に、 s_t を月中の日次終値の標準偏差に置き換えたもので TOPIX 月次の実験では $v_t = (d_t, s_t)$ とした。また参考まで CAC, DAX, NKY, SPX, および UKX の月次データの結果を表 1 に示す。いずれも図 1 のアルゴリズムの予測に基づく収益がインデックス投資法を上回っている。

図 4 は本論で議論する学習期間調整の重要性を示すヒートマップである。横軸は時期、縦軸は学習期間を表し、色は該当する時期の予測に該当する学習期間を用いた場合の予測結果と実データの相関（黄色が強い相関、青が弱い相関）を示す。赤で示した線が図 3 の結果を得る途中で図 1 のアルゴリズムが選択した学習期間を示している。図 4 は 1954 年から 2016 年にかけて何度か大幅に選択された学習期間が短くなる時期（1975, 1980, 1996, 2012 年）があり、結果として得られるモデルが変化したことを示している。

図 5 は各部分図のタイトルに示した時点（たとえば図 5 最上段左は 1989 年 1 月）で GBDT が作成したモデルが翌月価格が下降すると判断した領域（横軸が d_t 、縦軸が s_t ）を青で示した図で、1996 年以降は

- 価格が暴落 ($d_t < -0.05$ かつ $s_t > 0.029$) した翌月の価格は下がる。
- 調整局面 ($|d_t| < 0.05$ かつ $s_t < 0.029$) の翌月の価格は反転する。
- それ以外では上昇する。

と比較的安定したモデルであることを示している。

また、このモデルは上記 3 つの状況に対応した分布の混合分布を表している。すなわち、分析の対象とした TOPIX

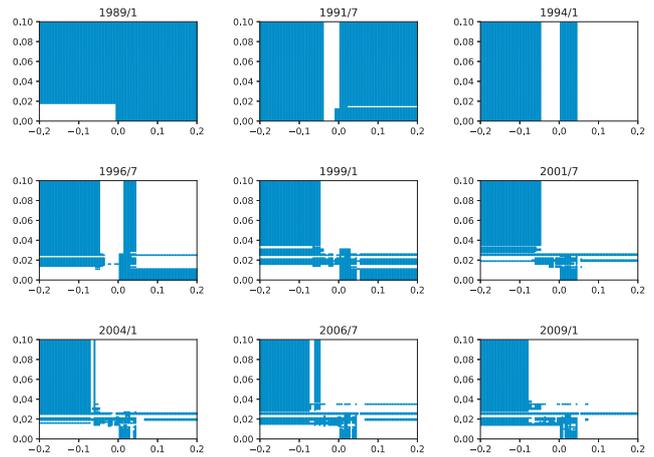


図 5 TOPIX 月次：売買規則

Fig. 5 Rule of Monthly TOPIX.

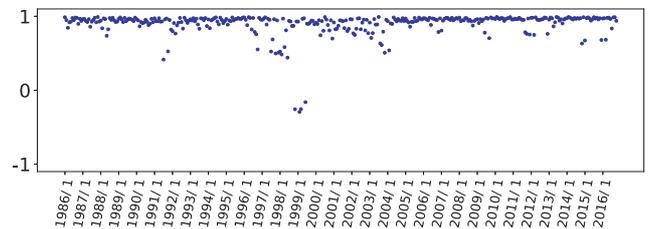


図 6 TOPIX 月次：相関の変化

Fig. 6 Correlation of Monthly TOPIX.

