経済論叢（京都大学）第 197 巻第 1 号，2023 年 2 月　　https://doi.org/10.57475/keizaironso.197.1.4

経済学研究科優秀修士論文賞受賞論文

Hawkes 過程と共和分による コロナウイルス感染状況の分析

中　村　裕　貴*

概　要

新規感染者数と入院・療養患者数は，確保すべきベット数やホテルの部屋数の見積もりを含む，様々な政策決定に重要な指標である。本論文では，2021 年の東京都における新規陽性者数と入院・療養患者数の日別データを用い，それらの累積数の先 1 ヵ月に関する予測精度を 3 つのモデル間で比較する。パラメトリック Hawkes 過程を用いたモデル，ノンパラメトリック Hawkes 過程を用いたモデル，VECM の 3 つである。また，後者の予測には，直接 Hawkes 過程を用いるだけのコンセンサスがないため，2 つのデータの共和分関係を利用し，前者の予測と組み合わせることで，その予測を行った。その結果，共和分関係を用いることへの一定の有効性と，モデルの選択基準の 1 つを得た。それはどちらの予測においても，VECM が先 1 ヵ月の急上昇を予測した場合は Hawkes 過程を用いたモデル，その他の場合は VECM を選択するというものである。

Ⅰ　はじめに

本論文では，コロナウイルスによる累積新規感染者数と累積入院・療養患者数の先 1 ヵ月に関する予測について，3 つのモデルを比較する。新規感染者数と入院・療養患者数は政府や企業の意思決定に重要な指標である。特に後者は，確保すべきベット数やホテルの部屋数の見積もりに有用である。本研究では，2021 年の東京都における新規陽性者数と入院・療養患者数の日別データを用いた。PCR 検査の誤判定など，厳密には陽性者と感染者は異なるが，本論文では陽性者を感染者と考えることとする。累積新規感染者数の予測には，指数型のカーネル関数を持つパラメトリックな Hawkes 過程を用いたモデル，ヒストグラム型のカーネル関数を持つノンパラメトリックな Hawkes 過程を用いたモデル（Marsan and Lengliné［2008］），vector error correction model（VECM）を使用した。累積入院・療養患者数の予測には，VECM については同様のモデルを用いた。一方，累積入院・療養患者数に Hawkes 過程を直接フィットさせる十分な先行研究がない。そこで，新規感染者数と入院・療養患者数の共和分関係と，累積新規感染者数の予測を組み合わせることで予測を行った。その結果，以下の 3 つのことがわかった。1 つ目は，少なくとも本分析で用いたデータと Hawkes 過程のフィットに関する検定において，指数型のカーネル関数を持つパラメトリックな Hawkes 過程は，ヒストグラム型のカーネル関数を持つノンパラメトリックな Hawkes 過程と同程度のフィットの良さを示していながら，予測精度がより高いことである。ただし，Park et al.［2022］で指摘されているように，感染症のデータに対するカーネル関数ついての経験的な知見が浅い。そのためカーネル関数の選択ついては，茅根・白石［2021］で注意されているように，

＊　京都大学大学院経済学研究科修士課程 2 年

慎重に行わなければならない。2 つ目は，累積入院・療養患者数の予測に共和分関係を利用することに一定の有効性が確認されたことである。2 つの累積数のグラフの形状に大きな差がないことから，適切な予測モデルを用いれば，予測精度に大きな差が出ないことが推測される。そして，累積新規感染者数と累積入院・療養患者数の予測に関して，3 つのモデル間の予測精度と，予測値が過大または過小評価されているか否かの両方に差異がほとんど見られなかったことが理由である。3 つ目は，3 つのモデル間の選択基準である。基本的に VECM を用いて予測を行い，それが 1 ヵ月先の急上昇を予測する場合のみ Hawkes 過程を用いたモデルにより予測を行うというものである。ただし，先 1 ヵ月の予測ではいづれの場合も Hawkes 過程を用いたモデルが優れていたが，VECM による予測の急上昇度合いにより，先 1 週間や先 2 週間の予測に関しては VECM のほうが優れている場合があることが確認された。この急上昇度合いの境界については，本研究で用いたデータだけでは明らかとはならなかった。また，急上昇局面の予測に用いるモデルは，その他の局面において VECM 以上の精度がない限り，Hawkes 過程によるもの以外のモデルを用いた場合でも，同様の議論が成り立つと考えられる。以下で挙げる先行研究に対する本研究の新規性は，今挙げた 2 つ目と 3 つ目の結果，すなわち急上昇局面における Hawkes 過程と共和分を用いた累積入院・療養患者数の予測に対する一定の有効性と，局面ごとの分析により，各局面におけるモデルの選択基準が 1 つ与えられたことにある。一方で，以下で挙げる先行研究を踏まえて注意しなければならないことが 2 つある。1 つ目は Hawkes 過程のモデル選択である。繰り返しになるが，カーネル関数に対する経験的知見が浅いため，検定やその他の検証により，常に慎重に選択しなければならない。また，これは本研究にとっては今後の課題となることであるが，新規感染者数や入院・療養患者数の日別データは離散的なため，本研究のような恣意的な学習データの生成をすることなく，直接離散的なデータを扱う手法を選択することが望ましいと考えられる。2 つ目は，新規感染者数と入院・療養患者数の日別データが何階差分を取れば定常となると仮定するかということである。Nguyen, Turk, and McWilliams［2021］では 1 階差分が定常であると仮定されているが，本研究では 2 階差分が定常であると仮定している。これは分析対象となるデータに対して毎回検証すべき項目であると言える。

点過程の中でもクラスター性を持つものを自己励起過程という。クラスター性とは，過去のイベント発生に起因して，短期間にイベントが集中的に発生する性質を指す。このような性質から，地震や神経細胞のスパイク発火，金融市場における取引・注文などの分析に用いられている。単純な自己励起過程の 1 つに Hawkes 過程（Hawkes［1971］）と呼ばれるものがある。Hawkes 過程はイベントが瞬間的に起こる条件付確率により特徴づけられる。その確率は，イベントが瞬間的に起こる平均的な確率と，過去のイベント発生が現在のイベント発生へもたらす影響度によってモデル化されている。その影響度合いを表す関数はカーネル関数と呼ばれている。（近江・野村［2021］）

Hawkes 過程は地震学に応用される中での発展も多い。漸近理論（Ogata［1987］）はその 1 つである。また Marsan and Lengliné［2008］は，カーネル関数の形を強く仮定しないヒストグラム型のカーネル関数を持つ Hawkes 過程を用いて，地震の発生に関する分析を行っている。そうすることで，地震間の因果構造を特定する困難さの解消を試みている。このノンパラメトリック手法を用いて，Park et al.［2022］は西アフリカにおけるエボラ出血熱の感染状況を分析している。そこでは，この Hawkes 過程を用いたモデルと，感染症分野で用いられている SEIR モデルによる予測精度の比較が行われている。そこで用いらているギニア南東部，シエラレオネ東部，リベリア北西部

の 2014 年のデータでは，2 週間先までの累積新規感染者数の予測に関して，Hawkes 過程を用いたモデルの予測精度が SEIR モデルに比べて高かったことが報告されている。また茅根・白石［2020］は，新規感染者数という離散的なデータを直接扱うことができる Kirchner［2017］による手法と Hawkes グラフ表現を用いて，関西圏と関東圏における第 1 波・第 2 波の感染拡大の影響構造を分析している。

Nguyen, Turk, and McWilliams［2021］はノースカロライナ州のある地域におけるコロナウイルスによる患者数（hospital census）と新規感染者数の日別データを VECM により分析し，患者数の予測精度を auto regressive integrated moving average（ARIMA）モデルと比較している。そして，先 7 日の予測においては VECM のほうが高い予測精度をもつことが報告されている。そこでは，2020 年 6 月 16 日から 2020 年 11 月 28 日のデータと，評価基準 mean absolute persentage error（MAPE）を用いて，交差検証を行うことで，VECM と ARIMA を比較している。また，新規感染者数には通常の対数変換を行い，hospital censusn にはデータに 1000 という上限があるため，$log(x(1000-x)^{-1})$ という変換を行っている。更に，これらの系列が I (1) 系列，すなわち 1 階差分が定常であると仮定としている。

