

# 金融時系列データ生成に関する研究の調査

## 金融時系列データ生成の手法と評価：最新研究レビュー

### 金融時系列に共通するスタイライズドファクトの整理

金融市場データには、多くの市場や資産クラスに共通して現れる\*\*経験的な統計的特徴（スタイライズドファクト）\*\*があります ([Generation of synthetic financial time series by diffusion models](#))。主要なスタイライズドファクトとして以下が挙げられます。

- **自己相関の欠如（線形予測不可能性）**：価格リターン（対数収益率）の時系列は過去との線形な自己相関がほとんど見られず、統計的にほぼホワイトノイズに近い挙動を示します ([Microsoft Word - 改Jiawei\\_He copy.docx](#))。日次リターンの自己相関関数は1日を超えるラグではほぼゼロに減衰し、過去の価格変動から将来の方向を線形には予測できません。この特徴により、単純なAR（自己回帰）モデルなどでは価格系列を予測不能なランダムウォークと見なすほかなくなります ([Microsoft Word - 改Jiawei\\_He copy.docx](#))。
- **肥厚な分布尾（ファットテール、重尾分布）**：金融リターンの分布は正規分布に比べて裾が厚い（leptokurtic）ことが知られています ([Generation of synthetic financial time series by diffusion models](#))。実務的には、極端な値動き（大損失・大利益）の生起確率が正規想定よりはるかに高く、しばしばパレート型のべき乗則に従うことも報告されています ([Generation of synthetic financial time series by diffusion models](#))。統計的指標では過剰な尖度（excess kurtosis）が正の大きな値を示します。例えば日次リターン系列の尖度は、米国株式（Microsoft）の場合約12.8、原油価格で約7.1、主要為替ペア（EUR/USD）で約4.8と、いずれも正規分布の尖度3を大きく超えています ([Microsoft Word - 改Jiawei\\_He copy.docx](#))。これらはいずれも裾の厚い分布（heavy-tailed distribution）であり、特に株式指数などでは極端値の頻度が高いことを意味します。
- **ゲイン・ロス非対称性（分布の非対称性）**：リターン分布の形は左右対称ではなく、上昇局面と下落局面で統計的性質が異なる傾向があります ([\[PDF\] Stylized Facts - Financial Data - Portfolio Optimization Book](#))。典型的には株式市場指

数では分布が左に長い裾を持ち（負の歪度を示す）、急落のリスクが急騰より高い**負のスキュー**が観測されます (Microsoft Word - 改Jiawei\_He copy.docx)。一方で、個別株式や新興市場では正の歪度（大きな上昇の可能性）を持つケースもあり、資産クラスによって非対称性の方向と程度は異なります (Stylised Facts for Cryptocurrencies: a sector analysis - Medium)。例えば前述の例では、米株式（MSFT）と原油はわずかに負の歪度を示したのに対し、主要為替（EUR/USD）はほぼ対称でした (Microsoft Word - 改Jiawei\_He copy.docx)。このような**ゲインとロスの非対称**は投資家心理や市場メカニズムの偏りを反映しています。

- **ボラティリティ・クラスタリング（変動クラスタリング）：価格変動の大きさ**  
（ボラティリティ）が時間的にクラスタを形成する現象です (Generation of synthetic financial time series by diffusion models)。すなわち、大きな変動が起きる時期は集中的に続き、安定した時期もまとまって現れます (Microsoft Word - 改Jiawei\_He copy.docx)。技術的には**絶対リターン**や**二乗リターンの自己相関**が有意に正であり、しかもゆっくりと減衰することとして測定されます (Microsoft Word - 改Jiawei\_He copy.docx)。この特徴により「大きな価格変化は時間的に連鎖し、小さな変化は小さな変化同士でまとまる」傾向があるといえます。ボラティリティ・クラスタリングの存在下では、単純なランダムウォークや定常的な正規誤差ではデータを説明できず、時変ボラティリティを組み込んだモデル（後述のARCH/GARCHなど）が必要となります (Microsoft Word - 改Jiawei\_He copy.docx)。
- **レバレッジ効果**: 主に株式市場で顕著な現象で、**価格下落が将来のボラティリティ上昇を引き起こす負の相関関係**です ([PDF] Gain/loss Asymmetry and the Leverage Effect - ETH Zürich)。企業の株価が急落すると財務レバレッジ（負債比率）が上がりリスクが増大するという経済メカニズムや、投資家心理の萎縮によるボラティリティ増大が背景にあり、**過去のリターンと将来の分散の間に負の相関**が観測されます。この効果により、悪材料による急落はその後もしばらく市場の不安定性を高める一方、同程度の急騰がもたらすボラティリティ上昇は相対的に小さい傾向があります ([PDF] Gain/loss Asymmetry and the Leverage Effect - ETH Zürich)。レバレッジ効果は株式指数や株価で典型的ですが、為替や商品、仮想通貨では明確な構造的理由がないため弱かったり対称的だったりすると報告されています (A comparison of cryptocurrency volatility-benchmarking new and ...)。
- **集計正規性（Aggregational Gaussianity）**: 短期のリターン分布は大きく歪んでいても、**リターンを長い期間で集計すると正規分布に近づく傾向**があります (Have any new stylized facts of asset returns been discovered since 2001? - Quantitative Finance Stack Exchange) (Have any new stylized facts of

asset returns been discovered since 2001? - Quantitative Finance Stack Exchange)。これは中心極限定理に基づく現象で、例えば日次リターンは裾が厚く偏っていても、月次リターンに集約すると分布の尖度が低下し正規型の形状に近づきます。ただしこの収束速度や形状変化も資産によって異なり、金融市場では**特定のスケールで自己相似的な分布**（スケーリング現象）を示唆する研究もあります。集計正規性はCont (2001) によりスタイライズドファクトの一つとして提唱されました (Have any new stylized facts of asset returns been discovered since 2001? - Quantitative Finance Stack Exchange)。

