

世代間・地域間の時系列相互相関に着目した COVID-19の分析

村上 大輔[†]・松井 知子[†]

(受付 2021年6月29日；改訂 8月28日；採択 9月17日)

要 旨

新型コロナウイルス感染症(COVID-19)が猛威を振るう昨今、陽性者数を減らすための対策が国・地方自治体で検討されている。そこで本研究では対策立案に向けた第一歩として陽性者数の世代間・地域間の相互相関に着目した分析を行う。分析には vector autoregressive (VAR) モデルを用いる。VAR モデルを用いれば世代間・男女間の相互相関関係が推定できる。しかし同モデルには次の短所がある：(i)パラメータが時間不変であり時々刻々と変化する状況が捉えづらい；(ii)パラメータが多く擬似相関を招きやすい。そこで本研究では(i)に対処するために局所回帰を、(ii)に対処するために非負制約を、それぞれ VAR モデルに導入した。(i)と(ii)を導入することが陽性者数のモデル化の推定精度を大幅に改善することを従来モデルとの比較によって確認した。次に、同モデルを東京都内の陽性者数データに適用した。それにより世間一般である程度予想されていたと思われる次の事柄をデータから確認した：陽性者数の時間変化パターンが、男性は労働世代かどうかで、女性は20代かどうかで異なること；20-50代男性の感染が増えた後に20代男女の感染が増える傾向があること；特に人流が回復した2020年8~10月頃は20-50代男性の影響が顕著であること。次に、東京都市圏を対象とした分析を実施した。それにより都内の労働世代の男性の影響が大きいことなどを確認した。

キーワード：COVID-19, vector autoregressive モデル, 非負制約, 局所回帰, 相互相関。

1. 背景

昨今、新型コロナウイルス感染症(coronavirus disease 2019; COVID-19)への対策が各国・地方自治体で検討されている。COVID-19に関する陽性者数・死者数や病床数などの幅広いデータが公開されており(例：東京都新型コロナウイルス感染症対策サイト：<https://stopcovid19.metro.tokyo.lg.jp/>)、それらを活用した対策の検討が可能となっている。

陽性者数・死者数などの関連データは、無症状感染者や感染からのタイムラグ、報告遅れ・漏れなどの影響を受けるため、それらの活用の際にはデータの背後にある要因やノイズを推定・識別するための数理モデルや統計モデルが用いられてきた。例えば代表的な数理モデルである SIR (Susceptible-Infected-Recovered) モデルやその拡張モデルは、感染者1人あたりの2次感染者数を表す実効再生算数の評価や将来予測などに幅広く用いられてきた(例えば Cooper et al., 2020, Chen et al., 2020)。また数理モデルに基づく統計的推論により、感染リスクの不

[†] 統計数理研究所：〒190-8562 東京都立川市緑町10-3

確実性評価や状態推定が行われてきた(例えば Kobayashi et al., 2020, Koyama et al., 2021)。

一方、要因分析や政策評価などを目的として、回帰や時系列モデルなどの統計モデルも用いられた。例えば Wood (2021)は、イギリスではロックダウン以前から陽性者数が減少に転じていたことを一般化加法モデルを用いて示した。また Murakami and Matsui (2021)は滞留人口密度が一定の閾値以下にならない限り陽性者数は減少に転じないことを同様のモデルを用いて示した。それら統計モデルは、感受・感染から回復・死亡に至るまでのメカニズムを記述するものではなく、解釈性の面で数理モデルに劣る。一方、統計モデルには、人流、世代、地域といった各種要因の影響が柔軟に推定でき、感染症対策の検討に資するような仮説探索がしやすいという利点がある。COVID-19 に対する対策検討は喫緊の課題であることから、本研究では統計モデルに着目することとする。