本論文は，以下のように構成されている。 Ⅱ節では，分析対象となるデータと，Hawkes 過程のパラメータ推定を行う際のデータの前処理について説明する。 Ⅲ節では，モデルの概要と，予測手法について述べる。 Ⅳ節では，各々のモデルのパラメータ推定と，モデルや仮定に関する検定を行う。 Ⅴ節では，各々のモデルによる予測について考察する。 Ⅵ節では，本論文のまとめと今後の課題を述べる。 Ⅶ節では，補足的な図表や，今回の分析で使用したプログラムのコードを掲載している。

Ⅱ　データ

本論文では，厚生労働省が発表している 2 つのデータを用いる。1 つ目は，新規陽性者数の推移（日別）に関するデータである。ここには，HER-SYS データをもとに集計された，各都道府県の日別の新規陽性者数が記録されている。2 つ目は，入院治療等を要する者等の推移に関するデータである。ここには，各自治体が公表した，入院中（調整中を含む）・宿泊療養中・自宅療養中等の者の数が記録されている。本論文では，これを入院・療養患者数と呼ぶことにする。以下の分析においては，これらのデータの中でも特に，東京都の 2020 年 11 月から 2021 年 12 月のデータを分析対象とする。これらのデータを用いて，2021 年 1 月から 2021 年 12 月までの各月の累積新規感染者数と累積入院・療養患者数を予測し，その考察を行う。VECM の推定には，2020 年 11 月から予測月の前月末までの新規感染者数と入院・療養患者数の日別データを用いた。一方，Hawkes 過程の推定には，予測月の前 2 ヵ月についての新規感染者数のデータを用いた。このとき，Park et al.［2022］に倣い，各日の新規感染者数を 1 日を分割した各時間に一様分布を用いて発生させた。その際，1 日を分単位に区切り，新規感染者数の 1 の位を四捨五入して 10 で割ったものを発生させた。このような区切り方と丸め方をしたのは，Ⅴ節で扱う予測精度と Ⅳ節で行う検定結果がともに，その他の場合と比べて良かったからである。予測精度については，例えば表 2.1 から見て取れる。これはパラメトリック Hawkes 過程を用いて，2021 年 1 月の累積新規感染者数を予測した結果である。この予測精度は式 9 で表されたものである。ここから，データを丸めないほうが予測精

表 2.1 パラメトリック Hawkes 過程を用いた 2021 年 1 月の累積新規感染者数に関する予測の RMSE を示している。1 行目は 1 日を秒単位に区切り，四捨五入をしないデータをフィットさせた場合である。2 行目は 1 日を秒単位に区切り，1 の位を四捨五入して 10 で割ったデータをフィットさせた場合である。1 week，2 weeks，3 weeks，4 weeks という項目は各々，先 1 週間予測，先 2 週間予測，先 3 週間予測，先 4 週間予測を表している。各値は小数点第 1 位を四捨五入したものである。

区切り方・四捨五入 ＼ 予測期間	1 week	2 weeks	3 weeks	4 weeks
秒・なし	3051	16837	38734	65156
秒・あり	1729	11030	25296	41067

表 2.2 100 通りのランダムシードから生成した学習データに対する，Hawkes 過程を用いたモデルによる 2021 年 1 月の累積新規感染者数に関する予測の RMSE の標準偏差を示したものである。ここで，学習データの生成は 1 の位を四捨五入して 10 で割り，1 日を分単位に区切ることで行った。各値は小数点第 5 位を四捨五入したものである。1 week，2 weeks，3 weeks，4 weeks という項目は各々，先 1 週間予測，先 2 週間予測，先 3 週間予測，先 4 週間予測を表している。

Hawkes 過程 ＼ 予測期間	1 week	2 weeks	3 weeks	4 weeks
パラメトリック	27.5710	50.1222	64.3424	75.6064
ノンパラメトリック	8.4410	8.4791	8.1923	8.6517

表 2.3 100 通りのランダムシードから生成した学習データに対する，Hawkes 過程を用いたモデルによる 2021 年 1 月の累積新規感染者数に関する予測の RMSE の標準偏差を示したものである。ここで，学習データの生成はデータを丸めず，1 日を秒単位に区切ることで行った。各値は小数点第 5 位を四捨五入したものである。1 week，2 weeks，3 weeks，4 weeks という項目は各々，先 1 週間予測，先 2 週間予測，先 3 週間予測，先 4 週間予測を表している。

Hawkes 過程 ＼ 予測期間	1 week	2 weeks	3 weeks	4 weeks
パラメトリック	1.0943	1.6118	1.8291	2.0503

度が悪くなることがわかる。検定結果については，区切り方を秒単位にした場合に，パラメトリック Hawkes 過程ついて検定を行ったところ，すべての予測月に対して，元のデータがフィットさせた Hawkes 過程から来るものであると統計的に有意に言えないことがわかった。ただし，本分析では以上のようなデータの区切り方や丸め方をしたが，フィットさせるデータの 1 日の最大イベント発生数に応じて，その都度適切な区切り方や丸め方を選択する必要がある。また，ランダムシードの選択によって，Hawkes 過程のパラメータの推定値が大きく変わることはないことも確認された。この結果，予測の誤差も考察に影響がないほど小さくなることも確認された。実際，2021 年 1 月の累積新規感染者数の予測について，100 通りのランダムシードに対して学習データを生成した場合の，式 9 で表された予測精度の標準誤差は表 2.2 の通りである。これは，5 節の表 5.1 内の予測月が 1 月の項目を見ると，モデルの比較に影響を及ぼさない大きさであることが分かる。一方，

パラメトリック Hawkes 過程については丸めない場合の方が RMSE の誤差が小さいことが確認された。これは表 2.3 の通りである。

Ⅲ　モデルと手法

3.1　Hawkes 過程

Hawkes 過程 N_t は以下の条件付き強度関数 λ によって特徴づけられる（近江・野村［2021］）：

$$\lambda(t \mid H_t) = \mu + \sum_{t_i<t} g(t - t_i) \tag{1}$$

ここで，μ は非負の定数で，g は $t < 0$ のとき $g(t) = 0$ を満たす非負の関数である。g はカーネル関数と呼ばれ，過去のイベントからの影響を表現している。t_i はイベント発生時刻を表す確率変数である。H_t は時刻 t までのイベントの発生履歴である。また，十分小さい $\Delta > 0$ に対して，λ は次のような式を満たす（近江・野村［2021］）：

$$P(N_{t+\Delta} - N_t = 1 \mid H_t) = \lambda(t \mid H_t)\Delta \tag{2}$$

$$P(N_{t+\Delta} - N_t = 0 \mid H_t) = (1 - \lambda(t \mid H_t))\Delta \tag{3}$$

この式から，λ はイベントが発生する瞬間的な条件付確率と解釈できる。

3.1.1　パラメトリック Hawkes 過程

本論文の分析では，カーネル関数が $g(t) = abexp(-bt)$ であるモデルを扱う。a と b は，$0 < a < 1$，$b > 0$ を満たす。最尤法でパラメータ $\theta = (\mu, a, b)$ を推定する。対数尤度関数 $L(\theta|t_1, t_2, \ldots, t_n)$ は次で与えられる（近江・野村［2021］）：

$$\begin{aligned} L(\theta|t_1, t_2, \ldots, t_n) = \sum_{i=1}^{n} log(\mu + \sum_{j<i} abexp(-b(t_i - tj)) \\ - (\mu T + \sum_{i=1}^{n} a(1 - exp(-b(T - t_i)))) \end{aligned} \tag{4}$$

ここで，$t_1, t_2, \ldots, t_n$ は時刻 T までに観測したイベントの発生時刻である。Hawkes 過程の漸近理論については Ogata［1987］で展開されている。漸近分散の推定値は Ogata［1987］の THEOREM 3, THEOREM 5 より対数尤度関数のヘッセ行列とした。また，このモデルのカーネル関数は減少関数であるため，予測のシュミレーションには近江・野村［2021］のアルゴリズム 7.16 を用いた。