月次価格の推移データは 1 つの分布から生成されているデータではなく、3 つの状況それぞれに対応する分布の混合分布から生成されていると考えるのが素直である。

図 6 は $-0.2 \leq d_t \leq 0.2$, $0 \leq s_t \leq 0.1$ の範囲 (d_t は 0.002 刻み, s_t は 0.001 刻み) で翌月の株価の変動を予測した値に関して、ある月に作成したモデルで予測した値と次の月に作成したモデルで予測した値の相関を計算したもので、モデルの作成時点が翌月になっても予測値（すなわち学習結果の予測モデル）は大きく変わらないことを示している。これについては日次データのところで再度議論する。

4. 日次データの分析結果

4.1 TOPIX 日次

図 7, 図 8, 図 9, 図 10 は図 3~図 6 を TOPIX の日次データを使って書き直した図である。TOPIX の日次データは株式会社日本取引所グループから提供いただいた 2015 年 1 月から 2019 年 12 月までの秒速データを用いた。図 7 では比較の対象としてインデックス投資を用い、標準偏差 s_t を計算するときに 1 秒ごとのデータを $p_{t,m}$ として使用したもの（点線）、1 分ごとのデータを使用したもの（実線）、2 分間隔のデータを使用したもの（破線）を比較した。また月次と同じく $v_t = (d_t, s_t)$ とした。

図 7 に示したように、1 秒ごとと 1 分ごとのデータを使用した場合、赤で示したインデックス投資に対して超過収



図 7 TOPIX 日次：収益比較
Fig. 7 Results of Daily TOPIX.

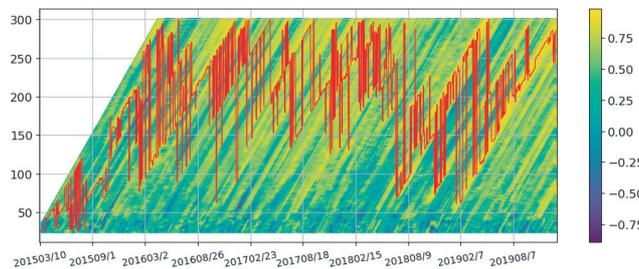


図 8 TOPIX 日次：学習期間
Fig. 8 Learning Period of Daily TOPIX.

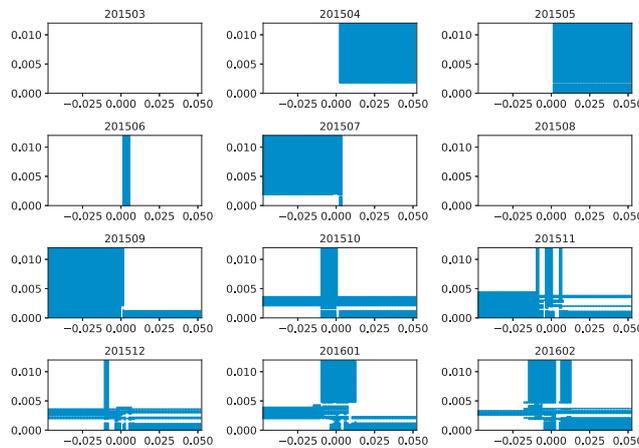


図 9 TOPIX 日次：売買規則
Fig. 9 Rule of Daily TOPIX.

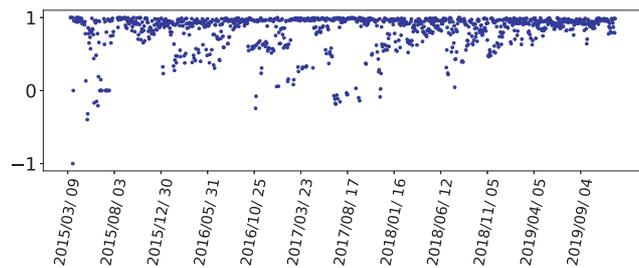


図 10 TOPIX 日次：相関の変化
Fig. 10 Correlation of Daily TOPIX.

表 2 TOPIX 日次のシャープ・レシオ*

Table 2 Sharpe Ratio of Daily TOPIX*.