以上のようなスタイライズドファクトは、株式・債券・為替・コモディティ・暗号資産など多様な資産クラスで広く確認されています。ただし**その強度や具体的な数値特性は資産クラスごとに異なります** (Microsoft Word - 改Jiawei\_He copy.docx)。例えば、株式指数は一般に強い肥尾（高い尖度）と負の歪度（ゲイン・ロス非対称）を示すのに対し、主要為替や国債金利などは尖度がやや低めで分布もほぼ対称的です (Microsoft Word - 改Jiawei\_He copy.docx)。暗号資産（仮想通貨）は歴史が浅いものの、**伝統的資産と同様に厚い分布尾とボラティリティクラスタリングを示し、多くの場合に負のスキュー（急落リスクの偏り）も観測される**と報告されています ([PDF] Comparing Cryptocurrency Indices to Traditional Indices) ([PDF] Risks and Returns of Cryptocurrency)。このような市場間差異を踏まえ、金融データの生成モデルは**各市場のスタイライズドファクトを適切に再現することが重要**となります。

## 金融時系列データ生成手法の概要と学術的貢献

金融時系列を生成するモデルには、大きく分けて古典的な確率モデルと近年のデータ駆動型の深層生成モデル、およびエージェントベースモデルなどが存在します。ここでは代表的な手法の目的、アーキテクチャ、および学術的貢献点を概観します。

### 古典的な確率的時系列モデル（ARIMA, GARCH など）

**ARIMA（自己回帰和分移動平均）モデル**は、時系列の自己相関構造を表現する伝統的手法です。ARIMAは**過去値（AR項）と過去の誤差（MA項）の線形結合**で系列をモデル化し、必要に応じて差分をとることで非定常なトレンドを除去します。目的は主に短期的な予測ですが、金融リターン系列に適用すると、先述の線形予測不可能性ゆえにARやMAの係数はしばしば小さくなり**ほぼランダムウォーク**となるケースが多いです。それでも、例えば**市場マイクロストラクチャによる微小な自己相関**（リターンのわずかなモメンタムや逆張り効果）をモデル化するのにARIMAが利用されることがあります。またARIMA残差に非正規の分布（例えばt分布）を仮定することで肥尾への対応も可能ですが、ボラティリティの変動は捉えられません。

**ARCH/GARCHモデル**は、時系列の**分散（ボラティリティ）の時間変化**を捉えるために開発された確率モデルです (Generation of synthetic financial time series by diffusion models)。ARCH（自己回帰的条件付ヘテロスケダスティシティ）モデルでは、当期の分散を過去の残差の二乗に依存させます。後に提案されたGARCH（Generalized ARCH）モデルでは**分散の自己回帰項**も導入され、現在広く使われる標準的モデルとなりました (Microsoft Word - 改Jiawei\_He copy.docx)。GARCHモデルの目的は、金融リターンの**ボラティリティ・クラスタリング**を再現し予測することにあります。学術的貢献として、GARCH(1,1)モデルの定常解は過剰尖度を持つことが示され、これにより**肥尾分布の再現**にも一定の成功を収めました (Microsoft Word - 改Jiawei\_He copy.docx)。実際、標準的なGARCHモデルでも定常分布は尖度が無限大となりうるため、ある程度の重尾を説明できます (Generation of synthetic financial time series by diffusion models)。さらに、残差分布に $t$ 分布などを仮定したり、GARCHの拡張（EGARCHやGJR-GARCH）によって**非対称なショック効果（レバレッジ効果）の組み込み**も可能となりました。例えばNelson (1991) のEGARCHは対数分散方程式内に非対称項を導入し、悪材料の方がボラティリティへ与える影響を大きくモデリングしています。このように古典的モデルは**少数のパラメータで主要な統計特性を説明する**という学術的簡潔さがあり、金融工学（リスク管理やデリバティブ価格モデル）の基礎として長年発展してきました。

しかし古典的モデルには限界もあります。例えば**複雑な自己相関構造や多峰性分布**、異常な逸脱（ブラックスワンの事象）などを十分捉えるにはモデルが複雑化しすぎる傾向があります (Can GANs Learn the Stylized Facts of Financial Time Series?)。パラメータ推定も高次元になると困難で、モデル誤差の影響が大きくなります。こうした背景から、**データ駆動型で柔軟に分布を学習する生成モデル**への期待が高まり、近年は機械学習・深層学習を用いた生成手法が台頭しています (Generation of synthetic financial time series by diffusion models)。

## 深層生成モデル: GAN (Generative Adversarial Networks)

GAN（敵対的生成ネットワーク）は、画像生成で画期的成果を上げた深層学習モデルであり、2010年代後半から時系列データの生成にも応用されています (Generation of synthetic financial time series by diffusion models)。GANは生成者（ジェネレータ）と判別者（ディスクリミネータ）の二つのニューラルネットワークを競合的に訓練する仕組みです。ジェネレータはランダムノイズからデータを生成し、判別者はそれが本物データか偽物データかを識別します。学習が進むと、ジェネレータは判別者を欺くようなリアルなデータパターンを習得し、最終的に元データと同じ分布を持つサンプルを生成できると期待されます。

金融時系列へのGAN適用の目的は、市場データの複雑な分布や時間依存構造をモデルから直接学習し、リアルな疑似データを得ることにあります ([Can GANs Learn the Stylized Facts of Financial Time Series?](#))。特に過去の学術研究では、GANが価格変動の重要なスタイライズドファクト（無自己相関、肥尾、ボラティリティクラスタ等）を再現できるかが盛んに検証されました ([Can GANs Learn the Stylized Facts of Financial Time Series? - arXiv](#)) ([Can GANs Learn the Stylized Facts of Financial Time Series?](#))。例えばTakahashiら(2019)の**FIN-GAN**は、畳み込みニューラルネットワーク（CNN）と多層パーセプトロン（MLP）を組み合わせた生成器アーキテクチャを採用し、**S&P500指数の系列を学習して主要なスタイライズドファクトを再現可能であることを示しました** ([Fin-GAN: forecasting and classifying financial time series via ...](#)) ([PDF] [Option pricing using Generative Adversarial Networks](#))。FIN-GANはリターンの自己相関が無いこと（線形予測不可能）、肥尾分布、ボラティリティクラスタリング、さらにある程度のレバレッジ効果まで捕捉できたと報告されています ([PDF] [Generative Adversarial Networks: A Novel Approach to Predictive ...](#)) ([PDF] [Option pricing using Generative Adversarial Networks](#))。またWieseら(2020)の**QuantGAN**は、ジェネレータと判別器に時系列用のテンポラル・コンボリューションネット（TCN）を採用し、長期の依存構造や極端値の分布をうまく学習する工夫を行いました ([Quant GANs: deep generation of financial time series - ResearchGate](#)) ([Quant GANs: Deep Generation of Financial Time Series](#))。彼らは特にリターン分布の両端（テール）やボラティリティの自己相関構造に着目した評価を行い、GANにより極端なイベントの頻度を含めた分布をかなり正確に再現できることを示しています。