3.1.2　ノンパラメトリック Hawkes 過程

Marsan and Lengliné［2008］が提案したカーネル関数の形を強く仮定しないモデルを扱う。これはカーネル関数を，幅を固定したヒストグラムとしたモデルである。パラメータは近似的な尤度関数を最大化することで推定される（Marsan and Lengliné［2010］）。λ のモデルは次で与えられる（Fox, Schoenberg, and Gordon［2016］）：

$$\lambda(t \mid H_t) = \mu + K \sum_{t_i<t} g(t - t_i) \tag{5}$$

ただし，K は $0 < K < 1$ を満たす，過去のイベントからの影響の規模を表すパラメータである。

g は区間 $[0,T]$ 上のヒストグラムである。本分析では，ヒストグラムの幅を 3 分の 1 日とした。Park et al.［2022］は半日程度にしていたため，少し細かく設定した。幅を 1 日とした場合と比較したとき，Ⅴ節で述べる予測精度に大きな変化が見られなかった。これはデータが 1 日毎にしか観測されておらず，その間には一様分布で発生させたためであると考えられる。このときパラメータは，$(\mu, K, g_1, g_2, \ldots, g_{3n})$ である。ここで予測月の日数を n とした。$g_k (k = 1, 2, \ldots, 3n)$ はヒストグラムの高さを表している。パラメータは Fox, Schoenberg, and Gordon［2016］の ALGORITHM 1 を用いて推定した。また，その標準誤差は Fox, Schoenberg, and Gordon［2016］の ALGORITHM 3 を用いて推定した。このモデルのカーネル関数は一般に減少関数とは限らないため，予測のシュミレーションには茅根・白石［2021］の Algorithm 1 を用いた。

3.2 VECM モデルと共和分

ラグ $k-1$ の VECM モデルは次で与えられる（沖本［2013］）：

$$\Delta X_t = \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Pi X_{t-1} + \mu + \epsilon_t \tag{6}$$

ここで，X_t は $p \times 1$ ベクトル，ϵ_t（$t = 1, \ldots, T$）は平均が 0，分散行列が Λ である独立な p 次元正規分布である。また，共和分ランクが r であるとき，Π は $\alpha\beta^\top$ と書け，β は各列が共和分ベクトルの $p \times r$ 行列となる。(6) の各パラメータを推定するためには，共和分ランクとラグを決める必要がある。ラグは，k 次のラグをもつ vector auto regressive（VAR）モデルに対して，AIC を用いて選択する。そのラグに対して，Johansen［1991］の Theorem 2.1 にある最大固有値検定とトレース検定を行い，共和分ランクを決定する（沖本［2013］）。その後，その共和分ランクに対する最尤推定により，パラメータを推定する（Lütkepohl［2005］）。また，標準誤差は Lütkepohl［2005］の Proposition 7.4 に関する Remark 3，Remark 4 に従って計算した。

3.3 予測の手順

3.3.1 VECM による予測

a_t を入院・療養患者数（日別）の時系列，b_t を新規感染者数（日別）の時系列とする。a_t，b_t は正であるため，対数を取ったものを扱う。$log(a_t)$，$log(b_t)$ が共に $I(2)$ であると仮定する。(6) の X_t は $(\Delta log(a_t), \Delta log(b_t))^\top$ となる。共和分ランクの決定後，パラメータ $(\Gamma_1, \Gamma_2, \ldots, \Gamma_{k-1}, \alpha, \beta, \mu, \Lambda)$ を推定する。最後に，VECM を VAR モデルに変形し，その点予測を求める（沖本［2013］）。本分析では，それを予測値として用いる。

3.3.2 Hawkes 過程を用いたモデルによる予測

まず，Hawkes 過程のパラメータ推定を行い，そのモデルのシュミレーションを 1000 回行う。シュミレーションのアルゴリズムは先に述べたものを用いて，学習期間の最後のイベント発生時刻から続きのシュミレーションを行う。その際，学習期間の最終時刻までに発生したイベントについてはカウントしないものとする。i 回目のシュミレーションにより求めた値を $N_s^i(s > mT)$ とする。ここで，T は予測月の前 2 ヵ月すなわち学習期間の総日数で，m は 1 日を分単位で表す 60×24 と

する。Park et al.［2022］に倣い，累積新規感染者数の予測月第 t 日目の予測値を，次で定める：

$$\frac{1}{1000}\sum_{i=1}^{1000} N^i_{(t+T)m} \tag{7}$$

次に，共和分による定常過程をモデル化し，そのシュミレーションを 1000 回行う。上で求めた共和分ベクトルの 1 つを $(1,c)^\top$ とする。$\Delta log(a_t) + c\Delta log(b_t)$ を auto regressive moving average（ARMA）モデルにフィットさせ，シュミレーションを行う。推定とシュミレーションの手順は沖本［2013］に倣った。その i 回目のシュミレーションにより求めた値を $z^i_t(t > T)$ とする。ここで，ARMA モデルの時刻 t は学習期間の初日を 1 としている。連続時間モデルである Hawkes 過程と，離散時間モデルである ARMA を区別するために，あえて時間を表す変数に s と t という異なる文字を用いた。最後に，これらを用いて累積入院・療養患者数の予測値を定める。ARMA モデルにフィットさせた定常系列を z_t とすると，$a_t = a_{t-1}e^{z_t}(b_{t-1})^c(b_t)^{-c}$ と書ける。これより，累積入院・療養患者数の予測月第 t 日目の予測値を次で定める：

$$\frac{1}{1000}\sum_{i=1}^{1000}\sum_{j=1}^{t} a^i_{j+T-1}e^{z^i_{j+T}}(b^i_{j+T-1})^c(b^i_{j+T})^{-c} \tag{8}$$

ここで，日々の新規感染者数の予測値 b^i_t は $t > T$ に対して $N^i_{tm} - N^i_{(t-1)m}$ とした。ただし，$b^i_T = b_T$，$N^i_{mT} = N_{mT}$ である。また，$t > T$ に対して a^i_t は，b^i_t と上の漸化式を用いて，$a^i_t = a^i_{t-1}e^{z^i_t}(b^i_{t-1})^c(b^i_t)^{-c}$ により求めた。だだし，$a^i_T = a_T$ である。

Ⅳ　推定結果

4.1 Hawkes 過程

Hawkes モデルについて，パラメータの推定後，近江・野村［2021］ アルゴリズム 8.3 で紹介されている 2 つの検定を行う。観測されたイベント発生時刻を $\{t_1, t_2, \ldots, t_n\}$，$\Lambda(t) = \int_0^t \lambda(s|H_s)ds$ とする。このとき $\{t_1, t_2, \ldots, t_n\}$ が $[0,T]$ で条件付強度 $\lambda(t|H_t)$ をもつ Hawkes 過程に従うならば，$\{\Lambda(t_1), \Lambda(t_2), \ldots, \Lambda(t_n)\}$ が $[0, \Lambda(T)]$ で強度 1 のポアソン過程に従う（近江・野村［2021］ 定理 2.5）。この事実から，強度 1 のポアソン過程が満たす 2 つの性質について検定する。1 つ目は，$\{\Lambda(t_1), \Lambda(t_2), \ldots, \Lambda(t_n)\}$ が $[0, \Lambda(T)]$ 上で一様分布に従っているかを検定する（以下，検定 1 と呼ぶ）。2 つ目は，$\tau_i = \Lambda(t_{i+1}) - \Lambda(t_i)$ として，$\{exp(-\tau_1), exp(-\tau_2), \ldots, exp(-\tau_{n-1})\}$ が $[0,1]$ 上の一様分布に従っているかを検定する（以下，検定 2 と呼ぶ）。その際，ともにコロモゴロフ・スミノルフ検定（以下，KS 検定）を用いる。

4.1.1　パラメトリックモデル

パラメータの推定結果は表 4.1 の通りである。また，KS 検定の p 値は表 4.2 のようになる。KS 検定の帰無仮説は，2 つの分布が一致することである。表 4.2 から，両方の検定において，有意水準 5% で帰無仮説が棄却されない予測月は 1 月，2 月，6 月，9 月とわかる。この結果は， Ⅴ節で述べる予測についての考察と，概ね整合的である。例えば，予測月が 2 月の場合の学習データは 2020 年 12 月と 2021 年 1 月である。この場合に検定が通っていることは，この 2 ヵ月間の観測値が Hawkes 過程に由来すると考えられることを意味する。これは，予測月が 1 月の場合に，長期的には，パラメトリック Hawkes 過程を用いたモデルの予測精度が比較的良いことと整合的である。

表 4.1 パラメトリック Hawkes モデルのパラメータを推定した。各パラメータについて，1 行目にその推定値，2 行目に標準誤差を掲載している。各値は小数点第 5 位を四捨五入したものである。

予測月	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
μ	0.0035	0.0061	0.0038	0.0081	0.0043	0.0103	0.0092	0.0029	0.0058	0.0034	0.0013	0.0010
	0.0009	0.0009	0.001	0.0005	0.0007	0.0046	0.0028	0.0005	0.0015	0.0008	0.0003	0.0003
a	0.9152	0.9131	0.9396	0.6657	0.8797	0.7766	0.7738	0.9806	0.9740	0.9816	0.9666	0.6700
	0.0305	0.0011	0.0176	0.0021	0.0243	0.0022	0.0764	0.0012	0.0097	0.0111	0.015	0.1098
b	0.0013	0.0017	0.0018	0.0013	0.0011	0.0018	0.0019	0.0019	0.0029	0.0038	0.0032	0.0010
	0.0002	0.0002	0.0002	0.0002	0.0002	0.0007	0.0003	0.0001	0.0003	0.0003	0.0003	0.0004