手法	1分ごと	1秒ごと	2分ごと	インデックス投資
シャープ・レシオ	0.069756	0.050229	0.010728	0.005343

*：近年の金利は 0 に近く、シャープ・レシオ計算のための安全資産の金利は 0 とした。

益を得ている。またシャープ・レシオ (表 2) と期間全体での収益は 1 分ごとのデータを使用した場合に最大となっている。

先に述べたように、提案手法は「投資家が終値の変化や標準偏差で表現される株価の推移を見て判断した結果が将来の株価の推移に影響を及ぼしている」という仮説をベースに、その判断基準のモデル化を試みている。仮説ではあるが、我々は 1 秒ごとのデータを用いるより 1 分ごとのデータを使用した場合の超過収益が大きかったことと、2 分ごとのデータでは超過収益が得られなかったことを、「目視で株価の推移を見るときに、秒単位のデータは粒度が細かすぎる。一方 2 分ごとのデータでは必要な情報が欠落する」と考えている。投資家が株価チャートを目視しながら投資判断を行っているのであれば、秒単位のデータの粒度が細かすぎるという実験結果は自然な結果と理解できる。

図 8 は月次と同じく日次においても学習期間により予測値と実測値の相関に違いが生じ、学習期間の調整が必要であることを示している*1。

一方提案手法が各日付で取り出した売買規則は月次と異なり安定しない*2。たとえば 2015 年 4 月に取り出した売買規則 (図 9 最上段、真ん中) と 2015 年 7 月に取り出した売買規則 (その下) は真逆である。

この不安定性は 1 日ごとに作成したモデルで予測した翌日の株価の予測値の、作成日を 1 日ずらした場合の相関を示した図 10 (対象の日付において、図 10 の範囲で d_t は 0.0005 刻み、 s_t は 0.0001 刻みで予測値を計算した後に、その相関係数を計算し、横軸を日付、縦軸を相関係数として、各日付の相関係数をプロット) にも表れている。月次について示した図 6 が 1 近辺で比較的高い値に集中しているのと対比的に図 10 はマイナスの値をとる場合も多い。これは翌日には著しく異なる予測モデルが作成されたことを示している。

また学習結果のモデルには、図 9 最下段、真ん中のように混合分布を示唆するモデルが多く含まれており、TOPIX 日次においても混合分布の扱いが重要であることは TOPIX 月次と同じであった。

*1 予備実験で学習期間の調整を行わない実装でも実験を行った。結果を付録に示す。

*2 図 9 において、各部分図上端の 6 桁の数字は最初の 4 桁は年、後ろの 2 桁が月を示している。各部分図は 6 桁の数字が示す月で最初に取引のあった日に対して学習した予測モデルを用いて作成した。

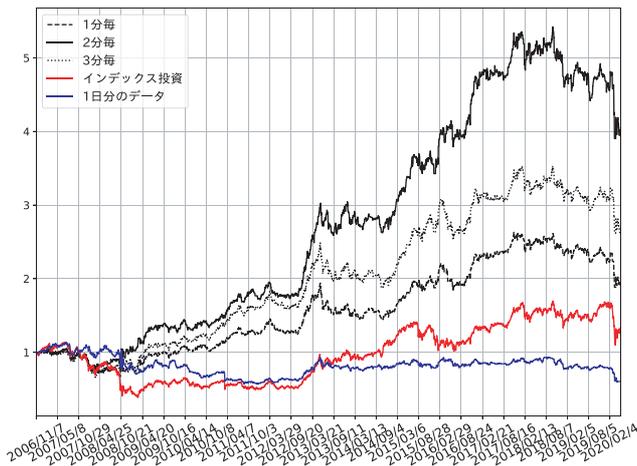


図 11 日経 225 先物：収益比較
Fig. 11 Results of Nikkei 225 Futures.

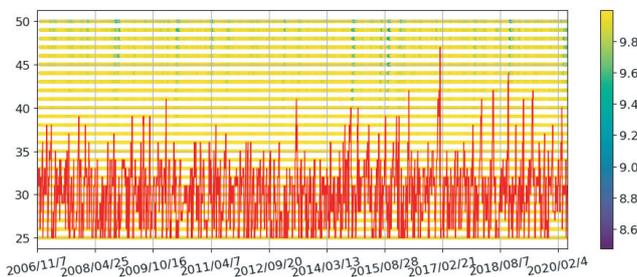


図 12 日経 225 先物：学習期間
Fig. 12 Learning Period of Nikkei 225 Futures.

