

ウェーブレット解析を用いた株式市場の構造解明

Clarification of stock market structure using wavelet analysis

瀬之口潤輔¹ 小畑崇弘² 倉橋節也²

Junsuke SENOGUCHI¹, Takahiro OBATA², and Setsuya KURAHASHI²

¹名古屋商科大学

¹Nagoya University of Commerce & Business

²筑波大学大学院ビジネス科学研究科

² Graduate School of Business Sciences, University of Tsukuba

Abstract: Although predicting stock price by machine learning is an important theme both in the academic and business fields, it is also extremely important to elucidate the structure of the background stock market behind the predict model. In this research, we aim to elucidate what kind of factors are in the background of the stock price prediction by using wavelet analysis. From the historical stock price technical index and various financial market data, we create a model to predict the sign of the wavelet coefficient. We also visualize the mechanism of stock price forecasting by elucidating in time series the types of indicators contributing greatly in the prediction model.

1. はじめに

機械学習により株価を予想する際、予想精度を向上することは重要な課題であるが、背景にある市場の構造を解明することも極めて重要である。本研究では、ウェーブレット解析を用いて、将来の株価は現在の金融市場のどのような特徴により影響を受けるかを解明することを目的とする。

ウェーブレット解析は、任意の時系列データを、時間の流れとともに生まれては消えてゆく、周波数の異なる複数の波の和として表現するものである。株価の時系列特性は時間を通じて変化し（非定常データ、トレンド系列を含む）、不規則な変動を繰り返す。このようなデータに対しては、ウェーブレット解析が威力を発揮すること考えられる。この仮説を証明するため、各種金融指標を一時的な複数の波（ウェーブレット係数）に分解し、それぞれの係数が株価予想の重要な要因となっていることを確認することが、本研究の大きな目的である。

本研究では、将来の日経平均株価の動向を予想するため、日経平均株価および各種金融指標から周波数（フィルターレベル）の異なるウェーブレット係数を抽出し、これらを目的変数として、将来の日経平均株価の騰落を予想するモデルを作成する。日経平均株価のウェーブレット係数については、その背後にある観測されない状態が潜んでいる可能性があるという仮説をたて、隠れマルコフモデルにより抽

出した隠れ状態も説明変数として用いる。

またウェーブレット係数と比較するため、一般的に用いられているテクニカル指標も説明変数として用いる。

これらの説明変数を用いると、高い精度で将来の株価動向が予想できることを確認するとともに、どのような説明変数が株価予想に寄与しているかを抽出することにより、株価予想のメカニズムを可視化する。

2. 先行研究

ウェーブレット係数を用いて金融市場の過去の特性を解明する試みはすでに行われている。例えば、[1]は、日本、ドイツ、英国、米国各市場の総合株価指数および業種指数の月次リターンに対してウェーブレット解析を行い、ドイツ、英国、米国の3市場に強い共変動の傾向がみられること、特に英米間でその傾向が強いことを報告している。またウェーブレット係数を用いて株価を予想する試みもすでに行われている[2]。[3]は株価予測について、テクニカル指標のみを使った場合、ウェーブレット解析のみを使った場合、両方を組み合わせて使った場合についてそれぞれ取引シミュレーションを行い、組み合わせた場合の収益が最も安定していたと報告している。[4]は株価指数を含む数種類の時系列データに対してARIMAモデルまたはSARIMAモデルとウェーブ

レット解析を組み合わせることで予測誤差が小さくなることを示した。[5]も米国株価指数の予測についてウェーブレット多重解像度分解とARモデルまたはARMAモデルを組み合わせることで予測誤差が減少したと報告している。一方で、金融市場の動向を予想し、さらに予想の背景にある要因解明を行った研究は存在しない。

3. データ

これらの先行研究を受けて、本研究では、直近の各種金融指標から抽出した異なるレベルのウェーブレット係数、株価ウェーブレット係数の隠れ状態、および各種テクニカル指標から、将来の日経平均株価の騰落を予想するモデルを作成し、さらに株価予想の背景にある構造を可視化する。モデル作成に使用するデータを以下に示す。

3.1 目的変数

本研究では、日経平均株価を予想の対象とする。日経平均株価はTOPIX（東証株価指数）などに比較して古くからデータの提供がある。特に日次の始値に関するデータはTOPIXに比較して約10年古くから取得することができるため、多くのサンプル数を使用することができる。