表 4.2 パラメトリックモデルについて KS 検定を行った。各値は小数点第 5 位を四捨五入したものである。

予測月	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
検定 1	0.0789	0.1378	0.0017	0.0035	0.0495	0.3673	0.0005	0.0077	0.3824	0.0399	0.0198	0.0001
検定 2	0.4566	0.2446	0.6358	0.4106	0.3602	0.6124	0.4258	0.9511	0.4770	0.9564	0.7744	0.2786

表 4.3 ノンパラメトリック Hawkes モデルのパラメータ（μ，K）を推定した。各値は小数点第 5 位を四捨五入したものである。各パラメータについて，1 行目にその推定値，2 行目に標準誤差を掲載している。ここでは計算時間の都合上，予測月が 1 月から 8 月までの場合のみ掲載している。

予測月	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
μ	0.0057	0.0065	0.0039	0.0088	0.0059	0.0105	0.0107	0.0055	0.0076	0.0031	0.0011	0.0007
	0.0055	0.0167	0.0223	0.0034	0.0057	0.0081	0.0048	0.0107				
K	0.8278	0.9024	0.9358	0.6313	0.8112	0.7693	0.7298	0.9167	0.9609	0.9832	0.9698	0.7657
	0.1485	0.2123	0.282	0.1374	0.1721	0.1638	0.1126	0.1526				

その他，予測月が 4 月，8 月の場合に，パラメトリック Hawkes 過程を用いたモデルの予測精度が比較的良いこととも整合的である。

4.1.2 ノンパラメトリックモデル

パラメータの推定結果は表 4.3・図 7.2 の通りである。また，KS 検定の p 値は表 4.4 のようになる。表 4.4 より，両方の検定において，有意水準 5% で帰無仮説が棄却されない予測月は 2 月，6 月，9 月であることがわかる。1 月についても有意水準 10% では棄却されないことを踏まえると，概ねパラメトリックモデルの場合と同様の結果が得らていることがわかる。Park et al.［2022］によれば，一般的に感染症のデータに対する適切なカーネル関数に関する経験的知見が浅いため，ノンパラメトリックモデルを用いるのが望ましいとされている。また，Park et al.［2022］において，先 2 週間程度の累積新規感染者数の予測について，ノンパラメトリックモデルは SEIR モデルより高い予測精度を示すと結論付けられている。これらの点からノンパラメトリックを用いることが適切であると考えられる。一方，検定結果が類似している点と，パラメータ推定の精度や V 節で述べる予測精度が高い点を踏まえ，比較のためにパラメトリックモデルも用いることとする。

4.2 VECM モデルと共和分

まず，入院・療養患者数（日別）と新規感染者数（日別）の対数系列に対する argumented Dickey-Fuller（ADF）検定を行う。各々の系列を，log(患者数)，log(感染者数) と書く。また，推

表 4.4　ノンパラメトリックモデルについての KS 検定を行った。各値は小数点第 5 位を四捨五入したものである。

予測月	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
検定 1	0.0047	0.1005	0.0011	0.0019	0.0021	0.3364	0.0001	0	0.1159	0.0017	0.0007	0.0012
検定 2	0.2488	0.0739	0.9066	0.7215	0.5138	0.9124	0.5892	0.3756	0.6417	0.5685	0.8688	0.1789

表 4.5　VECM モデルの推定に関連する ADF 検定を行った。各値は小数点第 5 位を四捨五入したものである。また，各値の肩に乗っている*は有意水準 5% で棄却されることを意味する。共和分系列に対する ADF 検定は最大ラグを 7 として AIC によりラグを選択して行った。その他の時系列に対する ADF 検定は最大ラグを 14 として AIC によりラグを選択して行った。共に最大ラグは自己相関，偏自己相関を考慮して選択した。

予測月	1 月	2 月	3 月	4 月	5 月	6 月
log(感染者数)	-0.8099	-1.7617	-1.5437	-2.0041	-2.2159	-2.3228
log(感染者数) の 1 階差分	-8.7494*	-2.9162	-1.6522	-1.4262	-1.7212	-1.9794
log(感染者数) の 2 階差分	-5.1176*	-5.8721*	-7.0739*	-7.5741*	-8.4201*	-9.251*
log(患者数)	-1.5676	-3.8959*	-2.2227	-2.7572	-3.156	-2.9949
log(患者数) の 1 階差分	-0.254	-0.9573	-1.9958	-1.8947	-1.8377	-2.1608
log(患者数) の 2 階差分	-8.8263*	-4.6983*	-8.3028*	-5.2155*	-4.087*	-6.4914*
共和分系列	-3.2277*	-3.6735*	-5.6837*	-5.2408*	-5.7089*	-8.3619*

予測月	7 月	8 月	9 月	10 月	11 月	12 月
log(感染者数)	-2.6279	-1.3933	-2.8091	-3.0917	-2.4716	-2.2972
log(感染者数) の 1 階差分	-2.0607	-1.3309	-2.1447	-1.5629	-1.9796	-2.0711
log(感染者数) の 2 階差分	-9.976*	-10.4189*	-10.7599*	-11.3983*	-12.0864*	-12.4475*
log(患者数)	-3.2908	-2.2879	-3.6794*	-4.3665*	-3.7829*	-3.6189*
log(患者数) の 1 階差分	-2.2841	-1.2432	-2.121	-1.5095	-1.8974	-1.9398
log(患者数) の 2 階差分	-4.8406*	-3.9773*	-4.3975*	-4.6584*	-4.8281*	-4.9275*
共和分系列	-9.0358*	-9.2540*	-9.9804*	-8.4629*	-8.7874*	-9.5953*

定した共和分ベクトルを $(1, c)^t op$ とし，時系列 log(患者数) + $clog$(感染者数)（以下，共和分系列と呼ぶ）に対しても ADF 検定を行う。検定統計量は表 4.5 の通りである。また，図 7.1，図 7.3 は表 4.5 にある上から 6 つの系列をプロットしたものである。図 7.1 を見ると，新規感染者数（日別）の対数系列の 1 階差分系列は定常のように思われる。一方，表 4.5 を見ると，1 階差分系列に対しては単位根過程であるという帰無仮説が，予測月が 1 月の場合を除いて棄却されないことがわかる。また，これを I (1) と仮定して共和分を推定すると，共和分系列が定常とはならない場合があることが確認できた。これらを踏まえ，直観的な予想とは異なるが，新規感染者数（日別）を I (2) 系列であることを仮定する。また，入院・療養患者数（日別）については，図 7.1，表 4.5 のどちらの観点からも，I (2) 系列であると考えられる。ただし，表 4.5 から，予測月が 9 月，10 月，11 月，12 月の場合，入院・療養患者数（日別）の対数系列に対する検定の帰無仮説が棄却されていることが見て取れる。しかし，図 7.1 から定常とは言い難いので，この系列についても同様に I (2) 系列であることを仮定する。次に VECM のパラメータを推定する。表 4.5 は VAR モデルに対して最大ラグを 14 として AIC を計算した結果である。これによりラグを選択し，Johansen の最大固有値検定とトレース検定を行った。表 4.6 はその結果を示したものである。

ここから，予測月が 1 月の場合のトレース検定以外においては，すべて有意水準 5% で，共和分

表 4.6 表の左側が最大固有値検定，右側がトレース検定の結果である。各予測月の 1 行目は多くとも 1 つの共和分関係しか存在しないといという帰無仮説に対する検定の結果，2 行目は共和分関係が存在しないという帰無仮説に対する検定の結果を意味している。各値は小数点第 2 位を四捨五入したものである。