対策を検討する上で重要な視点に、例えば「移動の多い労働世代は感染しやすい」、「高齢世代は死亡リスクが高い」のような世代差や、「就業率の高い男性は女性に比べて感染しやすい」のような男女差がある。これまでも、我が国を対象として、世代や男女の違いが感染に及ぼす影響が分析されてきた。例えば Mizumoto et al. (2020)は世代毎の発症率や発症件数を分析し、小児の発症率の少なさを指摘した。Griette et al. (2020)は世代毎の異質性を推定した上で、未報告分も含む全陽性者数を世代毎に推計した。また Murakami and Matsui (2021)は労働世代とそれ以外では人口あたり陽性者数が大きく異なることを示した。一方、例えば「親から子に感染する」、「若者から高齢者に感染する」のような世代間・男女間の相互影響関係を明らかとすることもまた、対策を検討する上で重要である。しかしながら、筆者らの知る限り、陽性者数に着目して、そのような世代別・男女別の相互相関関係を時期毎に評価した研究は存在しない。

そこで本研究では、2020年1月から2021年6月までを対象に、世代毎・男女毎の陽性者数の相互相関関係とその時間変化を分析する。分析には、標準的な多変量時系列モデルである vector autoregressive (VAR) モデルを用いる。VAR モデルは複数の時系列間の相互相関関係を捉えられる一方で次の短所がある：(i)パラメータが時間不変であり時々刻々と変化する状況が捉えづらい；(ii)パラメータが多く擬似相関を招きやすい。特に、目まぐるしく状況が変化する COVID-19 の解析に際しては、両方に対処することで、できる限りモデルの精度を高める必要がある。

以上を踏まえ、第2章では(i)と(ii)に対処した VAR モデルを提案する。その上で、第3章では東京都を対象とした分析を実施する。その中で提案モデルの精度も検証する。第4章では神奈川県、埼玉県、千葉県も含む一都三県を対象とした分析をそれぞれ行う。第5章では以上の結果を踏まえた考察を行う。なお、本研究で実施するのはあくまで仮説探索のための相関分析であり、因果関係を評価するものではない点に注意されたい。

2. モデル

世代別・男女別・週別の陽性者数 $y_{g,t}$ (g は世代・男女別グループの添字, t は週の添字) を分析する。陽性者数に対する基礎的な VAR モデルは下式で与えられる：

$$(2.1) \quad y_{g,t} = a_g + \sum_{g'=1}^G \beta_{g,g'} y_{g',t-1} + e_{g,t}, \quad E[e_{g,t}] = 0, \quad \text{Var}[e_{g,t}] = \sigma^2,$$

a_g は定数項, $\beta_{g,g'}$ は $y_{g',t-1}$ と $y_{g,t}$ の相互相関の強さを表すパラメータ, σ^2 は分散パラメータである。上式は最小二乗法や最尤法で推定できる。

VAR モデルには前章で述べたような短所(i)や(ii)があるが、それぞれ対処法が検討されている。(i)に関しては例えば、係数 $a_g, \beta_{g,g'}$ が時変関数に従うことを仮定する VAR モデル(例：

Christopoulos and León-Ledesma, 2008; Bringmann et al., 2018) や、時点の近い標本により大きな重みを与えた上で推定を行う局所 VAR モデル (例えば Casas and Fernandez-Casal, 2019) がある。今回のような比較的小さな標本の場合は後者の精度がより良いとの指摘があることから (Haslbeck et al., 2021), 本研究では局所 VAR モデルを用いることとする。

(ii) 擬似相関への対処に関しても例えば、相互相関と共和分を識別しようという vector error correction モデルや、縮小推定を用いる方法などが提案されてきた (例えば Gefang, 2014)。一方、本研究に着目すると、「世代 A の陽性者が増えるほど世代 B の陽性者は減少する」のような負の相互相関は考えづらいため、相互相関係数 $\beta_{g,g'}$ は非負と仮定することが自然である。また、相互相関が存在しない場合 ($\beta_{g,1} = \beta_{g,2} = \dots = \beta_{g,G} = 0$) の陽性者数を表す係数 a_g もまた非負と仮定することが自然である。一般に、回帰係数に非負制約を導入すると、符号の反転などに伴う擬似相関が自然に抑制され、安定的で疎な推定結果が得られやすいことが知られている (Kim and Park, 2007)。そこで本研究でも、係数に非負制約を導入することで擬似相関を抑制する。

