

2022年1月19日提出

論文題目

移動エントロピーを用いた大規模金融時系列データの分析

小林 照義 研究室

学籍番号 1712012E

氏名 森田 光春

## 要旨

株式市場などの証券市場では、平常時は価格はランダムウォークすることが暗黙の前提とされているが、何らかの外部要因などによって市場にショックがもたらされたとき市場全体がまるで単一の銘柄のように共鳴しランダムでは説明しきれない暴落現象を起こすことがある。本稿ではそのようなショック時に株式市場でどのような現象が起きているのかを分析するために 2020 年 2 月後半から 3 月後半ごろに掛けて発生したコロナショックを時間的に内包する 2019 年 12 月 2 日から 2020 年 5 月 22 日を期間とし、NASDAQ 市場に上場する株式のうち NASDAQ100 指数に採用されている 102 銘柄および NASDAQ100 指数に連動するもので世界最大の上場投資信託『Invesco QQQ Trust Series 1：米, Invesco 社』（QQQ）の合計 103 個の株価データのそれぞれのペアについて一定時刻おきに移動エントロピー（Transfer Entropy）と呼ばれる量を計算することでペアごとの結びつきの強さを定量化し、株式市場を各銘柄をノードとする時系列の重み付き有向ネットワークとして表現した。

得られたネットワークの各種統計量を検討する中で、特に QQQ の変動とネットワーク全体の移動エントロピーの分布の変化に重要な関係があることが分かった。そこでコルモゴロフ-スミルノフ検定 (Kolmogorov-Smirnov test) を用いてコロナショック時に移動エントロピーの分布に有意な変化があったことを確認した。次に、そのような分布の変化の程度をカルバックライブラー情報量 (Kullback-Leibler divergence) を用いて定量化し、QQQ の変化率を説明することを目的とした時系列モデル (VAR モデル) を作成した。モデルによってインパルス応答、グレンジャー因果性などを検証しカルバックライブラー情報量から QQQ への因果性があるらしいことを発見した。

最後にカルバックライブラー情報量などのネットワーク情報と株価データを組み合わせた予測モデルを作成し、株価データのみを使用する予測モデルより予測精度が改善することを見た。

また、同様の分析を暗号通貨市場についても行った。時価総額上位 100 の暗号通貨のうちデータの入手可能であった 30 の通貨を対象に 2021 年 4 月 1 日から 2021 年 5 月 31 日を期間として時系列の移動エントロピーネットワークを作成した。得られたネットワークからカルバックライブラー情報量などの統計量を計算し、ビットコイン価格の変化率を被説明変数とした VAR モデルを用いてインパルス応答やグレンジャー因果性などを検証した結果、株式市場の場合と同じくカルバックライブラー情報量から

ビットコイン価格の変化率に因果性があるらしいことを発見した。

おわりに予測モデルを作成しネットワーク情報を加味したモデルの方がやはり、暗号通貨価格データのみを説明変数とするモデルよりも予測精度が改善することを見た。

種類の異なる二つの市場について同様の結果が得られたことは多数の時系列データをネットワークに変換し、そこから有用な情報が抽出できるということを説得的に主張する結果になったと言えるだろう。

# 目次

1	序論	1
1.1	はじめに . . . . .	1
1.2	先行研究および理論 . . . . .	2
2	方法	4
2.1	データの入手 . . . . .	4
2.2	前処理 . . . . .	4
2.3	ネットワークの構築 . . . . .	5
2.4	基本統計量 . . . . .	9
3	分析	11
3.1	株式市場 . . . . .	11
3.2	VAR モデル . . . . .	14
3.3	暗号通貨市場 . . . . .	27
4	予測	39
4.1	目的 . . . . .	39
4.2	株価予測 . . . . .	39
4.3	暗号通貨価格予測 . . . . .	42
5	結論	46
5.1	本稿のまとめ . . . . .	46
5.2	本稿の課題 . . . . .	47
	参考文献	51

# 1 序論

## 1.1 はじめに

株価の予測は一般的に難しいとされる。効率的市場仮説 (Fama 1970) によれば株価はその時点で得られるあらゆる情報を反映して適切に決定されているので、利用可能な情報をいくら加味しても追加的な知見は得られないとされ、そのため株価は過去の動向に関係なくランダムに変動し、未来の株価を予測することは不可能であるとされる。しかしながらこれは理想化された市場において成り立つべき仮説であるので、実際は投資家は利用可能な膨大な情報の中から質の高い情報を抽出し、株価予測に役立たせることができる可能性がある。そのため株式市場の分析、予測についてはこれまで伝統的なファンダメンタルズ分析やファイナンス理論、最近ではニューラルネットワークなどの機械学習の方法を駆使したものや自然言語処理によってテキストデータなどの情報をモデルに組み込んだものなど、様々な方向から多くの研究がなされている (Abarbanell and Bushee 1997; Dechow et al. 2001; Ding et al. 2015; Chong, Han, and Park 2017; Akita et al. 2016).

さて、前述の通り株価は基本的にはランダムウォーク同様にそれぞれの銘柄が独立に変動することが期待されるが、株式市場では時として2020年のコロナショック時などのように市場全体がまるで一つになったかのように連動し、ランダムの域を超えた急落を起こすことがある。そのような現象の背後には株式の間の『繋がり方』に何らかの変化があったからではないかと考えるのは無理な発想では無いだろう。株式市場をそのような何らかの意味での繋がりが複雑に絡み合ったネットワークだと捉え、例えばその中心性やコミュニティなどの構造的特性をを分析することで何らかの知見を得ようとする研究もこれまで少なからずなされている (Ahern 2013; Zhao et al. 2018; Heiberger 2014; Namaki et al. 2011)。その中で本稿は (Kim and Sayama 2017) を先行研究とする。詳細については次節で述べるが、彼らの研究では株式市場を株式どうしの繋がりの深さ (リンク) を相互情報量と呼ばれる量で定義したネットワークとして捉え、ネットワークから得られる情報を用いた株価予測モデルがそれを用いないモデルに比べて予測精度が改善したことを報告している。本稿も同じように最終的には予測によって株式市場をネットワークとして捉えることの有効性を主張する形になっ

ているが、先行研究との相違点はリンクの決定に移動エントロピーと呼ばれる別の量を用いていること、予測のみでなく VAR モデルを用いてグレンジャー因果性やインパルス応答を検証するなどの計量的分析を行っていること、そして株式市場のみならず暗号通貨市場についても同様の分析をなすことで結果の再現性を確かめていることにある。

2 章では本稿で用いた手法の説明をし、3 章では株式市場と暗号通貨市場について VAR モデルを用いて実際に分析をする。4 章では株価と暗号通貨価格の予測を検証し、5 章はまとめの章になっている。また、参考文献の後に付録の章がある。

## 1.2 先行研究および理論

本稿は (Kim and Sayama 2017) を先行研究とする。彼らの研究では S&P500 構成銘柄を対象に株価データの各ペアについて一定時刻ごとに相互情報量 (Mutual Information) と呼ばれる量を計算し株式市場を重み付き無向ネットワークとして表現することで、相互情報量の分布の変化が株価の変動と重要な関係にあることを発見し、ネットワークから得られる情報を加味した株価の予測モデルとそれを用いない予測モデルを比較した結果、ネットワークから得られる情報を加味したモデルの方が予測精度が改善したことを報告している。

本稿では彼らのネットワークの作成方法を参考にしているが、相違点はリンクの重みの決定に相互情報量ではなく移動エントロピーを用いている点にある。ここで二つの離散型確率変数  $X, Y$  とその確率分布  $p(x), p(y)$  および同時分布  $p(x, y)$  に対して相互情報量  $I(X; Y)$  は次で定義される。

$$I(X; Y) \equiv \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (1.1)$$

相互情報量は 2 つの変数  $X$  と  $Y$  の非線形な関係も含めた相関を定量化する量としてよく用いられ、その値が大きい程 2 つの変数の間の相関が高いとされる。

次に離散型確率変数列  $X = \{X_t\}, Y = \{Y_t\}$  に対して  $X_{t-1:L}$  を、 $X_{t-1:L} \equiv \{X_{t-1}, X_{t-2}, \dots, X_{t-L}\}, Y_{t-1:L}$  に関しても同様に定義するものとする、 $X$

から  $Y$  への移動エントロピー  $T_{X \rightarrow Y}$  は条件  $Y_{t-1:L}$  の下での次の相互情報量の期待値,

$$\begin{aligned}
T_{X \rightarrow Y} &\equiv \mathbf{E}_{Y_{t-1:L}} [I(Y_t; X_{t-1:L} | Y_{t-1:L})] \\
&= \sum \left[ \left( p(y_t, x_{t-1:L} | y_{t-1:L}) \log \frac{p(y_t, x_{t-1:L} | y_{t-1:L})}{p(y_t | y_{t-1:L}) p(x_{t-1:L} | y_{t-1:L})} \right) p(y_{t-1:L}) \right] \\
&= \sum p(y_t, y_{t-1:L}, x_{t-1:L}) \log \frac{p(y_t, y_{t-1:L}, x_{t-1:L}) / p(y_{t-1:L})}{p(y_t, y_{t-1:L}) p(y_{t-1:L}, x_{t-1:L}) / p(y_{t-1:L})^2} \\
&= \sum p(y_t, y_{t-1:L}, x_{t-1:L}) \log \frac{p(y_t, y_{t-1:L}, x_{t-1:L})}{p(y_{t-1:L}, x_{t-1:L}) p(y_t, y_{t-1:L}) / p(y_{t-1:L})} \\
&= \sum p(y_t, y_{t-1:L}, x_{t-1:L}) \log \frac{p(y_t | y_{t-1:L}, x_{t-1:L})}{p(y_t | y_{t-1:L})} \tag{1.2}
\end{aligned}$$

で定義される。直感的にはこれは  $X_{t-1:L}$  と  $Y_t$  の相互情報量だと考えて相当である。時間差を設けて相互情報量を計算することで  $X$  の過去の履歴  $X_{t-1:L}$  と  $Y$  の現在の値  $Y_t$  との関係の深さを定量化しているのである。すなわちこれは  $X$  から  $Y$  への因果性を定量化する量である。計算に当たって  $Y_{t-1:L}$  で条件付けするのは、 $X_{t-1:L}$  に含まれているであろう  $Y$  の過去の履歴  $Y_{t-1:L}$  の影響をキャンセルするためである。

相互情報量は定義式から  $X$  と  $Y$  について対称であるが移動エントロピーは非対称である。株式市場においては例えば親会社と子会社の関係のように一方の株価は他方に大きな影響を与えるが逆はそうではないという状況は十分考えられる。したがって移動エントロピーの方が株式市場を正確に表現できる可能性がある。また先行研究では主として予測によってネットワーク指標の有効性を主張しているが、本研究ではそれに加えて VAR モデルによる統計的な分析と暗号通貨市場という別市場に関する同様の分析がなされたことによってより強固な主張が可能となった点が新しい点である。

## 2 方法

### 2.1 データの入手

本研究で使用するデータは全て、データサイト Tiingo(<https://www.tiingo.com>) より入手した。株価データに関しては 2021 年 5 月 1 日時点で NASDAQ100 指数に採用されていた 102 の銘柄およびデータの入手可能性の観点から NASDAQ100 指数の代替指標とした米、Invesco 社が運営する NASDAQ100 指数に連動するものの中で世界最大の上場投資信託 Invesco QQQ Trust Series 1 (証券コード, QQQ) の合計 103 の銘柄に対して 2019 年 12 月 2 日から 2020 年 5 月 22 日までのおよそ半年間を対象期間とした 5 分ごとの価格データを入手した。

また、暗号通貨については 2021 年 4 月 1 日から 2021 年 5 月 31 日を対象期間とし、2021 年 10 月 1 日時点で時価総額上位 100 であったの通貨のうちデータの入手可能であった 30 通貨についてこちらは 10 分ごとの価格データを入手した。

### 2.2 前処理

#### 2.2.1 欠損処理および標準化

得られた株価データのうち一部の欠損のあるものについては直前の値で埋めることで欠損処理とした。なお、2 ステップ以上連続で欠損のあったデータはなかったためデータの正確性にはほとんど影響を与えていないはずである。

次に、データのスケールを統一し移動エントロピーの計算を安定化するためにそれぞれのデータについて標準化を施した。すなわち、時系列データ  $x_{i,t}$  に対して期間全体の平均を  $\mu_{x_i}$ 、標準偏差を  $\sigma_{x_i}$  として変換

$$z_{i,t} = \frac{x_{i,t} - \mu_{x_i}}{\sigma_{x_i}} \quad (2.1)$$

を  $i = 1, \dots, 103$  とした 103 個すべての銘柄について施した。元データの値が必要な場合は適宜逆変換を施した。

## 2.2.2 窓化

次にそれぞれのデータを一定時間間隔の窓に分割する窓処理を施した. 具体的には窓の大きさを 108 ステップ (9 時間) とし, 窓の間隔を 6 ステップ (30 分) とする窓処理を施した. すなわち一つの株価データ  $\{z_{i,t}\}_{t=0,1,\dots}$  に対して長さ 108 で 6 ステップ間隔の小時系列データの列

$$D_{i,t} = \{\{z_{i,t}, \dots, z_{i,t+107}\}\}_{t=6\tau, \tau=0,1,\dots} \quad (2.2)$$

を得る処理を  $i = 1, \dots, 103$  としたすべての銘柄について施した.

## 2.3 ネットワークの構築

### 2.3.1 移動エントロピー

前節で得られた 103 銘柄分の窓の列  $D_{i,t}$  ( $i = 1, \dots, 103$ ) の自分以外の各ペアの順番付き組み合わせ 10506 組について (1.2) 式を  $L = 3$  とし, 各時刻  $t = 6\tau, \tau = 0, 1, \dots$  において移動エントロピーを計算することでそれぞれの組み合わせに対する時系列の移動エントロピーデータを得た. ただし移動エントロピーの計算にはかなりの時間がかかるため計算に当たっては (Ikegwu et al. 2020) による近似アルゴリズムを彼らの公開する Python ライブラリによって利用した.

### 2.3.2 フィルタリング及びネットワークの構築

最後に移動エントロピーデータの平滑化のため離散ウェーブレット変換によるフィルタリング処理を施した. 具体的にはウェーブレットに db30, Daubechies ウェーブレットを選択しウェーブレット係数が次で定義される Universal threshold,  $T$  を下回る場合には係数を 0 として逆変換した.

$$T = \sigma \sqrt{2 \log N}, \quad \sigma = \frac{\mathbf{E}[|W - \mathbf{E}(W)|]}{0.6745} \quad (2.3)$$

ただし  $N$  は元データの長さを表し,  $W$  はウェーブレット係数の列を表し,  $\mathbf{E}[\cdot]$  は平均を取る操作を表す. なお計算には PyWavelets という Python ライブラリを使用した (Lee et al. 2019).

こうして, 10506 個の 30 分間隔の時系列の移動エントロピーデータから同じく 30 分間隔の時系列の移動エントロピーネットワークを得ることになる. なお, NASDAQ 証

券取引所の 1 日当たりの取引時間は 7.5 時間であるため 30 分間隔の時系列データではちょうど 15 ステップ分が 1 営業日の長さに相当する。以上のことを模式的に示した図を図 2.1 に示す。また、参考のために Apple 社から QQQ への移動エントロピーの元データ (Original) およびフィルタリング後のデータ (Filtered) のグラフを図 2.2 に、2020 年 4 月 1 日におけるネットワークの上位 1% の重みををもつリンクのみを表示した図を図 2.3 に示す。

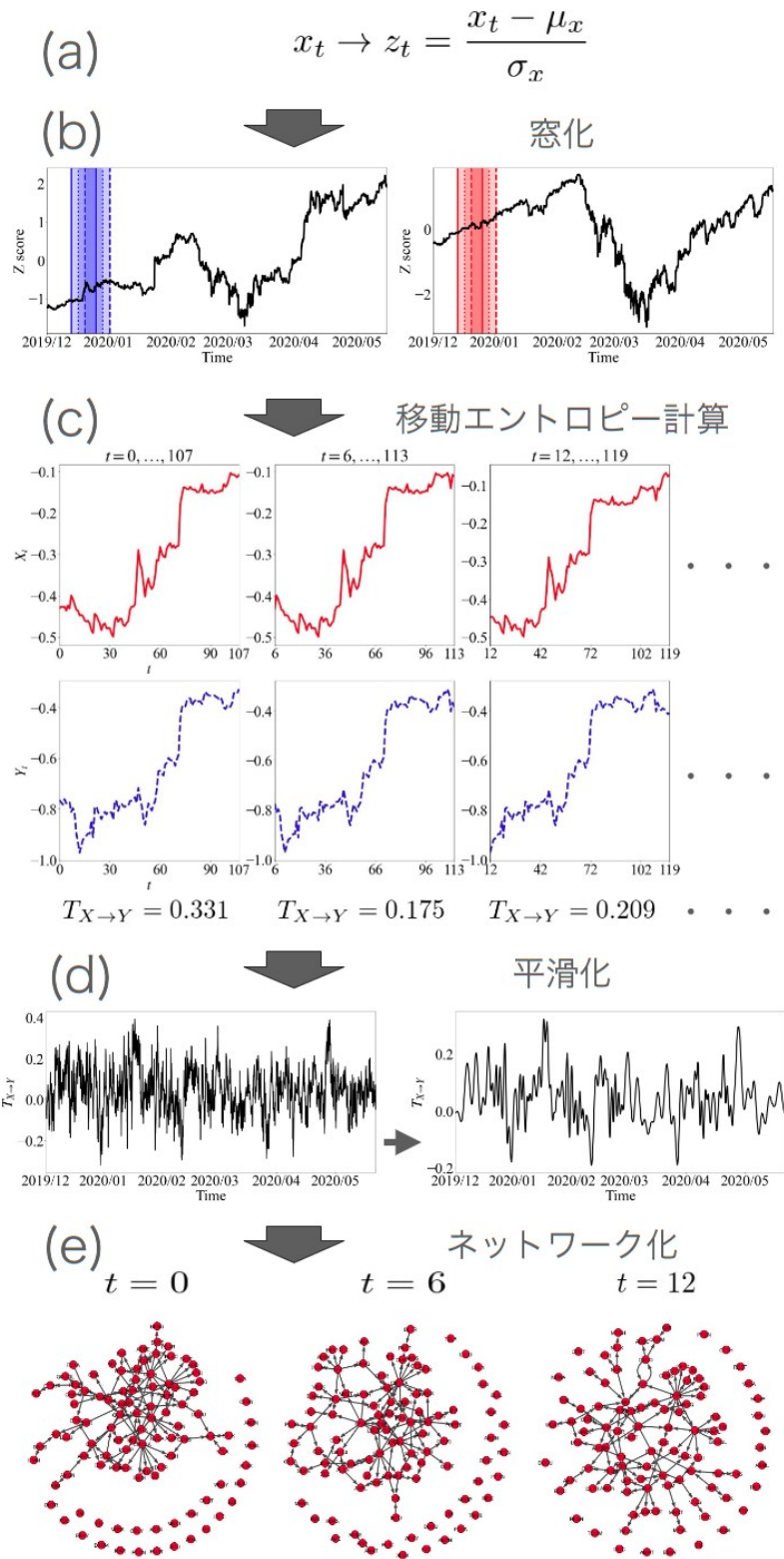


図 2.1

図 2.1 (a) 103 個の株価データのそれぞれに対してデータのスケールを統一するために標準化を行う. (b) 得られた 103 個の標準化済み株価データのそれぞれに対して一定間隔ごとに窓処理を行う. 窓の大きさは 108 ステップ (9 時間), 隣り合う窓の間隔は 6 ステップ (30 分) である. (c) 得られた 103 個の窓化された株価データの自分以外の各ペアの順番付き組み合わせに対して, 対応する窓ごとに移動エントロピーを計算する. 従って各ペアの組み合わせごとに 6 ステップ (30 分) 間隔の時系列移動エントロピーデータが合計 10506 個得られることになる. (d) 得られた移動エントロピーデータのそれぞれに対して, ウェーブレット変換による平滑化を行う. (e) このようにして各時刻  $t = 0, 6, \dots$  においてノード数 103, リンク数 10506 の時系列の重み付き有向ネットワークを得ることができる. このネットワークを対象にして分析を行う.

