

金融経済要因の波及効果を考慮した投資ストラテジー構築による自動売買への適用

2120028 戸田真聡

情報基盤工学講座

指導教員 António Oliveira Nzinga René

要約

1996年の外国為替（Foreign Exchange: FX）取引の完全自由化によりFX取引が誕生してから、年々金融市場の規模は拡大している。現在では多くのトレーダーが、コンピュータを駆使することで自動的にルールに従い取引をするシステムトレードを導入している。また昨今は人工知能を導入することで価格の予測、戦略を獲得するという研究もおこなわれている。本研究では、リアルタイムで取得したTickデータを使用して、LSTMによる機械学習を行い、さらには他市場が円ドル為替市場に与える影響を考えることで未来の値動きを予測し自動的に取引に用いる分析手法を考える。

キーワード：外国為替取引，VAR-LINGAM，RNN

1 はじめに

1996年の外国為替（Foreign Exchange: FX）取引の完全自由化によりFX取引が誕生してから、年々金融市場の規模は拡大している。現在では多くのトレーダーが、コンピュータを駆使することで自動的にルールに従い取引をするシステムトレードを導入している。また昨今は人工知能を導入することで価格の予測、戦略を獲得するという研究もおこなわれている。

従来の為替予測手法では現在から数年遡った期間のデータを使用することが多かった。そのことにより算出した指標が現在の市場の動きに合わないことがある。また、金融市場への他市場からの影響を調べている研究 [1] は存在するものの、為替市場においてそれらを用いて自動売買をする研究は少ないように見受けられる。本研究では、リアルタイムで取得したTickデータを使用して、LSTMによる機械学習を行い、さらには他市場が円ドル為替市場に与える影響を考えることで未来の値動きを予測し自動的に取引に用いる分析手法を考える。

2 波及効果の分析と自動売買の戦略

2.1 金融経済の要因と因果関係

為替市場とは各国通貨間、各国経済力の強さを示す指標であり、米国ドルを基準に自国レート（通貨）がどれくらいの比率になっているかを示したものである。金融市場の拡大や国際通貨の多様化が進むにつれ、国際競争力を示す為替市場は、ビジネスを行うにあたって無視できない重要なファクターの一つになっている。

本研究では因果探索の手法として時系列を考慮した因果性を求めることができるVAR-LiNGAMを採用する。Pythonのモジュール「LiNGAM」を用いてVAR-LiNGAMによる時系列を考慮した因果探索を行う。「LiNGAM」を用いた因果探索ではブートストラップ法を用いることで信頼区間の評価も行い、比較的评价が高いものを選び出すことが可能となる。

ここでは、円ドル為替市場に対して、金、日経平均株価、ダウジョーンズ、WTI原油、ハイグレード銅、英FTSE、英ポンド/日本円、ユーロ/米ドル、プラチナ、米ドル指数の10個の市場が因果性を持つのかを確認する。

