

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, 投資戦略, 自動売買

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 波及効果と自動売買—

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。

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。

2.2 時系列を考慮した因果探索

VAR-LiNGAM は因果探索手法の一つであり、時系列データに対する因果関係の推定を行うためのモデル化手法である。VAR-LiNGAM は 2010 年に提唱され、様々な分野で活用されている。VAR-LiNGAM は VAR モデルと Linear Non-Gaussian Acyclic Model(LiNGAM) を組み合わせたものである。

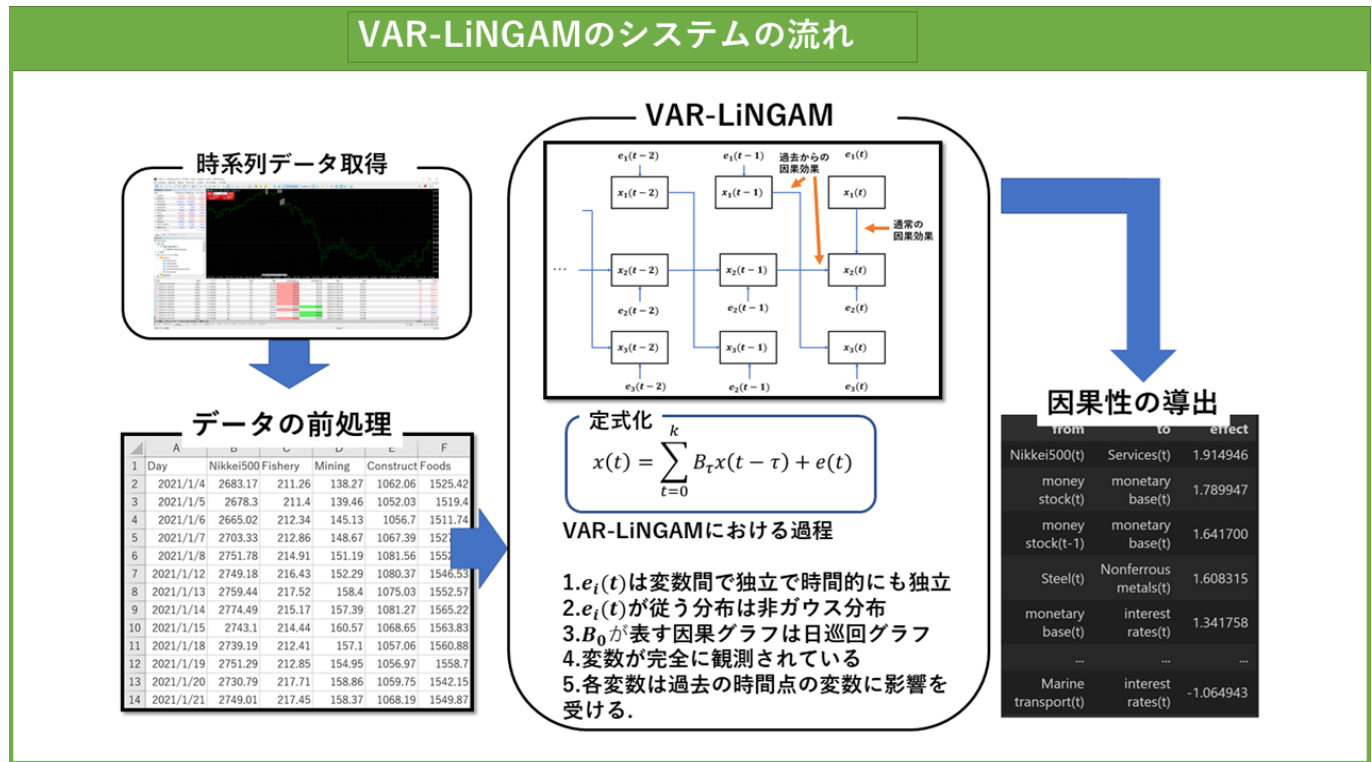


図1 VAR-LINGAMのシステムの流れ

2.3 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 因果関係を組み込んだ予測

今回のLSTMモデルによるドル円予測では特徴量は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、VAR-LiNGAMによる因果探索で求めたeffectの値の絶対値が0.5以上の変数の終値である..