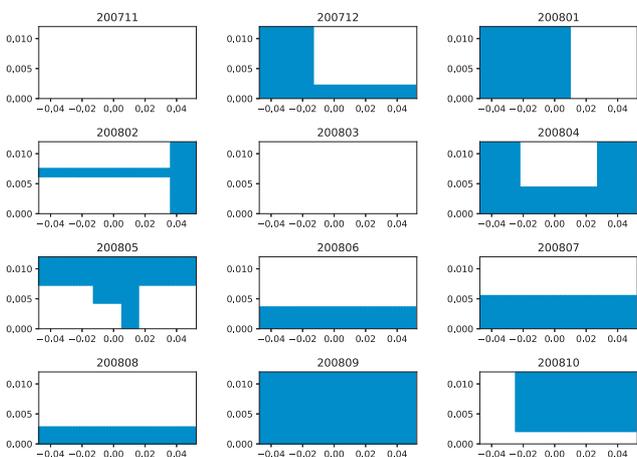


図 13 日経 225 先物：売買規則
Fig. 13 Rule of Nikkei 225 Futures.

4.2 日経 225 先物

同様に図 11, 図 12, 図 13, 図 14 は図 3~図 6 を日経 225 先物の日次データを使って書き直した図である。日経 225 先物の日次データは、2006 年 09 月から 2020 年 4 月のデータを <http://db-ec.jpx.co.jp/category/C410/C410508.html> から入手して用いた。また、先物価格は限月の違いで異なる価格が存在するが、当該日において取引量の多い限月に対する価格を用いた。

日経 225 先物の日次データの場合、標準偏差は 2 分ごと

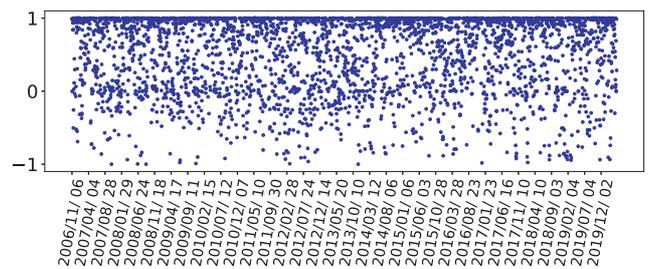


図 14 日経 225 先物：相関の変化
Fig. 14 Correlation of Nikkei 225 Futures.

表 3 日経 225 先物のシャープ・レシオ
Table 3 Sharpe Ratio of Nikkei 225 Futures.

手法	1分ごと	2分ごと	3分ごと	インデックス投資	1日分のデータ
シャープ・レシオ	0.018305	0.031632	0.023520	0.006952	-0.012656

の価格変動から求めた結果 (図 11 実線) の超過収益が最も大きく、3 分ごと (点線)、1 分ごと (破線) の標準偏差を用いた超過収益が、それに続いてインデックス投資の収益 (赤の実線) を上回った。シャープ・レシオも 2 分ごとのデータを用いた場合に最大になっている (表 3)。

TOPIX 日次と異なり、日経 225 先物は 1 日分のデータだけでは翌日を予想できず過去 3 日分のデータを用いた。すなわち $v_t = (d_t, s_t, d_{t-1}, d_{2,t}, s_{t-1}, d_{t-2}, d_{3,t}, s_{t-2})$ とした。図 11 で青の実線は 1 日分のデータだけ、すなわち $v_t = (d_t, s_t)$ (ここで s_t は 1 分ごとの標準偏差) を用いた結果である。

また学習期間 (図 12) も大きく異なり TOPIX の日次の学習期間が 100~300 日であった (図 8) のに対して日経 225 先物の日次では 25~40 日と短かな期間で予測値と実際の値が高い相関を示した。図 12 以外に根拠を持たない仮説ではあるが、先物取引は取引の期限を制約する「限月」(日経 225 先物の場合 3, 6, 9, 12 月) が設けられており、限月を目安に投資家の行動が変化している (限月より前の情報の重要性は低い) 可能性がある。