騰落の予想期間は、説明変数のウェーブレット係数のサポートに合わせて、2営業日間、4営業日間、8営業日間、16営業日間とした。

騰落の判断は、それぞれの予想期間における始値と終値の方向で行う。

3.2 説明変数（金融指標）

説明変数として使用する金融指標を表1に示す。これら全ての金融指標からウェーブレット解析により各レベルのウェーブレット係数を抽出する。ウェーブレットフィルターはハールウェーブレットを使用する。

ウェーブレット係数のフィルターレベルは、騰落の予想期間を最大として全て用いる。つまり翌4営業日間における日経平均株価の騰落を予想する場合、使用するウェーブレット係数は、サポート期間が2営業日及び4営業日の2種類とする。また翌16営業日間における日経平均の騰落を予想する場合、使用するウェーブレット係数が、サポート期間が2営業日、4営業日、8営業日、16営業日の4種類とする。

サポート期間が予想期間より短いものについては、予想期間に対して直近のウェーブレット係数をひと

つだけ用いる。

また観測されない隠れた状態を抽出し、株式市場への影響を確認するため、隠れマルコフモデルにより、日経平均株価のウェーブレット係数から隠れ状態を算出し、説明変数として用いる。使用するウェーブレット係数のサポート全てに対して、隠れ状態を算出し説明変数として用いる。

データのインターバルは、日経平均株価の騰落の予想期間に合わせ、予想期間が終了した後に次の予想期間が始まるものとした。

表1：分析に使用する金融指数

No.	項目名	基準
1	日経平均株価	水準
2	MSCIバリュー株指数	対TOPIX
3	MSCIグロース株指数	対TOPIX
4	S&P500指数	対TOPIX
5	ダウ平均株価	対S&P500
6	ナスダック指数	対S&P501
7	米ドル指数	水準
8	ドル円	水準
9	ユーロ米ドル	水準
10	豪ドル米ドル	水準
11	米10年債利回り	水準
12	米5-30年債利回り差	水準
13	MSCI国債指数	水準
14	金スポット価格	水準
15	S&P500指数	水準
16	S&P500益利回り-米10年債利回り	水準
17	TOPIX RSI	水準

3.2 説明変数（テクニカル指標）

使用するテクニカル指標を表2に示す。日経平均株価の日次終値を用いて、これらのテクニカル指標を作成する。テクニカル指標に対しては、ウェーブレット解析は行わない。

表 2 : 分析に指標するテクニカル指標

No.	項目名	基準
1	RSI_ind	指標値
2	RSI_sig	30以下:-10、70超:+10
3	BB_pct	指標値
4	BB_sig	下振れ:-10、上振れ:+10
5	MA_ind	指標値
6	MA_sig	10日<30日:-1、10日>30日:+1
7	MACD_ind	指標値
8	MACD_sig	下振れ:-1、上振れ:+1

4. 分析手法

4.1 使用したモデル

前項で示した説明変数を使用して、将来の日経平均株価の騰落を予想するモデルを作成する。

モデル作成に使用する手法として `xgboost` を用いた。`xgboost` を用いた理由は、先行研究で日本株の騰落予想には `xgboost` を用いたときの最も高い精度を示したことである[6]。他の手法に比較して `xgboost` が高い精度を示す理由は、日本の株式市場の構造が、複数の要因の線形結合で表すことが難しく、決定木のように複数の条件が重なったときに典型的なパターンを示す傾向があるからであろう。

4.2 予想精度の算出

予想モデルに用いる説明変数は、予想期間の直近のもののみとする。直近以外の情報はすでに株価に織り込まれているという前提を用いる。

まず最も古い 250 個の訓練データを用いて予想モデルを作成し、この予想モデルによる 251 個目の日経平均株価予想し、実績との正誤を記録する。次に訓練データの最初は固定し最後を 1 期分ずらし、251 個の訓練データを用いて予想モデルを作成し、この予想モデルにより 252 個目の日経平均株価を予想し、実績との正誤を記録する。これを最も新しいデータまで繰り返し、全ての予想に対する正誤率を算出する。