	最大固有値検定					トレース検定			
予測月	10pct	5pct	1pct	teststat		10pct	5pct	1pct	teststat
1 月	6.5	8.18	11.65	1.11		6.5	8.18	11.65	1.11
	12.91	14.9	19.19	15.2		15.66	17.95	23.52	16.31
2 月	6.5	8.18	11.65	1.41		6.5	8.18	11.65	1.41
	12.91	14.9	19.19	20.8		15.66	17.95	23.52	22.2
3 月	6.5	8.18	11.65	2.02		6.5	8.18	11.65	2.02
	12.91	14.9	19.19	34.73		15.66	17.95	23.52	36.75
4 月	6.5	8.18	11.65	2.65		6.5	8.18	11.65	2.65
	12.91	14.9	19.19	49.05		15.66	17.95	23.52	51.7
5 月	6.5	8.18	11.65	4.78		6.5	8.18	11.65	4.78
	12.91	14.9	19.19	54.35		15.66	17.95	23.52	59.13
6 月	6.5	8.18	11.65	5.89		6.5	8.18	11.65	5.89
	12.91	14.9	19.19	41.17		15.66	17.95	23.52	47.06
7 月	6.5	8.18	11.65	7.76		6.5	8.18	11.65	7.76
	12.91	14.9	19.19	46.79		15.66	17.95	23.52	54.55
8 月	6.5	8.18	11.65	2.55		6.5	8.18	11.65	2.55
	12.91	14.9	19.19	53.73		15.66	17.95	23.52	56.28
9 月	6.5	8.18	11.65	7.19		6.5	8.18	11.65	7.19
	12.91	14.9	19.19	58.45		15.66	17.95	23.52	65.64
10 月	6.5	8.18	11.65	3.45		6.5	8.18	11.65	3.45
	12.91	14.9	19.19	56.68		15.66	17.95	23.52	60.13
11 月	6.5	8.18	11.65	4.51		6.5	8.18	11.65	4.51
	12.91	14.9	19.19	60.35		15.66	17.95	23.52	64.86
12 月	6.5	8.18	11.65	5.66		6.5	8.18	11.65	5.66
	12.91	14.9	19.19	58.72		15.66	17.95	23.52	64.38

関係が存在しないという帰無仮説が棄却されていることがわかる。また，同様の場合において，多くとも 1 つの共和分関係しか存在しないという帰無仮説は有意水準 5% で棄却されていないことがわかる。更に，共和分ランクを 1 として共和分を推定し，それに対応する共和分過程に対する ADF 検定の結果が表 4.5 の 7 行目である。これを見ると，有意水準 5% で定常といえることがわかる。以上を踏まえて，本分析では共和分ランクを 1 とする。それを用いて VECM のパラメータを推定した。

V　予測と考察

本節では，累積新規感染者数と累積入院・療養患者数の予測精度を，Ⅳ節で推定したモデル間で比較する。図 5.1 は日別の新規感染者数と入院・療養患者数をプロットしたものである。上昇または下降局面の把握のためにここに示した。表 5.1 と表 5.2 は，各々累積新規感染者数と累積入院・療養患者数について，各モデルの予測精度を RMSE で表したものである。先 1 週間予測の RMSE は，次のような式で計算した：

図 5.1　上から新規感染者数，入院・療養患者数の日別データをプロットしたものである。横軸の開始地点は 2020 年 11 月 1 日である。

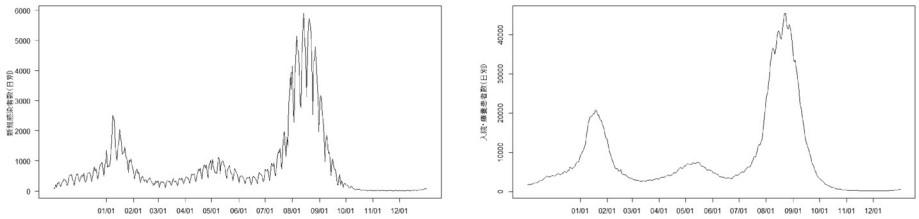

表 5.1　各々のモデルによる累積新規感染者数の予測の RMSE を示したものである。HE はパラメトリック Hawkes 過程を用いたモデルを，HN はノンパラメトリック Hawkes 過程を用いたモデルを意味する。1 week，2 weeks，3 weeks，4 weeks という項目は各々，先 1 週間予測，先 2 週間予測，先 3 週間予測，先 4 週間予測を表している。各値は小数点第 1 位を四捨五入したものである。

予測月	1 月			2 月			3 月			4 月		
	HE	HN	VECM	HE	HN	VECM	HE	HN	VECM	HE	HN	VECM
1 week	959	1828	810	1048	950	63	1472	1440	86	239	292	198
2 weeks	9099	14085	3042	6255	5699	220	7029	6624	264	1495	1689	1217
3 weeks	24371	34453	23192	16865	15389	1378	17345	15997	1143	5040	5399	4214
4 weeks	43224	59272	96325	33698	30659	3721	32389	29633	3411	12126	12669	10326

予測月	5 月			6 月			7 月			8 月		
	HE	HN	VECM	HE	HN	VECM	HE	HN	VECM	HE	HN	VECM
1 week	563	1085	404	595	573	111	143	169	55	1708	7823	8949
2 weeks	3670	5524	603	3345	3243	216	1694	1906	406	8029	40511	83353
3 weeks	8430	12172	1939	8330	8050	1006	8347	8779	3227	31106	105070	370476
4 weeks	13691	19862	7034	14764	14231	4314	25223	25868	11844	68605	198082	1280270

予測月	9 月			10 月			11 月			12 月		
	HE	HN	VECM	HE	HN	VECM	HE	HN	VECM	HE	HN	VECM
1 week	2782	2212	2168	3180	2278	142	1186	789	20	105	75	12
2 weeks	18839	15457	10184	19220	13530	396	5777	3902	118	416	305	37
3 weeks	55467	46312	24904	53730	38558	736	14374	10159	295	877	657	49
4 weeks	116597	98591	46125	109204	80438	1166	27238	20080	547	1407	1055	126

$$\sqrt{\frac{1}{7i}\sum_{j=1}^{7i}\Big((\text{モデルによる第}j\text{日目の累積数の予測値})-(\text{第}j\text{日目の累積数の観測値})\Big)^2} \tag{9}$$

以上の予測結果を考察する。累積新規感染者数，累積入院・療養患者数について，同様の結果が得られた。予測精度については，表 5.1 と表 5.2 を比較すればわかる。予測値が過大，過小評価のいづれとなっているかは，図 5.2 と図 5.3 の比較や，図 5.4 と図 5.5 の比較から見て取れる。この結果から，共和分を用いて累積入院・療養患者数の予測を行うことに一定の有効性があると考えられる。図 5.1 から，2 つの累積数のグラフの形状に大きな差がないことが見て取れる。そのため適切な予測モデルを用いれば，予測精度に大きな差が出ないことが推測されるからである。またその予測精度は，新規感染者数の予測に用いたモデルの予測精度に大きく依存すると考えられる。

次に，パラメトリック Hawkes 過程を用いたモデルと，ノンパラメトリック Hawkes 過程を用いたモデルを比較する。もちろん，Hawkes 過程がイベントの感染的な拡大を表現するモデルである

表 5.2 各々のモデルによる累積入院・療養患者数の予測の RMSE を示したものである。HE はパラメトリック Hawkes 過程と共和分を用いたモデルを，HN はノンパラメトリック Hawkes 過程と共和分を用いたモデルを意味する。各値は小数点第 1 位を四捨五入したものである。

予測月	1 月			2 月			3 月			4 月		
	HE	HN	VECM	HE	HN	VECM	HE	HN	VECM	HE	HN	VECM
1 week	12173	10932	4320	16408	14476	2843	21078	20352	713	903	1390	1413
2 weeks	54992	110031	22445	117761	106710	12551	116581	108908	2285	5872	7959	7449
3 weeks	216689	331803	167367	330750	301432	19114	304973	279156	3369	20131	24131	22682
4 weeks	472882	657364	767016	672217	611243	23176	586920	536345	12971	54509	60603	57849

予測月	5 月			6 月			7 月			8 月		
	HE	HN	VECM	HE	HN	VECM	HE	HN	VECM	HE	HN	VECM
1 week	4102	9395	704	7912	7591	724	2887	2492	486	28169	75669	31027
2 weeks	25601	44401	1090	47980	46621	2725	4535	4406	2218	183678	394679	391614
3 weeks	67535	105226	8780	126094	122324	10117	28280	31930	4614	527527	1003222	2151264
4 weeks	118360	180599	41626	240233	232981	32405	116650	122998	33126	1096801	1926724	8416949

予測月	9 月			10 月			11 月			12 月		
	HE	HN	VECM	HE	HN	VECM	HE	HN	VECM	HE	HN	VECM
1 week	121647	109104	1412	62962	45314	1315	19988	13562	11	1273	926	71
2 weeks	518852	450152	29008	377539	264668	5148	107462	74295	328	5437	4126	427
3 weeks	1304507	1119043	103532	1055641	752093	11511	275126	197502	1156	12302	9678	1030
4 weeks	2531652	2169382	224639	2162180	1580385	19987	527834	394781	2464	20915	16677	1372

