

短期価格予測における連続値パターンの効用

Effect of Continuous Patterns in the Short-term Price Forecast

田中 美栄子

Mieko TANAKA-YAMAWAKI

鳥取大学工学研究科 情報エレクトロニクス専攻

Department Information and Electronics, Graduate School of Engineering,
Tottori University

Abstract: We examine the effect of introducing continuous variables into forecasting of the short-term price changes and compare the result with the case of using discrete signals. Though the price fluctuation is highly random, it has been observed that many arbitrage chances exist in the financial market and they are indeed used by many investors. We first present our result of forecasting the price movement by adaptively selecting the optimum set of various indicators. Although these indicators by themselves continuous, the signal extracted from the technical analysis by means of those indicators is intrinsically discrete. Namely, the question asked in the technical analysis is usually the price range compared to some moving averages. In this paper, we ask a question whether the detailed information such as the slope (first derivative) or the degree of the slope (second derivative) on top of the price changes give us better insight for the short-term price forecast.

1. はじめに

株価や外国為替のような金融価格の変動が全くのランダムでないことは現に多くの投資家が様々なテクニックを使って利益を得ていることから窺い知れるにもかかわらず、金融市場は一応、万人に対して公平なゲームである、もしくはそうであるべきだ、と信じられてきた。然るに金融市場の取引に関する詳細なデータベースが作成され、超短期価格変動を記録した tick データまでが解析対象になってくると、株価や為替など金融価格のもつ様々なパターンが存在することが明らかになってきた。これらの tick データは大まかに言って 2 種の動きをすることが知られている。価格が安定している時はジグザグの動きをするが、不安定になってくるとポジティブ・フィードバックの動きをするため、暴騰や暴落を引き起こす。この二つの相の移り変わる時期を予言できれば、短期予測を行うことが可能である。もちろん社会現象である以上は、9.11 事件のように外因によって価格が大きく動く場合も多い。一方、バブルの崩壊に見られるように、内因による価格の大きな動きは、その原因を作るパラメータを観測していれば力学的説明が可能となり、予測可能なはずである。問題はそのようなパラメータとして何が適当かを見極めることであるが、それが何であるにせよ、価格を測る単位に依らないものであることは必要条件であるはずである。このような考えから、Sornette 等は流体力学の F 値に相当するパラメータを用いて市場の状態を測ることを試みた[1]。

一方、著者等は近年、短時間ごとの、テクニカル指標の最適な組合せとそれに用いるパラメータの最適値を一定の時間間隔ごとに算出しつつ、短期予測を行うシステムを金融 tick データに適用したところ、数分先の価格上下に関して、高い精度で予測ができることを見出した[2,3]。しかしここで用いたテクニカル指標は何種類もの移動平均と現在価格との

上下関係の組合せであり、価格の導関数等の情報を無視したものであった。

そこで本論文では、文献[2]の予測戦略生成法を用いながら、そこでパラメータとして用いたテクニカル指標の組合せの代わりに、上記の F 値ともう一つの独立な、無次元の動的パラメータ T の組合せとを用いて市場の状態を記述することで、予測精度の向上が見られるかどうかを検討する。

2. 予測可能性

市場が完全に効率的であれば、条件付確率が偏ることはないはずである。条件付確率とは、ある事情 B が起こるという条件下での別の事象 A の起こる確率をいい、これを $P(A|B)$ と書く。tick データの 1-tick 間の価格差に対して条件付確率を求めるとある種の安定な規則性が見える。価格が上昇/下降/不変を $+/ - / 0$ として条件付確率を算出する。例えば、3 回連続で価格が上昇した条件下で次に価格が上昇する確率は $P(+|+++)$ と表され、逆に下がる確率は $P(-|+++)$ となる。

これを 1-tick 先の条件付確率を $P_{1\text{-tick}}$ 、10-tick 先の条件付確率を $P_{10\text{-ticks}}$ として、価格の動きの偏りを調べる。このとき、本研究で用いる円ドル為替 tick データ 900 万点を 5 万点ずつのデータセットに分割して条件付確率を求めた結果を図 1、図 2、図 3 に示す。