以上を踏まえ、本研究では局所 VAR モデルの係数に非負制約を課した下式で解析を行う：

$$(2.2) \quad y_{g,t} = a_{g,t} + \sum_{g'=1}^G \beta_{g,g',t-1} y_{g',t-1} + e_{g,t}, \quad E[e_{g,t}] = 0, \quad \text{Var}[e_{g,t}] = \sigma_t^2$$

$$s.t. \quad a_{g,t} \geq 0, \quad \beta_{g,g',t-1} \geq 0.$$

週次の定数項 $a_{g,t}$ と係数 $\beta_{g,g',t-1}$ は局所回帰で推定する。具体的には、 $y_{g,t}$ をモデル化するために、指数カーネル $w_{t,t'} = \exp(-|t - t'|/h)$ で時点の近い標本により大きな重みを与えることで、週 t に対する当てはまりが良くなるように $a_{g,t}$ と $\beta_{g,g',t-1}$ を推定する。同様の手順を週毎に全週について行うことで、各週の係数を推定した。時変係数 $a_{g,t}$ は SIR モデルなどでも考慮される時間経過による陽性者数の増減(時間トレンド)を捉える係数であり、共和分などの影響を吸収することが期待できる。 $\beta_{g,g',t-1}$ は相互相関関係の時間変化を捉えることが期待できる。

予備的検討の結果、指数カーネルのバンド幅 h は 4 とした。これは、重みの 95% が消失する時差である有効バンド幅が $4\sqrt{3} \approx 6.93$ 週であり、概ね前後 50 日以内のデータを用いて (2.2) 式を推定することを意味する。この推定は、重み $w_{t,t'}$ を用いた局所加重最小二乗推定を非負制約の下で行うことで実施する。

(2.2) 式を用いると週毎のインパルス応答が評価できる。インパルス応答は、世代 g の誤差項 $e_{g,t}$ を 1 単位変化させた時の l 週先の陽性者数 $y_{g,t+l}$ の変化量 $IR_{g,t}(l) = \partial y_{g,t+l} / \partial e_{g,t}$ で定義される。週毎に係数 $\beta_{g,g',t-1}$ が変化することに注意しながら通常の VAR モデルと同様に $IR_{g,t}(l)$ を求めると、各世代に対するインパルス応答 $\mathbf{I}_{g,t}(l) = [IR_{1,t}(l), \dots, IR_{G,t}(l)]'$ は下式となる：

$$(2.3) \quad \mathbf{I}_{g,t}(l) = \mathbf{B}_{t+l}^l \mathbf{u}_g$$

\mathbf{B}_{t+l} は $\beta_{g,g',t+l}$ を (g, g') 要素に持つ行列、 \mathbf{u}_g は第 g 要素を 1、それ以外を 0 とするベクトルである。(2.3) 式を用いることで、 t 週目に世代 g の陽性者が一人増えることが、 l 週先の世代別陽性者数に及ぼす影響が推定できる。次章では世代 g の影響力の指標として、 L 週先までのインパルス応答を足し合わせた累積値 $\mathbf{I}_{t,g} = \sum_{l=1}^L \mathbf{I}_{t,g}(l)$ を用いる。