もっとも、**全てのスタイライズドファクトを完全に学習できるGAN構造は未だ確立されていないのも事実です** ([Generation of synthetic financial time series by diffusion models](#))。Kwon & Lee (2023) の包括的実験では、様々なGANの設計を比較した結果、**アーキテクチャの選択によって時系列特性の再現性能が大きく異なり、安易な設計では金融時系列の複雑な特徴を捉えきれないことが示されています** ([Can GANs Learn the Stylized Facts of Financial Time Series?](#))。すなわちGAN自体は強力な枠組みですが、適切なネットワーク構造（例えばRNNやTCNの活用）、損失関数の工夫、条件付け手法（条件付きGANで市場状態を指定するなど）を取り入れないと、発振やモード崩壊などにより**一部の統計特性が劣化した系列を生成してしまうことがあります** ([Can GANs Learn the Stylized Facts of Financial Time Series?](#))。学術的には、この問題に対処するため**金融分野特有の拡張**が提案されています。例として、生成器の損失関数に経済的な指標を組み込んで**統計特性の再現を直接ペナルティ化**したり、**複数のGANを市場レジーム（相場状態）ごとに使い分けるレジーム条件付きGANの研究** ([Fin-GAN: Forecasting and Classifying Financial Time Series via ...](#))、生成器を自己回帰的に設計して長期記憶を持たせる工夫などがあります ([Quant GANs: deep generation of financial time series -](#)

ResearchGate)。これらの発展により、近年の研究は**金融データ用GAN**でより**高次の統計構造まで再現する**方向に進んでいます。総じて、GANの金融応用の学術的貢献は、**データ主導でスタイライズドファクトを学習・再現する新しいパラダイム**を示した点にあり、十分なデータがあれば伝統的モデルを超える柔軟性を発揮しうることが示唆されています ([Generation of synthetic financial time series by diffusion models](#))。

## 深層生成モデル: 拡散モデル (Diffusion Models)

**拡散モデル (Diffusion Model)** は、GANやVAEに続く新しい深層生成アプローチで、近年画像生成で顕著な性能を示しています。**Denoising Diffusion Probabilistic Model (DDPM)** などに代表され、逐次的にノイズを付加・除去するプロセスを通じてデータの分布を学習します ([Generation of synthetic financial time series by diffusion models](#)) ([Generation of synthetic financial time series by diffusion models](#))。2020年代に入り、金融時系列データ生成にも拡散モデルの適用が検討され始めました。**\*\*Takahashi & Mizuno (2024)\*\***は、**拡散モデルを用いた金融時系列の生成**を提案し、既存手法ではすべて再現できていなかったスタイライズドファクトを高水準で満たせる可能性を示しました ([Generation of synthetic financial time series by diffusion models](#)) ([Generation of synthetic financial time series by diffusion models](#))。

彼らの手法の特徴は、**時間系列を画像に変換してから拡散モデルを適用する点**です ([Generation of synthetic financial time series by diffusion models](#))。具体的には複数の時系列（株価・取引高・スプレッドなど）を**ウェーブレット変換によるスペクトログラム画像**にマッピングし、カラー画像（RGBの3チャンネル）としてDDPMに入力します ([Generation of synthetic financial time series by diffusion models](#)) ([Generation of synthetic financial time series by diffusion models](#))。拡散モデルは画像内のパターンを学習することで、元の時系列に内在する**多変量の時系列パターンや相関**を獲得します。そして学習後、生成した画像を逆ウェーブレット変換することで**合成された金融時系列**を得ます。学術的貢献として、このアプローチにより**複数の関連する時系列を同時に生成**でき、価格だけでなく出来高やスプレッドといった他系列も含めた**総合的なリアリズム**を追求しています ([Generation of synthetic financial time series by diffusion models](#))。結果として、提案モデルは**肥尾分布・ボラティリティクラスタリング・周期的季節性パターン**など、金融データの持つ様々なスタイライズドファクトを高い精度で再現することに成功したと報告されています ([Generation of synthetic financial time series by diffusion models](#)) ([Generation of synthetic financial time series by diffusion models](#))。特に季節性（例: 時間帯や曜日による変動パターン）や複数変数間の同時依存といった高度な特徴も捉えられる点が新規性です。

拡散モデルの利点は、**GANに比べ安定した学習**（モード崩壊が起きにくい）と**高い表現力**にあります (Generation of synthetic financial time series by diffusion models)。一方で計算コストは大きく、また結果の解釈は依然難しい面があります。しかし金融時系列への応用は始まったばかりで、DDPMのような手法がGANやVAEを凌駕して**全スタイルドファクトを満たす合成データ**を生み出せるか、今後の研究の焦点となっています (Generation of synthetic financial time series by diffusion models) (Generation of synthetic financial time series by diffusion models)。拡散モデルの導入は、金融データ生成手法の多様化と精度向上に大きく寄与しており、学術的にも「画像・言語に続き**時系列分野でも生成モデルを高度化する**」という潮流の一端を担うものです。