3.3 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。

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。

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。

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。

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 \uparrow , \downarrow の適合率を $P \uparrow$, $P \downarrow$, ラベルの個数を $n \uparrow$, $n \downarrow$ 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。

混合行列											
正解率(Accuracy)	再現率(Recall)	混合行列									
$\frac{TP + TN}{TP + FP + FN + TN}$	$\frac{TP}{TP + FN}$	<table border="1"> <tr> <th></th><th>実際は正 (Positive)</th><th>実際は負 (Negative)</th></tr> <tr> <th>予測が正 (Positive)</th><td>TP(真陽性) True Positive</td><td>FP(偽陽性) False Positive 第1種の誤り</td></tr> <tr> <th>予測が負 (Negative)</th><td>FN(偽陰性) False Negative 第2種の誤り</td><td>TN(真陰性) True Negative</td></tr> </table>		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) False Negative 第2種の誤り	TN(真陰性) True Negative
	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) False Negative 第2種の誤り	TN(真陰性) True Negative									
適合率(Precision)	F値(F-measure)										
$\frac{TP}{TP + FP}$	$\frac{2 * Precision * Recall}{Precision + Recall}$										

图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。

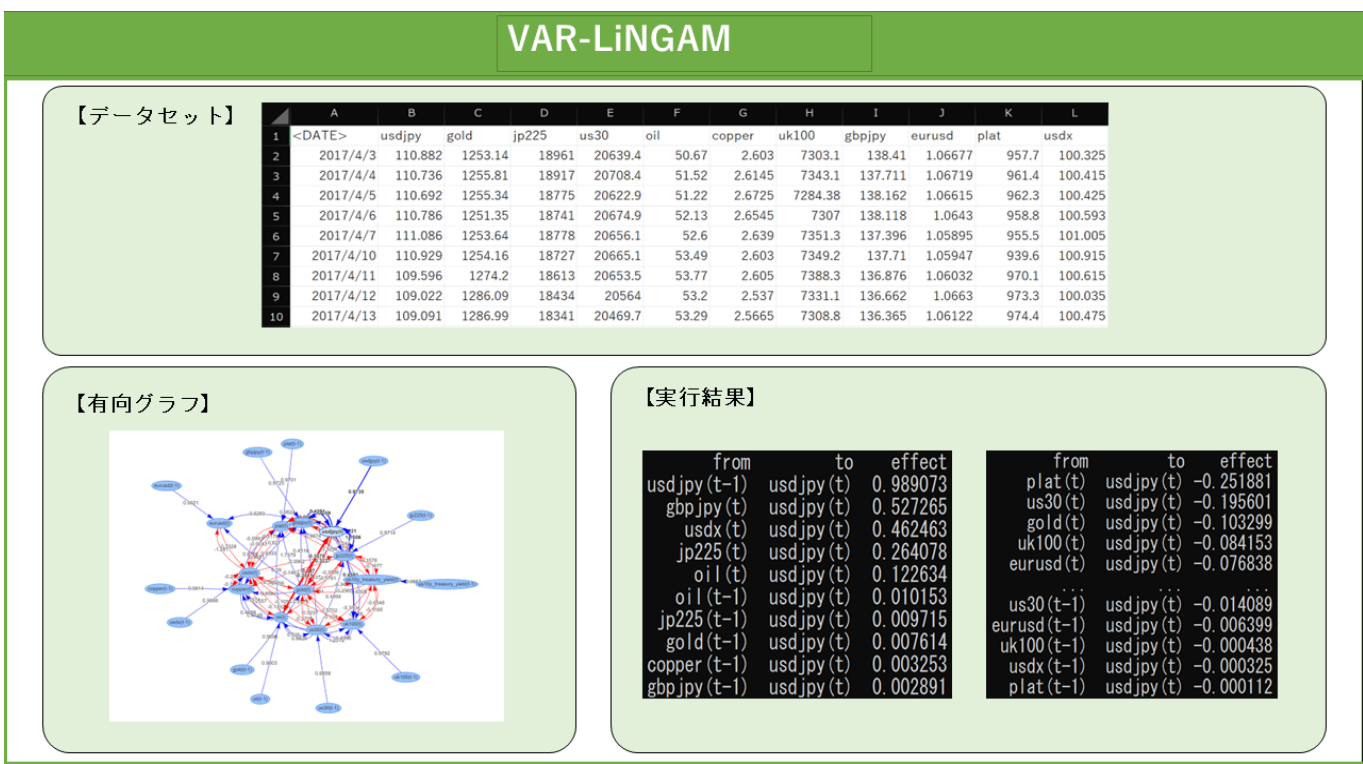


図3 VAR-LiNGAM

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。

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、特徴量がドル円のための 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。