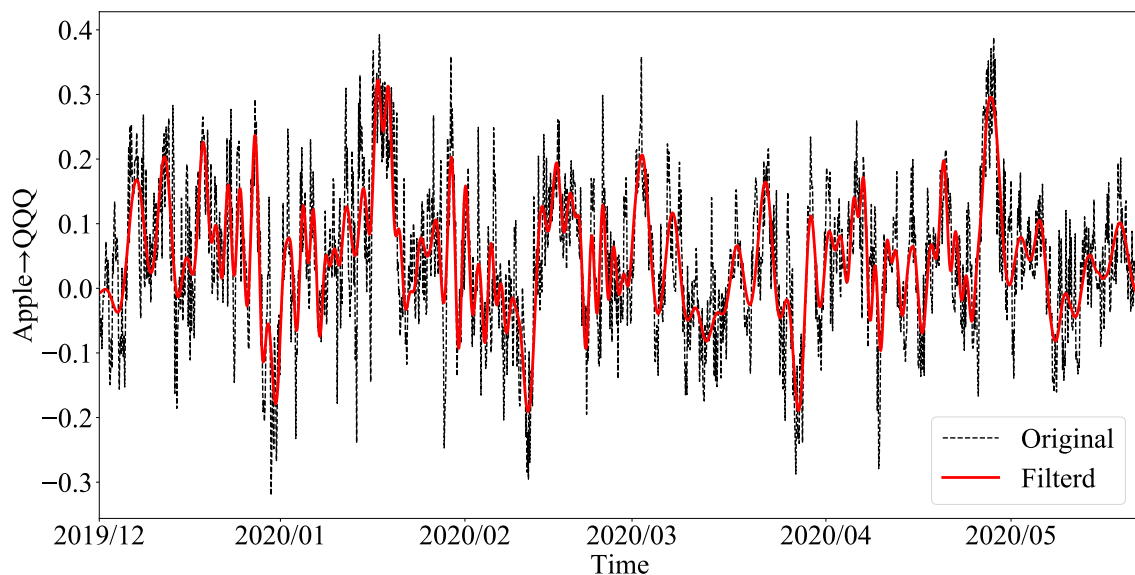


図 2.2 黒い破線が元データ, 赤い実線がフィルタリング後のデータを表す. 横軸は時間, 縦軸は移動エントロピーの値である.

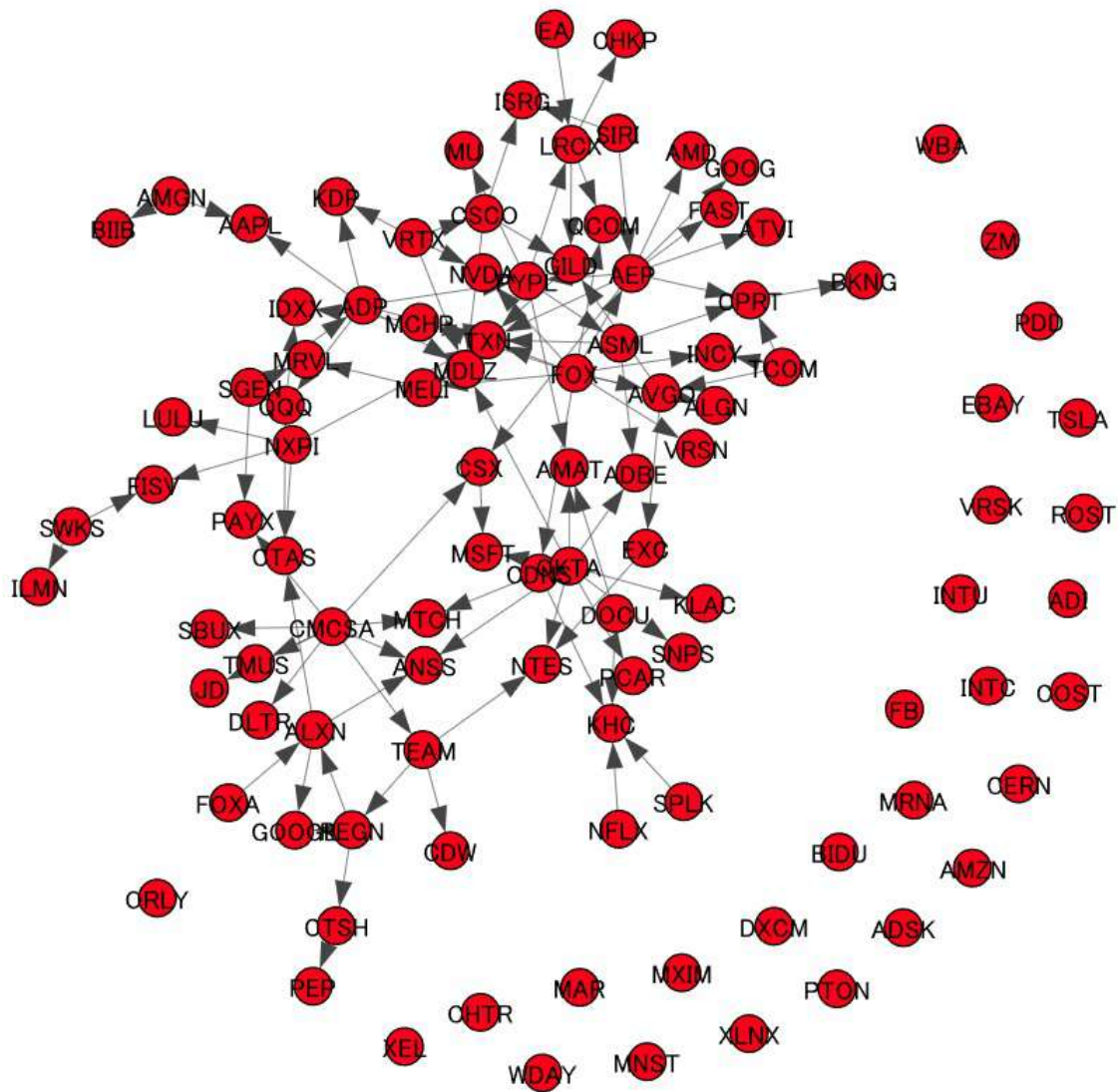


図 2.3 矢印がノード間の有向リンクを表している。ノード名は株式の証券コードを表す。

## 2.4 基本統計量

図 2.4 は期間中の QQQ 価格と各時刻における移動エントロピーの平均 (Mean), 分散 (Variance), 歪度 (Skewness), 尖度 (Kurtosis) について表示したものである。ただし変数間のスケールを統一するために式 2.1 による標準化を施している。

直感的に感じられることは 2020 年 2 月後半から 3 月後半ごろのコロナショック期間中に各種統計量のステージが変わったということだろう。したがって移動エントロ

ピーの分布の変化とショックとの間には何らかの重要な関係があることが示唆される。特に統計量の変化の仕方を見ると、あえて言うならばショック時にネットワークは『均質になる』あるいは『解体される』と表現できるであろう。平均や分散、歪度などが減少する方向に動き、ノードどうしの繋がり強度である移動エントロピーの分布は、歪みがなくばらつきも少なくなり、さらに全体的に値が減少する方向に変化する。結果、各ノードは特徴を失い、まるで同一のノードのようにふるまい、ノード間の繋がりも弱くなるという意味である。このことは移動エントロピーが変数間の影響力や因果性に相当する効果を定量化する量であるため、外因性のショック時にはそのような変数間の因果性が見かけ上失われてしまうからではないかと推察される。

そこで次章では移動エントロピーの分布の変化と株価の変動との関係についてより詳細な検証をする。

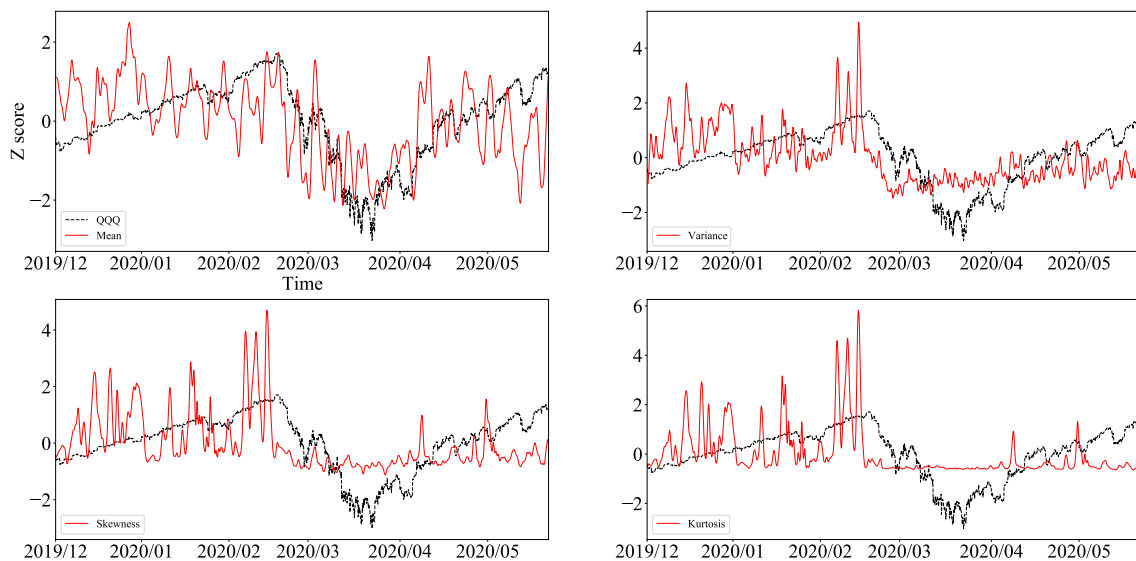


図 2.4 Z score は標準化後の変数の値を表す。コロナショック期間中に各変数のふるまいが変化したことが見て取れる。分散、歪度、尖度については特に顕著である。

## 3 分析

### 3.1 株式市場

#### 3.1.1 分布の変化

図 3.1 は 2020 年 1 月 15 日からおよそ 15 日ごとの移動エントロピーの分布のヒストグラムを表したものである。ただし、縦軸 (Density) については対数スケールで表示されている。図の中段はコロナショックの過渡期であった 2020 年 2 月末から 3 月末の期間に相当する。

最初、分布は左右非対称で右側の裾が厚い形状をしていた。すなわち、一部ではあるが他の銘柄について重大な影響を及ぼすものが存在していたということである。しかし、ショックの間中は裾が消失し、ばらつきも減少し、正規分布系の左右対称な形状に変化しその後少しではあるが元の形状に戻っていったことが窺える。移動エントロピーの分布がばらつきが少なく左右対称であるということはネットワークは均質で特別な場所が存在しないということの意味し、各銘柄は似たようなふるまいをして、ほかの銘柄の株価について重大な影響力を持つ銘柄が存在しないということの意味する。このような現象が起こるのは、通常時であれば例えば Google の株価が上がったから Apple の株価が上がったというような株式市場内での因果関係で説明できるような株価変動が、コロナショックのような外因性の強いショック下では市場内の因果性では説明できない外部からの力によって起こるため、銘柄間の因果関係が見えなくなってしまふからだと考えられる。すなわち、因果性の消滅が移動エントロピーの分布を情報が失われる方向に変化させているということである。

とはいえ、このような分布の変化は全体から見れば小さな変化であるので視覚的な情報のみでは意味のある変化であるのか判断しかねる。従って次節ではこのような分布の変化について統計的に有意であるのかどうかを検証する。

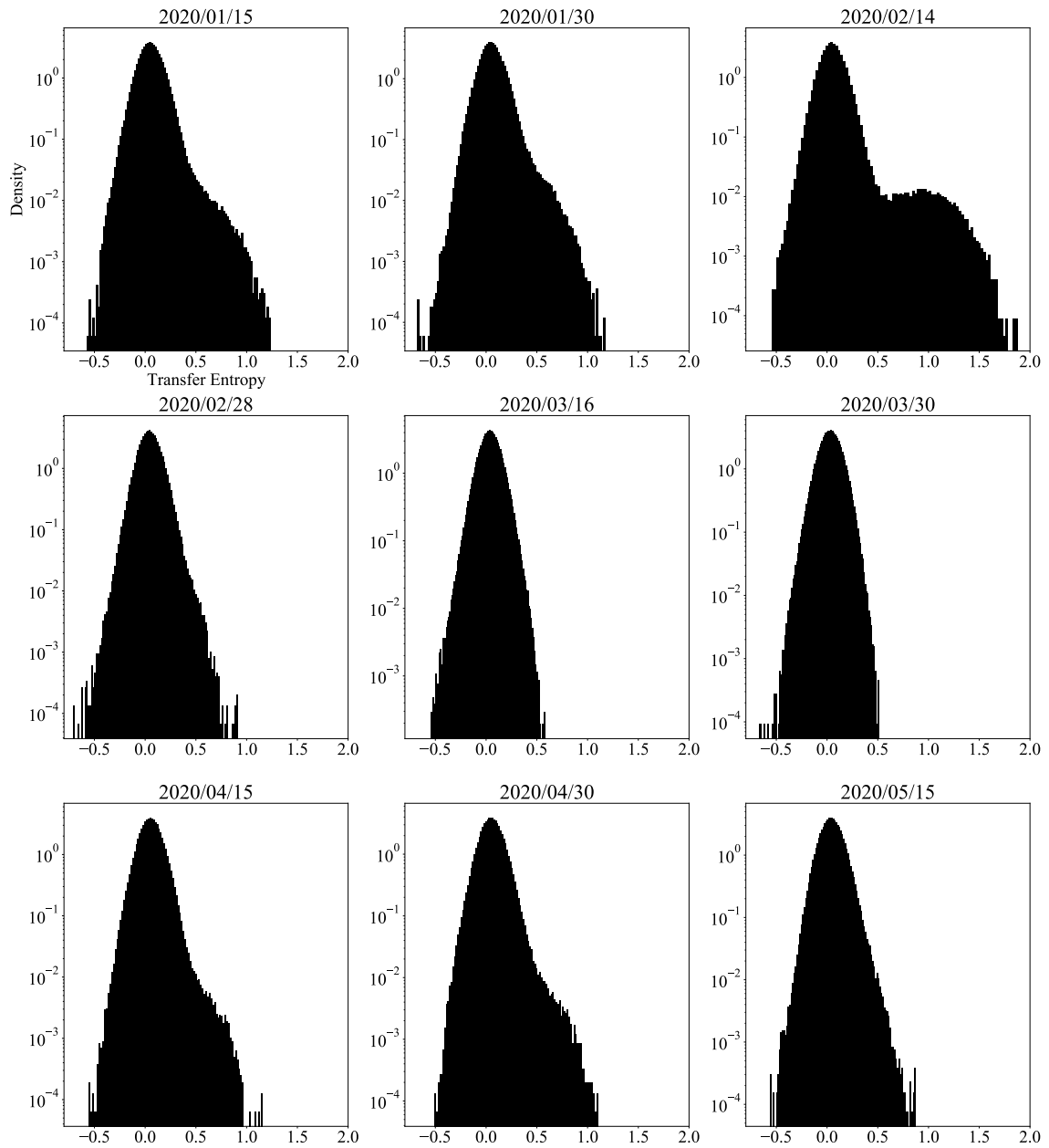


図 3.1 横軸は移動エントロピーの値 (Transfer Entropy), 縦軸は各階級の度数が確率密度になるように規格化した値 (Density) を対数スケールで表示したものを表す. ショック前に健在であった右に厚い裾が中段のショック期間中にはほぼ消失し, その後少しではあるが回復しているように見える.

### 3.1.2 コルモゴロフ-スミルノフ検定

前節で見た移動エントロピーの分布の変化が統計的に有意であるのかどうかを調べるために2標本コルモゴロフ-スミルノフ検定 (Two sample Kolmogorov-Smirnov test, K-S 検定) による検定を行う。

$m, n$  個のデータからなる経験分布  $f(x), g(x)$  の累積分布  $F(x), G(x)$  に対して分布の差異の程度, K-S 検定量  $D$  を,

$$D = \max_x |F(x) - G(x)| \quad (3.1)$$

で定義する。2つの分布は等しいという帰無仮説の下で K-S 検定量  $D$  がある閾値より大きい確率  $\mathbb{P}$  は  $m, n$  が十分大きければ,

$$\mathbb{P} \left( D \sqrt{\frac{mn}{m+n}} \geq c \right) \simeq 2 \sum_{k=1}^{\infty} (-1)^{k-1} \exp(-2k^2 c^2) \quad (3.2)$$

であることが知られている。この関係を用いて2つの分布に有意な差異があるかどうかを検定する手法を2標本コルモゴロフ-スミルノフ検定という。式(3.2)から計算される  $p$  値が有意に小さければ2つの分布は有意に異なるということになる。

今回は0ステップ目から149ステップ目までの150ステップ(10営業日分)の移動エントロピーの分布からランダムに10506個(1ステップ分)をサンプリングし、従って間接的に平均化したデータを平常時の分布  $f$  とみなして固定し、 $t - 149$  ステップ目から  $t$  ステップ目までの150ステップ分の移動エントロピーの分布からランダムに10506個をサンプリングしたものを時刻  $t$  における現在の移動エントロピーの分布  $g_t$  とみなして K-S 検定を行った。

図 3.2 は異なる有意水準 (0.05, 0.01, 0.001) について  $p$  値が有意水準を下回った期間を色付きで示したものである。ただし、実線はそれぞれの時刻における QQQ 価格を表す。やはりコロナショックの期間中に移動エントロピーの分布に有意な変化が起こっていたことがわかる。従って株価変動と移動エントロピーネットワークの間には何らかの重要な関係があることが示唆され、ネットワークが株価の変動に対して何らかの情報を有している可能性がある。しかも、図 3.2 において色が付き始める時点やあるいは図 2.4 において各種統計量の変化が始まる時点は株価変動に若干先行しているように見える。このことはネットワーク情報が未来の株価変動に関する情報を含んで

いるのかどうかということの意味するため、興味深い対象である。また、コロナショックのような数年に一度程度の大規模で稀な現象だけでなく、より小規模で高頻度の価格変動についてもネットワークが何らかの情報を有しているかということにも興味がある。そこで次節以降ではコロナショックから離れて、これらのことについてより深く調べることにする。

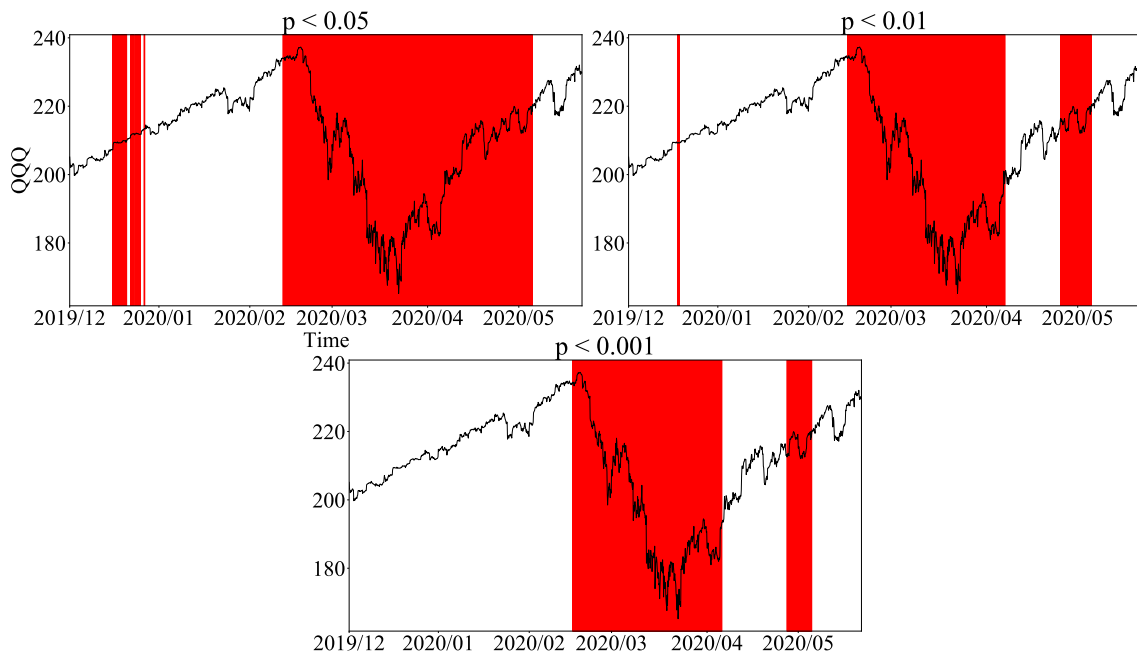


図 3.2 実線は QQQ の価格を表し、色付きの部分は  $p$  値が有意水準を下回っている部分を表す。2 月後半ごろから 3-4 月にかけて株価が大きく下落したところに  $p$  値も低く出ていることがわかる。

## 3.2 VAR モデル

### 3.2.1 目的

この節では QQQ 価格の変化率を被説明変数、各種ネットワーク指標などを説明変数とした時系列モデルを作成することでネットワーク指標が株価について有意に説明力があるか、グレンジャー因果性や、インパルス応答などを検証する。モデルにはベクトル自己回帰モデル (Vector autoregression model, VAR モデル) を使用する。

### 3.2.2 カルバックライブラー情報量

3.1 節でみたような移動エントロピー分布の変化の程度を量的に評価しモデルに登場させるために次で定義するカルバックライブラー情報量 (Kullback–Leibler divergence, KLD) を用いる.