## エージェントベースモデル (Agent-Based Models)

エージェントベースモデル (ABM) は、金融市場を構成する個々の参加者（エージェント）の行動ルールを定め、その相互作用から市場全体の価格時系列を生成しようとするアプローチです。ABMは伝統的には経済物理学や計算経済学の文脈で発展し、特に**市場のスタイルドファクトの起源を再現するシミュレーション手法**として注目されてきました (Generation of synthetic financial time series by diffusion models)。LuxとMarchesi (1999) やLeBaronらのモデルなど、黎明期の代表的な株式市場ABMでは、エージェントに**ファンダメンタル志向**と**トレンド追従志向**の2タイプを設け、その間の集団行動のスイッチング（例: **羊群行動**による急激な需要偏重）によって**価格に肥尾分布やボラティリティクラスタリングが自然発生**することが示されました (Generation of synthetic financial time series by diffusion models)。これは学術的に重要な成果で、従来は外生的に仮定していた厚い分布尾や変動クラスタが、**エージェントの異質性と相互作用の帰結として内生的に説明**できることを明らかにしたのです (Generation of synthetic financial time series by diffusion models)。

ABMのアーキテクチャは機械学習モデルとは異なり、**エージェントの意思決定ルール**（例: 価格が一定以上上がれば利益確定売り、移動平均ゴールデンクロスで買いなど）と、**市場メカニズム**（注文簿による価格決定や需給曲線の単純モデルなど）をコード化したシミュレーション環境から成ります。研究者はエージェント数やパラメータを調整し、シミュレーションを何度も走らせることで**統計的性質が現実の市場と似通うか**検証します (Generation of synthetic financial time series by diffusion models)。特に**出来高とボラティリティの相関**や**取引頻度分布**など、価格以外の市場統計もABMから得られる利点があります。また近年では、ABMに進化計算や学習アルゴリズムを組み合わせ、エージェントの戦略を適応的に進化させる研究も進んでいます。このような適応的ABMにより、市場が**\*\*構造変化（レジームシフト）\*\***する様子や新規参入者の影響なども再現し、より現実的な長期系列を生成しようという試みもあります。

ABMの学術的貢献は、何と言っても\*\*「市場現象のメカニズム理解と再現」にあります。スタイライズドファクトを単なる統計的な与件ではなく、合理的・非合理的なエージェント行動の帰結として説明できる点が魅力です (Generation of synthetic financial time series by diffusion models)。例えば、先行研究によりボラティリティクラスターリングはエージェント間の模倣行動とフィードバックによって生まれることや、肥尾分布はエージェントの期待形成の揺らぎ（確率的意思決定）により生じることが示唆されています。これは単なる数学モデルでは得がたい洞察です。またABMから生成されたデータは、統計的には現実と区別困難なほどリアルになりつつあります。近年のある研究では、エージェントの行動パターンを大量の市場データから校正（キャリブレーション）し、実データのスタイライズドファクトを高い精度で再現する人工市場を構築した例も報告されています。もっともABMは計算負荷が大きく、またモデル自由度が高いためパラメータ調整が難しいという課題もあります。汎用的な深層学習モデルとは異なり、市場構造ごとにルール設計をカスタマイズする必要があるため、汎用的な適用には手間がかかります。しかしその分、「なぜそのようなデータ特性が現れるのか」を因果的に考察できる\*\*メリットがあり、学術研究としては金融時系列生成の一つの重要なアプローチです。

## その他の生成アプローチ (VAE など)

他にも金融時系列生成のためのさまざまなアプローチが提案されています。**変分オートエンコーダー (VAE)** は確率的生成モデルの一種で、エンコーダ・デコーダのネットワークによりデータを潜在変数空間に写像し学習するものです。Dogariuら (2022) はVAEを金融時系列に応用し、**パラメトリックな時系列モデル（例えば GARCHやHestonモデル）のパラメータ分布**を学習の潜在空間に組み込む工夫を行いました (Generation of synthetic financial time series by diffusion models)。具体的にはエンコーダが実データから推定した「モデルパラメータ」を潜在ベクトルに対応付け、デコーダがそこからデータを再構成することで、従来モデルの知見と学習の柔軟性を組み合わせています (Generation of synthetic financial time series by diffusion models)。このアプローチによって、ある程度**既知の金融モデルの振る舞いを継承しつつ**データ駆動で不足部分を補完する生成が可能となり、従来モデルが苦手とする複雑な分布形状の再現性向上が報告されています。

さらに、時系列特有の生成モデルとしては、**時系列GAN (TimeGAN)** や時系列用フロー変換モデル、RNNを用いた自己回帰生成ネットなども研究されています。Yoonら (2019) のTimeGANはGANとオートエンコーダ、教師あり学習を組み合わせ、時系列の潜在表現と時間的一貫性を同時に学習する枠組みを提案しました。これは金融専用ではありませんが、**高次元のマルチバリエーション時系列データを時間の順序を保って生成**できる手法として注目され、実際に金融分野（例えば異なる経済指標の同時時系列生成など）への適用例もあります。また\*\*正規化フロー (Normalizing Flow)



\*\*を用いた手法では、金融リターン分布の複雑な形状を可逆変換でモデル化し、そこに時系列依存を組み込む試みも見られます。

総じて、「その他」のカテゴリに属する生成手法は、GANや拡散モデルほど大規模には研究されていないものの、**既存モデルの長所を活かしたハイブリッド型**が多い点が特徴です。学術的には新規の深層生成モデルを金融データにそのまま適用するだけでなく、**金融特有の知識（経済モデルや経験的事実）を統合した独自手法**が提案されている状況と言えます (Generation of synthetic financial time series by diffusion models)。これらは個別には実証段階ですが、最終的な目標はいずれも**現実の金融時系列と見分けがつかないほど高品質なシミュレーションデータ**を得ることにあります。

## 生成データの評価指標と検証方法

金融時系列生成モデルの評価では、「**生成データが実際の金融データと同様の性質を持つか**」が最重要となります。学術研究では以下のような指標や検証方法が用いられています。

