慶應義塾大学学術情報リポジトリ
Keio Associated Repository of Academic resouces

Title	ティック・データによる市場価格の短中期予測
Sub Title	Short term prediction of market prices using tick data
Author	櫻井, 彰人(Sakurai, Akito) Serjam, Chanakya( )
Publisher	
Publication year	2019
Jtitle	科学研究費補助金研究成果報告書 (2018.)
JaLC DOI	
Abstract	金融市場の資産価格の短中期 (分～時間) 予測の高精度化を目的とする。特に外国為替交換比率 (EUR/USD, USD/JPY) の分足とtickについて詳細な検討を行った。(1) 1分～90分足について予測モデルを作成し、長期に渡る予測可能性には収益反転が関わっていること、また収益反転にはフラクタル的であることを示した。(2) 分足およびtickそれぞれに、収益反転を説明し収益分布を回帰するモデルを作成した。2001年～2015年の実データに対し、よい回帰となっていることを数値実験により確認した。 The purpose of the research is to improve accuracy of short and medium horizon (minutes to hours) prediction of asset prices in financial markets, specifically focused on minutely and tick movement of exchange rates in foreign exchange markets (mainly EUR/USD and USD/JPY). (1) We successfully modeled 1 to 90 minutely returns of the rates and also showed that the long-term predictability of the model is based on the fractal-like property of return reversals. (2) We successfully built models to explain the return reversals and regress the distribution of the minutely and tick returns, which is supported by simulations for 2001 to 2015 data.
Notes	研究種目：挑戦的萌芽研究 研究期間：2016～2018 課題番号：16K12492 研究分野：機械学習
Genre	Research Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KAKEN_16K12492seika

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　元 年　 ６ 月　 ５ 日現在

機関番号：３２６１２
研究種目：挑戦的萌芽研究
研究期間：2016～2018
課題番号：１６Ｋ１２４９２

研究課題名（和文）ティック・データによる市場価格の短中期予測

研究課題名（英文）Short term prediction of market prices using tick data

研究代表者
櫻井　彰人（Sakurai, Akito）

慶應義塾大学・理工学部（矢上）・教授

研究者番号：００３０３３３９

交付決定額（研究期間全体）：（直接経費）　　2,700,000円

研究成果の概要（和文）：金融市場の資産価格の短中期(分～時間)予測の高精度化を目的とする。特に外国為替交換比率（EUR/USD, USD/JPY）の分足とtickについて詳細な検討を行った。
(1) 1分～90分足について予測モデルを作成し、長期に渡る予測可能性には収益反転が関わっていること、また収益反転にはフラクタル的であることを示した。
(2) 分足およびtickそれぞれに、収益反転を説明し収益分布を回帰するモデルを作成した。2001年～2015年の実データに対し、よい回帰となっていることを数値実験により確認した。

研究成果の学術的意義や社会的意義
金融市場における資産価格変動の重要な例である外国為替市場におけるbid値の変動に関し、収益反転の存在を明確に示し、それを反映した、価格変化である収益の分布をこれまでにない精度で記述できるモデルが構築できたことは、為替変動のリスク評価の精緻化、他の資産の価格変動のモデル化の研究に、大きく寄与すると考えられる。

研究成果の概要（英文）：The purpose of the research is to improve accuracy of short and medium horizon (minutes to hours) prediction of asset prices in financial markets, specifically focused on minutely and tick movement of exchange rates in foreign exchange markets (mainly EUR/USD and USD/JPY).
(1) We successfully modeled 1 to 90 minutely returns of the rates and also showed that the long-term predictability of the model is based on the fractal-like property of return reversals.
(2) We successfully built models to explain the return reversals and regress the distribution of the minutely and tick returns, which is supported by simulations for 2001 to 2015 data.