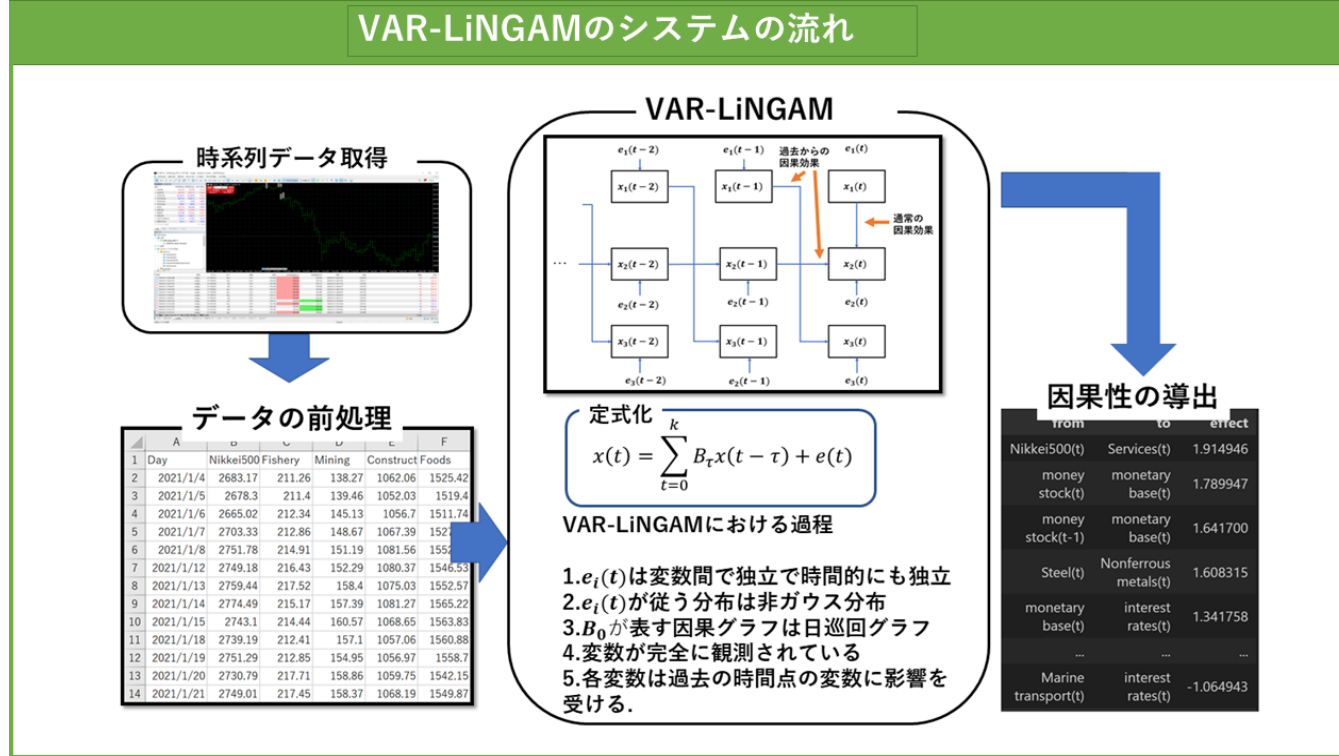


図1 VAR-LiNGAMのシステムの流れ

2.2 投資戦略と自動売買

今回の実験において用いる取引プラットフォームはMetaTrader 5（MT5）である。MT5はFXにおけるテクニカル分析及び取引業務を行うトレーダー向けの無料アプリケーションである。MT5は現在世界で最も利用されているトレードツールであり、様々な機能が存在する。またMT5はデモ口座を開き非常にリアルな取引を行うことができる。Pythonを使用することでMT5からTickデータの取得や取引のオーダーを送ることで実際にMT5で取引を行うことができる。

3 因果関係と時系列予測

3.1 RNNによる因果関係を組み込んだ時系列予測

機械学習(Machine Learning)とは、コンピュータが膨大なデータをもとにして、複数のルールやパターンを学習し、分類や予測する技術である。機械学習のモデルに

は様々なものがある。例えば、時系列解析モデル、回帰モデル、深層学習モデル、アンサンブルモデルなどがある。RNN(Recurrent Neural Network)とは、ニューラルネットワークの一種であり、特に時系列データやシーケンスデータの解析に特化したモデルである。LSTM(Long Short-Term Memory)とは、RNNの一種であり、過去の情報を長期記憶しておく記憶セルの導入により、RNNが持っていた「長期記憶の消失」というデメリットをある程度改善したものである。本研究では、時系列解析モデルに含まれるLSTMモデルを利用する。

3.2 すべての要因と取捨選択した要因での予測比較

特徴量がドル円の終値のみのLSTMによる時系列予測で、↑の予測精度が悪いときはVAR-LINGAMから正の因果性が強いものを5つ選び、↓の予測精度が悪いときはVAR-LINGAMから負の因果性が強いものを5つ選ぶ。