2つの離散型確率分布  $f(x), g(x)$  に対し  $f$  の  $g$  に対するカルバックライブラー情報量  $D(f||g)$  を,

$$D(f||g) \equiv \sum_x f(x) \log \frac{f(x)}{g(x)} \quad (3.3)$$

で定義する.  $D(f||g)$  は  $f, g$  について対称ではないことに注意するべきだが, その値が大きい程分布間の差異が大きいことを表す指標としてよく用いられる.

時刻  $t$  において  $t-14$  ステップから  $t$  ステップまでの 15 ステップ (1 営業日分) の移動エントロピーの分布からランダムに 10506 個 (1 ステップ分) をサンプリングした分布  $f_{t:t-14}$  を時刻  $t$  における現在の分布とみなし, 現在の分布  $f_{t:t-14}$  から見て 2 営業日前の分布に当たる  $t-44$  ステップから  $t-30$  ステップまでの分布から同様にランダムサンプリングした分布  $f_{t-30:t-44}$  を時刻  $t$  における過去の分布とみなして, 時刻  $t$  における移動エントロピー分布の変化の程度を  $f_{t-30:t-44}$  の  $f_{t:t-14}$  に対するカルバックライブラー情報量

$$\text{KLD}_t = D(f_{t-30:t-44} || f_{t:t-14}) \quad (3.4)$$

で定量化する. この場合,  $\text{KLD}_t$  はコルモゴロフ-スミルノフ検定で見たような分布が正常か異常かを判断するような量ではなく, あくまで直近の分布からどの程度分布の変化があったかを評価する量であることに注意する必要がある. これは 3.1 節で捉えようとした移動エントロピー分布の変化が大局的な変化であることに対して, こちらではより高頻度で小規模な変化をモデルの中に登場させることを意図しているからである.

図 3.3 はこのようにして得られた  $\text{KLD}_t$  のグラフである. 見ると,  $\text{KLD}_t$  は多くの時間はおよそ 0.03 近辺の値に留まっているように見えるが, 時折瞬間的にそこから 2 倍から 3 倍程度の値までパルス信号的な増加の仕方をしていることがわかる. 従ってこのパルス信号的な増加のタイミングで移動エントロピー分布に何らかの変化が起こっているのであろうことがわかる.

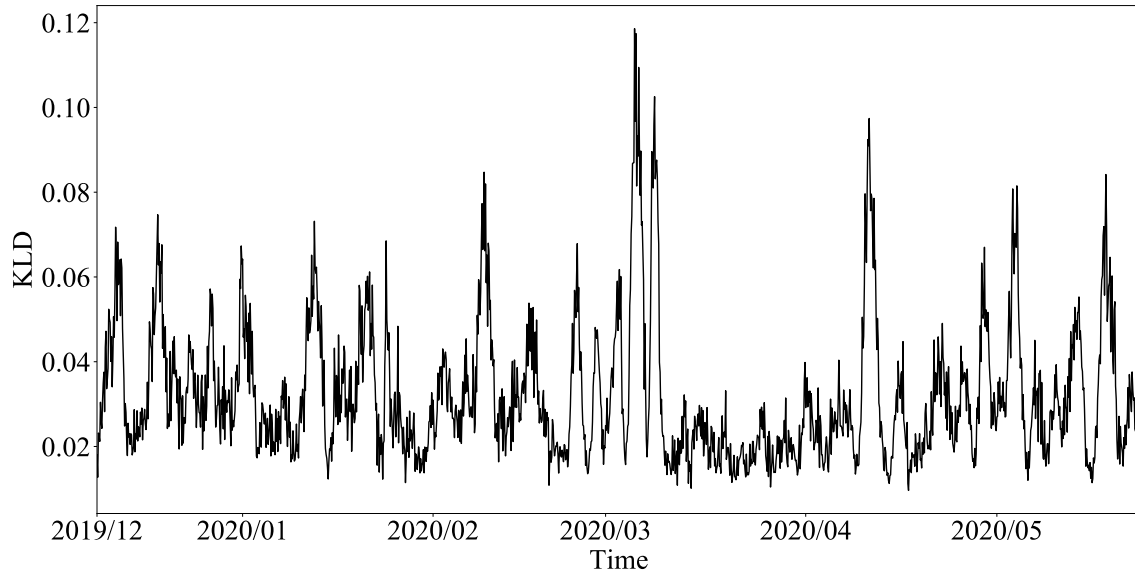


図 3.3 縦軸は  $KLD_t$  の値, 横軸は時間を表す. 多くの時間はおよそ 0.03 近辺で定常しているように見えるが, 時折パルス信号のようなパターンを発しているように見える.

### 3.2.3 VAR モデル

$\mathbf{y}_t = (y_{1,t}, \dots, y_{n,t})^\top$  を  $n$  変量の時系列データとし,  $\Phi_1, \dots, \Phi_L$  をそれぞれ  $n \times n$  行列,  $\mathbf{c} = (c_1, \dots, c_n)^\top$  を定数項ベクトル,  $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \dots, \varepsilon_{n,t})^\top$  を誤差項ベクトルとするとき,

$$\mathbf{y}_t = \mathbf{c} + \sum_{l=1}^L \Phi_l \mathbf{y}_{t-l} + \boldsymbol{\varepsilon}_t \quad (3.5)$$

で表されるモデルを次数  $L$  の VAR( $L$ ) モデルという. 係数の推定には最小二乗法か最尤法が用いられるが, ここでは最小二乗法を用いる.

今回は QQQ の  $t-K$  ステップから  $t$  ステップ間の変化率,

$$\text{Change}_t = \frac{QQQ_t - QQQ_{t-K}}{QQQ_{t-K}} \quad (3.6)$$

を第 1 変数とし, 第 2 変数  $x_t$  をその他の株価データについて式 3.6 同様に計算した変化率や時刻  $t$  における移動エントロピー分布の平均, 分散, 歪度, 尖度の基本統計量および式 3.4 による  $KLD_t$  のいずれかとして,  $\mathbf{y}_t = (\text{Change}_t, x_t)^\top$  とした 2 変量の VAR モデルを作成する. ただし, 変数のスケールを統一し回帰係数の推定を安定化するため

各変数について式 2.1 による標準化を施しておく。それぞれのモデルについて、次数  $L$  の決定には赤池情報量基準 (Akaike's Information Criterion, AIC) の最小化を用いる。式 3.5 を用いて具体的に書き下すと、

$$\text{Change}_t = c_1 + \phi_{1,11}\text{Change}_{t-1} + \phi_{1,12}x_{t-1} + \cdots + \phi_{L,11}\text{Change}_{t-L} + \phi_{L,12}x_{t-L} + \varepsilon_{1,t} \quad (3.7)$$

$$x_t = c_2 + \phi_{1,21}\text{Change}_{t-1} + \phi_{1,22}x_{t-1} + \cdots + \phi_{L,21}\text{Change}_{t-L} + \phi_{L,22}x_{t-L} + \varepsilon_{2,t} \quad (3.8)$$

という 2 本の回帰式が同時に推定されることになるが、実は VAR モデルの場合は式 3.7, 式 3.8 のそれぞれについて個別に最小二乗推定を行えばよいことが知られている。なお、今興味があるのは式 3.7 の方である。

変化率をとる期間である  $K$  については  $K = 6$ (3 時間) や  $K = 45$ (3 営業日) などの値を検討したが最終的に  $K = 225$ (15 営業日) を採用した。

### 3.2.4 AIC による説明変数の比較

式 3.7, 式 3.8 において  $x_t$  に時刻  $t$  における移動エントロピー分布の平均 (Mean), 分散 (Var), 歪度 (Skew) 尖度 (Kur), 式 3.4 による  $\text{KLD}_t(\text{KLD})$  や式 3.6 と同様に定義した QQQ 以外の株価変化率などをとり、探索する次数  $L$  の上限を 150 としたうえでそれぞれの場合で AIC 最小化によるモデル選択を行い、そのときの AIC 値を比較した。

図 3.4 はそれらのうち上位 20 について表示したものである。5 つのネットワーク指標が上位 5 に位置しており、ネットワーク指標が株価に関する情報を上手く捨象し、情報としての質を向上させている可能性が示唆される。ただし、各銘柄の株価情報を QQQ は部分的に内包しているため、株価データから QQQ への説明力はそもそも低い可能性がある。そのためフェアな比較ではないだろう。従ってこのような分析はあくまで例示的であり、より詳細な分析が必要である。

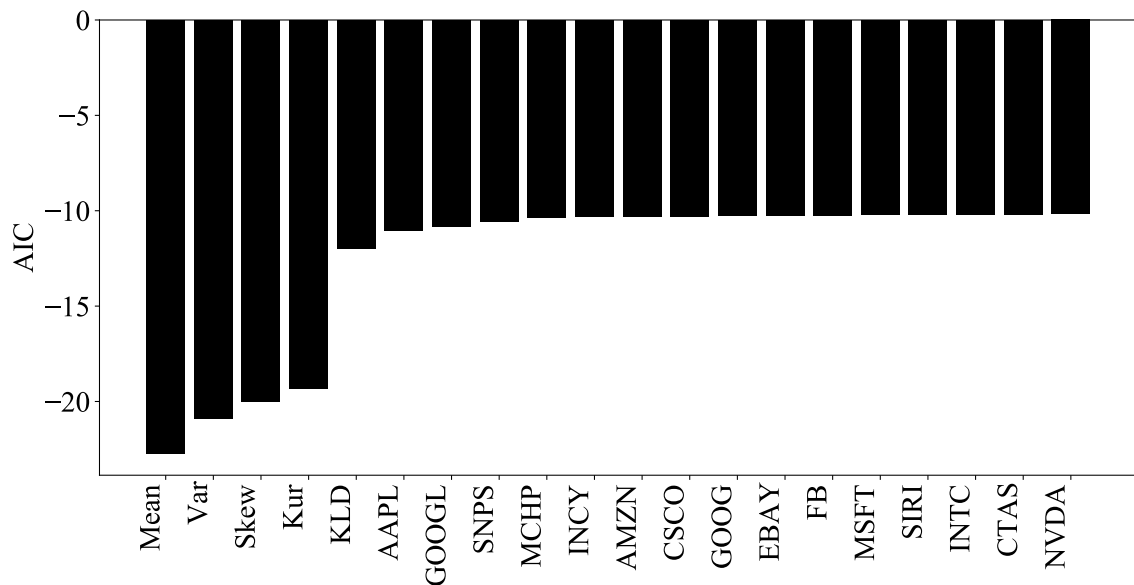


図 3.4 縦軸はそれぞれのモデルについて次数  $L$  の決定時における最も低い AIC の値を表しており、小さい程当てはまりが良いとされる。AAPL などは株式の証券コードを表す。ネットワーク指標 (Mean, Var, Skew, Kur, KLD) が軒並み上位に位置していることがわかる。

### 3.2.5 単位根検定

VAR モデルを用いて  $t$  検定やグレンジャー因果検定など統計的な検定を行いたい場合には各変数がトレンドのない定常過程であることが理論的に要求される。しかしながら株価データを含む多くの経済データは普通定常過程ではない。そのような場合でも各変数について差分をとる処理を施せば定常過程とみなせる場合がほとんどである。式 3.6 による処理は差分を取る処理に相当し、実際、小さい  $K$  については  $\text{Change}_t$  は定常過程であるとみなすことができるであろう。しかし大きい  $K$  については振動が緩やかになり限られた期間ではトレンドが存在してしまう可能性がある。

与えられた時系列データが定常過程とみなせるかどうかについては単位根検定を用いる。今回は検定方法に拡張ディッキー-フラー検定 (Augmented Dickey-Fuller test, ADF test) を用いている。検定の結果、 $K = 225$  の場合は多くの株価データが有意水準 5% で単位根を持つことが棄却されなかった。そこでネットワーク指標を含むすべてのデータに対して 1 階の差分を取る処理、

$\Delta y_t = y_t - y_{t-1}$  を施した。この場合すべての変数は有意水準 5% で単位根をもつこと

が棄却された。従って式 3.5 によるモデルは

$$\Delta \mathbf{y}_t = \mathbf{c} + \sum_{l=1}^L \Phi_l \Delta \mathbf{y}_{t-l} + \boldsymbol{\varepsilon}_t \quad (3.9)$$

と変更され、式 3.7, 式 3.8 は、

$$\begin{aligned} \Delta \text{Change}_t = c_1 + \phi_{1,11} \Delta \text{Change}_{t-1} + \phi_{1,12} \Delta x_{t-1} + \cdots + \phi_{L,11} \Delta \text{Change}_{t-L} \\ + \phi_{L,12} \Delta x_{t-L} + \varepsilon_{1,t} \end{aligned} \quad (3.10)$$

$$\begin{aligned} \Delta x_t = c_2 + \phi_{1,21} \Delta \text{Change}_{t-1} + \phi_{1,22} \Delta x_{t-1} + \cdots + \phi_{L,21} \Delta \text{Change}_{t-L} \\ + \phi_{L,22} \Delta x_{t-L} + \varepsilon_{2,t} \end{aligned} \quad (3.11)$$

に変更される。

なお、単に予測がしたいだけならこのような変更は必要ない。また今回の場合は差分の差分をとるという処理を施しているため結果の解釈性は低い。従って式 3.9 によるモデルを用いるのはあくまで定常性の過程が必要な部分だけとし、インパルス応答や次章における予測については式 3.5 によるモデルを用いる。

### 3.2.6 回帰結果

ネットワークの 4 つの基本統計量、平均、分散、歪度、尖度を説明変数としたモデルは AIC のスコアこそよかったが、有意水準 5% で効いている回帰係数の数は高々 2 つしかなかった。一方カルバックライブラー情報量 KLD を説明変数としたモデルは次数  $L$  は  $L = 109$  で、ラグ 14, 20, 38, 70, 85, 92, 107, 108 の 8 個の回帰係数が 5% 有意であった。そのため KLD を用いたモデルがネットワーク指標を用いたモデルの中では最も有望であるように思われる。

一方、株価データを使用するモデルについては例えば Apple 社 (AAPL) を説明変数とするモデルでは次数  $L$  は  $L = 143$  でラグ 11, 13, 25, 30, 33, 40, 44, 53, 54, 73, 77, 83, 93, 99, 111, 118, 121, 124, 131, 136, 141, 143 の 22 個の回帰係数が 5% 有意であった。他の株価データでも同じく有意な係数の数は株価データを用いたモデルの方がネットワーク情報を用いたモデルよりも多い傾向にあったが、これは当然といえば当然の結果であってこのことをもってネットワーク指標の非有効性を結論するべきではないだろう。

詳細な回帰結果については付録に示し,ここではKLDを説明変数とするモデルについて回帰結果の一部のみを表3.1に示す.

表 3.1 カルバックライブラー情報量 KLD を説明変数とするモデルについて, 得られた回帰結果の一部のみを表示したものである. 例えば L1.KLD などはラグ 1 の  $KLD_{t-1}$  の項を表す. その他の変数名についても同様である. また †, \*, \*\*, \*\*\* はそれぞれ  $p$  値が 10, 5, 1, 0.1% を下回っていることを表す.

	coefficient	std error	$t$ stat	$p$ value
Const.	0.000	0.003	0.085	0.933
L1.QQQ	-0.156	0.032	-4.833	0.000***
L1.KLD	0.103	0.132	0.781	0.435
L2.QQQ	0.009	0.033	0.260	0.795
L2.KLD	-0.140	0.146	-0.956	0.339
L3.QQQ	-0.066	0.033	-2.032	0.042*
L3.KLD	0.026	0.150	0.176	0.860
L4.QQQ	-0.006	0.033	-0.193	0.847
L4.KLD	-0.106	0.147	-0.721	0.471
L5.QQQ	0.007	0.033	0.219	0.827
L5.KLD	-0.026	0.148	-0.176	0.860
L6.QQQ	0.054	0.033	1.648	0.099†
L6.KLD	0.033	0.149	0.222	0.824
L7.QQQ	0.019	0.033	0.575	0.565
L7.KLD	0.257	0.148	1.736	0.082†
L8.QQQ	0.018	0.032	0.561	0.575
L8.KLD	0.003	0.149	0.017	0.986
L9.QQQ	0.059	0.032	1.813	0.070†
L9.KLD	-0.096	0.149	-0.645	0.519
L10.QQQ	-0.073	0.032	-2.249	0.025*
L10.KLD	0.259	0.149	1.738	0.082†
L11.QQQ	0.008	0.033	0.247	0.805
L11.KLD	-0.118	0.149	-0.792	0.428
L12.QQQ	0.212	0.033	6.490	0.000***
L12.KLD	-0.115	0.149	-0.771	0.441
L13.QQQ	-0.059	0.033	-1.761	0.078†
L13.KLD	0.036	0.149	0.243	0.808

### 3.2.7 グレンジャー因果性

次に、各ネットワーク指標から QQQ 価格変化率  $\text{Change}_t$  へのグレンジャー因果性、およびその逆向きのグレンジャー因果性を検定する。式 3.10 において  $x_t$  から  $\text{Change}_t$  へのグレンジャー因果性がないということは次の帰無仮説、

$$\phi_{1,12} = \dots = \phi_{L,12} = 0 \quad (3.12)$$

すなわち  $x_t$  の関わる項の係数が全て 0 ということで定式化される。この帰無仮説の下で  $F$  検定をすることでグレンジャー因果性を検定できる。具体的にはまず帰無仮説の制約の下で推定されたモデルにおいて残差二乗和  $\text{SSR}_0$  を計算し、制約なしのモデルにおいても同様に残差二乗和  $\text{SSR}_1$  を計算する。ただし残差二乗和  $\text{SSR}$  とは  $y_t$  を時刻  $t$  における観測値、 $\hat{y}_t$  をその予測値とするとき  $T$  をデータの長さとして、

$$\text{SSR} = \sum_{t=0}^T (y_t - \hat{y}_t)^2 \quad (3.13)$$

で定義される。このとき次の  $F$  統計量は自由度  $(L, T - 2L - 1)$  の  $F$  分布に従うことが知られているため、これによって検定する。

$$F = \frac{(\text{SSR}_0 - \text{SSR}_1)/L}{\text{SSR}_1/(T - 2L - 1)} \quad (3.14)$$

表 3.2 は検定結果である。変数名 QQQ, Mean, Var, Skew, Kur, KLD はそれぞれ各時刻  $t$  における QQQ 価格変化率  $\text{Change}_t$ , 移動エントロピー分布の平均, 分散, 歪度, 尖度, およびカルバックライブラー情報量  $\text{KLD}_t$  を表す。他のすべての結果は有意では無かったが、カルバックライブラー情報量から QQQ ( $\text{KLD} \rightarrow \text{QQQ}$ ) へは有意水準 5% でグレンジャーの意味での因果性があった。このことはカルバックライブラー情報量が将来の株価変動について重要な情報を持っていることを示唆する。すなわち、QQQ 価格のようなマクロ的な指標に変化が起こることに先立って株式市場には何らかの変化が起こっていて、その変化をカルバックライブラー情報量がネットワーク構造の変化という形で捉えているのだと解釈できる。これは 3.1.2 節で述べたネットワークの統計量や移動エントロピー分布のステージの変化がコロナショックに先立って起こっているように見えるということに一定の説得力をもたらすものであり意味のある結果である。また、以上のことはカルバックライブラー情報量が株価予想のための重要な