研究分野：機械学習

キーワード：外国為替証拠金取引　収益分布　ティックデータ　金融時系列

１．研究開始当初の背景

金融市場における価格予測は、我々の研究も含めて過去多くの研究がなされてきたものの、長期間使用した場合の実際の予測精度はさほど高くない。変動方向の予測では50%をわずかに越える程度である。我々の研究では、予測式を適応的に変更する、特に進化計算手法等の大域的探索に基づき予測式を変更する方法では、仮想的な取引において収益を上げることができる一方、予測式を固定して用いる方法では、収益を上げることが難しいことが分かっていた。

予測式を適応的に変更する方法では、予測式が比較的短い時間で変更されるため、すなわち、当該予測式を継続的に用いると収益が得られなくなるため、言い換えれば、当該予測式の精度がよいのは比較的短い期間に限られるため、当該予測式の精度がよい理由、またそれにより収益を上げることができる理由を解明することができなかった。

一方、超短期（秒～分程度）での方向予測では60%以上の予測精度が得られることが分かっている。この場合には、同一の予測式が長期間使用できた。

こうしたことから、長期に渡る短中期(分～時間)予測を可能とする方法の存在が推測され、その発見を通じて、予測の高精度化が可能と考えられた。

なお、予測ができるということは、市場が効率的ではない可能性を意味する。予測可能性を研究することは、市場の効率性の議論に重要な検討材料を与える。

２．研究の目的

市場価格の短中期(分～時間)予測の高精度化が本研究の目的である。また、それを通じて、市場の非効率性の改善に寄与することを目的とする。

３．研究の方法

市場の超短期の動きを解析するには、売買取引が成立した価格（約定価格）だけの分析では不充分であり、より高頻度に発生する売買注文の価格の分析が必要である。また、短期間の予測可能性を研究するためには、個別の注文一件（tick と呼ばれる）ごとの価格変動を対象とする必要がある。

しかし、個別注文での着目する特徴量である価格やその差分である収益は離散量であり、さらに個別注文の発生間隔、すなわち tick 間の時間間隔は、分布する。不定間隔で発生する離散事象の解析は困難である。

本研究においては、組織的な探索方法をとる。一つの方法は、tick データに対し、短時間（分単位。1 分～60 分程度）の「足」を作成し、通常の時系列解析手法を適用し、時間枠の長い方から短い方へ（60 分、30 分、20 分、等）と分析を進める方法である。得られた結果から推測される性質の内、tick 列（時間間隔は考えない）について成立するものを選び、類似する性質の有無を調べ、次に時間間隔の時系列との依存関係を調べるという方法である。

他の一つには、注文データから考えられる特徴量を様々選択し、その時系列に対しデータ圧縮手法や予測手法を適用し、分析すべき特徴量・表現方法を絞り込んでいく方法である。

なお、売買注文の発生状況を表すデータは、公開されておらず、購入可能であっても高価なことがある。これに対しては、従来購入してきたデータに加え、新たに、購入することにより対応する。

４．研究成果

以下論文発表未了の結果については定性的に記載する。データ数が十分あるユーロ/米ドル、米ドル/日本円に関する結果を中心に記す。k 分足(k=1～120)は tick データから作成した。成果はこの分足に関するものと、tick データそのものに基づくものとに分かれる。いずれも、新規の結果が多く含まれるが、特に後者に多いので、後者に重点をおいた記述となる。ただし、論文としての公表はされていないため、数値的記述はさけ、定性的記述またグラフを用いた記述が中心となる。

用語の説明を、使用したデータの説明とともに記す。株価等の価格を有する資産（asset）の価格変動を考察する時への類推から、通貨対の交換比率を価格と考える（事実、価格である。例えば、米ドル/日本円の場合の交換比率は、1 米ドルを日本円で購入・売却する時の価格である)。通貨を取引する（交換する）市場は、電子的にのみ存在する。こうした市場は外国為替市場と呼ばれる。外国為替市場は数多くあるため、交換比率は一意には決まらないが大きく異なることはない。我々は ICAP と呼ばれる市場のものを用いた。

外国為替市場においては、market maker と呼ばれる業者が、売値と買値の希望値（米ドル/日本円の場合、米ドルの売値と買値）を同時に注文として提示する。業者の希望買値は bid、売値は ask と呼ばれる。注文ごと、bid＜ask である。ある時点で最も高い bid、最も低い ask を best bid/best ask と呼ぶ。本稿では、best bid を対象とする。なお、実際の取引価格は、多くの場合、その時点での best bid/best ask のいずれかになるが、market maker の出す注文よりはるかに発生頻度は低い。以下では、best bid/best ask を単に bid/ask と呼ぶ。