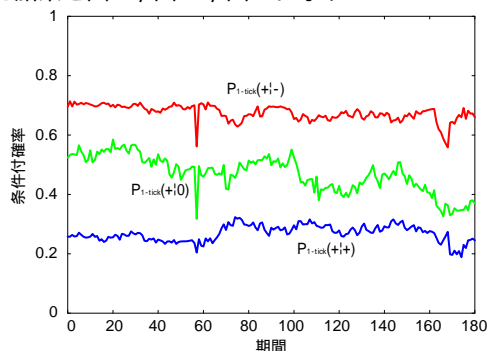


図 1 1-tick 先の価格が上昇する条件付確率

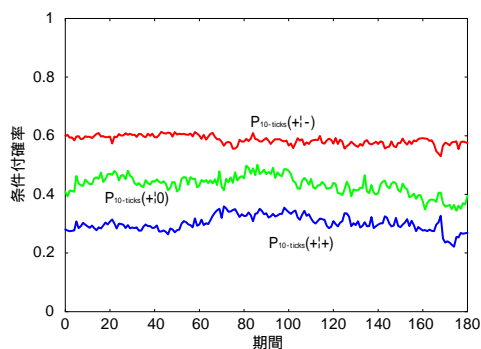


図2 10-ticks 先が上昇する条件付確率

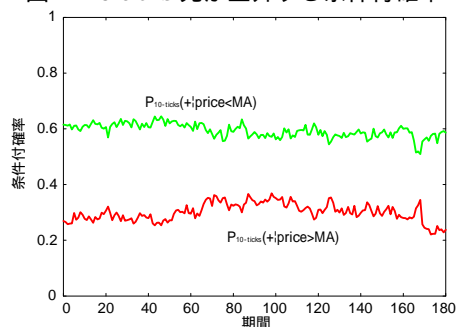


図3 現在価格と移動平均との大小関係を条件として 10-ticks 先の価格が上昇する条件付確率

図1は, down, flat, upの条件下で1-tick先の価格が上昇する確率を表している。この図により, downの条件下では70%の確率で上昇していることが分かる。つまり, 下がった後は上がりやすいという性質が tick データには見られる[4]。

図2は 10-ticks 先の条件付確率を表す。これにより, 現在の価格が 1-tick 前と比較して下降したという条件下では60%の確率で 10-ticks 後は上昇することを意味している。つまり, tick データは過去 1-tick の価格の履歴が 10-ticks 先の価格に影響を与えていることになる。

図3は現在価格とその 50-tick-移動平均の大小関係を前提条件として用いた条件付確率である。前提条件の違いによって 10tick 先で価格上昇する確率と下降する確率の平均値が異なることにより, 移動平均を用いた短期予測が可能となることが示唆される。

以上の結果から, tick データには 10-tick 先以降にも影響があり十分に予測する可能性があると考えられる。以降に, この tick データの定常性を利用した予測手法を試みる[2,3]。

3. 予測プログラムの設計

テクニカル・チャートは数多く提案されており, 大別して, 以下の3カテゴリーがある。

- 1) 価格トレンド(上昇/下降)
- 2) テンド転回(上昇 下降)
- 3) 市場の強度の指標(モメンタム, 出来高, 等)

上記の1)に属するものは何らかの移動平均(MAと略), すなわち時系列をある期間(T)にわたって平均したもの, に関するものが多く, 価格が上がり調子であるか下がり調子であるかを示す。

2)に属するものはトレンドの転換点に関係するものが多

く, 異なる期間 T の移動平均の上下関係などがよく用いられる。例えば, SLMA というチャートは期間 T の短いもの(SMA)と長いもの(LMA)を元の価格時系列に比較して, これらの3時系列の上下関係を判断材料とする。この変形として, SLEMAがある。これは遠く離れるほど指数関数的に減少するような重みを用いた移動平均, すなわち指数平均(EMAと略)の, 期間 T の長いほうの指数移動平均 LEMA と短いほうの SEMA を比較して元の価格時系列との上下関係を判断材料とする。一般にはどれか一つで判断することは少なく, 異なるカテゴリーのチャートを複数組み合わせる用いることが多いようである。