説明変数になり得ることを示唆する。株価予想については次章で検討し、次節ではカルバックライブラー情報量などの統計量が QQQ にどのような意味で影響を与えるのかを検証する。

表 3.2 矢印の向きは因果の向きを表す。例えば QQQ → Mean は QQQ 価格変化率,  $\text{Change}_t$  から時刻  $t$  における移動エントロピー分布の平均, Mean へのグレンジャー因果性を表し,  $p$  value はその  $p$  値を表す。

	$p$ value
QQQ→Mean	0.983
QQQ→Var	1.000
QQQ→Skew	1.000
QQQ→Kur	1.000
QQQ→KLD	0.172
Mean→QQQ	0.972
Var→QQQ	1.000
Skew→QQQ	0.999
Kur→QQQ	1.000
KLD→QQQ	0.039*

### 3.2.8 インパルス応答

本節では VAR モデルの (直交化) インパルス応答関数を求めることによってカルバックライブラー情報量などのネットワーク指標が QQQ 価格変化率  $\text{Change}_t$  にどのような形で影響を及ぼしているのかを検証する。式 3.7, 式 3.8 によるモデルにおける  $x_t$  から  $\text{Change}_t$  への非直交化インパルス応答関数  $\text{IRF}_{x \rightarrow \text{Change}_t}(k)$  は,

$$\text{IRF}_{x \rightarrow \text{Change}_t}(k) = \frac{\partial}{\partial \varepsilon_{2,t}} \text{Change}_{t+k} \quad (k = 0, 1, \dots) \quad (3.15)$$

で定義される。ただし,  $\varepsilon_{2,t}$  は非確率変数として計算される。更に, 誤差項  $\varepsilon_{1,t}, \varepsilon_{2,t}$  の互いの影響を排除するために誤差項の分散共分散行列を三角分解を用いて互いに無相関になるように直交化した上で改めて式 3.15 を計算したものが直行化インパルス応答関数, あるいは単にインパルス応答関数と呼ばれるものであり, 本稿ではこれ以降単にイ

ンパルス応答あるいはインパルス応答関数といった場合は直行化インパルス応答関数を指すものとする。  $x_t$  から  $\text{Change}_t$  へのインパルス応答  $\text{IRF}_{x \rightarrow \text{Change}}(k)$  は  $x_t$  に起こったモデルでは予想できなかった未知の変動(ショック)が  $k$  ステップ先の  $\text{Change}_t$  にどのように伝わっていくかを表すものである。

図 3.5a, b, c, d はそれぞれ  $x_t$  に時刻  $t$  における移動エントロピー分布のカルバックライブラー情報量  $\text{KLD}_t(\text{KLD})$ , 平均 (Mean), 分散 (Var), 歪度 (Skew) を入れたときの 500 時間 (約 67 営業日) 間のインパルス応答を表したものである。ただし赤い破線は 90% 信頼区間の境界を表す。図 3.5a を見ると, カルバックライブラー情報量にモデルでは予測できない未知の 1% の変動が起きたとき, およそ 50 時間後に QQQ 価格変化率はそれとは逆方向におよそ 0.06% モデルの予測を超えて変動することがわかる。また, 図 3.3 を見てわかるように, カルバックライブラー情報量に起こるモデルでは予測できない変動とは, パルス信号的な増加方向への変動であろうことがわかる。従ってこれは移動エントロピーの分布に何らかの変化が起こりカルバックライブラー情報量が増大するとおよそ 50 時間後をピークとして株価は下落するということ意味している。

次に, 図 3.5b, c, d を見ると, これらは概形として似ているが, その解釈は移動エントロピー分布の平均, 分散, 歪度にモデルでは予測できない 1% の変動が起こるとおよそ 50 時間後をピークとして QQQ 価格変化率をモデルの予測を超えて同じ方向に 0.02% から 0.04% 程度変動させるということである。特に, 移動エントロピー分布が平均, 分散, 歪度が減少し分布の特徴が失われるような方向に変化するとき, 株価は下落することがわかる。これは 3.1.1 節で述べた移動エントロピー分布の変化と株価変動との関係を追認するものとなっている。

しかしながら, すべてのグラフにおいてほとんどの時間で信頼区間が正負両方の領域にまたがっているため統計的な信頼度は薄い。とはいえ, 今述べたことはこれまで見てきたものと整合的でありそのことによって一定の説得力がもたらされるべきであろう。また, 統計的な信頼性が薄いなかでも図 3.5a は比較的良い結果が得られたように見えるので, ネットワーク指標の中では QQQ 価格の変動に対して何らかの情報を持っているであろう最も有力な変数はやはりカルバックライブラー情報量であろう。

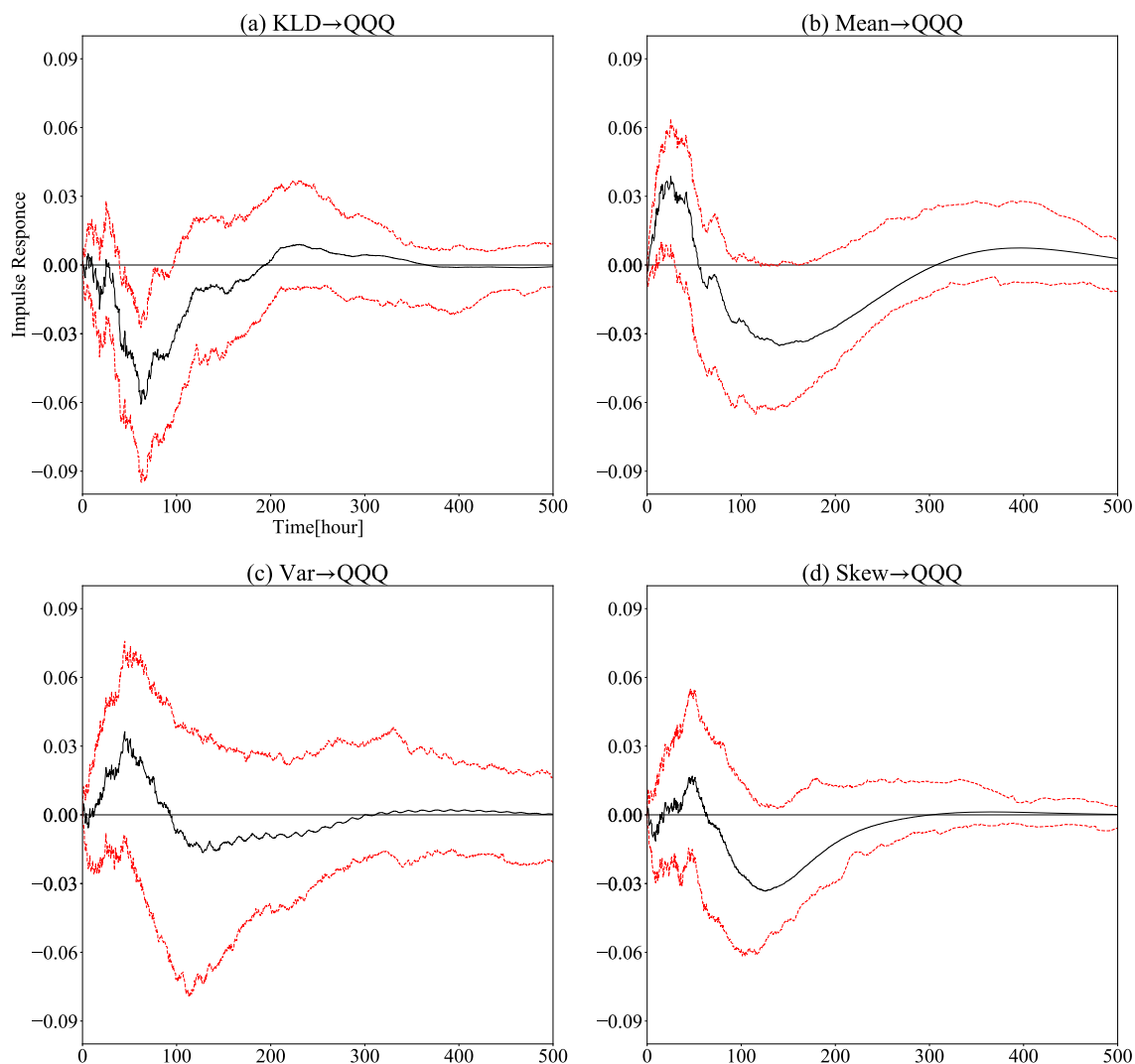


図 3.5 (a), (b), (c), (d) はそれぞれカルバックライブラー情報量  $KLD_t$  から QQQ へのインパルス応答, および時刻  $t$  における移動エントロピー分布の平均 (Mean), 分散 (Var), 歪度 (Skew) から QQQ へのインパルス応答を表す. 縦軸はインパルス応答の大きさ, 横軸はショックを与えた瞬間から経過した 1 時間ごとの時間を表す. また, 赤い破線は 90% 信頼区間の境界を表す.

### 3.2.9 小まとめ

ここまでは, 株式市場を対象とした分析を行ってきた. 得られたことは, QQQ 価格のような株式市場全体を代表するマクロな指標が何らかのショックなどによって下落するとき, 移動エントロピー分布は平均や分散, 歪度などが減少する方向に動き, 従っ

て移動エントロピー分布はばらつきや歪みが無い特徴を失う方向に変化すること。そのとき株式市場のネットワークは均質化され、特別なノードがなくなった状態にあると解釈できること。また、グレンジャー因果検定などの結果を踏まえると、そのような分布の変化と株価の下落はむしろ、分布の変化の方が先行している可能性があるということである。

なぜ株価の下落時にのみ移動エントロピーの分布にこのような変化が起きるのかを考えると、その理由は株式市場における株価の上昇と下落の非対称性に帰することができそうである。株式市場では買いは Long (長期)、売りは Short (短期) と別称されるように、株価の上昇は長い時間をかけてゆっくり起こるのに対し、下落は短期的で急激であることが多い。すなわち、株式市場は株価の上昇局面にある状態が通常状態で、株価の下落局面は異常状態だと考えられる訳であるが、そのような異常状態においては、大抵例えばコロナショックのような強い外部要因が存在していたり、あるいはそのような外部要因が見つけれなかったとしても株価の下落はそもそも連動するという株式市場本来の性質によって各銘柄の区別がなくなり、いずれにせよ結果的に株式市場内の因果性が見えなくなってしまうことによって、このような移動エントロピー分布の変化が起こるのではないかと考えられる。

次節では今回得られたことの再現性を試すため、株式市場を離れて暗号通貨市場について VAR モデルを用いた同様の分析をすることとする。

### 3.3 暗号通貨市場

#### 3.3.1 暗号通貨

暗号通貨 (Cryptocurrency) とは、分散的な認証システムによってその価値が担保される通貨のことである。

一般的な通貨はそれぞれの国における中央銀行によって発行され、その通貨に対する信用は、その国の政府に対する信用という形で保証される。また、このような従来型の通貨を用いて行われる取引は、特にインターネット上などの非対面型の取引の場合、その取引の信頼性を保証するためには皆が認める第三者、すなわち銀行などの金融機関の存在が必要である。それに対して暗号通貨はこのような中央集権的な構造を必要としない。その代わりに暗号通貨を使用するすべてのユーザーが共同してその価値を保証するブロックチェーン (Blockchain) と呼ばれる分散型の認証システムによって取引の信頼性や貨幣価値が維持される仕組みになっている (Delgado-Segura et al. 2018; Miraz and Ali 2018; Monrat, Schelén, and Andersson 2019; Yaga et al. 2019)。

暗号通貨の本質的な部分はブロックチェーンを用いた分散型の認証システムである。その概念は 2008 年に Satoshi Nakamoto という匿名の人物によって初めて世に出された (Nakamoto 2008)。また彼は、彼のアイデアを認める人々とともに世界初の暗号通貨となるビットコイン (Bitcoin) の開発に成功した。ビットコインに代表される暗号通貨は現在では多く存在しており、その市場規模も年々急速に拡大している (Hileman and Rauchs 2017)。

また現在では暗号通貨の根幹技術であるブロックチェーン技術の通貨としての利用にとどまらないより一般の集団認証システムとしての応用もなされるようになってきている (Abou Jaoude and Saade 2019)。例えば、アート作品などをインターネット上でデータとして取引する際にそのオリジナリティを保証するための手段としてブロックチェーンの技術が活用されたりしている (Zeilinger 2018)。

また近年では中国や日本、アメリカ、カナダ、ヨーロッパなどを含む多くの国々の中央銀行が中央銀行が発行する暗号通貨である中央銀行デジタル通貨 (Central Bank Digital Currency, CBDC) の発行可能性について真剣に研究し始めている (Boar, Holden, and Wadsworth 2020; Mancini-Griffoli et al. 2018)。特に中国は CBDC の導入についてその他の国に先んじており、既にいくつかの地域で実験的な運用を済ませ

ている (Náñez Alonso, Jorge-Vazquez, and Reier Forradellas 2021).

いずれにせよ暗号通貨はこれから先の将来において、より深くより広範囲に経済に浸透していくものと思われる。従って暗号通貨市場について分析をすることは経済学的に意味のあることである。

ただし、本節では株式市場の分析のために使用した手法の再現性を検証することが目的であって、暗号通貨市場について個別で込み入った分析はしない。使用する手法も株式市場の分析において試したのと同じものである。

### 3.3.2 データ

今回使用するデータは株式市場の場合と同じく Tiingo(<https://www.tiingo.com>) より入手した。具体的には 2021 年 9 月 23 日時点において時価総額上位 100 であった通貨のうちデータの入手可能であった 30 の通貨を対象とし、2021 年 4 月 1 日から 2021 年 5 月 31 日までを期間として 10 分ごとの価格データを入手した。ただし、価格はすべて米ドル基準で測られている。

なお、暗号通貨市場には株式市場における NASDAQ100 指数などの代表指数となるものが存在しない。そこで今回はビットコイン価格を暗号通貨市場の代表指標とみなすことでその代替とした。

### 3.3.3 前処理および移動エントロピーネットワークの構築

得られた暗号通貨価格データに対して 2 章において施したと全く同様の欠損処理、標準化、窓化、移動エントロピーの計算およびフィルタリング処理を行った。ただし、暗号通貨価格の元データが 10 分間隔のデータであることに起因して最終的に得られる移動エントロピーデータは株式市場の場合とは違い、1 時間間隔になっている。また、暗号通貨市場は 1 日 24 時間開いているため、1 時間ごとの時系列データではちょうど 24 ステップが 1 営業日に相当する。

参考のために、図 3.6 はこのようにして得られた暗号通貨市場における移動エントロピーネットワークの一例である。2021 年 4 月 1 日における移動エントロピーネットワークを全体の上位 5% の重みをもつリンクのみを表示したものである。ノード名はそれぞれの暗号通貨の通貨コードを表す。

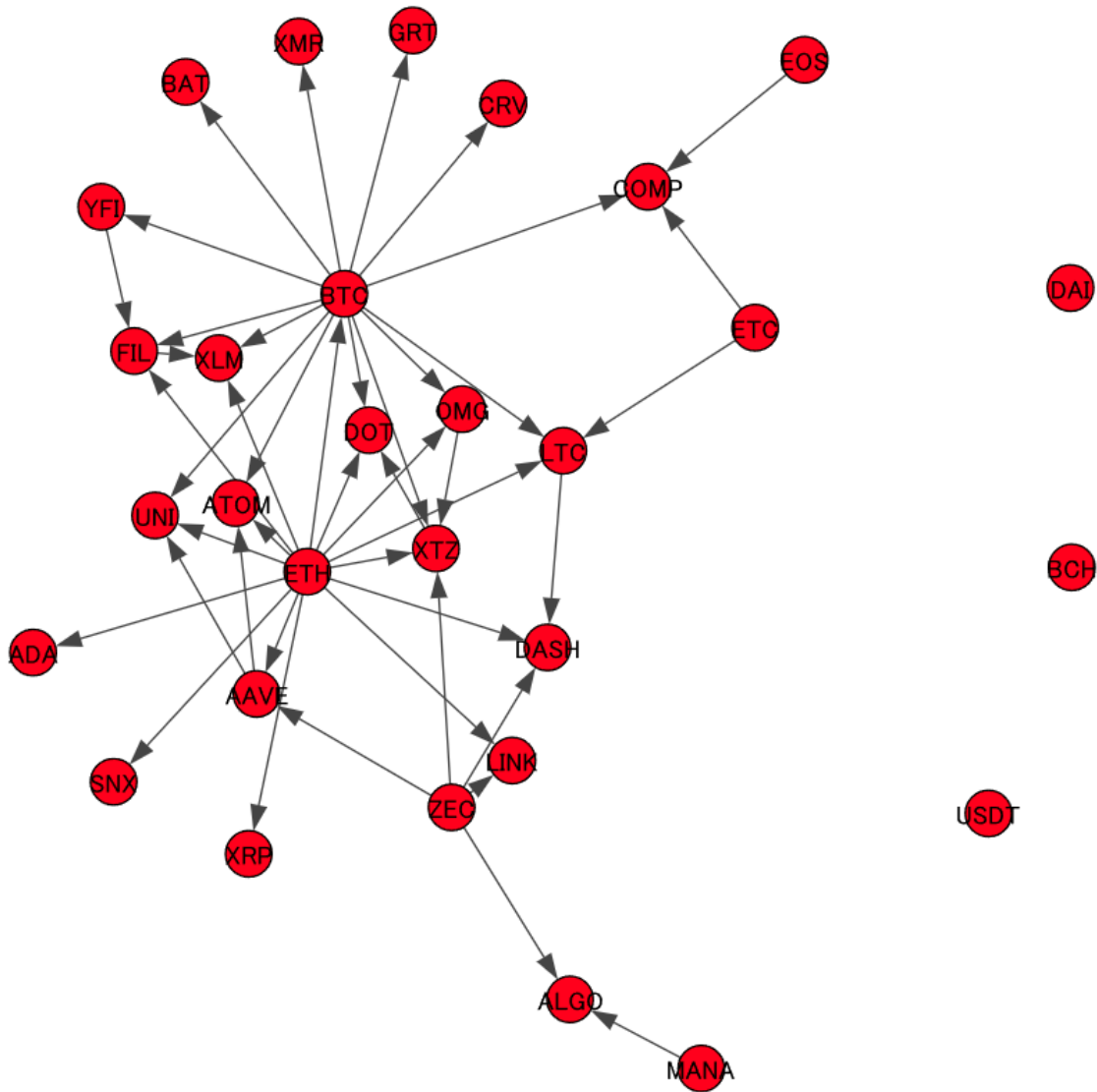


図 3.6 矢印が各ノード間の有向リンクを表し、ノード名は通貨コードを表す. 例えば、通貨コード BTC および ETH はそれぞれビットコインとイーサリアムのことである.