1 注文を tick と呼ぶ。Bid/ask 値が変わらない注文が入った場合は、その tick は ICAP のデータには含まれない。しかし、bid 値が同じであるが、ask 値が変わったために、当該データセットに含まれることが多くある。本研究ではどちらも対象としているが、本稿では、bid 値が変わった tick のみを対象とする。

tick 間の時間間隔は分布する。そのため多くの研究においては（株・商品等についても同様）tick をそのまま扱うことはせず、日足、時間足、分足等を扱う。ある一日、ある一時間、ある一分の時間の最後の注文の値を価格（終値）と考える（当該区間の最初の注文を考えることもある。これは始値と呼ばれる）。その時間内に注文がなかった場合、一時点前の価格をその時点の価格とする。本稿では、1、2、4、8～60 分足等をよく用いる。引き続く終値の差を収益（return）と呼ぶ。仮想的に買って売る、また売って買う取引をした場合の差益（差損は負の収益と考える）である。手数料は考えないが、外国為替の場合、bid/ask の差が手数料相当となる。

本稿においては、tick を扱う場合、tick 間の時間差を考えず、注文順に従い、すなわち一つ前の注文との収益を考える。時間間隔を考慮する研究も行っているが、本稿には、その一部を記す。

成果として、分足に関する成果と tick に関する成果とがある。まず分足に関する成果を記す。

(1)分足の時系列としての分析に関し、各分足には十分な依存性（マルコフ性）があり、かつそれが本質的であるため、その性質の成立要因を調べた。

多くの研究では、各足は相互に独立だと仮定する。外国為替の場合、日足は独立だとしても必ずしも不適切ではないが、分足は独立と考えるのは難しい。

実際過去の分足から次の分足の収益が予測できる。ただし、収益額の予測は難しいので、ここでは方向（価格が上がるか下がるか）の予測についての言明である。

本研究においては、k 分足時系列予測モデルを様々作成し適用し、その性質を調べた。予測というのは、ある時点から始まる k 分内の通貨対の交換比率の変化をその時点で終了する n 個の k 分足から予測するものであり、方向が予測可能であるとは、方向予測の正解率が 0.5 を有意に越えることとする。

① 収益値予測については、符号の予測可能性により、ある程度可能であるが、分散が大きく、有意な性質は得られ難かった。しかし、安定的に収益を得ることは可能であった（図 1 右）。

② 収益の符号は明確に予測可能である（図 1 左）。その予測方向を分析すると、参照する過去 n 個の収益符号の内、直前時点の収益符号を最も重視し、予測値はその逆であることが多いことが分かった。また、主要なテクニカル分析によって得られるテクニカル指標を、予測の特徴量として加え、予測を学習させたところ、符号の予測精度の有意な向上は観測されなかった。この性質は、広い範囲の k（1～90 程度）に対して確認できた。その際、学習した線形予測器（ただし、マージン最大化を図る、線形 SVM を用いた）を分析したところ、各予測器は類似していることが分かった。すなわち、収益方向の過去の収益方向への依存性にはフラクタル性が認められた。

次に実際にデータ、すなわち、収益の時系列そのものを分析したところ、収益反転が有意に発生していることが確認できた。収益反転とは、ある時点での収益の符号が、一時点前の収益の符号と逆になる傾向を言う。収益の符号が予測可能である一つの理由は、収益反転が発生していることであることが判明した。

価格の時系列において、収益反転が発生しやすいことは、報告されているが、広い範囲の時間枠（k 分足の k 分）に対して、収益反転が発生することを、すなわち、収益反転はフラクタル的な性質であることを明確に示した初めての報告である。

この収益反転のフラクタル性は、数理モデル的に自明なわけではない。例えば、1 分足と 2 分足で収益反転が起こるということは、ある 2 分の枠の収益と次の 2 分枠での収益は逆符号になることが多く、しかもそれぞれの 2 分枠の中の 2 個の 1 分枠でも収益反転を起こしている可能性が高いのである。これを説明する数理モデルの構築も行った。

この収益反転の存在は、収益値分布に関する統計的性質と矛盾するようにも見える。収益値の分布は、よく知られているように、fat tail になる。しかし、収益反転は fat tail に反するように見えるからである。この矛盾解消も行った。

bid 値で収益反転が観測されることは、意外でもある。実際に発生した取引における価格、約定価格が tick レベルで収益反転するのは自然である。それは、市場参加者が妥当だと考える価格の付近で売買が成立するとすれば、best bid/best ask 間の価格差のため、ある bid 値で約定した後は ask 値で約定し、次には bid 値で約定するというように、上下動をすると推定されるからである。bid 値において（実際には ask 値でも約定値でも）収益反転が観測できるというのは発見であり、数理モデル構築に関する制約かつヒントである。