以降、3 章では東京都を対象とした分析、4 章では首都圏を対象とした分析に、それぞれ応用する。

3. 東京都を対象とした分析

3.1 データ

本章では、JX Press 社 (<https://jxpress.net/>) 提供の都内陽性者数データ (2020 年 1 月 18 日～

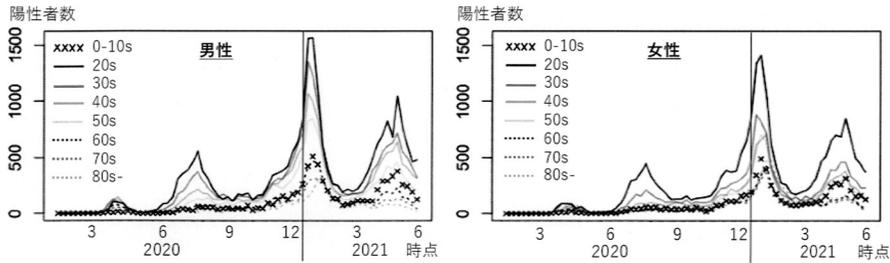


図 1. 東京都における男女別・世代別の陽性者数(左：男性，右：女性).

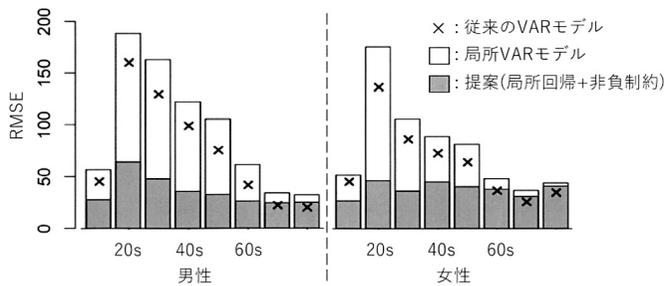


図 2. 世代別・男女別の交差検証による RMSE の評価結果.

2021年6月11日)を用いる. 同データは陽性者毎に年齢・男女別・判定日をまとめたものである. 同データを集計して得られる16グループ(男女8世代ずつ, 0–19歳(10s), 20–29歳(20s), ..., 70–79歳(70s), 80歳以上(80s)) \times 73週からなる多変量時系列データに前章のVARモデルを適用し, 分析を行った. 都内の男女別・世代別の陽性者数は図1の通りである. この図から, 特に第2波(2020年8月前後), 第3波(2021年1月前後), 第4波(2021年5月前後)においては, 20–50代男性の感染が特に多いことや, 女性の中では20代の感染が突出して多いことなどが確認できる.

3.2 提案モデルの精度検証

分析に先立って, 提案モデル(局所回帰+非負制約)の精度を, 従来のVARモデル(式(2.1)), ならびに局所VARモデル(局所回帰のみ)と比較する. 同比較は交差検証によって行う. この交差検証では「 t 週目以外のデータを用いて t 週目のデータを予測する」という操作を全週について繰り返すことで世代毎に評価されるRoot Mean Squared Error(RMSE)を比較する:

$$(3.1) \quad RMSE_g = \sqrt{\frac{1}{T} \sum_{t=2}^T (\hat{y}_{g,t} - y_{g,t})^2}.$$

$\hat{y}_{g,t}$ は陽性者数の予測値である.

RMSEの評価結果は図2に整理した通りである. 従来のVARモデルの結果を見ると, そのRMSEは労働世代で大きくなり, 特に20代のRMSEは大きい. 世代間の相互相関が時間不変であることを仮定する基礎的なVARモデルでは, 陽性者数の特に大きな労働世代についての精度の面で難があるとの示唆を得た. しかしながら相互相関を時間可変とする局所VARモデルの精度は従来型よりもむしろ悪化した. これはパラメータが多すぎて効果が識別できなかつ

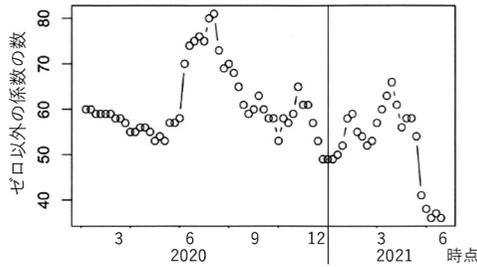


図 3. 提案モデルでゼロとならなかった係数の数. なお係数は全部で 272 個である.