このプロセスを、フィルターレベル 1~4 で行い、サポート期間ごとの予想精度を比較する。

4.3 寄与度の高い説明変数の推移

以上のように、時間とともに訓練データが変化するため、予想モデルも予想の度に少しずつ変化する。それぞれの予想モデルについて寄与度の高い説明変数を抽出し、それらを時系列で表示することにより、株式市場の予想にとってどのような指標が大きな要因となっているかを知ることができる。予想モデルの精度が高ければ、寄与度の高い説明変数の推移は、株式市場の参加者が意思決定をする際にどのような指標を注目しているかを明らかにすることができ、つまり株式市場の構造解明を行うことができる。

5. 分析結果

5.1 予想期間ごとの予想精度の比較

フィルターレベルごとの予想精度を表 3 に示した。フィルターレベル 1~3 では高い予想精度は確認できなかったものの、フィルターレベル 4 では高い正解率を示した。16 営業日のような比較的長い期間の株価騰落について高い精度を示すことは、ウェブレット解析のような非定常データを扱うことに優れている手法を用いることにより、株価の一時的なトレンド等が、株式市場の構造を決定する大きな要因となっていることが推測される。

この傾向は、目的変数として騰落の代わりに変化率を用いた場合にも、同様のことが確認された。

表 3 : 予想期間ごとの予想精度

サポート 期間	正解率	
	騰落	変化率
2営業日	50.8%	50.2%
4営業日	48.3%	52.0%
8営業日	51.1%	54.1%
16営業日	57.7%	56.2%
32営業日	53.7%	52.9%

5.2 寄与度の高い説明変数の推移

本研究で用いた予想モデルは、時間とともに訓練データが変化するため、構造も変化する。予想モデルの寄与度の高い説明変数を抽出して、それらを時系列で示すことにより、株式市場の構造変化を可視化することができる。

サポート期間 16 営業日の予想モデルが特に高い精度を示したため、表 4 に、サポート期間 16 営業日の予想モデルを作成する際に用いられた寄与度の高い説明変数を、一部の時期について示した。

表4：寄与度の高い説明変数の推移

	寄与度の順位									
	1	2	3	4	5	6	7	8	9	10
1997/5/6	53	87	43	1	23	17	71	77	34	75
1997/5/28	53	87	1	23	20	18	15	75	33	32
1997/6/19	53	32	85	23	75	76	18	1	36	40
1997/7/11	53	87	15	19	23	17	61	73	18	85
1997/8/5	87	53	19	23	34	1	15	17	43	58
1997/8/27	53	43	87	19	23	17	76	15	52	34
1997/9/19	1	52	87	53	77	40	23	19	15	76
1997/10/15	53	87	1	19	40	52	15	36	58	76
1997/11/7	52	1	40	15	58	77	19	11	36	71
1997/12/2	52	36	1	11	40	75	25	58	85	70
1997/12/25	1	40	52	36	11	85	77	58	20	25
1998/1/22	1	52	40	58	85	11	36	71	75	20
1998/2/16	52	1	75	11	58	36	85	25	40	17
1998/3/10	40	1	36	52	75	58	11	87	85	17
1998/4/1	1	52	40	87	58	75	11	36	43	15
1998/4/23	52	58	36	40	87	15	19	1	85	53
1998/5/20	53	1	52	87	40	19	15	58	18	77
1998/6/11	53	52	1	85	43	17	45	87	58	71
1998/7/3	87	53	52	85	1	43	40	58	14	18
1998/7/28	52	1	53	85	87	15	58	40	71	51
1998/8/19	53	87	1	43	52	40	36	58	71	18
1998/9/10	52	87	53	58	1	43	85	19	40	50
1998/10/6	87	1	53	19	52	43	58	69	50	40
1998/10/28	87	53	43	58	19	75	85	40	15	69
1998/11/20	53	87	43	40	58	15	85	69	36	77
1998/12/15	53	87	75	15	40	19	50	43	58	17
1999/1/11	53	87	75	15	19	43	69	36	58	32
1999/2/3	53	87	36	40	18	75	43	44	52	65
1999/2/26	53	36	87	43	40	69	45	52	14	30
1999/3/23	53	87	36	43	40	75	29	69	19	20
1999/4/14	87	53	36	43	45	40	78	52	17	71
1999/5/12	53	40	43	36	87	45	75	30	44	58
1999/6/3	31	43	75	40	36	52	18	53	45	44
1999/6/25	53	40	76	87	52	45	43	44	36	71
1999/7/19	53	75	43	40	36	87	45	31	20	11
1999/8/11	40	36	43	44	75	52	58	53	11	45
1999/9/2	53	40	43	75	45	36	44	52	31	22
1999/9/28	40	43	53	36	75	44	31	18	52	19
1999/10/21	40	53	43	44	31	11	52	45	36	80
1999/11/15	43	53	40	31	44	18	87	77	75	45