という特性上，減少局面における予測に対しては有用ではない。ここでいう減少局面とは，予測月の前月末あたりにかけて，日別の発生数が減少している局面を意味している。本分析で用いたデータでは，予測月が 2 月，3 月，6 月，9 月，10 月，11 月，12 月であるケースに該当する。このような場面では，Hawkes 過程を用いたモデルによる予測値は，過大評価される傾向があることがわかる。例えば，図 5.2 や図 5.3 から見て取れる。これは，Hawkes 過程を用いたモデルの平均累積発生数が，減少局面に入る前の学習データの累積発生数や規模のパラメータに依存していることに起因すると考えられる。一方，上昇局面に注目すると，Hawkes 過程を用いたモデルによる予測値は，過小評価される傾向があることがわかる。ここでいう上昇局面とは，下降局面以外の予測月を意味する。この結果は，Park et al.［2022］で，ノンパラメトリック Hawkes 過程を用いて得られていた結果と一致している。さらに，このような局面で，パラメトリックモデルとノンパラメトリックモデルを比較すると，ノンパラメトリックモデルによる予測値のほうが，過小評価されていることがわかる。これは例えば，図 5.4 や図 5.5 から見て取れる。これは，定常 Hawkes 過程 N_t の期待値 $E[N_t]$ が $\mu(1-(\text{規模のパラメータ}))^{-1}$ と書けることに起因すると考えられる。ここから，規模のパラメータが 1 に近いとき，μ よりも予測値の増加への影響が多いと推測される。ここで，規模のパラメータとは，パラメトリックモデルにおいては a，ノンパラメトリックモデルにおいては K に該当するものである。本分析での予測値は条件付期待値としていたので，これの期待値は上のものとなり，表 4.1 と表 4.3 から，この値は概ねパラメトリックモデルの方が大きく，特に規模のパラメータについては，ほぼすべての予測月でパラメトリックモデルの方が大きいことが分かる。また，1 月や 8 月のような状況では，パラメトリックモデルのほうが予測精度が良いことが 5.1 と 5.2 からわかる。しかし，一般の感染症データに対してパラメトリックモデルを用いるコンセンサス

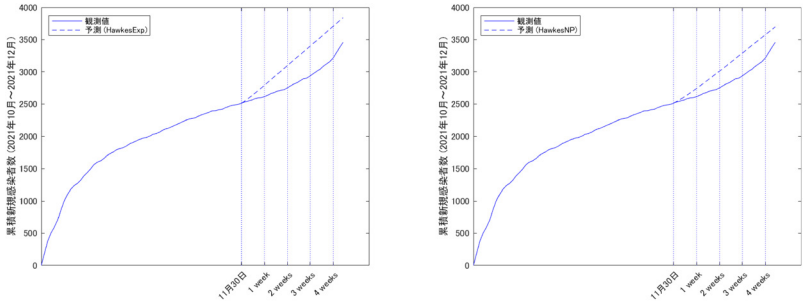

図 5.2　予測月が 12 月のときの累積新規感染者数の予測値と観測値をプロットした。左図がパラメトリック Hawkes モデル，右図がノンパラメトリック Hawkes モデルに関するものである。x 軸の 0 から 11 月 30 日と表示されている点までが学習機関を表す。すなわち，この場合は 10 月 1 日から 11 月 30 日までである。1week，2weeks，3weeks，4weeks と表示さている点はそれぞれ予測月の 7 日，14 日，21 日，28 日を表している。また，グラフの実線は観測値，破線は予測値を表わしている。

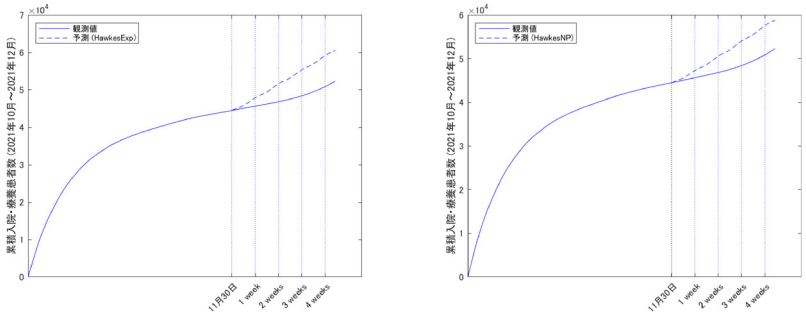

図 5.3　予測月が 12 月のときの累積入院・療養患者数の予測値と観測値をプロットした。左図がパラメトリック Hawkes モデル，右図がノンパラメトリック Hawkes モデルに関するものである。詳しい図の説明は図 5.2 を参照。

は，現在はないということに注意しなければならない。

次に，パラメトリック Hawkes モデルと VECM を比較する。結論から言えば，急上昇局面では，パラメトリック Hawkes モデルが最も精度が良く，その他の局面では VECM が最も精度が良いことがわかった。ここでいう急上昇局面とは，予測月が 1 月や 8 月であるようなケースを意味する。累積新規感染者数，累積入院・療養患者数共に，予測月の 1 ヵ月間で前 2 ヵ月の累積に対して，1 月では 2 倍以上，8 月では 4 倍以上になっている。このようなケースでは，VECM による予測は大きく上振れしている。これは，例えば図 5.6 や図 5.7 から見て取れる。一方，このような場面において，パラメトリック Hawkes モデルによる予測は過小評価される傾向が見られるものの，概ね，特に長期的には，VECM より精度が良いことが分かる。予測月が 1 月の場合，先 1 週間や先 2 週間の短期の予測においては VECM の方が優れており，先 3 週間の予測では大きな差はなく，先 4 週間の予測に関しては大きな差がみられる。予測月が 8 月の場合には，短期の予測でも大きな差がみられる。急上昇局面以外のケースにおいて VECM の精度が最も良いことは，表 5.1 と表 5.2 からわかる。

最後に，以上の考察を踏まえて，本分析で用いたモデルの選択基準を述べる。まず，基本的

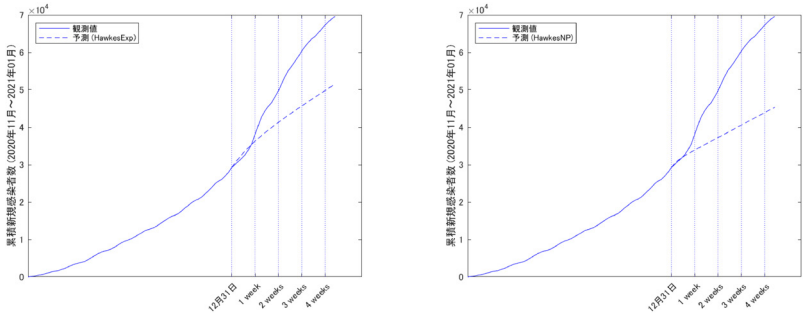

図 5.4 予測月が 1 月のときの累積新規感染者数の予測値と観測値をプロットした。左図がパラメトリック Hawkes モデル，右図がノンパラメトリック Hawkes モデルに関するものである。詳しい図の説明は図 5.2 を参照。

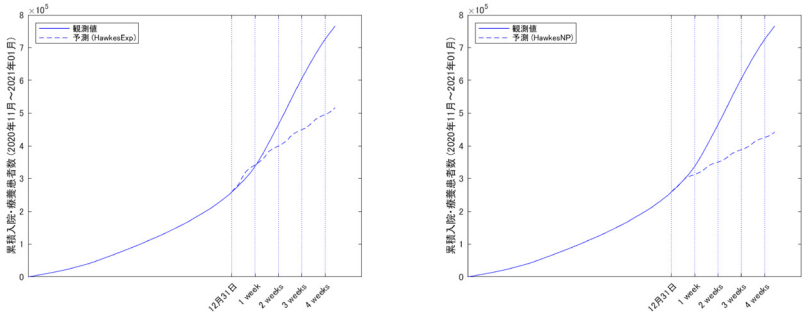

図 5.5 予測月が 1 月のときの累積入院・療養患者数の予測値と観測値をプロットした。左図がパラメトリック Hawkes モデル，右図がノンパラメトリック Hawkes モデルに関するものである。詳しい図の説明は図 5.2 を参照。