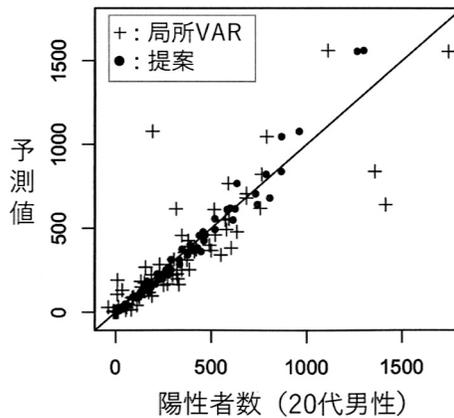


図 4. 陽性者数とその予測値の比較(提案モデルと局所 VAR モデル).

たための可能性がある。実際、各局所 VAR モデルでは陽性者数(世代別・男女別の 16 変数)を推定するために、定数項を含めて 272 もの係数を推定する必要があるため、その識別は困難を極める。その一方で、提案モデルの RMSE は世代を通して小さくなった。これは係数に対する非負制約を導入することで、多くの係数が 0 となったためと考えられる。実際、図 3 に示す通り、0 とならなかった係数の数は各週で概ね 60 程度である。このことは 1 世代・男女別あたり平均 3.75 個の係数で陽性者を推定する疎なモデルが得られたことを意味する。観測値と予測値を比較した図 4 から提案手法の精度の良さが確認できる。非負制約を導入することで時点毎の VAR モデルが安定的に推定できるという知見は、陽性者数や重症者数のような非負データの相互相関を分析する上で極めて有用といえよう。

3.3 結果

本節では、提案モデルのパラメータ推定結果を説明する。図 5 は式(2.2)の係数 $a_{g,t}$ の推定結果である。この時変係数は相互相関 ($\sum_{g'=1}^G \beta_{g,g',t-1} y_{g',t-1}$) で説明されないトレンドを捉える。図によると、20-50 代男性(左図実線)の係数値が第 2~4 波で急増しているのに対し、その他の世代(左図点線)ではそのような傾向が弱い。また 20 代女性(右図黒実線)の係数もまた波毎に急増している。この結果は、陽性者急増の時期においては、特に 20-50 代男性と 20 代女性への対策が必要という可能性を示唆するものである。

図 6 は式(2.2)の係数 $\beta_{g,g',t-1}$ の推定結果を 2 ヶ月おきに視覚化したものである。16 グループ(男性 8 世代+女性 8 世代)のうち、男性についての結果を左列に、女性についての結果を右

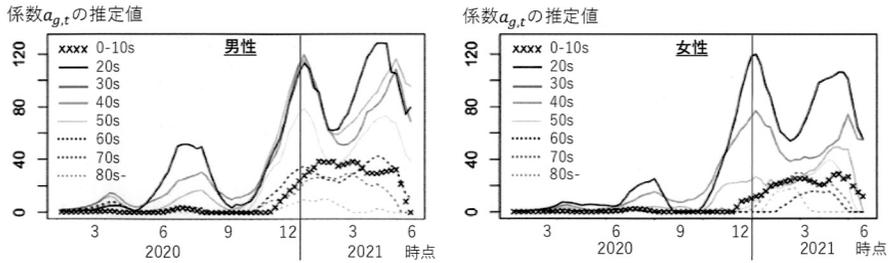


図5. 係数 $a_{g,t}$ の世代別・男女別の推定結果(左:男性, 右:女性).