3.3 F値によるモデルの評価

評価指標には、正解率、適合率、再現率、F値を用いた。ここで説明に用いる↑は翌日のドル円の終値が上がることを意味し、↓は翌日のドル円の終値が下がることを意味する。正解率(Accuracy)とは、事例数を分母、その中でモデルが正解した数を分子にした時の値であり、次の(1)式で求めることができる。

$$\frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

適合率(Precision)とは、モデルが↑と予測した数を分母、その中で実際に正解した数を分子にした値であり、次の(2)式で求めることができる。

$$\frac{TP}{TP + FP} \quad (2)$$

再現率(Recall)とは、正解データ中の↑の数を分母、その中でモデルが正解した数を分子にした値であり、次の(3)式で求めることができる。

$$\frac{TP}{TP + FN} \quad (3)$$

F値(F-measure)とは、適合率と再現率の調和平均であり、次の(4)式で求める。

$$\frac{2 * Precision * Recall}{Precision + Recall} \quad (4)$$

また、上の評価指標の補足として、次のような値も用いる。マクロ平均(macro avg)とは、各ラベルの評価値を合計して、単にラベルの種類数で割って平均をとった値である。また重み付き平均(weighted avg)とは、各ラベルの評価値とそのラベル数の積を合計して、事例数で割った値である。例えば、ラベル↑、↓の適合率をP↑、P↓、ラベルの個数をn↑、n↓とすると、適合率の重み付き平均は、次の式(5)で求めることができる。

$$\frac{P \uparrow * n \uparrow + P \downarrow * n \downarrow}{n \uparrow + n \downarrow} \quad (5)$$

混合行列		
	実際は正 (Positive)	実際は負 (Negative)
予測が正 (Positive)	TP(真陽性) True Positive	FP(偽陽性) False Positive 第1種の誤り
予測が負 (Negative)	FN(偽陰性) Flase Negative 第2種の誤り	TN(真陰性) True Negative

図3 混合行列

今回は交差検証を用いて、データセットを3分割し、それぞれで7割を訓練データ、3割を検証データとしてモデルの学習と評価を行い、得られた3つのモデルの評価指標の値を平均することでモデルの評価を行っている。

4 提案手法

本研究では、リアルタイムで取得したTickデータを使用して、LSTMによる機械学習を行い、さらには他市場が円ドル為替市場に与える影響を考えることで未来の値動きを予測し自動的に取引に用いる分析手法を考える。まず、最新のTickデータをMT5経由で取得し、データフレームに格納する。データフレームに格納されたTickデータをリサンブルし、日足のデータフレームとして保存し、VAR-LINGAMによる因果探索やRNNによる時系列予測に利用する。今回はモデルの精度を上げるため、2017年4月3日から2024年10月19日までの日足データを取得し、データフレームに追加した。これらのデータを用いて、VAR-LINGAMによる因果探索を行う。