表5：説明変数の項目番号

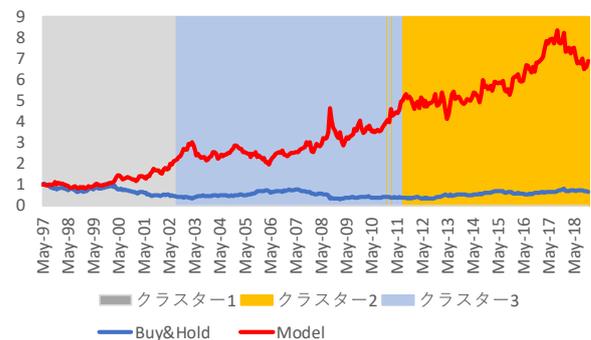
	サポート期間			
	2日	4日	8日	16日
日経平均株価	1	19	37	55
MSCIバリュー株指数	2	20	38	56
MSCIグロース株指数	3	21	39	57
S&P500指数	4	22	40	58
ダウ平均株価	5	23	41	59
ナスダック指数	6	24	42	60
米ドル指数	7	25	43	61
ドル円	8	26	44	62
ユーロ米ドル	9	27	45	63
豪ドル米ドル	10	28	46	64
米10年債利回り	11	29	47	65
米5-30年債利回り差	12	30	48	66
MSCI国債指数	13	31	49	67
金スポット価格	14	32	50	68
S&P500指数	15	33	51	69
S&P500益利回り-米10年債利回り	16	34	52	70
TOPIX RSI	17	35	53	71
日経平均ウェーブレット係数の状態	18	36	54	72

寄与度の高い説明変数の傾向を示すため、全ての説明変数について寄与度を算出し、BIC を使ってクラス数を算出しクラスタリングを行う x-means により、寄与度の高い説明変数が類似しているサンプルを分類した。

さらに分類されたサンプルごとに主成分分析を行い、寄与度が高い説明変数の特徴を抽出し、株式市場の構造解明を試みた。

また図1に各クラスターと、予想モデルによる株価パフォーマンスを示した。

図1：クラスターとモデルによるパフォーマンス



また図 2~4 に、クラスターごとの主成分を示した。

表 2 : クラスター1 の主成分の表示

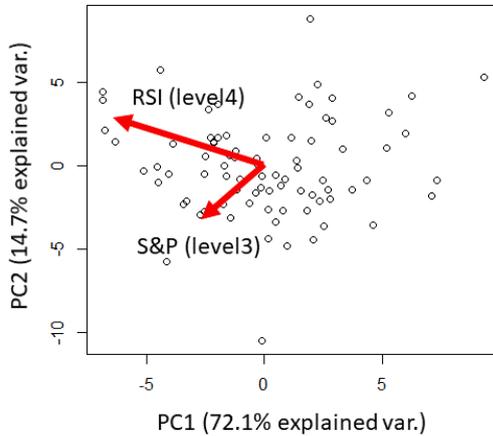


表 3 : クラスター2 の主成分の表示

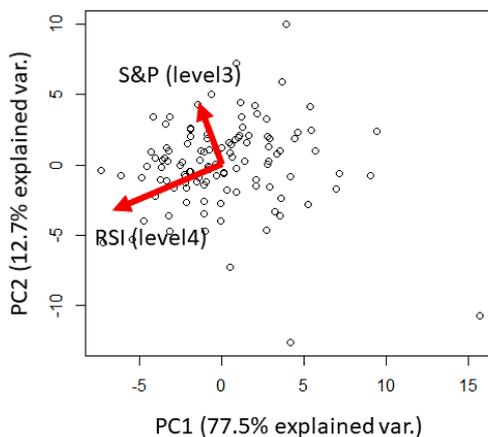
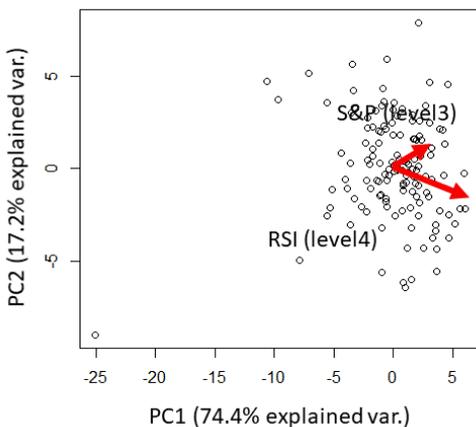