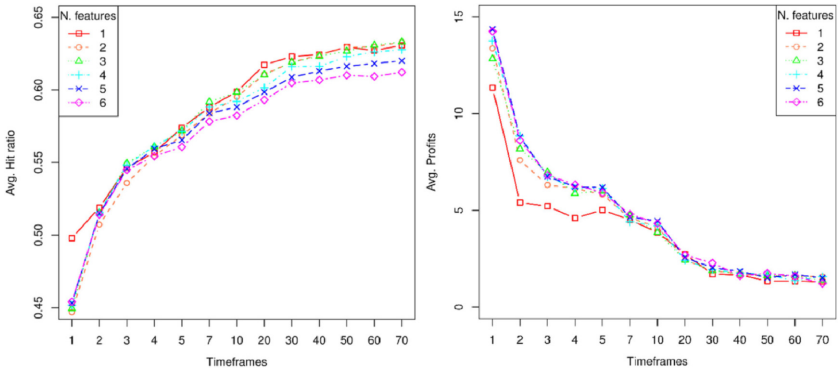

図 1 分足に対する方向予測正解率（左）と平均収益（右）

(2) 実収益分布を表現する分布関数を求めた。本課題は従前から追究を行ってきた課題である。Levy 分布や一般化双曲型分布を始め、いくつかの混合分布を試みたが、十分な回帰精度は得られなかった。今回、新たな分布関数を試み、良い回帰精度が得られることを確認した。

(3) 実収益分布の統計的性質を説明する混合モデルの構築を試みた。通貨交換率の分足の収益分布は、株価指数等と異なり、値が小さい部分（正規化後ドル円 4 銭相当程度）は Laplace 分布に、すなわち指数分布に近くしかも大きな尖度を持っている。また tail は指数分布よりゆっくりと低下する。これらと収益反転を同時に説明するモデルの構築に成功した。

従来金融時系列では、ランダムウォークのように、各注文の発生は独立であると仮定している。そしてその仮定のもと、実収益分布の発生を説明しようとしている。上にも記したように、分足では、そもそも各注文の発生は独立ではない。そこで、過去の試行への依存を適切に設定することにより、収益分布を実現するようなモデルとなる。

当該モデル（M1 と記す）は次のようなことを目指して作成した。M1 は、（収益と同じ）離散値をとる確率変数を m 個持ち、その和を値とする。これらは独立ではなく、適切に設定したある依存関係がある。それぞれの確率変数は、この M1 とほぼ同じ分布に従う。すなわち近似的に再生的である。

この確率変数が互いに独立であれば、正規分布で近似できよう。しかし、依存関係を適切に設定することにより所望の収益分布を作ることを考えたわけである。実際には、M1 とその要素のモデルとは同じではなく、またほぼ同じであることが理論的に示せるものでもない。しかし、実際に作成したモデルの数値実験により目標を達成していることが示せた（図 2 に例）。

つまり、M1 一つで 1 分足を表し、k 個用いれば k 分足が表現でき、一方、M1 のパラメータ（対称であれば 2 パラメータ）を適切に設定することにより k 分足の分布が生成できることになる。現在、2001 年から 2015 年（2011 年、2012 年を除く）の米ドル/日本円の交換比率に関する収益分布で、（数値的かつ近似的にではあるが）その正しさを示すことができる。

なお、回帰精度が不十分と思われる例もあった。2012 年の米ドル/日本円であり、尖度が高いうえに裾が厚く、M1 では、両方を精度よく回帰することができなかった。2011 年の米ドル/円にもその傾向がある。

(4) k 分足の収益時系列について、収益反転を起こす確率は、それ以前の収益反転の確率と密接かつ綺麗な関係があることを発見した。経済状況が大きく変わる 10 年以上のデータを併せても成立する性質であり、類似報告のない非常にユニークな結果である。

次に tick データ分析に関する成果を記す。

(5) 実際の tick のいくつかの統計的性質を説明するモデルを作成した。

① 前述の通り、tick は時間間隔を考慮せず、順序のみを考慮した離散値離散時間の時系列として考えた。この時系列においては、収益を 2 個の連続する tick の大きさの差とした。この時系列においても、収益反転が観測された。収益分布は、分足の分布より尖度が大きく、tail の落ちが遅く、分足のモデルである M1 では、十分には回帰できなかった。つまり、M1 のモデル