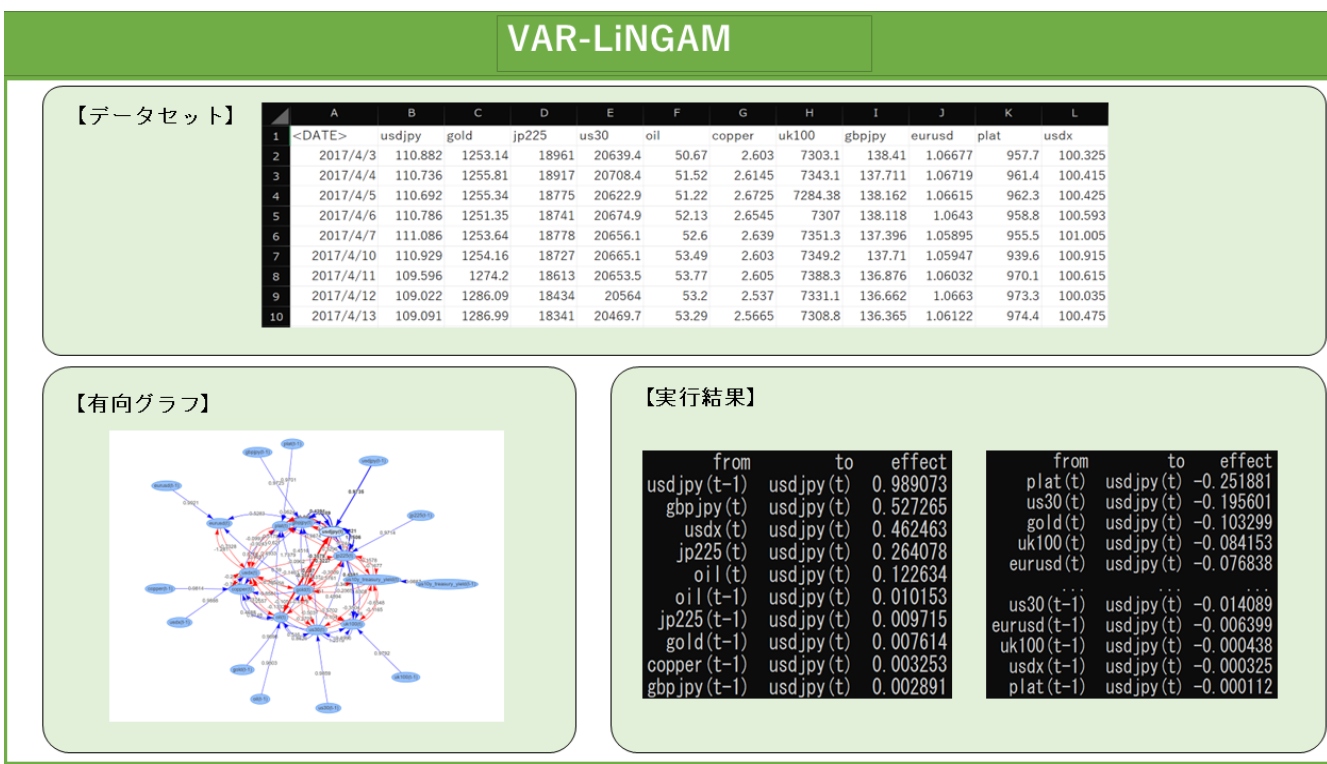


図3 VAR-LiNGAM

次にLSTMによる時系列予測を行う。データの前処理は次のように行った。2017年4月3日から2024年10月19日までの日足データを取得する。翌日のドル円の終値を目的変数とする。特徴量は3パターンあり、一つ目はドル円の終値のみで、二つ目はドル円、金、日経平均株価、ダウジョーンズ、WTI原油、ハイグレード銅、英FTSE、英ポンド/日本円、ユーロ/米ドル、プラチナ、米ドル指数の11個の変数の終値で、三つ目は〇〇である。またどれくらい前までの情報を含むかが重要となるが、今回は過去60日間のデータを用いて、翌日のドル円の終値が前日の終値よりも上がっているか(下がっているか)を予測する。今回データを取得する過程で、変数によって取引時間が異なるため欠損値が出現数する。そのため欠損値を埋める必要がある。欠損値の穴埋めには様々なやり方があるが、今回は11個の変数のうち、一つでも欠損値がある日は削除している。機械学習を行う際にはデータの正規化を行う必要がある。データの正規化にも様々なやり方があるが、今回は最小-最大スケーリング(Min-Max Scaling)というデータを0から1の範囲にスケーリングする方法を用いてデータの正規化を行う。全データをトレーニングデータとテストデータに分割する必要がある。今回は、全データを3分割して、それぞれで7割を訓練データ、3割を検証データとする。