また売買規則は TOPIX 日次以上に変化している (図 13, 図 14)。ここで図 14 は対象の日付において、図 13 の範囲で d_t は 0.0005 刻み、 s_t は 0.0001 刻みで変化させ、 $d_{t-1}, d_{2,t}, s_{t-1}, d_{t-2}, d_{3,t}, s_{t-2}$ は学習期間の平均値を使って相関係数を計算した。

図 14 に示されたように安定しないモデルが標準的に用いられているインデックス投資に対して大きな超過収益をもたらす (図 11) ことは不可解ではあるが、実験事実でもあり、学習期間調整の重要性を示している。

4.3 価格暴落期の分析

本研究は「投資家が目視で株価の推移を見て判断した結果が将来の株価の推移に影響を及ぼしている」という仮説をベースに、その行動のモデル化を試みている。もしその仮説が正しければリーマン・ショック (アメリカ合衆国の

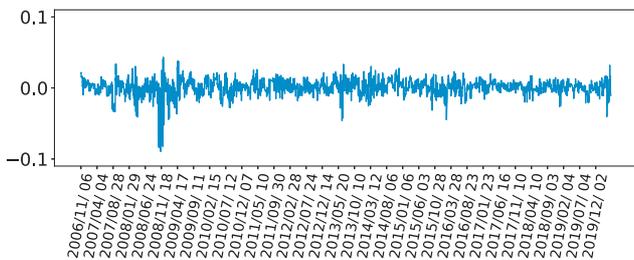


図 15 日経 225 先物：予測価格の平均値

Fig. 15 Average Prediction for Nikkei 225 Futures.

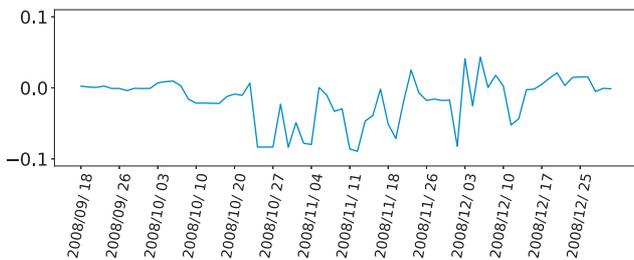


図 16 リーマン・ショック前後の予測価格平均値

Fig. 16 Average Prediction during Lehman Shock.

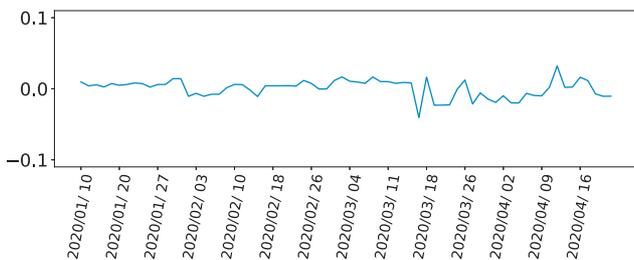


図 17 COVID-19 前後の予測価格平均値

Fig. 17 Average Prediction during COVID-19.

投資銀行であるリーマン・ブラザーズ・ホールディングスが 2008 年 9 月 15 日に経営破綻したことに端を発した金融不安) のような価格暴落期には不安感が投資家の行動変容を促すことが予想される。

日経 225 先物のデータは、リーマン・ショックおよび COVID-19 感染拡大が世界経済に深刻な影響を及ぼした時期を含んでいる。本節では 2 つの金融不安がモデルに与えた影響について分析する。

上記目的のため、日経 225 先物に関して図 1 の 2 行目で GBDT により作成した回帰モデルの入力パラメータ v を $-0.2 \leq d \leq 0.2$, $0 \leq s \leq 0.1$ の範囲 (d は 0.002 刻み, s は 0.001 刻み) で変えて計算した結果の平均値を図 15 に示す。これは、2 つの金融不安の期間の間は作成した回帰モデルの上記範囲での計算結果の平均値がマイナス側に偏る傾向を示すか、確認するためである。