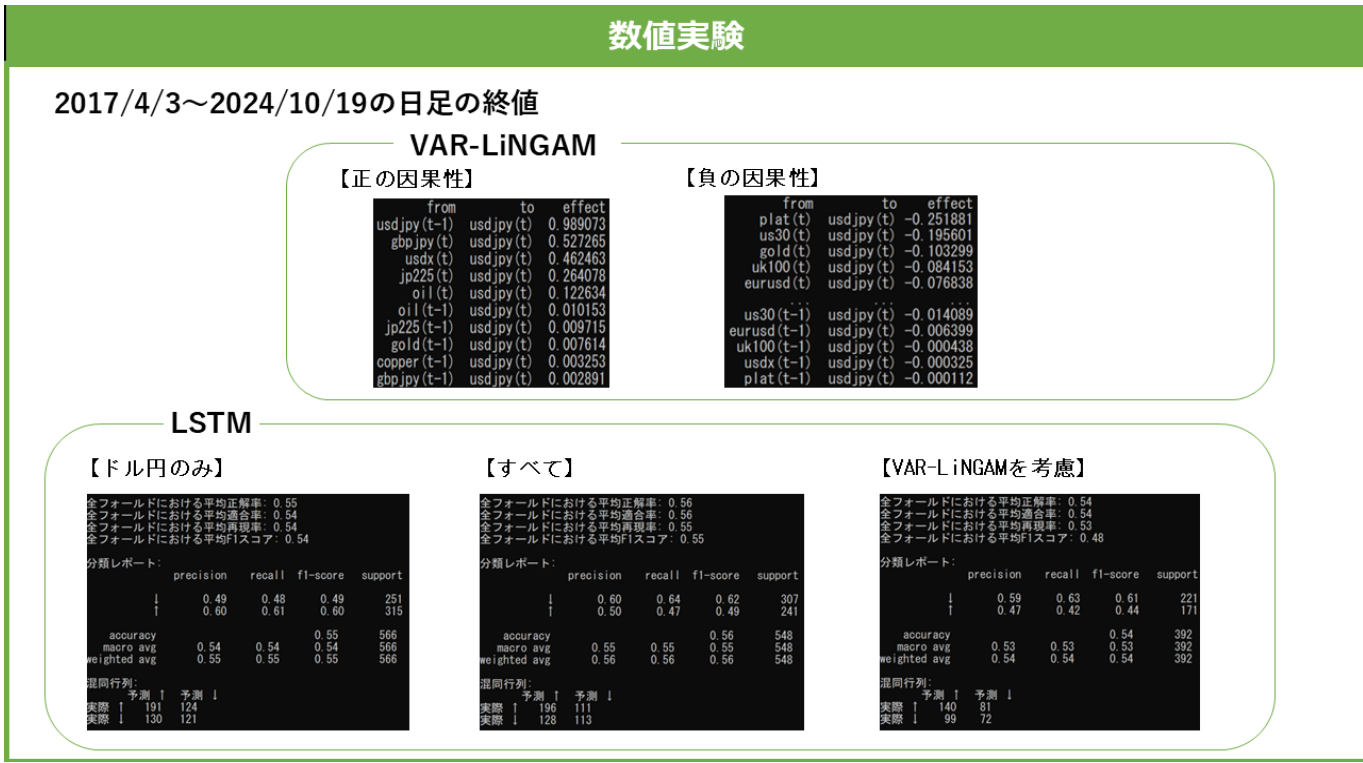


図5 数値実験の結果

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] 松永 諒, 于 海波, "機械学習を用いた暗号資産の収益率予測モデルの構築", 2024.

[2] 蒲田 涼馬, ”経済情報の波及メカニズムの分析による確率的グラフィカルモデルを適用した予測”, 2024	閲覧日, 2024. 10, 27	Vector Autoregression Model Using Non-Gaussianity ”, Journal of Machine Learning Reserch11 (2010) 1709-1731, 2010 年 5 月
[3] ”【sklearn】 Classification_reportの使い方を丁寧に”, https://gotutiyan.hatenablog.com/entry/2020/09/09/111740	引用元, 2020/09/09	Li, Z., Wang, S. Shimizu, P. O. Hoyer, “Estimation of a Structural