### 3.3.4 基本統計量およびカルバックライブラー情報量

株式市場の場合と同様に得られた時系列の移動エントロピー分布に対して各時刻  $t$  における平均 (Mean), 分散 (Variance), 歪度 (Skewness), 尖度 (Kurtosis) および式 3.4 によるカルバックライブラー情報量  $KLD_t$  を計算した. ただし, カルバックライブラー情報量を計算する際のランダムサンプリングのサンプル数については, 今回の場合は

870 個 (1 ステップ分) となっている。

### 3.3.5 VAR モデル

株式市場の場合と同じく 2 変量 VAR モデルを用いて分析を行う。今回はビットコイン価格の変化率を被説明変数とし、その他の暗号通貨価格の変化率や前節で定義した各種のネットワーク指標を説明変数とする。すなわち、式 3.6 において  $QQQ_t$  を時刻  $t$  おけるビットコイン価格  $BTC_t$  に代えたもの、

$$\text{Change}_t = \frac{BTC_t - BTC_{t-K}}{BTC_{t-K}} \quad (3.16)$$

を改めて  $\text{Change}_t$  とし、その他の暗号通貨についても同様に価格変化率を計算する。また、変化率を計算する期間である  $K$  には株式市場の場合と同じく  $K = 225$  を継承する。これらの変更によって再び式 3.7、式 3.8 による VAR モデルを作成する。モデルの次数  $L$  の決定には株式市場の場合と同じく探索する  $L$  の最大値を 150 として AIC の最小化を用いる。

### 3.3.6 AIC による説明変数の比較

株式市場の場合と同じく、まず AIC による各説明変数の比較を行う。それぞれのモデルにおいて AIC の最小化を行い最も結果の良かった上位 20 について示したものが図 3.7 である。ただし、変数名 Mean, Var, Skew, Kur, KLD は時刻  $t$  における移動エントロピー分布の平均, 分散, 歪度, 尖度, および  $KLD_t$  を表し、その他のものは暗号通貨の通貨コードを表す。

株式市場の場合とほぼ同じく、ネットワーク指標が上位 6 に位置する結果となった。しかも今回の場合は、株式市場の場合にあった  $QQQ$  の価格データが他の株価データの情報を内包してしまっているというような論理的な問題が存在しないため、結果には一定の説得力がある。従ってネットワーク指標は少なくとも単純な価格データよりは情報の捨象に成功し、情報としての質を向上させていると考えてよさそうである。

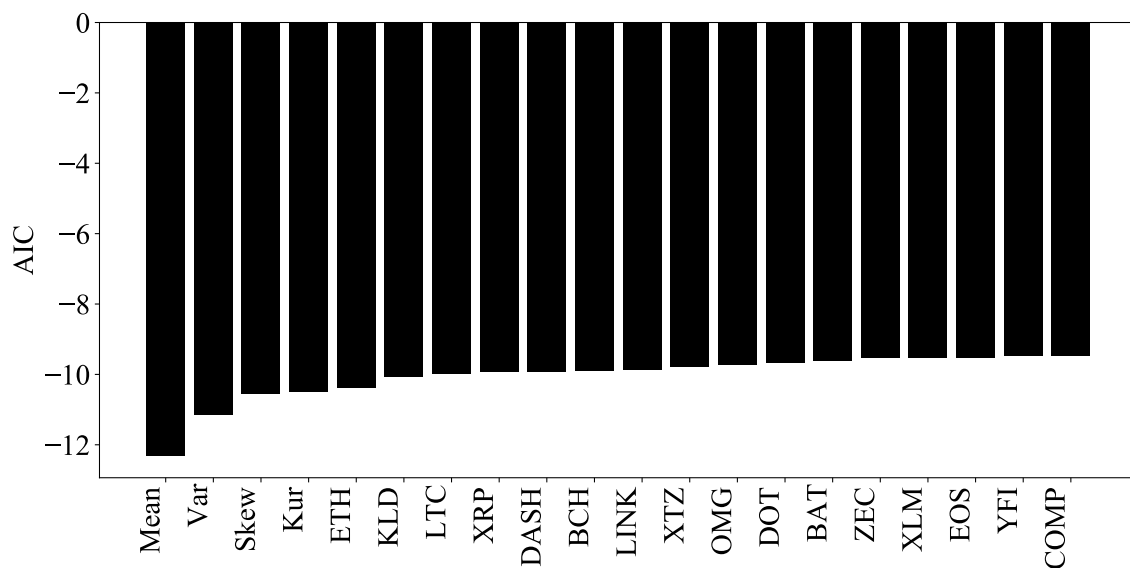


図 3.7 株式市場の場合と同じく、ネットワーク指標が軒並み上位に位置していることがわかる。ちなみに、ネットワーク指標に食い込んで上位に入った ETH はイーサリアムのことである。

### 3.3.7 単位根検定

株式市場の場合と同じく、統計的な検定を行うために必要な単位根検定を行う。検定の結果、やはり多くの暗号通貨データで有意水準 5% で単位根を持つことが棄却されなかったが、すべての変数に対し 1 階の差分を取る処理  $\Delta y_t = y_t - y_{t-1}$  を施した場合はすべての変数について有意水準 5% で単位根を持つことが棄却された。従って  $t$  検定とグレンジャー因果検定を行う部分については式 3.10, 式 3.11 によるモデルを用い、それ以外の場合には式 3.7, 式 3.8 によるモデルを用いる。

### 3.3.8 回帰結果

それぞれのモデルについて得られた回帰結果のうち、カルバックライブラー情報量 (KLD) を説明変数としたモデルは次数  $L$  が  $L = 62$  で、ラグ 4, 13, 14, 28, 29, 55, 58 の 7 つの係数が 5% 有意であった。これはその他のネットワーク指標や暗号通貨価格データを説明変数としたモデルの中で最多であった。その他のネットワーク指標では移動エントロピー分布の尖度 (Kur) や歪度 (Skew), 分散 (Var) を説明変数としたモデルが有意な回帰係数の数が 3-4 つで、これは全体の中で 3 番目または 4 番目に多い数

であった。

また、暗号通貨価格データを説明変数としたモデルでは5%有意な係数の数が3つを超えているものの数は5つだけでそれ以外については5%有意な係数は高々2つしかなかった。

このように、株式市場の場合と比較して全体的に有意な回帰係数の数は少なかった。特に暗号通貨価格データを説明変数としたモデルで有意な回帰係数の数が少なかったのは株式市場の場合と対称的であるが、これはQQQ価格データがその他の株価データの情報を内包していることに対して、今回の場合はそれぞれの暗号通貨価格データは理論的には独立であるはずなので自然な結果ではある。その中でKLDをはじめとしたネットワーク指標が相対的に有意な回帰係数の数が多かったということは一定の意味のある結果であるように思われる。

参考のために、図3.8は各モデルについて5%有意であった回帰係数の個数 (Significant Coefficients) の上位20について表したものである。

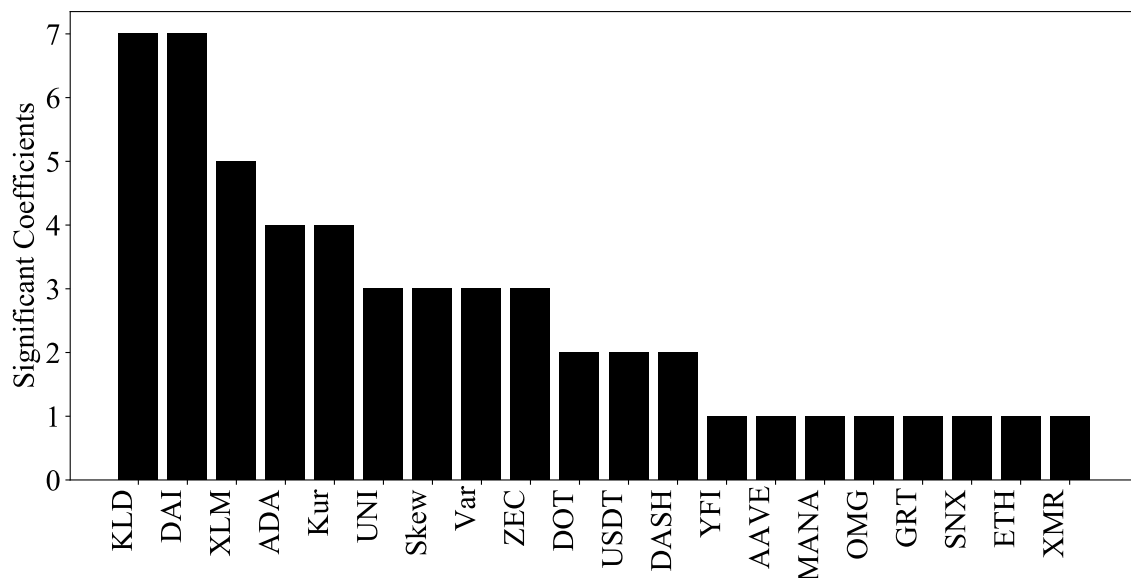


図 3.8 縦軸は各モデルにおいて5%有意であった回帰係数の個数を表す。全体として値は小さいものの、カルバックライブラー情報量 KLD をはじめとして、ネットワーク指標が相対的に上位に位置していることがわかる。

また、表3.3はカルバックライブラー情報量を説明変数としたモデルについて具体的な回帰結果の一部のみを示したものである。詳細な結果については付録にて示す。

表 3.3 カルバックライブラー情報量 KLD を説明変数とするモデルについて, 得られた回帰結果の一部のみを表示したものである. 例えば L1.KLD などはラグ 1 の  $KLD_{t-1}$  の項を表す. その他の変数名についても同様である. †, \*, \*\*, \*\*\* はそれぞれ  $p$  値が 10, 5, 1, 0.1% を下回っていることを表す.

	coefficient	std error	t stat	p value
Const.	-0.001	0.003	-0.309	0.757
L1.BTC	0.028	0.033	0.869	0.385
L1.KLD	0.061	0.059	1.042	0.297
L2.BTC	-0.089	0.033	-2.714	0.007**
L2.KLD	0.005	0.060	0.080	0.937
L3.BTC	-0.000	0.033	-0.002	0.998
L3.KLD	-0.003	0.058	-0.060	0.952
L4.BTC	-0.153	0.033	-4.660	0.000***
L4.KLD	-0.128	0.058	-2.191	0.028*
L5.BTC	-0.011	0.033	-0.332	0.740
L5.KLD	-0.078	0.059	-1.330	0.184
L6.BTC	0.060	0.033	1.801	0.072†
L6.KLD	0.025	0.059	0.431	0.666
L7.BTC	0.024	0.033	0.717	0.473
L7.KLD	0.050	0.059	0.837	0.403
L8.BTC	0.018	0.033	0.529	0.597
L8.KLD	0.010	0.059	0.162	0.872
L9.BTC	0.009	0.033	0.258	0.797
L9.KLD	0.031	0.059	0.517	0.605
L10.BTC	-0.016	0.033	-0.475	0.635
L10.KLD	0.086	0.059	1.447	0.148
L11.BTC	0.017	0.033	0.498	0.618
L11.KLD	-0.013	0.059	-0.225	0.822
L12.BTC	0.021	0.033	0.621	0.534
L12.KLD	-0.001	0.059	-0.013	0.989
L13.BTC	-0.031	0.033	-0.923	0.356
L13.KLD	0.128	0.059	2.158	0.031*

### 3.3.9 グレンジャー因果性

次に株式市場の場合と同じく、各ネットワーク指標からビットコイン価格へのグレンジャー因果性を検証する。その方法については3.2.7節と全く同じである。

表 3.4 は検定結果である。株式市場の場合と比較して、全体的に  $p$  値は低く出る傾向があった。その中でもやはり、株式市場の場合と同じくカルバックライブラー情報量 (BTC) からビットコイン (BTC) へのグレンジャー因果性が有意水準 5% で有意に検出された。そのためやはり、暗号通貨市場でも移動エントロピーネットワークがビットコイン価格の変動、すなわち相場変動に関して何らかの情報を持っていて、その情報をカルバックライブラー情報量が分布の変化という形で捉えているのだと思ってよさそうである。また因果の向きを考えるに、相場変動が起こることに先行して移動エントロピーネットワークには価格などのマクロ指標には表れない何らかの目に見えない変化が起こっていて、それに追随するようにして相場変動が起こっているのだと解釈できる。そのためカルバックライブラー情報量をはじめとしたネットワーク指標は将来の相場変動について何らかの重要な情報を含んでいて、価格予想をしたいときには重要な変数になり得ることが示唆される。

ここまで見てきたものは株式市場で得られた結果をおよそ追認するものである。そのため結果の再現性を補強するという意味で価値のある議論ができたと言うべきであろう。次節ではインパルス応答を検証し、具体的にどのようなメカニズムでこのような因果性が得られているのかを検証する。

表 3.4 各変数名 BTC, Mean, Var, Skew, Kur, KLD はそれぞれビットコイン価格変化率, 移動エントロピー分布の平均, 分散, 歪度, 尖度およびカルバックライブラー情報量を表す. また矢印の向きは因果の向きを表し, 例えば BTC→Mean はビットコイン価格変化率から移動エントロピー分布の平均へのグレンジャー因果性を表す.

	<i>p</i> value
BTC→Mean	0.526
BTC→Var	0.051
BTC→Skew	0.062
BTC→Kur	0.604
BTC→KLD	0.649
Mean→BTC	0.309
Var→BTC	0.284
Skew→BTC	0.265
Kur→BTC	0.301
KLD→BTC	0.015*

### 3.3.10 インパルス応答

株式市場の場合と同じく, 移動エントロピー分布のどのような変化が起こった結果ビットコイン価格変化率にどのような変化をもたらすのかをインパルス応答によって検証する.

時刻  $t$  における移動エントロピー分布の平均 (Mean), 分散 (Var), 歪度 (Skew), およびカルバックライブラー情報量  $KLD_t$  からビットコイン価格変化率 (BTC) に対する 250 時間 (約 33 営業日) にわたるインパルス応答を計算したグラフが図 3.9 である. 見ると, まずわかることは株式市場の場合と比較して全体的にインパルス応答の収束が速いということである. これは一般的に暗号通貨市場は株式市場と比較して相場の変動が激しい高いボラティリティを有しているため, ショックも比較的短時間で吸収されるからではないかと思われる. また, カルバックライブラー情報量からビットコイン価格変化率へのインパルス応答, 図 3.9a を見るとこれはちょうど株式市場の場合における対応物, 図 3.5a を反対にした様な形をしている. また信頼区間はより明確に正の領域に寄っている. 数値的に解釈するとカルバックライブラー情報量にモデルでは

予測できない 1% の変動が起こるとおよそ 40 時間後をピークとしてビットコイン価格変化率をモデルの予測を超えて 0.06% 程度同じ方向に変化させるということがわかる。すなわち、移動エントロピー分布に何らかの変化が起こりその結果カルバックライブラー情報量が増大すると約 40 時間後に、株式市場の場合とは反対に、ビットコイン価格は上昇するということである。これは、暗号通貨市場と株式市場との違い、すなわち暗号通貨市場では株式市場のような買いは Long (長期)、売りは Short (短期) というような概念が成立しにくく、暴落も暴騰も頻繁に起こっているという事実に起因しているのではないかと思われる。いずれにせよ、カルバックライブラー情報量の変動がおおよそ 40 時間後のビットコイン価格を有意に変動させているというのは意味のある結論である。また、図 3.9c を見ると、移動エントロピー分布の分散が (増加/減少) するとおよそ 40 時間後のビットコイン価格が (上昇/下落) することがわかる。こちらは株式市場の場合と同様の結果である。一方で図 3.9b, d は、信頼区間が正負両方の領域に大きく広がっており、全体的に値も小さいためあまり積極的に解釈をするべきではないように思われる。

以上のことをまとめると、移動エントロピー分布が変動し従ってカルバックライブラー情報量が増大すると、変動のあった最初の 40-50 時間程度は明確にビットコイン価格を上昇させ、またそのような分布の変化とは分散を増大させるような変化であろうこともわかる。ただしそれ以外の平均、歪度などについては明確な傾向は得られないであろうこともわかる。結局、株式市場の場合と同じくネットワーク指標の中でビットコイン価格の変動、すなわち相場変動に対して何らかの情報を持っているであろう最も有力な変数はカルバックライブラー情報量であろうということが本節の結論である。

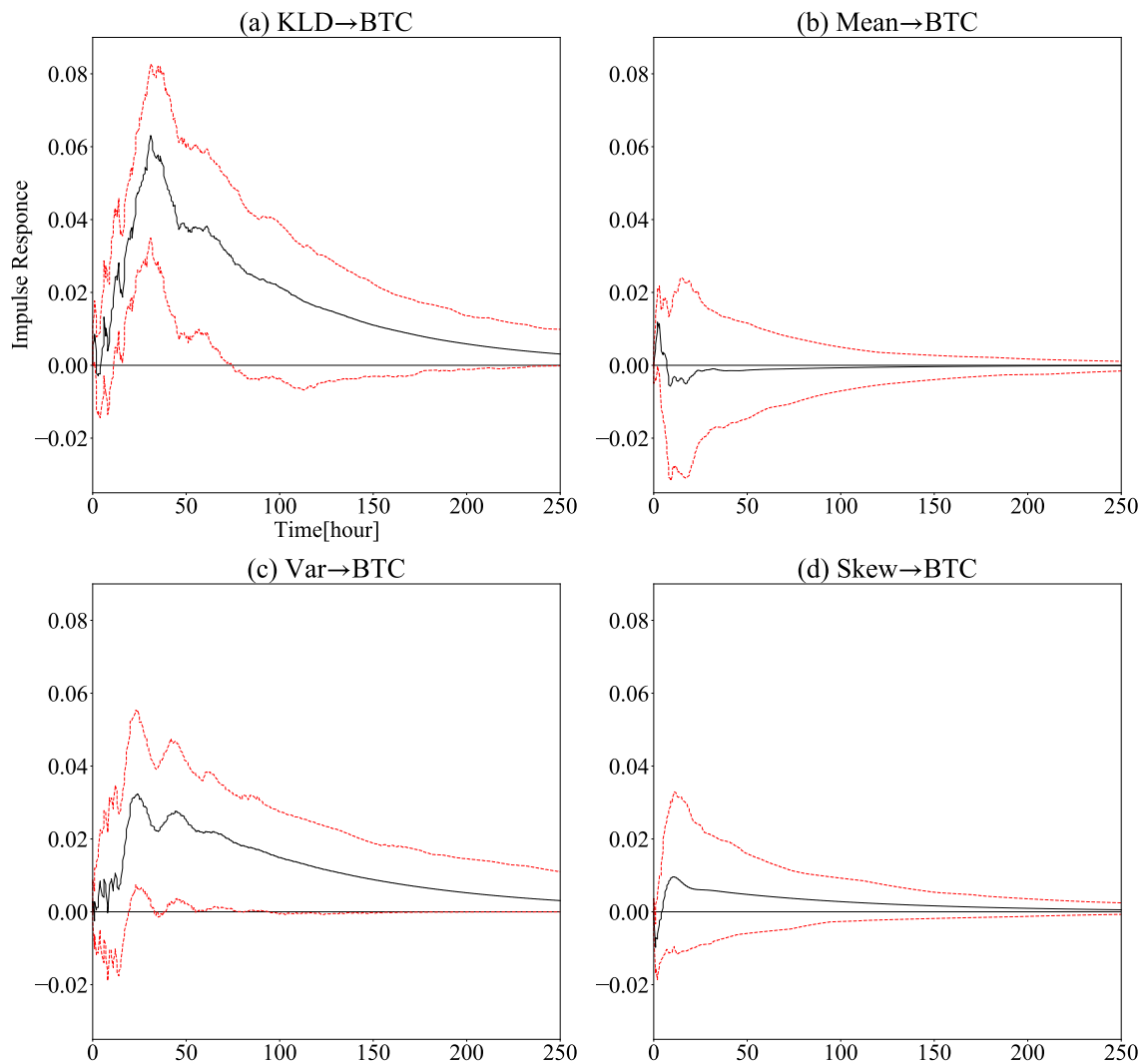


図 3.9 (a), (b), (c), (d) はそれぞれカルバックライブラー情報量  $KLD_t$  から  $QQQ$  へのインパルス応答, および時刻  $t$  における移動エントロピー分布の平均 (Mean), 分散 (Var), 歪度 (Skew) からビットコイン価格変化率 BTC へのインパルス応答を表す. 縦軸はインパルス応答の大きさ, 横軸はショックを与えた瞬間から経過した 1 時間ごとの時間を表す. また, 赤い破線は 90% 信頼区間の境界を表す.

### 3.3.11 本章のまとめ

本章でははじめに株式市場の移動エントロピー分布の変化とコロナショックのような大規模な相場変動の間に何らかの関係があるらしいことをコルモゴロフ-スミルノフ検定を用いて検証した.