図 16 はリーマン・ショックの前後、図 17 は COVID-19 感染拡大時期について図 15 を拡大した図である。図 16 に示したようにリーマン・ブラザーズ・ホールディングスの破綻 (9 月 15 日) の 1 月強後から約 2 月、予測の平均値はマイナスになり、我々の予測どおり金融不安時には回帰モ

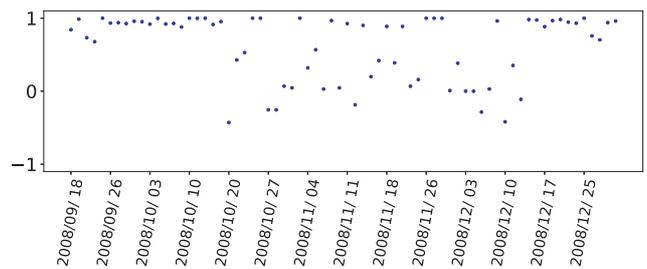


図 18 リーマン・ショック前後の相関変化

Fig. 18 Correlation during Lehman Shock.

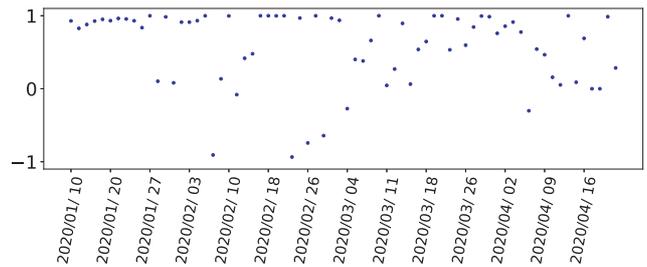


図 19 COVID-19 前後の相関変化

Fig. 19 Correlation during COVID-19.

デルの計算結果がマイナス側に偏る結果となった。学習期間の長さ (25~40 営業日) を考慮すると、モデル作成に直ぐさま実際の価格の変化が反映されなかったと考えるが、このモデルで超過収益が得られるということは、リーマンショックによる金融不安は投資家の行動変容に実際の価格変化より長く影響を及ぼしたと考えられる。

一方 COVID-19 の感染拡大は、図 17 から見る限り、投資家にそのような行動変容をもたらしていない。この結果は IMF の COVID-19 に関する報告「Investors seem to be betting that lasting strong support from central banks will sustain a quick recovery even as economic data point to a deeper-than-expected downturn, as shown in the June」[1] と整合している。

また図 18、図 19 はリーマン・ショックの前後と COVID-19 感染拡大時期について図 14 を拡大した図である。日が変わると異なるモデルが学習されて相関が低い現象は、これらの金融不安時期も同じであり、価格下降を示唆するモデルが毎日続いて学習されるというような現象は確認できなかった。

5. おわりに

本報では、日次の終値の変化と標準偏差のみを入力に用いた GBDT 法が代表的な株価インデックスである TOPIX や日経 225 先物の翌日の値を予測可能であり、代表的な取引手法であるインデックス投資と比較して超過収益が得られることを示した。予測に用いた方法は月次データの分析に有効性が示された手法を素直に日次に適用したものであり、混合分布が分析可能な GBDT 法に学習期間の調整手

順を組み合わせたものである。この結果、

- 月次データに関して報告されている代表的な株価インデックスが持つ特徴と同じ特徴が日次データにも存在することを示した。
すなわち、株価の分析における混合分布分析と学習期間調整の重要性が日次と月次という異なる時間の尺度において重要であることを示した。
- 実験結果が示唆する限り、日次データの方が月次データよりも学習期間の調整が重要であり、学習結果のモデルも日付により大きく変化することを示した。
- 日経先物の分析結果から、限月という先物特有の要因が有効な学習期間に影響している可能性を指摘した。

月次データより日次データの方が日ごとの学習モデルの変化が大きいことや、限月という先物特有の要因が学習期間に影響を及ぼしていることは、学習期間の調整に関する今後の研究の重要性を示していると考えられる。

謝辞 本研究は JSPS 科研費 19H04165 の助成を受け実施した。また TOPIX 日次の秒速データを株式会社日本取引所グループに提供いただきましたことを感謝します。

参考文献

[1] Adrian, T. and Natalucci, F.: Financial Conditions Have Eased, but Insolvencies Loom Large (2020), available from (<https://blogs.imf.org/2020/06/25/financial-conditions-have-eased-but-insolvencies-loom-large/>).