- **統計指標による分布の比較:** 生成系列の基本統計量（平均、分散、歪度、尖度）や分布形状を実データと比較します。特に**尖度 (kurtosis)** や**歪度 (skewness)** は肥尾性や非対称性の指標として重要です。例えば、あるモデルが生成した株価リターンの尖度が実データの値と近ければ、極端値の頻度が現実的と言えます (Microsoft Word - 改Jiawei\_He copy.docx)。Q-Qプロットで生成データと正規分布や実データ分布を重ね、尾部の逸脱具合を視覚的に確認することも一般的です。またHillプロットなどでべき分布の指数を推定し、実データの尾指数と比較する手法も採られます。
- **スタイライズドファクトの再現検証: Stylized facts** そのものを満たしているか直接確認します (Generation of synthetic financial time series by diffusion models)。具体的には、生成リターンの自己相関関数を計算し**有意な自己相関が無いこと**（または非常に短いラグのみ小さな自己相関）を確認します (Microsoft Word - 改Jiawei\_He copy.docx)。生成データの**絶対リターン**や**二乗リターンの自己相関**を調べ、ボラティリティクラスタリング（長いラグまで有意な正の自己相関）が存在するかを検証します (Microsoft Word - 改Jiawei\_He copy.docx)。さらには過去リターンと将来分散の相関を測り、レバレッジ効果（負の相関）が再現されているかを見ることもあります。例えば、QuantGANの研究では**自己相関関数 (ACF)** や**偏自己相関関数 (PACF)**、**リターンと分散のクロス相関**など複数の関数を評価し、モデル生成データが現実のスタイライズドファクトと整合するかを定量的に評価しています ()。

- **分布適合度や距離尺度:** 生成データと実データの分布の差を数値化するため、統計的距離（ダイバージェンス）を用いることがあります。例えばJensen-Shannon多様体距離やWasserstein距離を、生成系列と実系列のヒストグラム・分布関数間で計算し、値の小ささをもって「分布が類似している」とします ([Can GANs Learn the Stylized Facts of Financial Time Series?](#)) ([Can GANs Learn the Stylized Facts of Financial Time Series?](#))。これらの指標は**全体的な分布形状**を一括で評価できるため、複合的な判断に用いられます。ただし分布のどの部分で差異が生じているかは直接には示さないため、尖度やACFと併用して評価するのが一般的です。
- **判別モデルによるテスト:** もう一つのアプローチは、**実データ vs. 生成データ**を判別する機械学習モデル（例えば単純な分類器）を訓練し、その精度を測る方法です。理想的には、生成データが完全にリアルであれば判別器は**50%の精度（ランダム推測）**しか達成できないはずですが、この考えに基づき、例えばある研究では生成系列と実系列を混ぜたデータセットを用意し、ニューラルネットに時系列パターンから出所を当てさせる試みがなされています。判別精度が低ければ統計的に見分けがつかないことを示唆し、モデルの成功度合いを示す指標となります。ただし高品質な生成モデル（特にGAN）は判別器と戦う形で訓練されているため、単純な判別テストでは過学習の危険も指摘されています。
- **経済的・実務的なパフォーマンス検証:** 統計指標以外に、生成データを用いて何らかの**実務タスク**をシミュレーションし、その結果が**実データ使用時と近い**かを見る検証も行われます。例えば生成データ上でポートフォリオ最適化やリスク計測（VaR計算等）を行い、それが過去実データに基づく結果と整合するかを確認する方法です。Allenら(2024)の研究では、GANで生成した日次株価系列を用いて**リスク価値 (VaR) の推定**を試み、実データから直接計測した場合とほぼ同等の精度が出せることを示しました。このように**ダウンストリームタスクでの有用性**をもってデータ品質を評価する視点も、応用研究では重視されています。

以上のような評価を総合しても、**完全に実データと区別できない合成金融時系列**を得るのはなお難しい課題です ([Generation of synthetic financial time series by diffusion models](#))。Dogariuら(2022)は最新の複数手法を比較し、「**未だどのモデルも全てのスタイライズドファクトを網羅的には再現できていない**」と結論しています ([Generation of synthetic financial time series by diffusion models](#))。しかし各モデルは一部の指標では顕著な成功を収めており、例えば拡散モデルは分布尾と自己相関の双方で高い適合度を示すなど、着実な進歩も報告されています ([Generation of synthetic financial time series by diffusion models](#)) ([Generation of synthetic financial time series by diffusion models](#))。評価方法自体も進化しており、単一指標ではなく**複数の統計テストをバッテリー的に適用**してモデル性能を立体的に把握するアプローチが採られています。今後は、より洗練された評価基準

(例えば高次モーメントや極値理論に基づくテスト) や、経済理論と整合した検証(例えば市場効率性を乱さないか、裁定機会が生じないか等) も検討されるでしょう。総じて、「**実データ同様の性質**」を多面的に確認することが金融データ生成モデルの評価における肝要なポイントです (Generation of synthetic financial time series by diffusion models) (Can GANs Learn the Stylized Facts of Financial Time Series?)。

## 資産クラスごとに見られるスタイライズドファクトの違い

スタイライズドファクトは多くの資産で普遍的に観測されますが、その**程度や具体的パターンは資産クラスによって異なる**ことが知られています。各主要アセットクラスについて、特徴的な違いを整理します。