次に、そのことについて深く調べるために移動エントロピー分布の変化をカルバックライブラー情報量によって定量化し、VAR モデルを用いて詳細な分析を行った。得られたことは、カルバックライブラー情報量から QQQ 価格の変化率に 5% 有意でグレンジャー因果性が見られたこと、また移動エントロピー分布が変化し、従ってカルバックライブラー情報量が増大すると株価は下落するという、そしてそのような分布の変化とは分布の平均や分散、歪度などが減少し、従って分布の特徴が失われるような方向への変化であるということ、そのような状態はネットワークが特別な場所がない均質な状態にあって各銘柄の区別が失われてしまったような状態であると解釈できることである。

その次に、VAR モデルによる同様の分析を暗号通貨市場に対しても行った。結果はおおよそ株式市場の場合と整合的であったがカルバックライブラー情報量からビットコイン価格変化率へのインパルス応答に関しては、株式市場とは逆の結果が得られた。すなわち、暗号通貨市場においては移動エントロピー分布が変化しカルバックライブラー情報量が増大するとビットコイン価格は上昇するということである。このような違いは株式市場と暗号通貨市場の性質の違いに帰することが出来るだろうと思われる。すなわち、暗号通貨市場は株式市場に比べて相場の変動が激しく、株式市場とは違い、短期的な急騰も急落もよく起こっているという要因がもたらす差異ではないかと考えられる。

以上の分析を踏まえて、移動エントロピー分布から計算される各種ネットワーク指標、特にカルバックライブラー情報量などは、株式市場や暗号通貨市場の相場変動に関する重要な情報を持っていると結論してよいであろう。

するとこれらの情報が実際に将来の価格変動予測に役立つのかどうかという自然な疑問がわく。そこで次章では株価予測や暗号通貨価格予測に関する検証を行うこととする。

## 4 予測

### 4.1 目的

本章ではこれまで見てきたカルバックライブラー情報量などのネットワーク指標が株価予測や暗号通貨価格予測に有効であるかどうかを検証するために、ネットワーク指標を用いたモデルと用いないモデルとで予測の精度を比較する検証を行う。予測精度の評価には平均二乗偏差 (Root Mean Squared Error, RMSE) を用いる。ここで  $n$  を予測期間の長さとして、時系列データ  $y_t$  の予測期間  $t = t_0, t_0 + 1, \dots, t_0 + n - 1$  における RMSE は次で定義される。

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=t_0}^{t_0+n-1} (y_t - \hat{y}_t)^2} \quad (4.1)$$

ただし  $\hat{y}_{t_i}$  は観測されたデータ  $y_{t_i}$  の予測値を表す。

予測する対象は式 3.6 で定義される QQQ 価格またはビットコイン価格の変化率  $\text{Change}_t$  とし、予測には複数の変数を組み合わせた多変量の VAR モデルを用いる。

### 4.2 株価予測

株価予測に使うモデルは 5 変量の VAR モデルとする。まず、ネットワーク指標を用いず、株価データのみを変数とするモデルを株価モデルとし、使用する変数は QQQ 価格の変化率  $\text{Change}_t$  と図 3.4 において株価データの中で AIC スコアの良かった上位 4 つ (AAPL, GOOGL, SNPS, MCHP) の  $\text{Change}_t$  同様に計算した変化率とする。次に、株価データとネットワーク指標を組み合わせたモデルをネットワークモデルとし、使用する変数は QQQ 価格の変化率  $\text{Change}_t$  と株価データの中で AIC スコアの良かった上位 2 つ (AAPL, GOOGL) と時刻  $t$  における移動エントロピー分布のカルバックライブラー情報量  $\text{KLD}_t$  および時刻  $t$  における移動エントロピー分布の平均とする。これら 2 つの予測モデルによって QQQ 価格変化率  $\text{Change}_t$  の予測を行い、予測精度の違いを比較する。

それぞれの予測モデルについて与えられたデータの時刻  $t = 0, 1, \dots, t_0 - 1$  までを訓練データとし、訓練データからみて 105 ステップ (7 営業日) 分の未来のデータとな

る  $t = t_0, t_0 + 1, \dots, t_0 + 104$  までをテストデータとして、訓練データに対してモデルの推定を行い、テストデータに対する当てはまりの良さを式 4.1 による RMSE によって比較する。

具体的な例として、図 4.1a および b はそれぞれ 2020 年 4 月 1 日から 105 ステップ (7 営業日) の予測結果を表示したものである。黒い実線および破線 (Observed (train), Observed (test)) はそれぞれ観測されたデータのうち訓練データとして使用された部分とテストデータとして使用された部分を表し、赤い破線 (Forecast) は予測を表す。株価モデルにおける RMSE は 0.126 で、ネットワークモデルにおける RMSE は 0.056 であった。従ってこの予測期間においてはネットワークモデルの方が予測精度が高く、数値的には株価モデルが QQQ 価格変化率を約 13%(ポイント) の誤差で予測しているのに対し、ネットワークモデルでは約 6%(ポイント) の誤差で、従っておよそ倍の精度で予測していることがわかる。また、図 4.1a を見て明らかのように、株価モデルでは将来の相場変動を予測できていないが、図 4.1b を見る限りでは、ネットワークモデルではそれができているように見える。このように、特に大きな相場変動の近辺においてネットワークモデルの方が株価モデルよりも予測精度が良い傾向があった。その他の予測期間における予測についてはいくつかの例を付録にて示す。

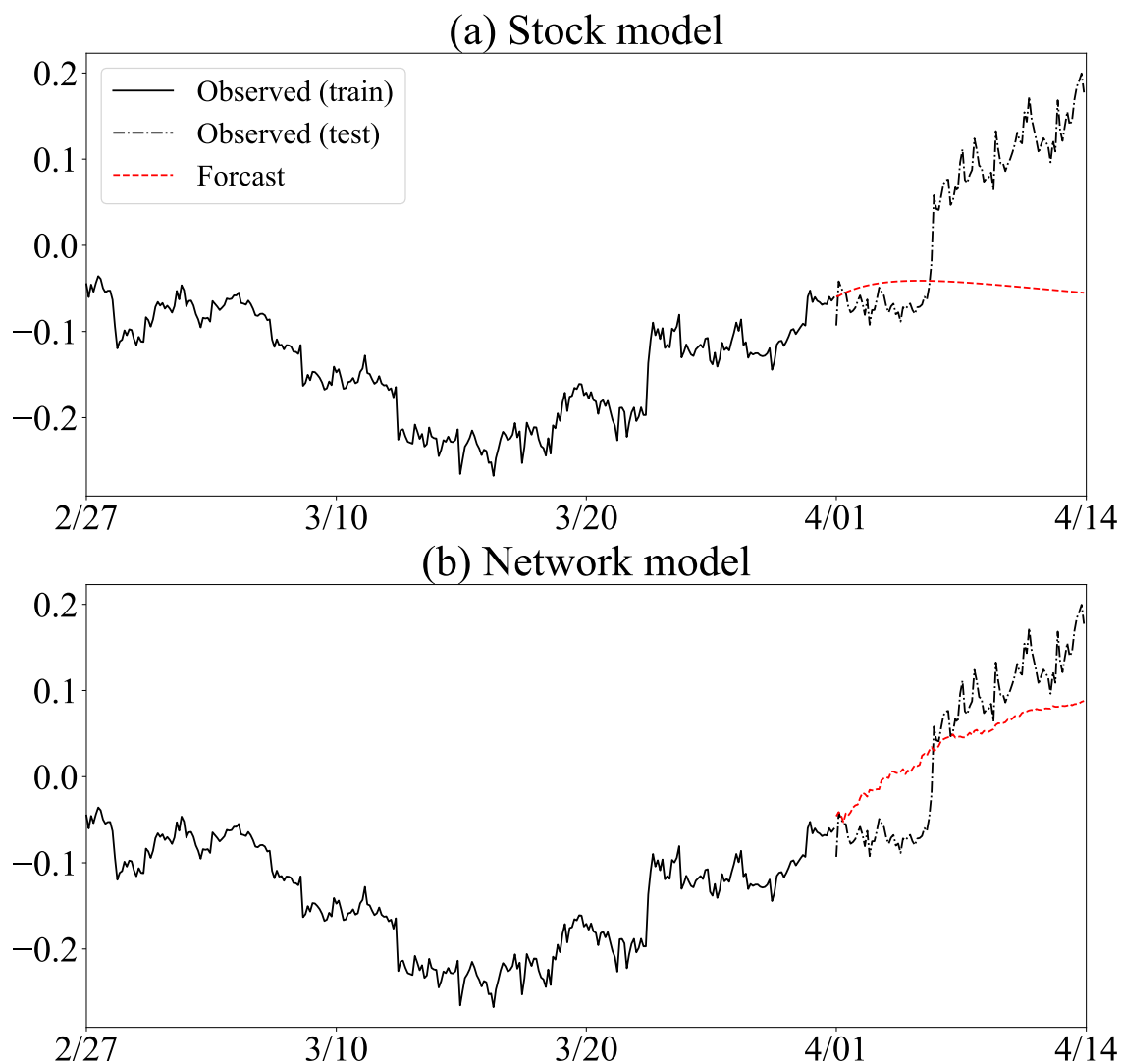


図 4.1 Observed (train), および Observed (test) はそれぞれ観測されたデータのうちの訓練データとテストデータを表し, Forecast は予測を表す.(a) 株価モデルでは将来の上昇相場への転換を予測できていないのに対し (b) ネットワークモデルではそれができているように見える.

また, 図 4.2 は予測の開始時点  $t_0$ , (Start Point of Forecast) を 2020 年 4 月 1 日から 2020 年 5 月 11 日までの各時刻にとったときの株価モデル (Stock model) とネットワークモデル (Network model) の予測における RMSE を表示したものである. 図 4.2 を見てわかるように, 全体的にネットワークモデルの方が株価モデルよりも RMSE が小さく, 従って予測精度が良い傾向にあった. 具体的には図 4.2 の全体の 68% でネット

ワークモデルの方が株価モデルよりも RMSE が小さかった。

以上のことから、ネットワーク指標が株価予測には有効であり、従って株価情報をネットワークに変換し、そこから有益な情報を得ようとする一連の方法には意味があったということが結論できるだろう。

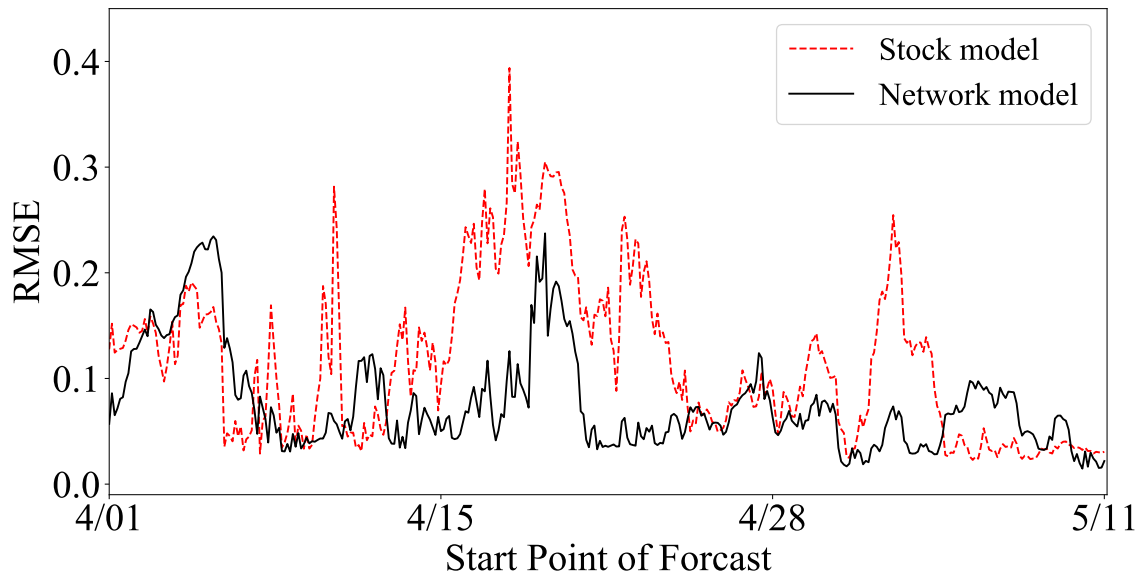


図 4.2 横軸は各予測の開始時点  $t_0$  を表し、縦軸は各予測開始時点  $t_0$  から 105 ステップ (7 営業日) 分の予測における RMSE を表す。

### 4.3 暗号通貨価格予測

株価予測と同様の方法で暗号通貨価格の予測も行う。モデルには 5 変量の VAR モデルを用い、式 3.16 で定義されるビットコイン価格の変化率  $\text{Change}_t$  を予測する。まず、ネットワーク指標を用いず暗号通貨価格データのみを変数とするモデルを暗号通貨モデルとし、使用する変数はビットコイン価格変化率  $\text{Change}_t$  と図 3.7 において暗号通貨価格データの中で AIC スコアの良かった上位 4 つの暗号通貨 (ETH, LTC, XRP, DASH) の価格変化率をビットコイン価格変化率  $\text{Change}_t$  同様に計算したものとする。次に、暗号通貨価格データとネットワーク指標を組み合わせたモデルをネットワークモデルとし、使用する変数はビットコイン価格変化率  $\text{Change}_t$  と暗号通貨価格データの中で AIC スコアの良かった 2 つの暗号通貨 (ETH, LTC) と時刻  $t$  における移動エ

ントロピー分布のカルバックライブラー情報量  $KLD_t$  および時刻  $t$  における移動エントロピー分布の分散とする。5 変量の VAR モデルで、使用する変数がビットコイン価格変化率  $Change_t$  および 4 つの暗号通貨指標であるものを暗号通貨モデルとし、使用する変数がビットコイン価格変化率  $Change_t$  , 暗号通貨指標の中で特に AIC スコアの良かった上位 2 つ (ETH, LTC) と 2 つのネットワーク指標であるものをネットワークモデルとし、暗号通貨モデルとネットワークモデルの予測精度を RMSE によって評価する。また、予測期間の長さは 96 ステップ (4 営業日) とする。それ以外の部分は前節の株価予測と同様である。

具体的な例として図 4.3a および b はそれぞれ、暗号通貨モデルとネットワークモデルにおける 2021 年 5 月 23 日から 96 ステップ (4 営業日) 分の予測結果を示したものである。図 4.3a, b どちらにおいても将来の上昇トレンドを大まかにとらえることは出来ているが、図 4.3b のネットワークモデルの方がより良い精度で予測できていることがわかる。暗号通貨モデルにおける RMSE は 0.102 でネットワークモデルにおける RMSE は 0.071 であった。従ってこの予測期間ではネットワークモデルの方が予測精度が高く、RMSE の減少率は約 30% であった。その他の予測期間における予測についてはいくつかの例を付録にて示す。

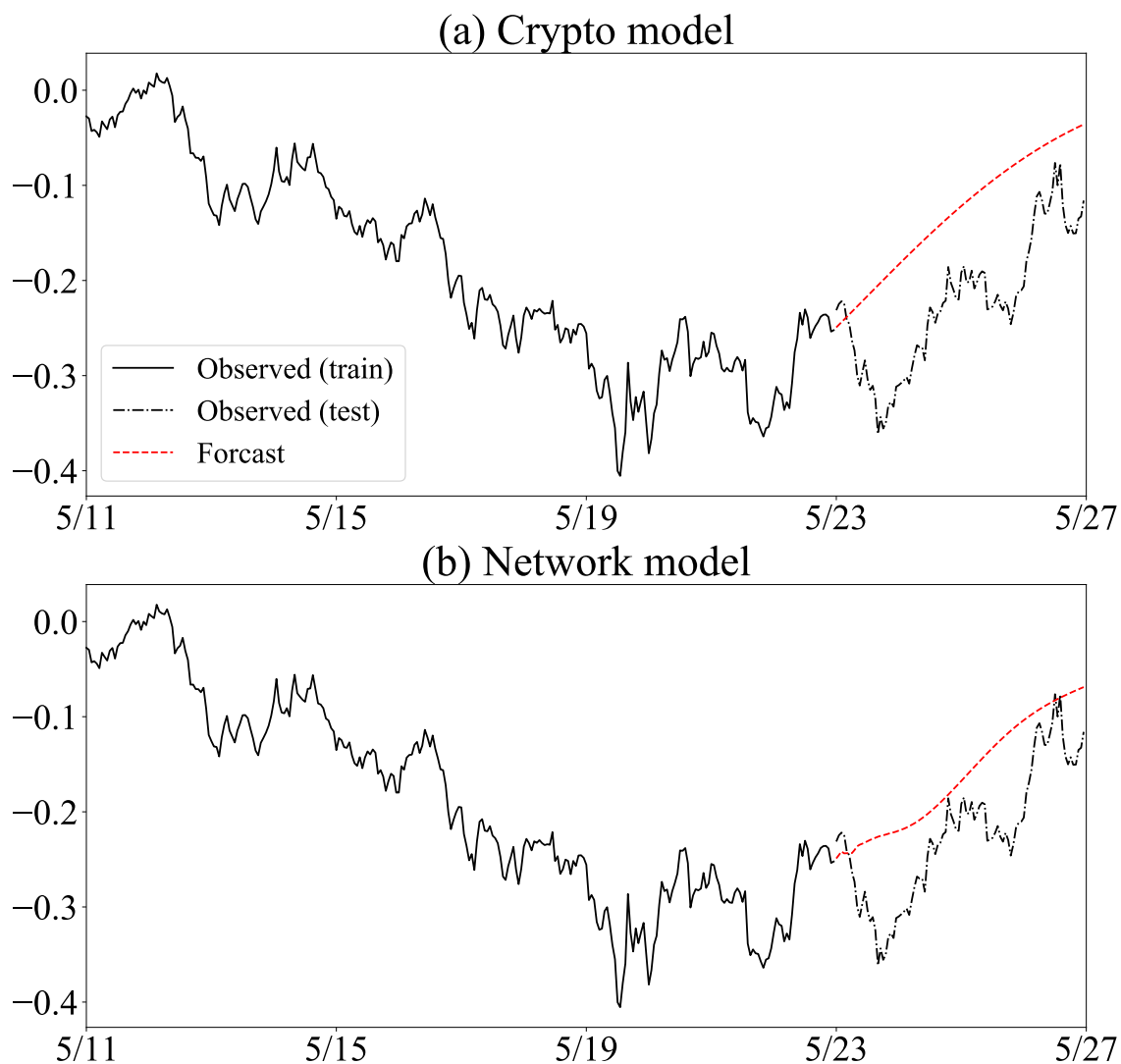


図 4.3 Observed (train), および Observed (test) はそれぞれ観測されたデータのうちの訓練データとテストデータを表し, Forecast は予測を表す. 両者とも大まかには正しい予測をしているように見えるが (b) のネットワークモデルの方が (a) の暗号通貨モデルよりも予測精度を高めていることがわかる.