表 4 : クラスター3 の主成分の表示



6. 考察

レベル 4 (サポート 16 営業日) までのウェーブレット係数を用いて、将来 16 日営業日の株価騰落の予想を行うモデルを作成すると、高い予想精度が得られることが確認された。株価やその他金融指標の比較的短い期間の変化は、ニュースフローや需給要因に大きな影響を受けるが、16 日営業日 (1 ヶ月弱) 程度の株価の騰落を対象とした場合、継続的に観測される株価の変動パターンを抽出することができると考えられる。

またレベル 4 までのウェーブレット係数を用いて株価騰落を予想する際、寄与度の高い説明変数の遷移を観測することにより、株式市場の局面が大きく変化していることも示された。具体的には、1997 年から 2002 年を一つの局面、2002 年から 2011 年を二つめの局面、2011 年から直近までを三つ目の局面ととらえると、それぞれの局面で TOPIX RSI のレベル 4 のウェーブレット係数が第一主成分に大きな影響を与え、米国 S&P のレベル 3 のウェーブレット係数が第 2 主成分に大きな影響を与えていることが共通の要因である一方で、主成分得点に与える方向は各局面によって異なっていることも示された。

株式市場の参加者は、長期間にわたりテクニカル指標や直近の株価騰落に注目して投資行動を決定するが、そのロジックは局面が変わると大きく変化することが示された。

本研究では、データのインターバルは、日経平均株価の騰落の予想期間に合わせ、予想期間が終了した後に次の予想期間が始まるものとした。このため、フィルターレベルが大きくなるにつれ、サンプル数は減少する。具体的には、もともと 1980 年 12 月 1 日から 2019 年 1 月 25 日までの 9,388 個の日次データが、フィルターレベルが 4 (サポート期間が 16 営業日) を使用すると、サンプル数は 333 個となり、安定した精度を保證するにはサンプル数が不足していると考えられる。

今後はサポート期間に拘わらず日次でウェーブレット係数を算出できる MODTW 等を用いて、多くのサンプル数を使用する必要があるだろう。

7. おわりに

本研究は、ウェーブレット解析を用いて、将来の株価は現在の金融市場のどのような特徴により影響を受けるかを解明した。株価の時系列特性は時間を通じて変化し、不規則な変動を繰り返すため、ウェーブレット解析が威力を発揮すること考えられた。

研究の結果、16 日営業日間の株価の騰落を予想することにおいて、各種金融指標のウェーブレット係

数を用いると高い精度を示すことが分かった。

また予想モデルにおいて高い寄与度を示す説明変数は、一定期間は同じものが継続し、主成分分析で特徴を抽出することにより、株式市場の構造を可視化することもできた。

一方でウェーブレット解析を行うことによりサンプル数が減少したため、十分に安定した結果を確認することが難しかった。今後は日次でウェーブレット係数を抽出することにより、本研究で示された結果の安定性を確保する必要があると考える。

参考文献

- [1] Antonio Rua, Luis C. Nunes: "International comovement of stock market returns: A wavelet analysis", *Journal of Empirical Finance* 16, pp. 632–639, (2009)
- [2] Jothimani, D., Shankar, R., Yadav, S.S.: *Journal of Financial Management and Analysis*, Vol. 28, No. 2, pp. 35-49, (2015).
- [3] 水野明哲,飯田明由,月成卓也, "Wavelet 解析に基づく株価のテクニカル分析", *可視化情報学会誌* 26, pp. 155~158, (2006)
- [4] Tae Woo Joo, Seoung Bum Kim: "Time series forecasting based on wavelet filtering", *Expert Systems with Applications* 42, pp. 3868–3874, (2015)
- [5] Keyi Zhang, Ramazan Gencay, M. Ege Yazgan: "Application of wavelet decomposition in time-series forecasting", *Economics Letters* 158, pp. 41–46, (2017)
- [6] 瀬之口潤輔,倉橋節也, "実数値 GA による変数選択を用いた株価予想モデル", *経営情報学会 2018 年春季是国研究発表大会*, (2018)