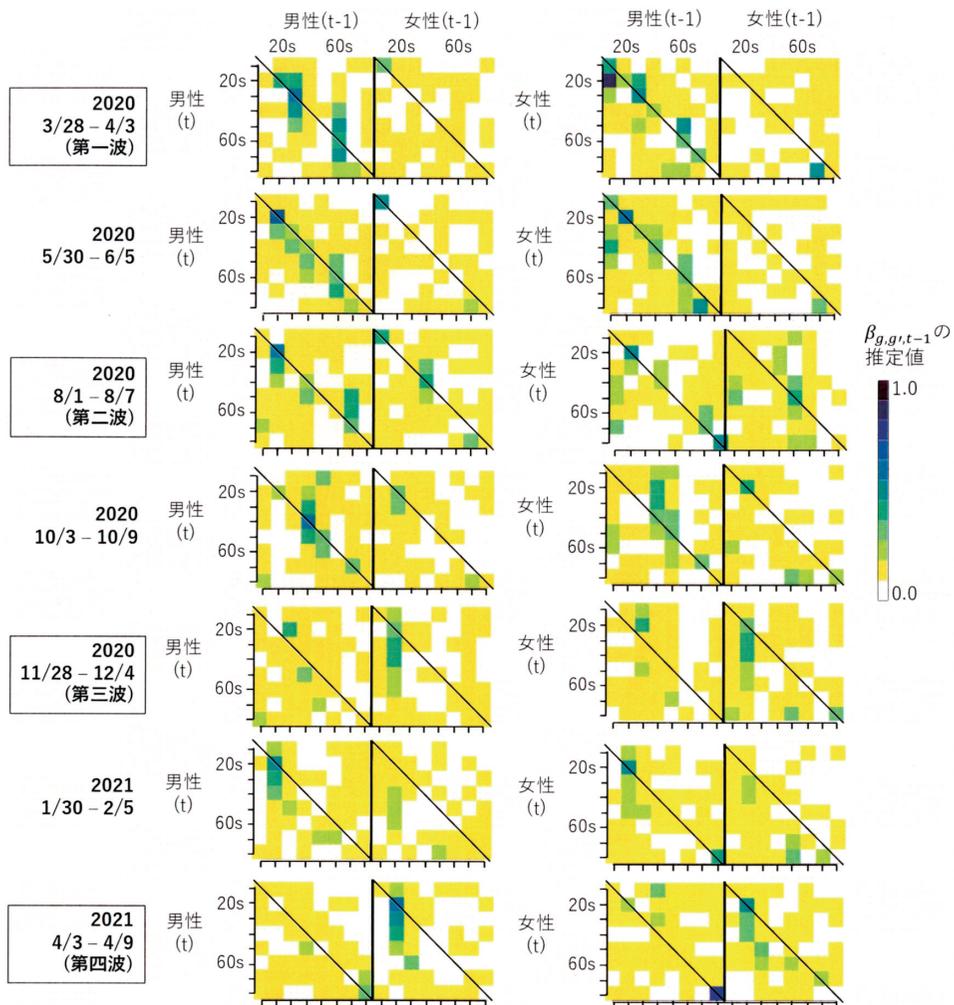


図6. 係数 $\beta_{g,g',t-1}$ の推定結果. 男性についての結果を左列に, 女性についての結果を右列に整理した. 左列は男性($t-1$ 週目; 8世代別)・女性($t-1$ 週目; 8世代別)から男性(t 週目; 8世代別)への影響の推定結果であり, 20sは20代, 60sは60代をあらわす. 右列は女性についての同様の推定結果である.