また, 図 4.4 は予測の開始時点  $t_0$ , (Start Point of Forecast) を 2021 年 5 月 3 日から 2021 年 5 月 26 日までの各時刻にとった時の暗号通貨モデル (Crypto model) とネットワークモデル (Network model) の予測における RMSE を表示したものである. 暗号通貨価格予測の場合でも株価予測の場合と同じくネットワークモデルの方が暗号通貨モデルよりも RMSE が小さい傾向があった. 具体的には図 4.4 の全体の 65% でネット

ワークモデルの方が暗号通貨モデルよりも RMSE が小さかった。

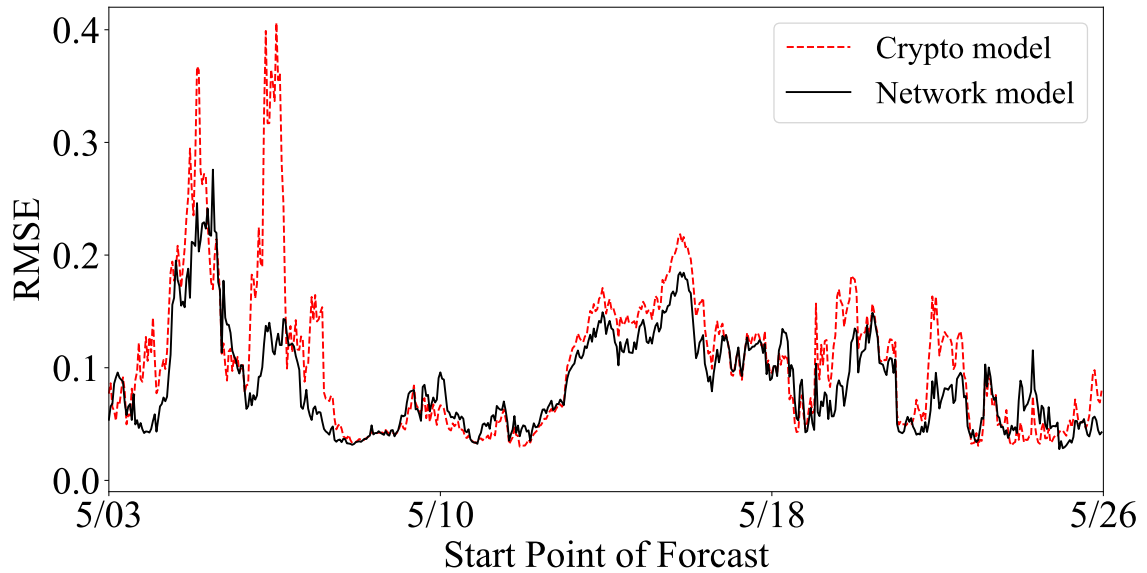


図 4.4 横軸は各予測の開始時点  $t_0$  を表し、縦軸は各予測開始時点  $t_0$  から 96 ステップ (4 営業日) 分の予測における RMSE を表す。

以上のことから暗号通貨市場においてもネットワーク指標を用いた予測は有効であり、2 つの異なる市場において同様の結果が得られたことは与えられた時系列データをネットワークに変換して、そこから有用な情報を得ようとする一連の方法に普遍的な効果があることを期待させる結果になったと言えるだろう。

## 5 結論

### 5.1 本稿のまとめ

本稿ではまず、株式市場を移動エントロピーのネットワークとして捉えたとき、コロナショックのような大規模なショック時には市場全体の移動エントロピー分布が平常時とは異なっていることをコルモゴロフ-スミルノフ検定によって見た。コロナショック時に起こっていた移動エントロピー分布の変化とは、分布の平均や分散、歪度などがともに減少する、分布の特徴が失われるような変化であった。そのような場合、ネットワークには特別な場所が存在しない均質な状態に近くなっていると考えられ、従って各銘柄はその区別を失い、似たようなふるまいをするようになるという現象がショック時に起こっていると考えられるのであった。ショック時に移動エントロピー分布の平均が減少するというのはネットワークのリンクの重みを移動エントロピーという因果関係の量で定義したからこそ起こる現象である。例えばリンクの重みを二つの銘柄の価格の一定時間ごとの相関係数や相互情報量などで定義するなら、ショック時には各銘柄は一様に下落するので見かけの繋がりの深さは高まり、従って相関係数や相互情報量は上昇するはずである。実際、(Kim and Sayama 2017) は相場に大きな変動があったとき、ネットワークのリンクの重みを相互情報量で定義した場合の相互情報量の分布は右方にシフトすることを報告している。しかしながら、因果関係まで含めて考えたとき、そのような繋がりの深さはむしろ失われるというのが本稿の発見である。その理由は、特にコロナショックのような相場の急落時には株価は市場内のメカニズムでは説明できない外的要因によって下落するため、市場内の因果関係が消失してしまうからではないかと考えられる。

次に、本稿では VAR モデルを用いてそのような移動エントロピー分布の変化が相場変動をどのように説明し得るかということを経済市場と暗号通貨市場を対象に検証した。その中で移動エントロピー分布の変化を量的に評価するためにカルバックライブラー情報量を用いた。顕著な結果はカルバックライブラー情報量から QQQ 価格変化率やビットコイン価格変化率の両方にグレンジャーの意味での因果性が見られたことである。このことは移動エントロピー分布の変化が株価や暗号通貨価格の変動に先立って起こっている可能性を示唆するため意味のある結果である。

最後に、本稿ではカルバックライブラー情報量などのネットワーク指標を用いたモデルとそうでないモデルとで株価予測として QQQ 価格変化率を、暗号通貨価格予測としてビットコイン価格変化率を予測し、その精度を比較した。結果は複数の異なる予測期間を試した中で株価予測、暗号通貨予測ともに全体のおよそ 65% でネットワーク指標を用いたモデルの方が予測精度が高かった。従って、カルバックライブラー情報量などのネットワーク指標は株式市場や暗号通貨市場の情報を上手く抽出し、質の高い情報を得られていると考えることができる。

全体を通じて、本稿の意義は大量の変数からなる多変量の時系列データをまずネットワークに変換し、そこからカルバックライブラー情報量などの意味のある情報を抽出できるという、時系列データの情報圧縮に関する方法を考察した点にあるだろう。一般的に言って、どのようなモデルでもモデルに入れる変数の数が多くなれば推定すべきパラメーターの数は飛躍的に多くなってしまう。例えば本稿で用いた VAR モデルの場合、 $n$  変量の VAR モデルではパラメーターの数は  $O(n^2)$  に比例し、株式市場の場合は  $n = 103$ 、暗号通貨市場の場合は  $n = 30$  など、大きな  $n$  に対しては巨大な計算コストがかかってしまう。しかも、このような大量の変数をそのまま載せただけのモデルは、往々にして係数の推定は不安定になり易く、おそらく良い分析は出来ないであろう。そのような情報圧縮の問題に対峙する一つの方法として、主成分分析や畳み込みニューラルネットワークなど、機械学習を利用した特徴量抽出などは考え得る (Yu, Chen, and Zhang 2014; Tsantekidis et al. 2017)。しかしながら、そのような機械学習的な方法はモデルの解釈性が著しく低いことが常であるため予測には用いることができても計量的な分析には用いられない。それに対して本稿で使用したカルバックライブラー情報量などの特徴量はネットワークという明確な意味が与えられるため、モデルの解釈は素直にできる。このように、多変量の時系列データをモデルの解釈性を損なわずに適切に情報圧縮することができるということが本稿で考察した手法の最たる意義であろう。

## 5.2 本稿の課題

本稿では株式市場や暗号通貨市場を移動エントロピーのネットワークとして捉え、そのネットワークから意味のある情報を抜き出そうということがテーマであった。し

かしながら、ネットワークとは言ったものの実際に使われたものは移動エントロピーの分布のみであり、ネットワークの構造についてより高度で踏み込んだ分析はなされていない。例えばネットワークの中心性や次数相関、クラスター係数などネットワークの構造についてより具体的な示唆を与えてくれる量を計算すれば、より豊かな分析が可能となるかもしれないし、予測の精度も上がるかもしれない。この点は十分に検討する価値のある課題である。

また、本稿で使用したモデルは VAR モデルのみであったが他のモデルでも検討してみて同様の結果が得られるのかどうかは試してみる価値のある課題である。例えば予測についてのみ興味があるのであれば、モデルの解釈性は犠牲になっても構わないと考えられるので再帰型ニューラルネットワーク (Recurrent neural network) など機械学習的なモデルを考えるのも一つである。

さらに言えば、本稿では扱う対象が株式市場や暗号通貨市場という完全に金融寄りのものであったが、必ずしも金融データにこだわる必要はないので、より一般の経済データに対象を拡大してみることも一考の価値がある課題である。例えば世界各国の期ごとの GDP 成長率や消費者物価指数などをデータとして同様の分析を行えばコロナ危機やリーマンショック時などに何らかの特有のパターンが表れるかもしれないし、そのようなことは十分検討しがいがある。

## 参考文献

- Abarbanell, Jeffrey S, and Brian J Bushee. 1997. “Fundamental analysis, future earnings, and stock prices.” *Journal of Accounting Research* 35 (1): 1–24.
- Abou Jaoude, Joe, and Raafat George Saade. 2019. “Blockchain Applications – Usage in Different Domains.” *IEEE Access* 7:45360–45381.
- Ahern, Kenneth R. 2013. “Network centrality and the cross section of stock returns.” *Available at SSRN 2197370*.
- Akita, Ryo, Akira Yoshihara, Takashi Matsubara, and Kuniaki Uehara. 2016. “Deep learning for stock prediction using numerical and textual information.” In *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, 1–6. IEEE.
- Boar, Codruta, Henry Holden, and Amber Wadsworth. 2020. “Impending arrival—a sequel to the survey on central bank digital currency.” *BIS paper*, no. 107.
- Chong, Eunsuk, Chulwoo Han, and Frank C Park. 2017. “Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies.” *Expert Systems with Applications* 83:187–205.
- Dechow, Patricia M, Amy P Hutton, Lisa Meulbroek, and Richard G Sloan. 2001. “Short-sellers, fundamental analysis, and stock returns.” *Journal of financial Economics* 61 (1): 77–106.
- Delgado-Segura, Sergi, Cristina Pérez-Solà, Jordi Herrera-Joancomartí, Guillermo Navarro-Arribas, and Joan Borrell. 2018. “Cryptocurrency Networks: A New P2P Paradigm.” *Mobile Information Systems* 2018.
- Ding, Xiao, Yue Zhang, Ting Liu, and Junwen Duan. 2015. “Deep learning for event-driven stock prediction.” In *Twenty-fourth international joint conference on artificial intelligence*.

- Fama, Eugene. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance* 25 (2): 383–417.
- Heiberger, Raphael H. 2014. "Stock network stability in times of crisis." *Physica A: Statistical Mechanics and its Applications* 393:376–381.
- Hileman, Garrick, and Michel Rauchs. 2017. "Global Cryptocurrency Benchmarking Study." *Cambridge Centre for Alternative Finance* 33:33–113.
- Ikegwu, Kelechi M, Jacob Trauger, Jeff McMullin, and Robert J Brunner. 2020. "PyIF: A Fast and Light Weight Implementation to Estimate Bivariate Transfer Entropy for Big Data." In *2020 SoutheastCon*, 1–6. IEEE.
- Kim, Minjun, and Hiroki Sayama. 2017. "Predicting stock market movements using network science: an information theoretic approach." *Applied Network Science* 2 (1): 1–14.
- Lee, Gregory, Ralf Gommers, Filip Waselewski, Kai Wohlfahrt, and Aaron O’Leary. 2019. "PyWavelets: A Python package for wavelet analysis." *Journal of Open Source Software* 4 (36): 1237.
- Mancini-Griffoli, Tommaso, Maria Soledad Martinez Peria, Itai Agur, Anil Ari, John Kiff, Adina Popescu, and Celine Rochon. 2018. "Casting light on central bank digital currency." *IMF staff discussion note* 8.
- Miraz, Mahdi H, and Maaruf Ali. 2018. "Applications of Blockchain Technology beyond Cryptocurrency." *arXiv preprint arXiv:1801.03528*.
- Monrat, Ahmed Afif, Olov Schelén, and Karl Andersson. 2019. "A Survey of Blockchain From the Perspectives of Applications, Challenges, and Opportunities." *IEEE Access* 7:117134–117151.
- Nakamoto, Satoshi. 2008. "Bitcoin: A Peer-to-Peer Electronic Cash System." *Decentralized Business Review*, 21260.
- Namaki, Ali, Amir H Shirazi, R Raei, and GR Jafari. 2011. "Network analysis of a financial market based on genuine correlation and threshold method." *Physica A: Statistical Mechanics and its Applications* 390 (21-22): 3835–3841.

- Náñez Alonso, Sergio Luis, Javier Jorge-Vazquez, and Ricardo Francisco Reier Forradellas. 2021. “Central Banks Digital Currency: Detection of Optimal Countries for the Implementation of a CBDC and the Implication for Payment Industry Open Innovation.” *Journal of Open Innovation: Technology, Market, and Complexity* 7 (1): 72.
- Tsantekidis, Avraam, Nikolaos Passalis, Anastasios Tefas, Juho Kannianen, Moncef Gabbouj, and Alexandros Iosifidis. 2017. “Forecasting stock prices from the limit order book using convolutional neural networks.” In *2017 IEEE 19th Conference on Business Informatics (CBI)*, 1:7–12. IEEE.
- Yaga, Dylan, Peter Mell, Nik Roby, and Karen Scarfone. 2019. “Blockchain Technology Overview.” *arXiv preprint arXiv:1906.11078*.
- Yu, Huanhuan, Rongda Chen, and Guoping Zhang. 2014. “A SVM stock selection model within PCA.” *Procedia Computer Science* 31:406–412.
- Zeilinger, Martin. 2018. “Digital Art as ‘Monetised Graphics’ : Enforcing Intellectual Property on the Blockchain.” *Philosophy & Technology* 31 (1): 15–41.
- Zhao, Longfeng, Gang-Jin Wang, Mingang Wang, Weiqi Bao, Wei Li, and H Eugene Stanley. 2018. “Stock market as temporal network.” *Physica A: Statistical Mechanics and its Applications* 506:1104–1112.

## 付録

### S1 回帰結果

表 3.1 および表 3.3 の完全な回帰結果をそれぞれ表 1.1 および表 1.2 に示す. 変数名について, 例えば L1.KLD などはラグ 1 の  $KLD_{t-1}$  の項を表す. その他の変数名についても同様である. また †, \*, \*\*, \*\*\* はそれぞれ  $p$  値が 10, 5, 1, 0.1% を下回っていることを表す.

付録表 1.1

	coefficient	std error	$t$ stat	$p$ value
Const.	0.000	0.003	0.085	0.933
L1.QQQ	-0.156	0.032	-4.833	0.000***
L1.KLD	0.103	0.132	0.781	0.435
L2.QQQ	0.009	0.033	0.260	0.795
L2.KLD	-0.140	0.146	-0.956	0.339
L3.QQQ	-0.066	0.033	-2.032	0.042*
L3.KLD	0.026	0.150	0.176	0.860
L4.QQQ	-0.006	0.033	-0.193	0.847
L4.KLD	-0.106	0.147	-0.721	0.471
L5.QQQ	0.007	0.033	0.219	0.827
L5.KLD	-0.026	0.148	-0.176	0.860
L6.QQQ	0.054	0.033	1.648	0.099†
L6.KLD	0.033	0.149	0.222	0.824
L7.QQQ	0.019	0.033	0.575	0.565
L7.KLD	0.257	0.148	1.736	0.082†
L8.QQQ	0.018	0.032	0.561	0.575
L8.KLD	0.003	0.149	0.017	0.986
L9.QQQ	0.059	0.032	1.813	0.070†
L9.KLD	-0.096	0.149	-0.645	0.519

次ページに続く

	coefficient	std error	t stat	p value
L10.QQQ	-0.073	0.032	-2.249	0.025*
L10.KLD	0.259	0.149	1.738	0.082 <sup>†</sup>
L11.QQQ	0.008	0.033	0.247	0.805
L11.KLD	-0.118	0.149	-0.792	0.428
L12.QQQ	0.212	0.033	6.490	0.000***
L12.KLD	-0.115	0.149	-0.771	0.441
L13.QQQ	-0.059	0.033	-1.761	0.078 <sup>†</sup>
L13.KLD	0.036	0.149	0.243	0.808
L14.QQQ	-0.019	0.033	-0.562	0.574
L14.KLD	-0.323	0.149	-2.169	0.030*
L15.QQQ	-0.106	0.033	-3.181	0.001**
L15.KLD	0.138	0.149	0.924	0.356
L16.QQQ	-0.015	0.033	-0.461	0.645
L16.KLD	0.107	0.149	0.719	0.472
L17.QQQ	-0.027	0.033	-0.809	0.418
L17.KLD	0.036	0.149	0.240	0.810
L18.QQQ	-0.043	0.033	-1.299	0.194
L18.KLD	-0.006	0.149	-0.039	0.969
L19.QQQ	0.012	0.033	0.374	0.709
L19.KLD	-0.237	0.149	-1.588	0.112
L20.QQQ	0.016	0.033	0.497	0.619
L20.KLD	0.332	0.149	2.222	0.026*
L21.QQQ	-0.000	0.033	-0.005	0.996
L21.KLD	-0.096	0.150	-0.640	0.522
L22.QQQ	-0.031	0.033	-0.949	0.342
L22.KLD	-0.144	0.150	-0.956	0.339
L23.QQQ	0.044	0.033	1.352	0.176
L23.KLD	-0.050	0.150	-0.331	0.741
L24.QQQ	0.031	0.033	0.955	0.340
L24.KLD	-0.073	0.149	-0.489	0.625

次ページに続く

	coefficient	std error	t stat	p value
L25.QQQ	-0.095	0.033	-2.899	0.004**
L25.KLD	0.258	0.149	1.725	0.085 <sup>†</sup>
L26.QQQ	0.050	0.033	1.502	0.133
L26.KLD	0.191	0.149	1.277	0.202
L27.QQQ	0.012	0.033	0.355	0.723
L27.KLD	-0.099	0.149	-0.661	0.509
L28.QQQ	-0.030	0.033	-0.912	0.362
L28.KLD	0.097	0.150	0.648	0.517
L29.QQQ	0.009	0.033	0.285	0.776
L29.KLD	-0.246	0.150	-1.637	0.102
L30.QQQ	0.042	0.033	1.260	0.208
L30.KLD	-0.087	0.150	-0.582	0.561
L31.QQQ	-0.029	0.033	-0.884	0.376
L31.KLD	-0.124	0.151	-0.822	0.411
L32.QQQ	-0.003	0.033	-0.084	0.933
L32.KLD	0.138	0.151	0.918	0.359
L33.QQQ	0.036	0.033	1.099	0.272
L33.KLD	0.147	0.151	0.977	0.329
L34.QQQ	-0.012	0.033	-0.358	0.720
L34.KLD	-0.086	0.151	-0.569	0.569
L35.QQQ	0.089	0.033	2.688	0.007**
L35.KLD	0.042	0.151	0.281	0.779
L36.QQQ	0.065	0.033	1.963	0.050 <sup>†</sup>
L36.KLD	-0.060	0.180	-0.336	0.737
L37.QQQ	0.009	0.033	0.260	0.795
L37.KLD	-0.353	0.184	-1.920	0.055 <sup>†</sup>
L38.QQQ	-0.015	0.033	-0.438	0.661
L38.KLD	0.395	0.185	2.141	0.032*
L39.QQQ	0.055	0.033	1.663	0.096 <sup>†</sup>
L39.KLD	0.161	0.171	0.941	0.346

次ページに続く

	coefficient	std error	t stat	p value
L40.QQQ	0.022	0.033	0.663	0.507
L40.KLD	-0.255	0.172	-1.483	0.138
L41.QQQ	-0.095	0.033	-2.871	0.004**
L41.KLD	0.186	0.172	1.077	0.281
L42.QQQ	-0.029	0.033	-0.872	0.383
L42.KLD	0.107	0.171	0.625	0.532
L43.QQQ	-0.014	0.033	-0.434	0.664
L43.KLD	-0.022	0.171	-0.130	0.896
L44.QQQ	-0.037	0.033	-1.120	0.263
L44.KLD	-0.144	0.172	-0.834	0.404
L45.QQQ	0.004	0.033	0.122	0.903
L45.KLD	0.188	0.172	1.089	0.276
L46.QQQ	-0.042	0.033	-1.253	0.210
L46.KLD	-0.024	0.172	-0.137	0.891
L47.QQQ	-0.146	0.033	-4.402	0.000***
L47.KLD	-0.058	0.172	-0.335	0.738
L48.QQQ	0.023	0.034	0.697	0.486
L48.KLD	-0.153	0.171	-0.899	0.369
L49.QQQ	0.048	0.034	1.429	0.153
L49.KLD	-0.060	0.170	-0.353	0.724
L50.QQQ	0.024	0.034	0.722	0.470
L50.KLD	0.224	0.170	1.316	0.188
L51.QQQ	0.020	0.033	0.608	0.543
L51.KLD	0.170	0.170	0.999	0.318
L52.QQQ	0.067	0.033	2.006	0.045*
L52.KLD	-0.213	0.170	-1.255	0.210
L53.QQQ	0.015	0.033	0.437	0.662
L53.KLD	-0.013	0.170	-0.078	0.938
L54.QQQ	-0.007	0.033	-0.195	0.845
L54.KLD	-0.006	0.170	-0.037	0.970

次ページに続く

	coefficient	std error	t stat	p value
L55.QQQ	0.064	0.033	1.897	0.058 <sup>†</sup>
L55.KLD	0.131	0.170	0.768	0.442
L56.QQQ	0.050	0.034	1.499	0.134
L56.KLD	-0.100	0.170	-0.588	0.557
L57.QQQ	0.069	0.034	2.053	0.040*
L57.KLD	-0.181	0.171	-1.064	0.288
L58.QQQ	0.039	0.034	1.165	0.244
L58.KLD	-0.058	0.171	-0.339	0.735
L59.QQQ	-0.004	0.034	-0.127	0.899
L59.KLD	-0.201	0.170	-1.180	0.238
L60.QQQ	0.029	0.034	0.862	0.389
L60.KLD	0.285	0.171	1.672	0.094 <sup>†</sup>
L61.QQQ	0.056	0.034	1.663	0.096 <sup>†</sup>
L61.KLD	0.244	0.171	1.425	0.154
L62.QQQ	0.007	0.033	0.216	0.829
L62.KLD	-0.183	0.171	-1.065	0.287
L63.QQQ	0.015	0.033	0.453	0.651
L63.KLD	0.183	0.172	1.065	0.287
L64.QQQ	0.071	0.033	2.135	0.033*
L64.KLD	-0.042	0.172	-0.246	0.806
L65.QQQ	-0.002	0.033	-0.060	0.952
L65.KLD	0.058	0.172	0.336	0.737
L66.QQQ	0.045	0.033	1.353	0.176
L66.KLD	-0.183	0.171	-1.074	0.283
L67.QQQ	-0.004	0.033	-0.114	0.909
L67.KLD	-0.055	0.171	-0.320	0.749
L68.QQQ	-0.046	0.033	-1.400	0.161
L68.KLD	0.050	0.171	0.289	0.773
L69.QQQ	-0.083	0.033	-2.498	0.012*
L69.KLD	-0.009	0.172	-0.051	0.960