しかしどのような条件下でどのチャートがいいかというような分析はあまりなく, 人ごとに自分の好きなチャートというのがある。気分によって適当に使い分けていたりすることが多い。

しかし実際に使ってみると, 期間 T を少し変えただけで結果が異なることがわかり, どんな場合にも有効な T というのを見つけるのは難しい。ここでは最適なチャートの選択と最適な期間 T の選択とを同時に行うようなシステムの構築を試みる。そのために, 頻用される 10 種類のチャートを選出し, その組み合わせとして最適なものをデータに合うように選出するとともに, 期間 T 等のパラメータの決定も同時に行えるようなプログラムを考案した。図4に全体の処理の流れを示す。

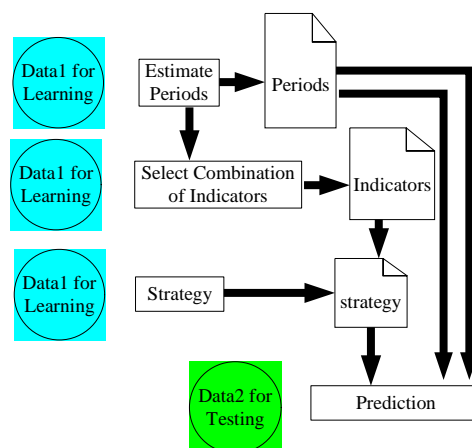


図4 予測システムの処理の流れ

表1 データから推定したパラメータ(円ドル為替 1999)

指標		パラメータ
MA		4
SLMA	SMA	4
	LMA	30
SLEMA	SEMA	3
	LEMA	4
MACD	SMA	2
	LMA	15
	PRICE	2
BB		4
MO1		1
MO2		1
RCI		3
RSI		11
PL		3

初期段階では学習データを用いて最適な各指標のパラメータを決定する。具体的には図4に示すシステムを実行する事により、使用データごとに一番良好な予測結果を出したパラメータを選出するのである。パラメータとは、移動平均に対しては期間 T であり、その他の指標に対しては分割数である。表1に1999年の円/ドル為替 tick データが、一番良好な予測結果を出したパラメータ値を示す。

次に指標の最適な組み合わせを学習データから決定する。表2に1999年の円/ドル為替データの10tick先を予測した場合に得られた最適な組み合わせの上位10個を示す。

表2 データの示す最適な指標の組合せ（円ドル為替1999）

順位	指標の組み合わせ
1	(MO1,MA1,MACD,RCI)
2	(MO2,MA1,MACD,RCI)
3	(MO1,MA1,RCI)
4	(MO2,MA1,RCI)
5	(MO2,MA1,MACD,BB,RCI)
6	(MO2,MACD,BB,RCI)
7	(MO1,MACD,BB,RCI)
8	(MO1,MA1,MACD,BB,RCI)
9	(MO1,BB,RCI)
10	(MO2,BB,RCI)

4. 予測結果

こうして得られた結果を用いて戦略を決定し、予測データに対して予測を行う。1995年から2000年にかけて、年毎に前半を学習データとし、その年の後半を予測データとした。こうして得られた10-tick先の予測結果を表3に示す。いずれの年についても65%-70%の正答率を得ていることがわかる。表3の上段は最上の戦略を用いた場合の結果であり、下段は指標の選び方を変えた場合の10種類の平均値である。実際には分散が非常に小さいことがわかる。

しかし予測的中率は1年間のあいだにかなり変動する。図5に2000年のデータに対する予測結果を示す。日次変化を実線で、週ごとの変化を点線で示してあるが、日次変化は57%位から74%位まで変動し、週次の場合は60%位から73%位までの範囲で変化する。

表3 円/ドル為替：10 tick 先を予測した場合の予測的中率

Year	1996	1997	1998	1999	2000
Best	69.9	66.5	64.9	66.0	67.7
average	69.9	66.5	64.8	66.0	67.7