に VECM で先 1 ヵ月の予測を行う。次に，本分析における予測月が 1 月や 8 月の場合のように，VECM による先 1 ヵ月の予測が大きく上振れしているときは，検定の結果を踏まえて，パラメトリック Hawkes モデル，またはノンパラメトリック Hawkes モデルを選択することが望ましいと考えらえる。だたし，先 1 ヵ月に関する予測は Hawkes 過程を用いたモデルが優れているが，VECM による予測の急上昇の度合いにより，先 1 週間や先 2 週間の予測については VECM のほうが良いことがある。本分析で用いたデータでは 1 月に該当する。ここでは VECM による累積新規感染者数の予測が前 2 ヵ月の累積の 8 倍程度となっていた。一方，8 月の場合は 15 倍程度となっていた。どの程度の急上昇度合いが境界となるかは，本分析で用いたデータだけでは明らかではない。また，ここでは急上昇局面の予測モデルとして，先行研究に倣って標準的な Hawkes モデルを用いたが，その他のより精度が良いモデルを用いても，そのモデルが急上昇局面以外の局面において VECM 以上の予測精度がない限り，同様の議論が成り立つことが推測される。

Ⅵ　おわりに

本論文では，累積新規感染者数と累積入院・療養患者数の先 1 ヵ月に関する予測について，3 つのモデルを比較した。前者の予測については，カーネル関数が指数型の Hawkes 過程，カーネル関数がヒストグラム型の Hawkes 過程，VECM を用いた。後者の予測については，日別の新規感染

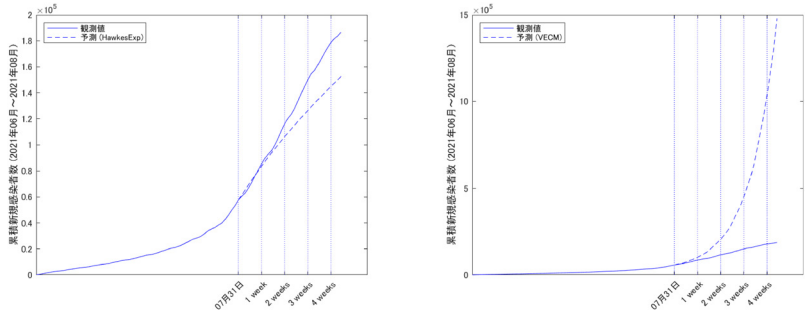

図 5.6　予測月が 8 月のときの累積新規感染者数の予測値と観測値をプロットした。左図がパラメトリック Hawkes モデル，右図が VECM に関するものである。詳しい図の説明は図 5.2 を参照。

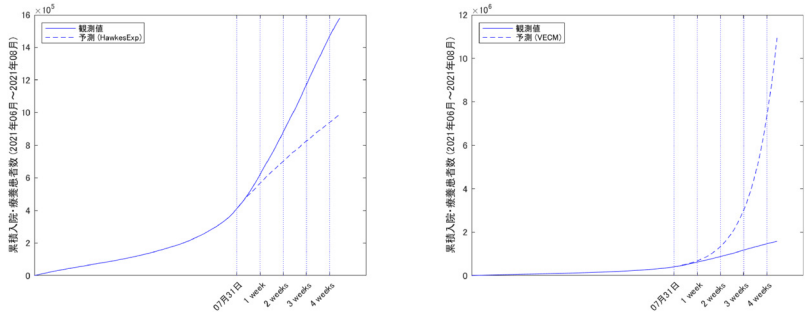

図 5.7　予測月が 8 月のときの累積入院・療養患者数の予測値と観測値をプロットした。左図がパラメトリック Hawkes モデル，右図が VECM に関するものである。詳しい図の説明は図 5.2 を参照。

者数と入院・療養患者数の対数差分系列がを I (1) であることを仮定して，Hawkes 過程を使ったモデルにおいては，前者の予測と共和分関係を用いて予測を行った。このような方法を選んだ理由は，累積入院・療養患者数に直接 Hawkes 過程をフィットさせるための十分な先行研究がないためである。累積入院・療養患者数を共和分関係を用いずに Hawkes 過程だけで予測した場合の議論は，7.2 小節にある。結論から言えば，Hawkes 過程を用いたモデルを選択するのが望ましいと考えられる予測月に関しては，共和分関係を用いた場合のほうが予測精度が良いことがわかった。以上の分析により，以下の 3 つの結果を得た。1 つ目は，少なくとも本分析で用いたデータと Hawkes 過程に関する検定においては，指数型のカーネル関数を持つパラメトリックな Hawkes 過程は，ヒストグラム型のカーネル関数を持つノンパラメトリックな Hawkes 過程と同程度のフィットの良さを示しており，予測精度はより高いことである。ただし，感染症のデータについてのカーネル関数の選択は，Park et al.［2022］で注意されているように十分な経験的知見がないため，慎重に行わなければならない。2 つ目は，累積入院・療養患者数に共和分関係を利用することに一定の有効性が確認されたことである。これは，累積新規感染者数と累積入院・療養患者数の予測に関して，Hawkes 過程を用いたモデルと VECM の間に大きな差がみられなかったことからわかる。なぜなら，2 つの累積数のグラフの形状に大きな差がないことから，適切な予測モデルを用いれば，予測精度に大きな差が出ないことが推測されるからである。3 つ目は，3 つのモデル間の選択基準

である。基本的に VECM を用いて予測を行い，1 ヵ月先の急上昇を予測する場合のみ，Hawkes 過程を用いたモデルにより予測を行うというものである。ただし，先 1 ヵ月の予測ではいづれの場合も Hawkes 過程を用いたモデルが優れていたが，VECM による予測の急上昇度合いにより，先 1 週間や先 2 週間の予測に関しては VECM のほうが優れている場合がある。この急上昇度合いの境界については本研究で用いたデータだけでは明らかとはならなかった。本研究で用いたデータでは，累積新規感染者数について前 2 ヵ月間の累積に対して 8 倍程度の急上昇を予測した場合は，先 1 週間や先 2 週間の予測に関しては VECM が優れており，先 3 週間や先 4 週間の予測については同程度か Hawkes 過程を用いたモデルが優れていることが分かった。また，累積新規感染者数について前 2 ヵ月間の累積に対して 15 倍程度の急上昇を予測した場合は，先 1 から 4 週間のいづれの予測についても Hawkes 過程を用いたモデルが優れていることがわかった。さらに，急上昇局面の予測に用いるモデルは，その他の局面において VECM 以上の精度がない限りでは，Hawkes 過程によるもの以外のモデルを用いた場合でも同様の議論が成り立つと考えられる。

本論文では，Park et al.［2022］に倣い，離散的なデータに恣意的な変形を加えることで，標準的な Hawkes 過程にフィットできるようにした。この恣意性をなくすために，茅根・白石［2021］で用いられている Kirchner［2017］による手法のような，離散的なデータを直接フィットさせる手法を用いることが今後の課題である。

謝　辞

本研究の遂行にあたり，指導教官として終始適切な助言を頂いた江上雅彦教授に，深く感謝の意を表します。また，研究室のメンバーには常に的確な指摘を頂き，感謝しております。

参考文献

Fox, E. W., F. P. Schoenberg, J. S. Gordon, “Spatially Inhomogeneous Background Rate Estimates and Uncertainty Quantification for Nonparametric Hawkes Point Process Models of Earthquake Occurrences,” *The Annals of Applied Statistics*, 10(3), 2016, pp. 1725–1756.

Hawkes, A. G. “Spectra of Some Self-exciting and Mutually Exciting Point Processes,” *Biometrika*, 58(1), 1971, pp. 83–90.

Johansen, S. “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models,” *Econometrica*, 59(6), 1991, pp. 1551–1580.

茅根脩司・白石　博「Hawkes 過程における 2 つの推定手法の比較と実データ解析への応用」『統計数理』69(2), 2021，181–207 ページ。

Kirchner, M. “An estimation Procedure for the Hawkes process,” *Quantitative Finance*, 17(4), 2017, pp. 571–595.

厚生労働省「入院治療等を要する者等推移」2022，（https://covid19.mhlw.go.jp/public/opendata/requiring_inpatient_care_etc_daily.csv，2022 年 10 月 18 日アクセス）

厚生労働省「新規陽性者数の推移（日別）」2022，（https://covid19.mhlw.go.jp/public/opendata/newly_confirmed_cases_daily.csv，2022 年 7 月 20 日アクセス）

Lütkepohl, H. *New Introduction to Multiple Time Series Analysis*, 2005, Springer.