[2] Basu, S.: Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis, *The Journal of Finance*, Vol.32, No.3, pp.663–682 (1977).

[3] Bhattacharjee, I. and Bhattacharja, P.: Stock Price Prediction: A Comparative Study between Traditional Statistical Approach and Machine Learning Approach, *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*, pp.1–6 (2019).

[4] Chong, E., Han, C. and Park, F.C.: Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies, *Expert Systems with Applications*, Vol.83, pp.187–205 (2017).

[5] Fama, E.F. and French, K.R.: Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics*, Vol.33, No.1, pp.3–56 (1993).

[6] Feng, F., He, X., Wang, X., Luo, C., Liu, Y. and Chua, T.-S.: Temporal relational ranking for stock prediction, *ACM Trans. Information Systems (TOIS)*, Vol.37, No.2, pp.1–30 (2019).

[7] Fischer, T. and Krauss, C.: Deep learning with long short-term memory networks for financial market predictions, *European Journal of Operational Research*, Vol.270, No.2, pp.654–669 (2018).

[8] Hamilton, J.D.: *Time Series Analysis*, Princeton university press (1994).

[9] Hatcher, W.G. and Yu, W.: A survey of deep learning: Platforms, applications and emerging research trends, *IEEE Access*, Vol.6, pp.24411–24432 (2018).

[10] Hu, Z., Liu, W., Bian, J., Liu, X. and Liu, T.-Y.: Listening to chaotic whispers: A deep learning frame-

work for news-oriented stock trend prediction, *Proc. 11th ACM International Conference on Web Search and Data Mining*, pp.261–269 (2018).

[11] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.-Y.: Lightgbm: A highly efficient gradient boosting decision tree, *Advances in Neural Information Processing Systems*, pp.3146–3154 (2017).

[12] Li, H., Shen, Y. and Zhu, Y.: Stock price prediction using attention-based multi-input LSTM, *Asian Conference on Machine Learning*, pp.454–469 (2018).

[13] McLean, R.D. and Pontiff, J.: Does academic research destroy stock return predictability?, *The Journal of Finance*, Vol.71, No.1, pp.5–32 (2016).

[14] Selvin, S., Vinayakumar, R., Gopalakrishnan, E., Menon, V.K. and Soman, K.: Stock price prediction using LSTM, RNN and CNN-sliding window model, *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp.1643–1647, IEEE (2017).

[15] Sharpe, W.F.: The sharpe ratio, *Journal of Portfolio Management*, Vol.21, No.1, pp.49–58 (1994).

[16] Yoshida, K.: Analysis of Investors' Behavior through Non-Time Series Analysis of Stock Prices, *JPX Working Papers*, Vol.28 (2019), available from (<https://www.jpx.co.jp/english/corporate/research-study/working-paper/index.html>).

[17] Yoshida, K. and Sakurai, A.: Short-term stock price analysis based on order book information, *Information and Media Technologies*, Vol.10, No.4, pp.521–530 (2015).

[18] Zhang, L., Aggarwal, C. and Qi, G.-J.: Stock price prediction via discovering multi-frequency trading patterns, *Proc. 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.2141–2149 (2017).

[19] Zhao, Y., Shen, Y., Zhu, Y. and Yao, J.: Forecasting wavelet transformed time series with attentive neural networks, *2018 IEEE International Conference on Data Mining (ICDM)*, pp.1452–1457, IEEE (2018).

[20] 中川 慧, 今村光良, 吉田健一: 価格変動パターンを用いた市場予測 k-Medoids Clustering with Indexing Dynamic Time Warping の株式市場への適用, *電気学会論文誌 C (電子・情報・システム部門誌)*, Vol.138, No.8, pp.986–991 (2018).