- **株式（現物株・株式指数）**：株式市場、とりわけ**株式指数**は典型的なスタイライズドファクトの強さが顕著です。日次レベルでは尖度が非常に大きく（肥尾が極端）(Microsoft Word - 改Jiawei\_He copy.docx)、例えばリーマンショックやCOVID-19ショック時のように正規分布では到底説明できない大幅変動が起こります。分布は**左に長い裾**を持つことが多く（**負のスキュー**）、急落リスクの偏り（ゲイン・ロス非対称性）が明確です (Microsoft Word - 改Jiawei\_He copy.docx)。ボラティリティクラスタリングも強く、**長期にわたる高ボラティリティ期**（例えば金融危機後の余震期間）が観測されます。またレバレッジ効果も**株式指数では顕著**で、指数が下落すると対応するボラティリティ指数（VIXなど）は急騰し、逆に上昇局面ではボラティリティ低下傾向があります ([PDF] Gain/loss Asymmetry and the Leverage Effect - ETH Zürich)。一方、**個別株式**では指数に比べて若干事情が異なり、銘柄によっては分布のスキューが正の場合もあります (Stylised Facts for Cryptocurrencies: a sector analysis - Medium)。特に小型成長株などは良いニュースで株価が跳ね上がる余地が大きいため**正の歪度（右裾が重い）**を示すこともあります。**ただしボラティリティクラスタリング現象は個別株でも普遍的**であり、売買高の増加と共に変動が荒くなるパターンが共通に見られます。株式全般にいえるのは、**金融市場の中でも最も研究が進んだ資産クラス**であり、多数のモデルが株価データのスタイライズドファクトを再現するよう校正されています。
- **外国為替（FX）**：為替相場のリターンも**明確に肥尾分布**を持ちますが、その尖度は株式指数ほど極端ではない場合が多いです。上の例では主要通貨ペア EUR/USDの日次リターン尖度は約4.8で、株式（12.8）に比べ低いものの、やはり正規分布（尖度3）より高く**厚い尾**を持っています (Microsoft Word - 改Jiawei\_He copy.docx)。為替は**分布の非対称性が小さい**ことが多く、政策金利差などによるわずかなトレンドはあるものの、上昇と下落のリスクは概ね対称

に近いです (Microsoft Word - 改Jiawei\_He copy.docx)。これは通貨には企業の倒産のような片方向の極端事象がなく、基本的に相対的価値であるためです。ただ、時期によっては新興国通貨の急落リスクや、ドル高トレンドなどで歪度が偏る場合もあります。ボラティリティクラスタリングはFXでも明確に見られ、経済指標発表や政情不安時には高ボラティリティ状態が続きます。レバレッジ効果については、**為替には株式のような資本構造が無い**ためクラッシュ時の機械的なレバ比上昇はありません。しかし**市場心理的な非対称性**は存在し、大きな下落（ある通貨の急落）は他通貨との波及的なボラティリティ上昇を招くケースがあります。総じてFX市場では**統計特性は株式と同様の方向だが程度は穏やか**と言えます (Microsoft Word - 改Jiawei\_He copy.docx)。また超短期の高頻度データでは**\*\*ミクロ構造特有の季節性（日中の活発時間帯と閑散時間帯）\*\***が顕著で、これは他市場にも共通する高頻度の特徴です (Stylised facts for high frequency cryptocurrency data - ScienceDirect)。

- **コモディティ（商品）**：金や原油などの商品先物も、リターン分布は**肥尾かつ自己相関なし**という基本事実は同じです。尖度は原油で7程度、金でも5前後など報告例があり (Microsoft Word - 改Jiawei\_He copy.docx)、これは株式とFXの中間くらいです。商品の場合、**需給バランスの急変**（戦争・天候・産油国政策など）で価格が急騰・急落することがあり、その際の変動は非常に大きな肥尾につながります。例えば原油価格は2008年に史上最高値から数ヶ月で1/3以下に急落し、その過程で日次変動が連日10%以上というクラスターを形成しました。分布の歪度は商品によって異なり、原油や工業用金属は**景気悪化時に需要減で急落することが多い**ため若干負の歪度傾向、一方で農産物や貴金属は**供給不安で急騰することがある**ため正の歪度になることもあります。ボラティリティクラスタリングも商品市場全般に強く、特に原油はOPEC会合や地政学リスクに絡む**不安定期に高ボラが持続**します。レバレッジ効果は企業株ほど明確ではないものの、**価格急落時に生産者のヘッジ行動が活発化して更なる変動を生む**など、下落局面でのボラ増幅は観察されています。商品市場の特殊要因としては**季節性や在庫統計の影響**があり、例えば農産物は季節サイクルで価格変動パターンが年次周期を持つなどの特徴もあります (Stylised facts for high frequency cryptocurrency data - ScienceDirect)。
- **債券（国債・社債）**：債券価格（または利回り）の変動も金融市場の一部ですが、他資産に比べ日次変動幅は小さく、**通常時はほぼ正規分布に近い**と言えます。ただし**危機時には例外的な肥尾**を示すことがあります。国債利回りは各国中銀の政策で安定していますが、例えば信用不安が高まった場合に社債スプレッドが急拡大する（価格が急落する）といった**尾部事象**が起こります。尖度は平常時には低くても、期間を通算するとリーマン危機や欧州債務危機などで非正規性が顕著になります。歪度は、市場によって異なりますが、低金利下では**金利**

上昇余地より低下余地が小さい（ゼロ下限）ため利回り変化に負の歪度がつくこともあります。一方、高金利から利下げ局面では逆に正の歪度になりえます。ボラティリティクラスタリングに関しては、債券市場では株式ほど長期に不安定が続くことは少なく、**政策イベント前後の一時的高ボラ**に留まるケースが多いです。ただ、信用不安が尾を引く場合は社債市場などで高ボラ状態が続くこともあります。レバレッジ効果というより、**金利とボラティリティの関係**では、一般に利回り低下（債券価格上昇）局面ではボラティリティも低下する傾向があり、これは株式のレバレッジ効果と似た**負の相関**として観測されます（債券の場合、将来の不確実性が減ると価格が上がるという解釈も可能です）。