次ページに続く

	coefficient	std error	t stat	p value
L70.QQQ	0.071	0.033	2.126	0.034*
L70.KLD	-0.336	0.171	-1.961	0.050 <sup>†</sup>
L71.QQQ	0.016	0.033	0.483	0.629
L71.KLD	0.103	0.184	0.558	0.577
L72.QQQ	-0.077	0.033	-2.292	0.022*
L72.KLD	-0.129	0.183	-0.701	0.483
L73.QQQ	-0.039	0.033	-1.156	0.248
L73.KLD	0.328	0.180	1.822	0.069 <sup>†</sup>
L74.QQQ	0.040	0.033	1.195	0.232
L74.KLD	0.090	0.150	0.599	0.549
L75.QQQ	0.029	0.033	0.876	0.381
L75.KLD	0.015	0.150	0.099	0.921
L76.QQQ	-0.016	0.033	-0.492	0.623
L76.KLD	0.051	0.150	0.341	0.733
L77.QQQ	0.032	0.033	0.949	0.343
L77.KLD	-0.126	0.150	-0.837	0.403
L78.QQQ	-0.040	0.033	-1.211	0.226
L78.KLD	0.109	0.150	0.727	0.467
L79.QQQ	-0.005	0.033	-0.143	0.886
L79.KLD	-0.209	0.150	-1.391	0.164
L80.QQQ	-0.016	0.033	-0.490	0.624
L80.KLD	-0.050	0.151	-0.335	0.738
L81.QQQ	-0.008	0.033	-0.227	0.820
L81.KLD	-0.060	0.150	-0.401	0.689
L82.QQQ	-0.040	0.033	-1.210	0.226
L82.KLD	0.049	0.150	0.324	0.746
L83.QQQ	-0.039	0.033	-1.172	0.241
L83.KLD	-0.008	0.150	-0.054	0.957
L84.QQQ	0.036	0.033	1.072	0.284
L84.KLD	-0.056	0.150	-0.372	0.710

次ページに続く

	coefficient	std error	t stat	p value
L85.QQQ	0.031	0.033	0.922	0.357
L85.KLD	0.369	0.150	2.467	0.014*
L86.QQQ	0.067	0.033	2.016	0.044*
L86.KLD	0.222	0.150	1.475	0.140
L87.QQQ	0.029	0.033	0.857	0.392
L87.KLD	-0.248	0.151	-1.647	0.100
L88.QQQ	-0.009	0.034	-0.253	0.800
L88.KLD	-0.061	0.151	-0.406	0.685
L89.QQQ	-0.045	0.034	-1.344	0.179
L89.KLD	0.233	0.151	1.545	0.122
L90.QQQ	-0.030	0.034	-0.903	0.366
L90.KLD	0.040	0.151	0.264	0.792
L91.QQQ	-0.056	0.034	-1.658	0.097 <sup>†</sup>
L91.KLD	-0.180	0.151	-1.196	0.232
L92.QQQ	0.121	0.033	3.613	0.000***
L92.KLD	-0.531	0.151	-3.522	0.000***
L93.QQQ	-0.029	0.034	-0.859	0.391
L93.KLD	-0.032	0.152	-0.208	0.835
L94.QQQ	-0.057	0.033	-1.709	0.087 <sup>†</sup>
L94.KLD	0.159	0.152	1.051	0.293
L95.QQQ	0.082	0.034	2.431	0.015*
L95.KLD	-0.010	0.152	-0.063	0.950
L96.QQQ	-0.018	0.034	-0.524	0.600
L96.KLD	0.000	0.151	0.000	1.000
L97.QQQ	-0.014	0.033	-0.422	0.673
L97.KLD	0.159	0.151	1.048	0.294
L98.QQQ	-0.017	0.033	-0.520	0.603
L98.KLD	0.203	0.151	1.345	0.179
L99.QQQ	0.003	0.033	0.083	0.934
L99.KLD	0.025	0.151	0.163	0.871

次ページに続く

	coefficient	std error	<i>t</i> stat	<i>p</i> value
L100.QQQ	0.044	0.033	1.357	0.175
L100.KLD	0.077	0.151	0.512	0.609
L101.QQQ	0.022	0.033	0.668	0.504
L101.KLD	-0.169	0.150	-1.125	0.261
L102.QQQ	0.040	0.033	1.233	0.217
L102.KLD	-0.099	0.150	-0.660	0.509
L103.QQQ	-0.023	0.033	-0.693	0.488
L103.KLD	-0.037	0.151	-0.248	0.804
L104.QQQ	-0.069	0.033	-2.097	0.036*
L104.KLD	-0.078	0.150	-0.522	0.602
L105.QQQ	0.025	0.033	0.769	0.442
L105.KLD	-0.098	0.149	-0.659	0.510
L106.QQQ	-0.114	0.033	-3.497	0.000***
L106.KLD	0.123	0.152	0.808	0.419
L107.QQQ	-0.023	0.033	-0.712	0.476
L107.KLD	-0.324	0.148	-2.186	0.029*
L108.QQQ	0.089	0.033	2.722	0.006**
L108.KLD	0.306	0.135	2.265	0.024*

付録表 1.2

	coefficient	std error	<i>t</i> stat	<i>p</i> value
Const.	-0.001	0.003	-0.309	0.757
L1.BTC	0.028	0.033	0.869	0.385
L1.KLD	0.061	0.059	1.042	0.297
L2.BTC	-0.089	0.033	-2.714	0.007**
L2.KLD	0.005	0.060	0.080	0.937
L3.BTC	-0.000	0.033	-0.002	0.998
L3.KLD	-0.003	0.058	-0.060	0.952
L4.BTC	-0.153	0.033	-4.660	0.000***
L4.KLD	-0.128	0.058	-2.191	0.028*
L5.BTC	-0.011	0.033	-0.332	0.740
L5.KLD	-0.078	0.059	-1.330	0.184
L6.BTC	0.060	0.033	1.801	0.072 <sup>†</sup>
L6.KLD	0.025	0.059	0.431	0.666
L7.BTC	0.024	0.033	0.717	0.473
L7.KLD	0.050	0.059	0.837	0.403
L8.BTC	0.018	0.033	0.529	0.597
L8.KLD	0.010	0.059	0.162	0.872
L9.BTC	0.009	0.033	0.258	0.797
L9.KLD	0.031	0.059	0.517	0.605
L10.BTC	-0.016	0.033	-0.475	0.635
L10.KLD	0.086	0.059	1.447	0.148
L11.BTC	0.017	0.033	0.498	0.618
L11.KLD	-0.013	0.059	-0.225	0.822
L12.BTC	0.021	0.033	0.621	0.534
L12.KLD	-0.001	0.059	-0.013	0.989
L13.BTC	-0.031	0.033	-0.923	0.356
L13.KLD	0.128	0.059	2.158	0.031*
L14.BTC	-0.023	0.033	-0.700	0.484

次ページに続く

	coefficient	std error	t stat	p value
L14.KLD	-0.156	0.059	-2.632	0.008**
L15.BTC	-0.100	0.033	-3.018	0.003**
L15.KLD	0.006	0.060	0.096	0.924
L16.BTC	-0.004	0.033	-0.124	0.902
L16.KLD	-0.078	0.060	-1.313	0.189
L17.BTC	-0.012	0.033	-0.353	0.724
L17.KLD	-0.044	0.060	-0.734	0.463
L18.BTC	-0.066	0.033	-1.972	0.049*
L18.KLD	0.003	0.060	0.047	0.963
L19.BTC	-0.017	0.033	-0.521	0.602
L19.KLD	-0.002	0.060	-0.026	0.979
L20.BTC	0.037	0.033	1.105	0.269
L20.KLD	0.094	0.060	1.575	0.115
L21.BTC	0.014	0.033	0.422	0.673
L21.KLD	-0.029	0.060	-0.492	0.623
L22.BTC	-0.040	0.033	-1.203	0.229
L22.KLD	0.054	0.060	0.907	0.364
L23.BTC	0.025	0.033	0.735	0.462
L23.KLD	0.042	0.060	0.702	0.483
L24.BTC	-0.101	0.033	-3.021	0.003**
L24.KLD	0.082	0.059	1.378	0.168
L25.BTC	-0.044	0.034	-1.311	0.190
L25.KLD	-0.062	0.060	-1.044	0.297
L26.BTC	-0.018	0.034	-0.528	0.597
L26.KLD	-0.028	0.060	-0.468	0.640
L27.BTC	-0.032	0.034	-0.934	0.350
L27.KLD	-0.091	0.060	-1.519	0.129
L28.BTC	-0.057	0.034	-1.694	0.09†
L28.KLD	0.138	0.059	2.325	0.020*
L29.BTC	0.038	0.034	1.117	0.264

次ページに続く

	coefficient	std error	t stat	p value
L29.KLD	-0.146	0.059	-2.465	0.014*
L30.BTC	0.032	0.034	0.948	0.343
L30.KLD	-0.110	0.059	-1.843	0.065 <sup>†</sup>
L31.BTC	0.024	0.034	0.697	0.486
L31.KLD	-0.005	0.070	-0.067	0.947
L32.BTC	-0.030	0.034	-0.879	0.380
L32.KLD	-0.011	0.070	-0.157	0.875
L33.BTC	-0.007	0.034	-0.209	0.834
L33.KLD	0.016	0.060	0.265	0.791
L34.BTC	0.075	0.034	2.193	0.028*
L34.KLD	0.018	0.059	0.302	0.762
L35.BTC	0.014	0.035	0.400	0.689
L35.KLD	0.081	0.060	1.354	0.176
L36.BTC	0.032	0.035	0.918	0.359
L36.KLD	0.082	0.060	1.362	0.173
L37.BTC	-0.036	0.035	-1.048	0.295
L37.KLD	0.059	0.060	0.983	0.325
L38.BTC	0.043	0.035	1.229	0.219
L38.KLD	0.102	0.060	1.703	0.089 <sup>†</sup>
L39.BTC	0.011	0.035	0.323	0.747
L39.KLD	-0.003	0.060	-0.057	0.955
L40.BTC	-0.035	0.035	-1.010	0.313
L40.KLD	0.034	0.060	0.559	0.576
L41.BTC	0.009	0.035	0.261	0.794
L41.KLD	-0.060	0.060	-0.993	0.321
L42.BTC	-0.037	0.035	-1.074	0.283
L42.KLD	0.035	0.060	0.585	0.558
L43.BTC	0.000	0.035	0.002	0.998
L43.KLD	0.015	0.060	0.248	0.804
L44.BTC	0.067	0.035	1.920	0.055 <sup>†</sup>

次ページに続く

	coefficient	std error	t stat	p value
L44.KLD	-0.060	0.060	-1.002	0.316
L45.BTC	-0.003	0.035	-0.074	0.941
L45.KLD	-0.042	0.060	-0.693	0.488
L46.BTC	0.057	0.035	1.643	0.100
L46.KLD	0.001	0.060	0.011	0.991
L47.BTC	0.025	0.035	0.706	0.48
L47.KLD	-0.021	0.060	-0.357	0.721
L48.BTC	0.033	0.035	0.929	0.353
L48.KLD	-0.081	0.060	-1.358	0.175
L49.BTC	0.070	0.035	1.989	0.047*
L49.KLD	-0.023	0.060	-0.387	0.699
L50.BTC	0.035	0.035	1.002	0.317
L50.KLD	0.113	0.060	1.898	0.058 <sup>†</sup>
L51.BTC	0.026	0.035	0.748	0.455
L51.KLD	-0.081	0.060	-1.353	0.176
L52.BTC	0.031	0.035	0.869	0.385
L52.KLD	0.113	0.060	1.903	0.057 <sup>†</sup>
L53.BTC	0.033	0.035	0.947	0.343
L53.KLD	0.066	0.060	1.114	0.265
L54.BTC	0.013	0.035	0.362	0.717
L54.KLD	0.090	0.060	1.502	0.133
L55.BTC	0.027	0.035	0.775	0.438
L55.KLD	-0.122	0.060	-2.032	0.042*
L56.BTC	0.037	0.035	1.063	0.288
L56.KLD	0.020	0.060	0.341	0.733
L57.BTC	0.036	0.035	1.036	0.300
L57.KLD	-0.093	0.060	-1.566	0.117
L58.BTC	0.072	0.035	2.030	0.042*
L58.KLD	0.121	0.059	2.043	0.041*
L59.BTC	-0.015	0.035	-0.421	0.674

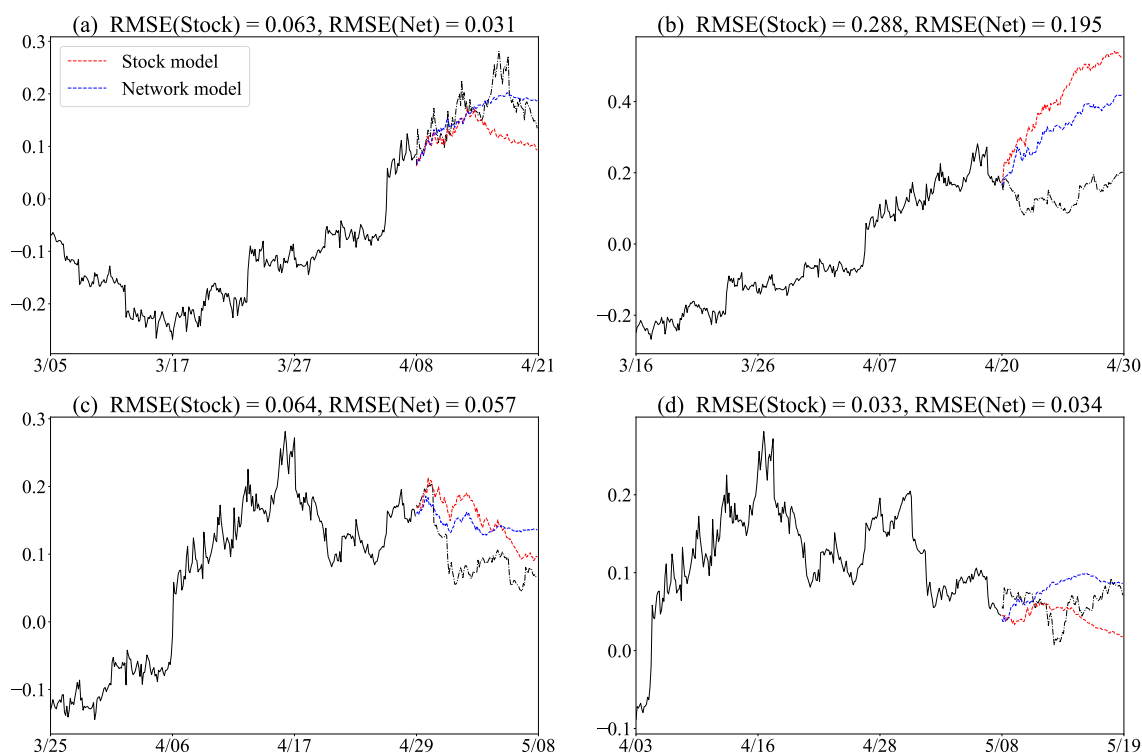
次ページに続く

	coefficient	std error	<i>t</i> stat	<i>p</i> value
L59.KLD	-0.046	0.059	-0.776	0.438
L60.BTC	0.015	0.035	0.423	0.672
L60.KLD	-0.003	0.059	-0.059	0.953
L61.BTC	0.001	0.035	0.030	0.976
L61.KLD	-0.072	0.060	-1.193	0.233
L62.BTC	0.064	0.035	1.827	0.068 <sup>†</sup>
L62.KLD	0.058	0.060	0.978	0.328

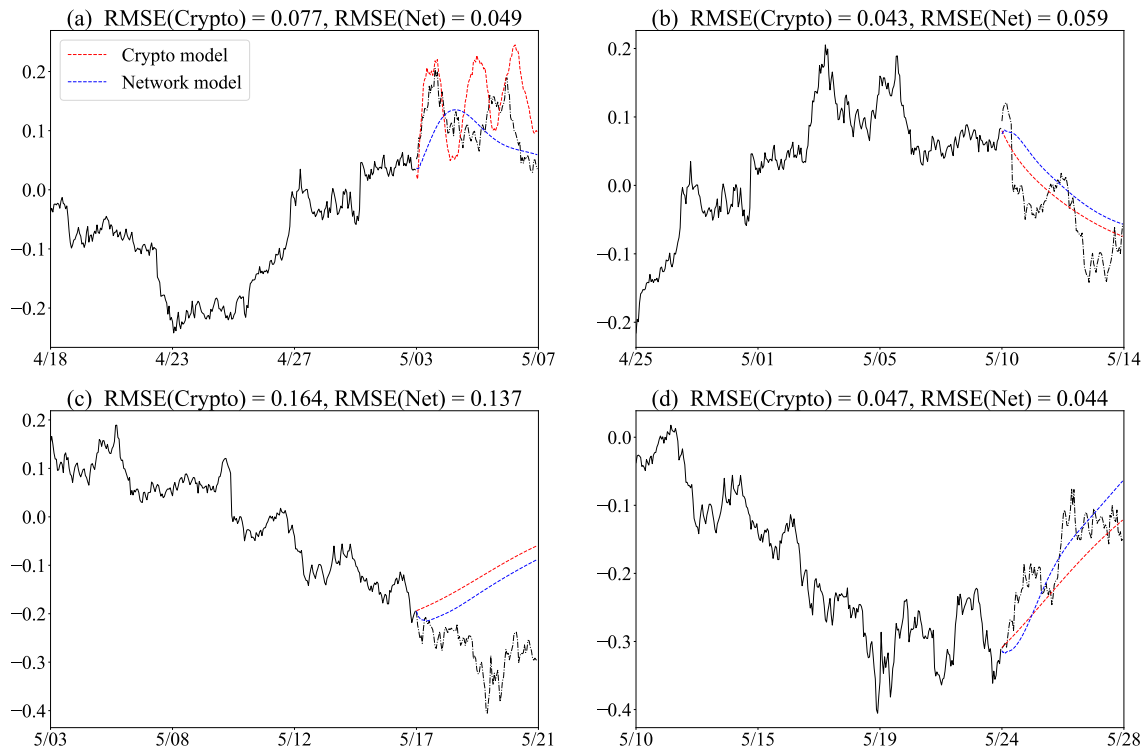
## S2 異なる予測期間における予測結果

4.3 節および 4.4 節で検証した株価予測と暗号通貨価格予測について、予測期間を変えたときの予測結果を示す。付録図 2.1 a, b, c, d はそれぞれ 4.3 節同様に、QQQ 価格変化率の 105 ステップ (7 営業日) 間の予測を、予測の開始時点をもとに 2020 年 4 月 8 日から 7 営業日ごとにとったときの株価モデルおよびネットワークモデルによる予測結果を示したものである。

また、付録図 2.2 a, b, c, d はそれぞれ 4.4 節同様に、ビットコイン価格変化率の 96 ステップ (4 営業日) 間の予測を、予測の開始時点をもとに 2021 年 5 月 3 日から 7 営業日ごとにとったときの株価モデルおよびネットワークモデルによる予測結果を示したものである。



付録図 2.1 赤い破線 (Stock model) は株価モデルによる予測を表し、青い破線 (Network model) はネットワークモデルにおける予測を表す。RMSE(Stock) および RMSE(Net) はそれぞれ株価モデルとネットワークモデルにおける予測の RMSE を表す。



付録図 2.2 赤い破線 (Crypto model) は暗号通貨モデルによる予測を表し, 青い破線 (Network model) はネットワークモデルにおける予測を表す.  $RMSE(Crypto)$  および  $RMSE(Net)$  はそれぞれ暗号通貨モデルとネットワークモデルにおける予測の RMSE を表す.