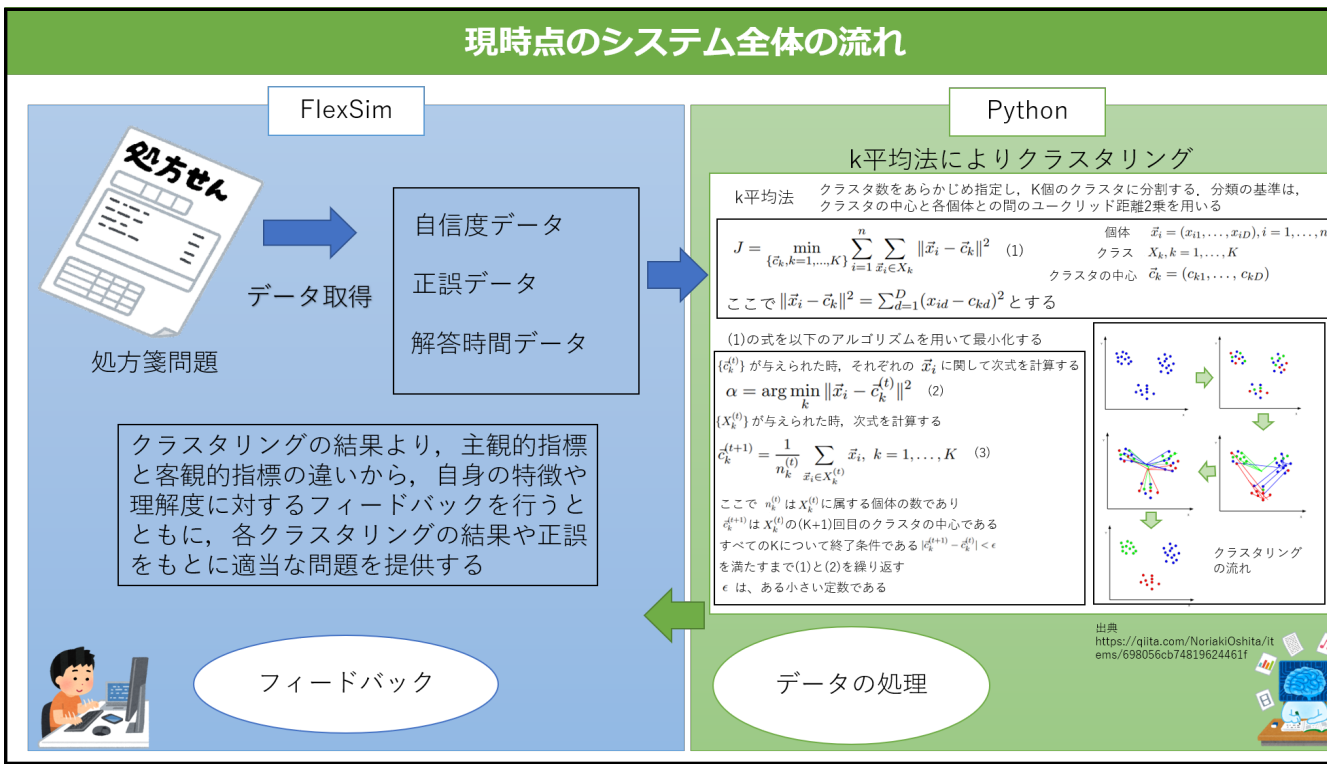


図3 提案手法の概要

モデルの作成、学習等について説明する。3つのLSTM層を構築し、各LSTM層は32ユニットを持つ。ドロップアウト層は、各LSTM層の後に2層ある。ドロップアウトは過学習を抑制するための手段で、ランダムなユニットを無効にする。ドロップアウト率は0.2に設定している。この値は状況に応じて変更できる。最後に全結合層があり、ユニット数は1で出力層となる。最適化アルゴリズムをAdamに設定し、学習率は0.0005に設定している。Adamは一般的に使われている最適化アルゴリズムで学習率の調整を自動で行う。損失関数をMSE(平均二乗誤差)に設定している。今回はエポック数を150に設定し、150回学習させた。ミニバッチサイズを64に設定し、学習データは全データを一度に使わず、小さなバッチに分割して学習させ、メモリの使用効率を向上させる。検証用データを指定し、モデルがトレーニングデータに過剰に適応していないかの確認を行う。

5 数値実験並びに考察

数値実験では、特徴量がドル円のみのLSTMによるドル円予測と、すべての特徴量を含んだLSTMによるドル円予測と、VAR-LINGAMによる因果探索結果を考慮したLSTMによるドル円予測とを比較する。