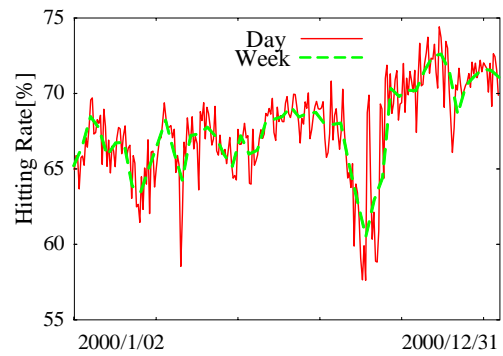


図5 予測的中率の年間変動(円ドル為替2000)

10tick よりずっと先まで予測しようとする、当然的中率は下がる。Xtick 先の予測的中率を図6に示す。1分先なら70%近くの予測率が得られても、10分先は60%程度まで落ちる。

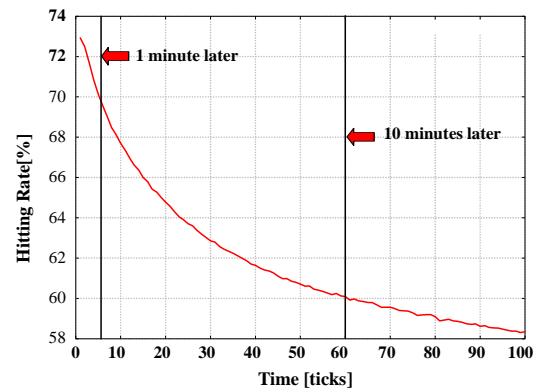


図6 予測的中率の推移 (2000年の円ドル為替：横軸は予測先 tick)

5. 動的パラメータの効用

価格時系列を長さ n の素片に分割し、各素片ごとに時刻 t を $0 \leq t < n$ の範囲に定めるものとする。そうして得た長さ n の時系列 $p(t)$ に対して、初期値、 $p(0)$ と1次導関数の初期値 $p'(0)$ 、2次導関数の初期値 $p''(0)$ を素片ごとに算出する。1次（線型）近似では2次導関数は考えないので

$$p(t) = A + Bt \quad (1)$$

となり、係数 A, B は以下の連立1次方程式の解として求められる。

$$\begin{bmatrix} 1 & t \\ t & t^2 \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} \bar{p} \\ t\bar{p} \end{bmatrix} \quad (2)$$

一方、2次近似では

$$p(t) = \alpha + \beta t + \frac{1}{2} \gamma t^2 \quad (3)$$

となり、係数 α, β, γ は以下の連立方程式の解として求められる。

$$\begin{bmatrix} 1 & \bar{t} & \bar{t}^2 \\ \bar{t} & \bar{t}^2 & \bar{t}^3 \\ \bar{t}^2 & \bar{t}^3 & \bar{t}^4 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma/2 \end{bmatrix} = \begin{bmatrix} \bar{p} \\ \bar{tp} \\ \bar{t}^2 p \end{bmatrix} \quad (4)$$

上記の係数 A , B , α , β , γ から無次元量を作ると、まず考えられるのが、文献 [1] で導入された F 値である。

$$F = \frac{\beta^2}{\alpha \cdot \gamma} \quad (6)$$

これと独立なもう一つの無次元量として

$$T = \frac{n \cdot B}{A} \quad (7)$$

が考えられるので F と T のパターンを利用して同様の予測を行い、これらの使用によって予測率が向上するかどうかを検討する。

1996 年から 2000 年の円ドル為替の tick データを予測対象データとし、離散手法と動的手法を用いて予測を行った場合に得られた最良戦略の予測的中率の平均値を 1-tick 先と 10-tick 先の 2 種類についてそれぞれ Table 2 と Table 3 に示す。離散の場合は、予測指標となる過去の価格差の履歴長 H を 1-5tick に変化させ、最良結果が得られた H における予測率を示す。この場合の予測の最適期間 H は過去 3-tick の価格差履歴を用いた場合である。そのため最適期間での予測戦略の総数は 3^{27} である。

過去 3-tick 区間の動的パラメータを使った場合の予測的中率が、1-tick 先の予測の場合も 10-tick 先の予測の場合に、全ての予測データに対して 0.5-2% 程度向上する結果が得られた。

さらに過去 5-tick 区間の動的パラメータを利用して予測した場合も、過去 3-tick と比べ予測率は下がるが、半分以上の予測データに対し離散の場合を上回る結果が得られた。