Marsan, D, O. Lengliné, “Extending Earthquakes’s Reach through Cascading,” *Science*, 319(5866), 2008, pp. 1076–1079.

Marsan, D, O. Lengliné, “A New Estimation of the Decay of Aftershock Density with Distance to the Mainshock,” *Journal of Geophysical Research: Solid Earth*, 115(B9) 2010.

Nguyen, H. M., P. J. Turk, A. D. McWilliams, “Forecasting Covid-19 Hospital Census: A Multivariate Time-series Model Based on Local Infection Incidence,” *JMIR Public Health and Surveillance*, 7(8), 2021, e28195

Ogata, Y. 1987, “The Asymptotic Behaviour of Maximum Likelihood Estimators for Stationary Point Processes,” *Annals of the Institute of Statistical Mathematics*, 30(1), 1987, pp. 243–261.

沖本竜義『経済・ファイナンスデータの計量時系列分析』2013，朝倉書店。

近江崇宏・野村俊一 『点過程の時系列解析』2021，共立出版。

Park, J., A. W. Chaffee, R. J. Harrigan, Schoenberg F.P “A Non-parametric Hawkes Model of the Spread of Ebola in West Africa,” *Jounal of Applied Statistics*, 49(3), 2022, pp. 621–637.

Ⅷ　付録

7.1　図や表の補足

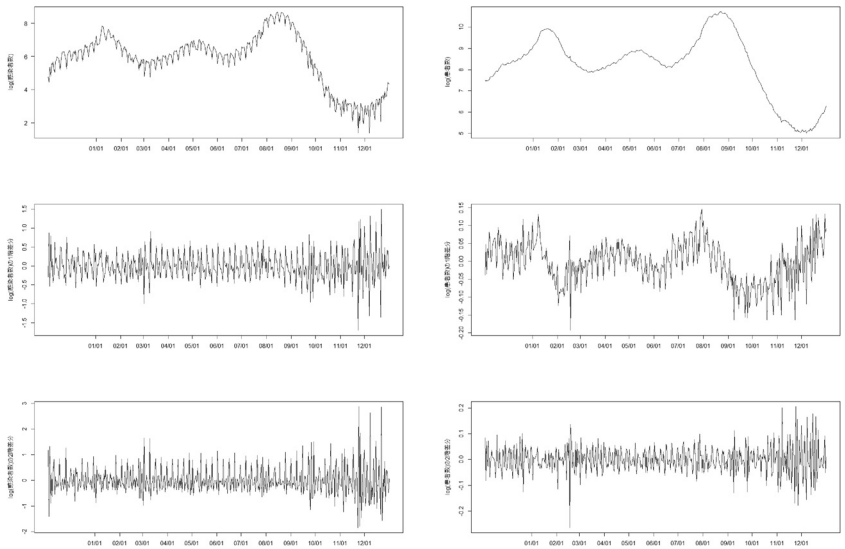

図 7.1　左，右の列は各々新規感染者数（日別），入院・療養患者数（日別の対数系列，その 1 階差分系列，その 2 階差分系列をプロットしたものである。1 階差分系列のグラフにおける 01/01 の目盛りでの値は，1 月 1 日と 12 月 31 日の差分である。2 階差分系列のグラフにおける 01/01 の目盛りでの値は，1 月 1 日と 12 月 30 日の差分である。

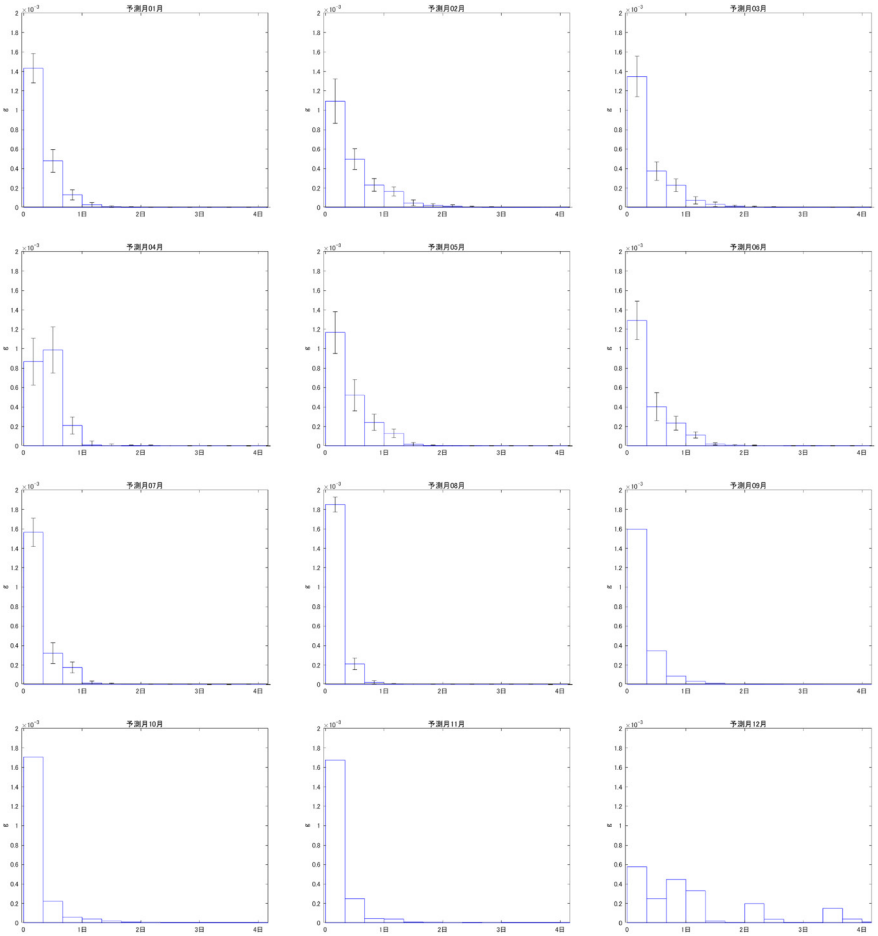

図 7.2　ノンパラメトリックモデルのパラメータ（g）推定値とその標準誤差を示した図である。標準誤差は Fox, Schoenberg, and Gordon［2016］の ALGORITHM 3 に従って求めた。ここでは計算時間の都合上，予測月が 1 月から 8 月までの場合のみ掲載している。

7.2　累積入院・療養患者数に直接 Hawkes 過程をフィットさせた場合

累積入院・療養患者数を共和分関係を用いずに Hawkes 過程だけで予測した結果を以下に示す。データは 10 の位を四捨五入して 100 で割ったものを用い，1 日は分単位に区切った。予測月が 2 月の場合には，検定 1・2 の p 値はそれぞれ 0.1424・0.8808 であった。予測月が 9 月の場合には，検定 1・2 の p 値はそれぞれ 0.4060・0.7936 であった。いずれの場合も有意水準 5% で帰無仮説が棄却されなかった。表 7.1 は予測月が 1 月と 8 月の場合ついての RMSE を掲載したものである。これと表 5.2 を比較すると，共和分関係を用いた場合のほうが予測精度が良いことがわかる。また，こちらの方法ではデータを比較的大きく丸めているため，RMSE の標準誤差の観点からも共和分関係を用いた場合のほうが良いことがわかる。

表 7.1　共和分関係を用いずに，パラメトリック Hawkes 過程だけを用いて，累積入院・療養患者数の 1 月と 8 月に関する予測を行った結果である。各値は RMSE を表している。各値は小数点第 1 位を四捨五入したものである。1 week，2 weeks，3 weeks，4 weeks という項目は各々，先 1 週間予測，先 2 週間予測，先 3 週間予測，先 4 週間予測を表している。

予測月＼予測期間	1 week	2 weeks	3 weeks	4 weeks
1 月	22172	129941	362586	703087
8 月	48125	216347	564303	1130030

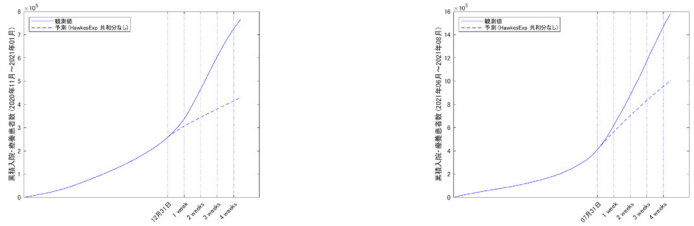

図 7.3　予測月が 1 月と 8 月のときの累積入院・療養患者数の予測値と観測値をプロットした。左図が予測月が 1 月の場合，右図が予測月が 8 月の場合である。