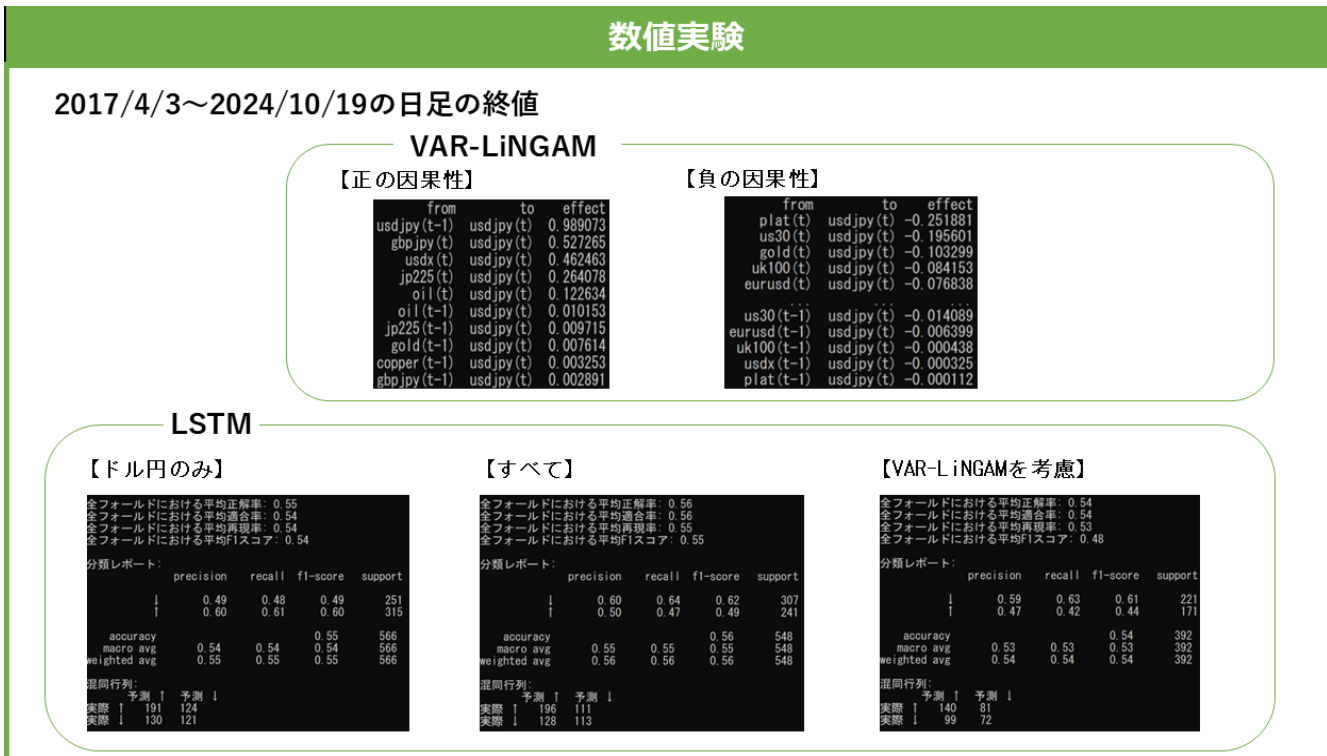


図5 数値実験の結果

6 おわりに

本研究では，リアルタイムで取得した Tick データを使用して，LSTM による機械学習を行い，さらには他市場が円ドル為替市場に与える影響を考えることで未来の値動きを予測し自動的に取引に用いる分析手法を考えた。

参考文献

- [1] 株式会社セールスフォース・ジャパン, “教育現場におけるデジタルトランスフォーメーションの課題と事例”, <https://www.salesforce.com/jp/blog/2021/04/dx-education.html>, 閲覧日, 2022. 11, 02
- [2] 石川 和信, 菅原 亜紀子, 小林 元, 奈良 信雄, “医学教育におけるシミュレータ活用に関する全国調査 2012 ”, 医療教育 44 巻 5 号, pp.311-314, 2013
- [3] 株式会社ゼネテック, “FlexSim ”, <https://flexsim.jp/>, 閲覧日, 2022. 11, 2
- [4] 福坂 祥基, 高木 正則, 山田 敬三, 佐々木 淳, “過去問題をリソースとする知識ベースを活用した問題自動生成システムの開発と評価”, 情報処理学会情報教育シンポジウム, pp. 39-46, 2016
- [5] 板垣 順平, 大坪 牧人, “「臨場感」の再現を試みた遠隔授業の試みとその学修効果 ”, 日本デザイン学会 第 68 回春季研究発表大会, pp. 108-109, 2021
- [6] 倉橋 和子, “分割・併合機能を有する K-Means アルゴリズムによるクラスタリング ”, 奈良女子大学学位論文 2007
- [7] 横内 文香, 齋藤 隆文, 宮村 浩子, “大規模試験での問題分析のための解答状況の可視化 ”, 第 68 回全国大会講演論文集 1 号, pp. 191-192, 2006
- [8] 廣瀬 雄真, 難波 道弘, “学習者の理解度診断に関する基礎的検討 ”, 情報科学技術フォーラム講演論文集 13 巻 3 号, pp. 397-398, 2014