しかし過去 10-tick 以上に遡った区間での動的指標によるパターンを利用して予測を行うと、予測率は下降していき、過去 20 - 30-tick 区間の動的指標を用いた予測では、ほぼランダムな結果となった。

表 4 円ドル為替 1-tick 先予測的中率(%)

区間/年	1996	1997	1998	1999	2000
従来手法	72.4	71.1	69.3	68.7	69.4
3-tick	72.8	72.0	69.9	69.6	70.9
5-tick	72.7	70.8	69.7	69.2	70.7
10-tick	68.7	69.0	60.7	62.7	62.4
20-tick	62.1	62.2	57.4	56.6	56.8
30-tick	56.5	58.9	54.4	54.7	53.0

表 5 円ドル為替 10-tick 先予測的中率(%)

区間/年	1996	1997	1998	1999	2000
従来手法	66.9	64.0	62.7	63.9	65.6
3-tick	67.5	65.2	63.7	64.9	66.5
5-tick	66.8	64.3	61.1	64.2	66.0
10-tick	60.4	61.4	53.9	59.4	60.8
20-tick	54.8	54.3	52.3	54.5	54.4
30-tick	52.2	51.8	50.8	53.5	52.0

次に、パターン算出区間を 3-tick から 30-tick まで変化させた場合の予測結果の変化を図 3 に示す。区間が 6-tick 以下の場合に離散の場合の最良値より良い結果が得られている。

予測先を 1-tick 先から 10-tick 先までの間で変化させた場合の予測率の変化を図 7 に示す。動的パラメータ利用により全区間での中率が上昇した。

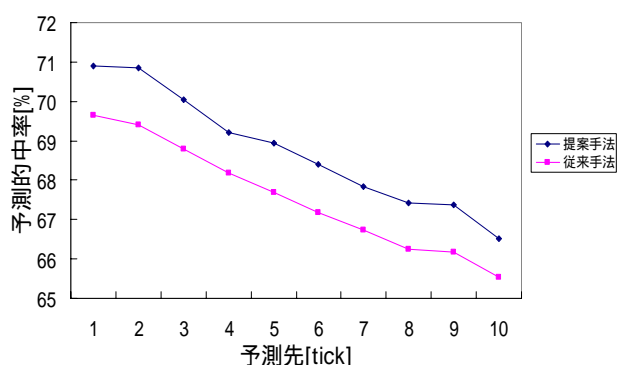


図 7 動的パラメータを指標とした提案手法とチャートの上限関係を利用した従来手法の比較。予測先が 1 - 10tick の全区間での中率が上昇した。

予測率が 1 年の間にどのような変化をするかを検討した結果を円ドル為替の 2000 年頭より年末までの場合について、図 8(1-tick 先予測の場合)と図 9(10-tick 先予測の場合)に示す。それぞれ、従来手法によるものを点線で表し、提案手法によるものを実線(3-tick 区間)、破線(5-tick 区間)、点破線(10-tick 区間)で示している。図 8 においては 3-tick 区間と 5-tick 区間によるものが従来手法による結果より上位にあり、図 9 においては 3-tick 区間によるものが圧倒的に上位にあり、5-tick 区間によるものは従来手法の結果と同程度である。これより区間が長くなると結果は悪くなり、10-tick 区間では図 8,9 とともに低い予測率しか得られない。

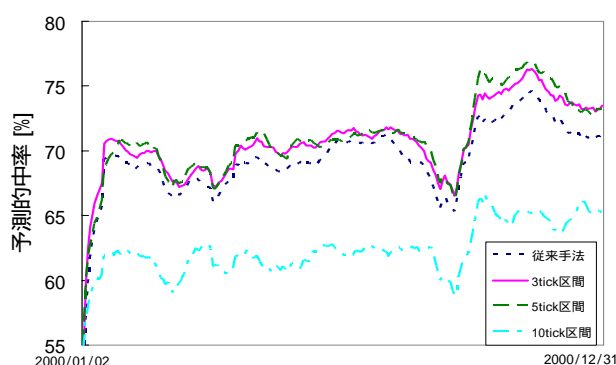


図 8 予測的中率(1-tick 先, 円ドル, 2000)