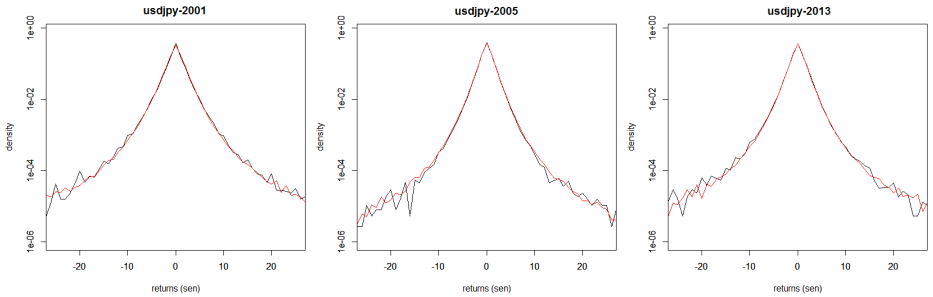

図 2　USD/JPY の 2001 年、2005 年、2013 年の best bid 1 分足（黒線）とモデル M1 による回帰（赤線）

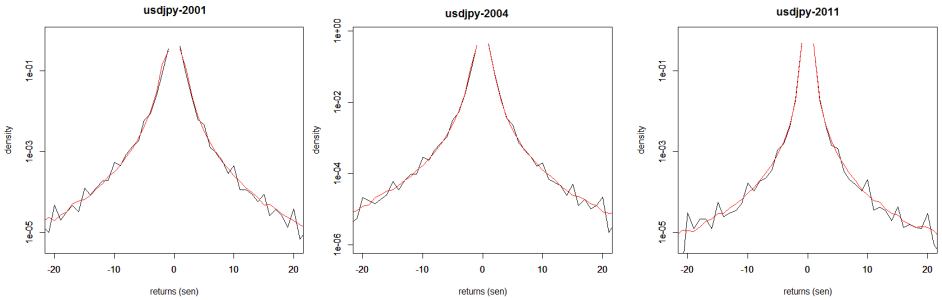

図 3　USD/JPY の 2001 年、2004 年、2011 年の best bid tick 分布（黒線）とモデル M2 による回帰（赤線）

が非常によくできていれば、それを例えば5個組み合わせれば（1分には平均5個のtickがあるとする）1分足のモデルになり、各M1がtickのモデルになったはずであるが、そうではなかった。そこで、異なるモデルM2（対称であれば2パラメータ）を作成した（図3に例）。

なお、実際のtickデータでは、

② tick 間の時間間隔の分布のモデル化を行った。分足の回帰に用いた関数が適用でき、非常に精度よい表現ができた。時間の単位は秒である。用いた実データにおいては、各注文は秒単位で集約している。そのため、1 秒間に複数の異なる注文があった場合、最後のもののみ、また約定については高々2件まで、表現できている。また、2秒、3秒のように短い間隔では、仮にモデルが正確であっても、実データとの差異が発生する。

以上記したように、tickデータを含む高頻度注文データのモデル化に進展が見られた。これらのモデルに基づけば、短中期の収益予測、リスク予測が現在より高精度にできることが予想される。今後、モデルの精緻化、評価、パラメータの推定方法の開発等を進め、実用化につなげたい。

５．主な発表論文等

〔雑誌論文〕（計　２　件）

[1] C. Serjam and A. Sakurai. Analyzing predictive performance of linear models on high-frequency currency exchange rates. Vietnam Journal of Computer Science, 5(2), 123-132, 2018.
DOI: 10.1007/s40595-018-0108-x　（査読あり）
[2] C. Serjam and A. Sakurai. Analyzing performance of high frequency currency rates prediction model using linear kernel SVR on historical data. Lecture Notes in Computer Science, vol. 10191, 498-507, Springer, 2017.
DOI: 10.1007/978-3-319-54472-4_47（査読あり）

〔学会発表〕（計　１　件）

[1] C. Serjam and A. Sakurai. Analyzing performance of high frequency currency rates prediction model using linear kernel SVR on historical data. Ninth Asian Conference on Intelligent Information and Database Systems, ACIIDS 2017, April 3-5, 2017.　（査読あり）

６．研究組織

(1)研究分担者

研究分担者氏名：

ローマ字氏名：

所属研究機関名：

部局名：

職名：

研究者番号（8桁）：

(2)研究協力者

研究協力者氏名：Chanakya Serjam

ローマ字氏名：Chanakya Serjam

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。