- **暗号資産（仮想通貨）**：ビットコインを代表とする暗号資産は、その**極端なボラティリティ**で知られます。統計的にもリターン分布の肥尾度合いは非常に高く、初期の研究では**株式よりも重い尾を持つ**と報告されたこともあります ([Relevant stylized facts about bitcoin: Fluctuations, first return ...](#))。もっとも最近の分析では、市場の成熟とともに尾の厚さ（べき指数）は伝統的資産と近い範囲に収まってきたとの指摘もあります。例えば2017-2019年の期間ではビットコインの尖度は3.46でS&P500の5.41より低かったが、2020年にはビットコインが急騰・急落を繰り返し尖度が大幅に上昇、2021-2022年は再びS&P500より高いが過去ほどではない、といった風に**時期により肥尾の程度が変動**しています。歪度については、ビットコインは長期上昇トレンドの中で暴落も経験しており、期間によって正にも負にも偏り得ます。過去数年を通すと**大暴落局面（例えば取引所破綻ニュース）の影響で全体としては負の歪度**が確認されています ([PDF] [Comparing Cryptocurrency Indices to Traditional Indices](#))。ボラティリティクラスタリングは暗号資産でも顕著で、**高ボラ期と低ボラ期が交互に訪れる**様子が統計的に捉えられています。これは株式以上に極端で、平時は主要通貨並みに穏やかな日もある一方、異常時は年率換算100%以上のボラティリティが続く、といった振る舞いです。レバレッジ効果に関しては企業価値の概念が無いものの、**投資家心理による非対称性**は存在しうるとされています。例えば市場全体が強気で一方向に動いている際は上昇が緩やかでボラ低下、一旦悪材料が出るとパニック売りで急落・ボラ上昇、という動きは観測されています。この点で、ビットコイン市場も株式市場と似た**下落時のボラティリティ急騰**を示す場合があります、実証分析でも弱いながらレバレッジ効果に似た負の相関が検出されたとの報告があります ([A comparison of cryptocurrency volatility-benchmarking new and ...](#))。

このように資産クラス間の差異は**程度的なもの**が多く、基本的なスタイライズドファクトの方向性は共通しています。暗号資産のような新興市場でも、重尾分布・自己相関構造・クラスタリングといった性質は「伝統的資産と同様に存在する」ことが確認されています ([PDF] [Comparing Cryptocurrency Indices to Traditional](#)

Indices)。ただし重要なのは、そのパラメータ（例えば尾の厚さを示すべき指数や自己相関の減衰スピードなど）は市場ごとに異なるため、生成モデルを訓練・評価する際には対象資産に合わせた指標値を意識する必要があるという点です (Microsoft Word - 改Jiawei\_He copy.docx)。また市場固有の現象（例: コモディティの季節性、暗号資産の週末効果、株式の取引停止ルールによるボラ抑制など）も存在するため、高度な生成ではそうした細部まで考慮されることが望まれます。

## 高品質な擬似データの可能な応用（ユースケース）

もし金融時系列のスタイライズドファクトを網羅的に再現できる高品質な擬似データが生成可能になれば、その応用範囲は広範囲に及びます。研究面・産業面で期待されている主なユースケースを整理します。

- **バックテストと投資戦略評価の高度化:** 金融機関やファンドは、新しいトレーディング戦略や金融商品の評価に過去データを用いたバックテストを行います。しかし歴史データだけではサンプルが限られ、未経験の相場状況での挙動は検証困難です (Can GANs Learn the Stylized Facts of Financial Time Series?)。高品質な生成データを用いれば、**仮想的に無数のマーケットシナリオをシミュレーション**できるため、戦略のロバスト性をより厳密にテストできます (Generation of synthetic financial time series by diffusion models)。例えば100年に一度級の暴落や急騰シナリオも人工的に作り出し、その中でアルゴリズムの収益やリスク挙動を検証する、といった**ストレステスト的バックテスト**が可能になります。これにより過去に依存しない**プロアクティブなリスク評価**ができ、ブラックスワンイベントへの耐性評価など実務上重要な分析が飛躍的に充実します。
- **リスク管理とシナリオ・ストレステスト:** 上記と関連しますが、特に**金融リスク管理**の分野では合成時系列データの価値が高いです。市場リスク計測（例えばVaRや期待ショートフォール算出）は通常過去一定期間のデータに基づきますが、もしそこに現れていない極端事象が将来起こればリスクは過小評価されます。生成モデルで**統計的にあり得る様々な極端パス**を大量に作り出し、その損失分布を評価することで、リスク指標の信頼区間を推定したり**潜在的なリスク要因の発見**につながります。また規制当局や金融機関が行う**ストレステスト**でも、過去の実例に囚われない多様なストレスシナリオを合成データから抽出できます (Generation of synthetic financial time series by diffusion models)。例えば「主要中銀が同時に利上げを行ったら」「〇〇ショック級のボラティリティが半年続いたら」といった**仮定シナリオ**を具体的な時系列パスとして生成し、ポートフォリオ影響を試算できます。生成モデルがスタイライズドファクトを備えていれば、そうしたシナリオも現実味のある相関構造や変動パターンを

備えており、リスク管理のシナリオ分析に信憑性をもたらします (Generation of synthetic financial time series by diffusion models)。

- **データ拡張と機械学習モデルの訓練:** 金融のAI・機械学習応用では**データ量の不足**がしばしばボトルネックになります。規制やプライバシーの問題でデータ共有が難しい場合もあります (

DEDOMENA • 10 use-cases for privacy-preserving synthetic data

) (

DEDOMENA • 10 use-cases for privacy-preserving synthetic data

)。高品質な疑似データは、こうした場面で**トレーニングデータの拡張**（データオーグメンテーション）に利用できます。例えばあるトレーダーの1年分の収益時系列しかない場合、生成モデルで類似のパターンを持つ多数の年度データを作り、機械学習モデルを訓練すれば、過学習を防ぎモデルの汎化性能を高められます。また個人情報を含むような金融取引データでも、匿名化のため**プライバシー保護した合成データ**に置き換えて分析を共有することが可能になります (Generation of synthetic financial time series by diffusion models)。実際、Brophyら(2023)の調査では**モデル訓練のためのデータ拡張および匿名化**に合成時系列を用いる動きが紹介されています (Generation of synthetic financial time series by diffusion models)。銀行間で実データを共有できなくても、スタイライズドファクトを維持した疑似データならノウハウを交換しやすくなる、といった**データ利活用促進**の効果も期待できます。