付 録

A.1 予備実験の結果

予備実験で学習期間の調整を行わない実装でも実験を行った。図 A.1 は TOPIX 月次データ, 図 A.2 は TOPIX 日次データ, 図 A.3 は日経 225 先物日次データの結果を示し, それぞれインデックス投資の収支と比較している。

「期間調整なし」の実験としての学習期間は「利用できるデータすべてを利用する, 各時期における最長期間」を選択した。これは学習期間に何かしらの固定値を使うと, その値が worst case や best case になり, 公平な比較ができなくなる危険を排除するために最長期間とした。具体的にはつねに最長期間で学習したモデルを選択するよう図 1 のアルゴリズムを修正 (より詳細には Select がつねに利用可能

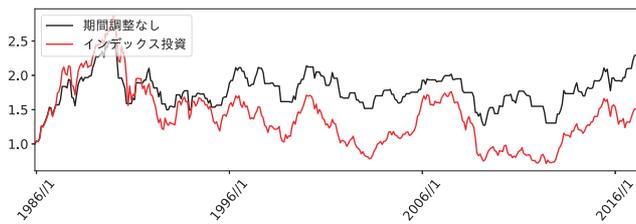


図 A.1 TOPIX 月次

Fig. A.1 Preliminary Experiment of Monthly TOPIX.

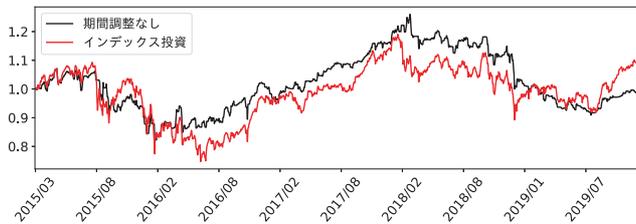


図 A.2 TOPIX 日次

Fig. A.2 Preliminary Experiment of Daily TOPIX.

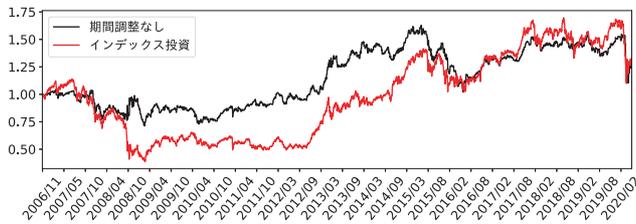


図 A.3 日経 225 先物日次

Fig. A.3 Preliminary Experiment of Nikkei 225 Futures.

な全データを意味する学習期間を選択するように) した実装で実験を行い, その他の条件は図 3, 図 7, 図 11 に結果を示した実験と同じとした. TOPIX 月次の説明変数 v_t は p_t を月末の終値, d_t を $(p_t - p_{t-1})/p_{t-1}$, s_t を月中の日次終値の標準偏差としたときの $v_t = (d_t, s_t)$ とした. TOPIX 日次の説明変数 v_t は上記で p_t を該当日の終値, s_t を該当日の 1 分ごとの値から求めた標準偏差としたもので, 日経 225 先物日次データでは $v_t = (d_t, s_t, d_{t-1}, d_{2,t}, s_{t-1}, d_{t-2}, d_{3,t}, s_{t-2})$ とした.

図に示したとおり, TOPIX 月次の場合, 期間調整なしでもインデックス投資に対して超過収益を得ているが, その値は期間調整した場合の方が大きい. 日次の場合, TOPIX も日経 225 先物も期間調整なしでは超過収益は得られない. 図 4 に示したように TOPIX 月次では 1986 年以降大きく学習期間が違っているのは 4 カ所のみである. 一方 TOPIX 日次と日経 225 先物日次の学習期間の変化はより頻繁である (図 8, 図 12). これらの結果も, 日次のデータを分析するうえで学習結果調整がより重要であることを示している.



吉田 健一 (正会員)

1980 年日立製作所入社. 1992 年大阪大学博士 (工学). 2002 年より筑波大学教授. インターネットと機械学習技術の応用研究に従事. 人工知能学会, 電子情報通信学会各会員.