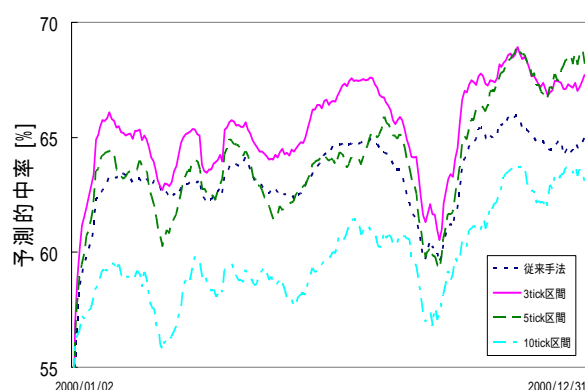


図 9 予測的中率(10-tick 先, 円ドル, 2000)

6. 結果の検討

実験結果より、過去の価格差の履歴、つまり価格変化の上下だけを考慮する相関に比べ、動的パラメータを用いた場合の相関のほうが、tick データの価格変化においては強いといえる。また従来手法で最も予測率が高かった最適期間である 3-tick 区間に対する戦略の総数が 3^{27} と膨大な数であるのに対し、過去 3-5tick 区間の動的パラメータを 9 段階でパターン化した提案手法は、戦略の総数が 3^9 と圧倒的に少なく、なおかつ従来手法よりも高い予測結果を得ることができた。したがって、tick データでの動的パターンを使用した結果が予測の精度・効率両方において上回っているといえる。また予測を行う中でパターンの段階数や閾値の数も変化させたが、最良の結果が得られたのは、F と T を 3 分割した 9 パターンで定義した場合であった。2 分割して 4 パターンで定義した場合や、4 分割以上にした場合、予測的中率は 9 パターンの場合に比べて減少した。

7. まとめ

本研究では、tick データ間に存在する相関に基づき、動的パラメータを用いて高頻度価格時系列の超短期的な予測を行った。その結果、従来手法での価格差の符号のみを予測指標とした従来手法に対し、過去 3-5tick の価格変化の傾向をパターン化して予測

を行うことで 1-2% 程度予測率を向上させることに成功した。また、予測する先を 10-tick とやや長くしても、従来手法に比較して良い予測結果を得ることができた。しかし、過去 10-tick 以上に遡った過去の情報に基づいて予測を行うと、予測率は激減し、過去 20-tick 前後の情報を使うと、価格の変化はほぼランダムになってしまう。この結果より、tick データの価格変化の相関が有効な区間は 3-5tick 程度であり、短期間の価格変化であっても、その相関は単に価格の上下だけでなく 1 次および 2 次の導関数まで考慮することが短期予測を行う上で有効であり、予測率の最大となる条件より、最も強い相関を引き出すパラメータが F と T を 3 分割した 9 パターンであることがわかった。

参考文献

- [1] J.V.Andersen, S.Gluzman and D.Sornette, Fundamental Framework for Technical Analysis, European Physical Journal B14, pp. 579-601, 1999
- [2] M.Tanaka-Yamawaki and S. Tokuoka, Adaptive Use of Technical Indicators for the Prediction of Intra-day Stock Prices, Physica A.383, pp.125-133, 2007
- [3] 徳岡聖二, 田中美栄子, 進化計算による tick 価格変動のトレンド予測, 情報処理学会論文誌：数理モデル化と応用, Vol. 48, SIG 19 (TOM19), pp. 68-74, 2007
- [4] 田中美栄子, 徳岡聖二, テクニカル指標の動的選択と tick 価格予測, 情報処理学会論文誌：数理モデル化と応用 Vol.49 NO. SIG 4 (TOM 20), pp. 88-92, 2008
- [4] M. Tanaka-Yamawaki, S.Tokuoka, and K. Awaji, Short-Term Price Prediction and the Selection of Indicators, Progress of Theoretical Physics, Supplement, Vol. 179, pp.17-25, 2009

連絡先

田中美栄子

E-mail: mieko@ike.tottori-u.ac.jp