- **アルゴリズム取引開発と市場シミュレーション:** 高頻度取引（HFT）などのアルゴリズム取引では、戦略のテストに**市場全体の板情報を再現**したシミュレーション環境が必要です。エージェントベースモデルや深層生成モデルで極めて現実的な**人工市場**を構築できれば、実際の取引に近い環境でアルゴリズムのテストやチューニングが可能になります。例えば人工市場で1年分のティックデータを生成し、その中でアルゴリズム売買を走らせて収益率やマーケットインパクトを測定するといった応用です。これにより、**本番稼働前に様々な市場状況下で戦略の検証**ができます。とりわけ**異常時のマーケットインパクト**や、アルゴリズム同士の相互作用による予期せぬ現象（フラッシュクラッシュなど）も疑似環境で研究でき、安定的なアルゴリズム取引の開発に貢献します。
- **新規商品の価格評価・リスク分析:** 金融工学ではデリバティブなど新しい商品の価格付けやリスク分析にモンテカルロシミュレーションを用います。従来はリスク中立確率測度の下でジャンプ拡散モデル等を仮定してシミュレーションしていましたが、今後**高度な生成モデルでリアルな経路を大量生成**し、これを直接プライシングやリスク計測に使う可能性があります。例えば伝統的なモデルでは

捉えにくい**ボラティリティの急変や相関の非定常**を、生成データなら自然に含めることができます ([Generation of synthetic financial time series by diffusion models](#))。そうすると、ある複雑な商品の損益分布もより現実に近い形で評価でき、プライシング精度やリスク管理精度が向上するでしょう。ただしこの応用では**リスク中立性の確保**（現実の確率測度から調整してリスク中立測度で生成する必要）といった課題もあるため、研究のハードルは高いですが将来的な方向性として考えられます。

- **学術研究への寄与（市場のメカニズム検証など）**：高品質な合成データは、**市場の構造や統計法則の研究**にも役立ちます。例えば経済物理学者は、合成データを使ってスタイライズドファクトの普遍性を検証したり、新たな統計的特徴を探索したりできます。また、因果推論の文脈では現実には観測できない「もしもシナリオ」をデータとして生成し、それを分析することで**市場ダイナミクスの仮説検証**が行えます。エージェントベースモデルと組み合わせれば、「もしリテール投資家が増えたらボラティリティはどう変わるか」「ショートセル禁止が継続されたら価格形成にどんな影響があるか」等をシミュレーションできます。こうした**仮想実験**は本来単一の現実しか辿れない歴史分析を補完し、金融理論の発展や政策立案にも寄与する可能性があります。

以上のように、実データと見分けがつかないほど質の高い金融時系列データが得られれば、**リスク管理、投資戦略、規制当局のシナリオ分析、学術実験**など様々な領域で革命的な進歩が期待できます ([Generation of synthetic financial time series by diffusion models](#))。特に金融のように**データが貴重な資産**でありつつ**プライバシーや極端事象の問題でデータ活用が制限**されがちな分野では、合成データのメリットは計り知れません。もっとも、そのためには生成モデルが本当に現実の統計特性を捉えていなければ逆効果にもなりえます。不十分な合成データで分析した場合、誤った安心感や偏った結果を招く恐れもあるため、引き続き**生成モデルの性能向上と精緻な評価**が求められます ([Generation of synthetic financial time series by diffusion models](#)) ([Can GANs Learn the Stylized Facts of Financial Time Series?](#))。しかし近年の深層生成モデルの急速な進歩を見ると、そう遠くない将来に\*\*「実データ同様の信頼性を持つシンセティック金融データ」\*\*が現実のものとなり、金融研究・産業界のスタンダードなツールとなっているかもしれません。

#### 参考文献（一部）：

- Rama Cont, "Empirical properties of asset returns: stylized facts and statistical issues," *Quantitative Finance*, vol. 1, 2001 – 金融リターンの経験的普遍性に関する古典的レビュー。 ([Have any new stylized facts of asset returns been discovered since 2001? - Quantitative Finance Stack Exchange](#)) ([Have any new stylized facts of asset returns been discovered since 2001? - Quantitative Finance Stack Exchange](#))



- Takahashi et al., "Modeling financial time-series with generative adversarial networks," *Physica A*, 2019 – FIN-GANモデルの提案。スタイライズドファクトの再現性に着目したGANアプローチ。 ([\[PDF\] Generative Adversarial Networks: A Novel Approach to Predictive ...](#)) ([\[PDF\] Option pricing using Generative Adversarial Networks](#))
- Wiese et al., "Quant GANs: Deep Generation of Financial Time Series," arXiv preprint 1907.06673, 2020 – TCNを用いたGANで極端事象の頻度も含めた金融系列生成。 ([Quant GANs: deep generation of financial time series - ResearchGate](#)) ([Quant GANs: Deep Generation of Financial Time Series](#))
- Eckerli & Osterrieder, "Generative Adversarial Networks in Finance: An Overview," arXiv 2106.06364, 2021 – 金融におけるGAN応用の包括的レビュー。
- Dogariu et al., "Generation of Realistic Synthetic Financial Time-Series," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 18(4), 2022 – 金融時系列生成モデルの比較研究。VAE的手法も含め、スタイライズドファクト網羅性の課題を指摘。 ([Generation of synthetic financial time series by diffusion models](#)) ([Generation of synthetic financial time series by diffusion models](#))
- Kwon & Lee, "Can GANs Learn the Stylized Facts of Financial Time Series?," arXiv 2410.09850, 2023 – GANの能力検証。複数GANでのスタイライズドファクト再現テスト。 ([Can GANs Learn the Stylized Facts of Financial Time Series?](#)) ([Can GANs Learn the Stylized Facts of Financial Time Series?](#))
- Takahashi & Mizuno, "Generation of synthetic financial time series by diffusion models," arXiv 2410.18897, 2024 – 拡散モデルによる時系列生成。ウェーブレット変換を活用し、既存研究の課題克服を主張。 ([Generation of synthetic financial time series by diffusion models](#)) ([Generation of synthetic financial time series by diffusion models](#))
- その他、各種スタイライズドファクトに関する文献 ([Microsoft Word - 改 Jiawei\\_He copy.docx](#)) ([\[PDF\] Stylized Facts - Financial Data - Portfolio Optimization Book](#)) ([\[PDF\] Comparing Cryptocurrency Indices to Traditional Indices](#))や、合成データ応用に関する調査 ([Generation of synthetic financial time series by diffusion models](#))など。