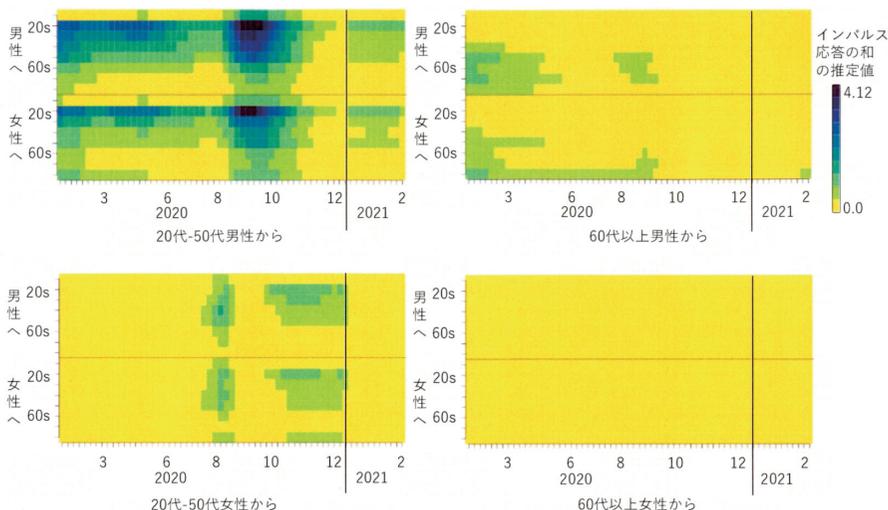


図 7. 男女別・世代別のインパルス応答(12 週後まで)の和の推定値. 横軸は時点, 縦軸は世代・男女別を表す. 色が濃いほど, その後 12 週先までに及ぼす影響が大きいことを意味する(左上: 20-50 代男性からの影響, 右上: 60 代以上男性からの影響, 左下: 20-50 代女性からの影響, 右下: 60 代以上女性からの影響).

列に示す. この図から 2020 年 10 月頃までは男性 ($t - 1$ 週目) から同世代の男女 (t 週目) への影響が強いことが, 2020 年 12 月以降は 20 代女性からの影響が強まる傾向があることが, それぞれ確認できる. 第 3 波の前と後では世代間の相互相関傾向が変化した可能性がある.

次にインパルス応答を評価する. 同応答の大きさは, 推定されたモデルの誤差項に 1 単位のショックを与えた時の $y_{g,t+l}$ の変化量を 12 週先 ($l = 12$) まで求めて足し合わせた累積値で評価する. 図 7 は, 20-50 代の 4 世代からのインパルス応答の平均値, ならびに 60 代以上の 3 世代からの応答の平均値を男女別にそれぞれ週毎に評価・視覚化したものである. 図中で黒いほど, その後 3 ヶ月 (12 週) 先周の影響が大きい. この図から, 20-50 代男性の感染が増えると 20 代男女の感染が増える傾向が全期間を通してみられた. また, 20-50 代男性が 1 人感染した時のその後の陽性者増加 ($IR_{g,t}(l) = \partial y_{g,t+l} / \partial e_{g,t}$) は 2020 年 8~10 月に特に大きくなった. 最も $IR_{g,t}(l)$ が大きくなったのは 9 月 5 日から 9 月 11 日の間であり, この期間に 20-50 代男性の陽性者が 1 人増えた場合, その後 3 ヶ月間で 4.22 人の陽性者数の増加があると推定された(週あたり 0.35 人増). Apple mobility trend (<https://covid19.apple.com/mobility>) によると同期時は公共交通機関ならびに自動車による移動が特に活発な時期である(図 8). 20-50 代男性の移動や交流が増えたことが, 2020 年 12 前後の第 3 波が深刻化した一因であったかもしれない. なお, 60 代以上男性, 20-50 代女性, 60 代以上女性の $IR_{g,t}(l)$ はいずれも小さく, 陽性者の 1 人増加に伴うその後 3 ヶ月間での陽性者数の増分は最大でも 1.17, 1.29, 0.31 であった.

図 9 は 20 代-50 代男性からのインパルス応答の統計的な有意性の評価結果である. なお 60 代以上の男性からの影響, および女性からの影響は期間を通して有意とならなかったのを割愛した. ここでは時系列データのような独立でない標本に対するリサンプリング法であるブロック・ノンパラメトリック・ブートストラップ法を用いたリサンプリングを 200 回反復し(ブロックサイズ: 8), それぞれから得られたインパルス応答を用いて各標準誤差を評価して t 値を評価した. 時点毎・世代毎の多重検定となることから Benjamini-Hochberg 法を用いた補正を行

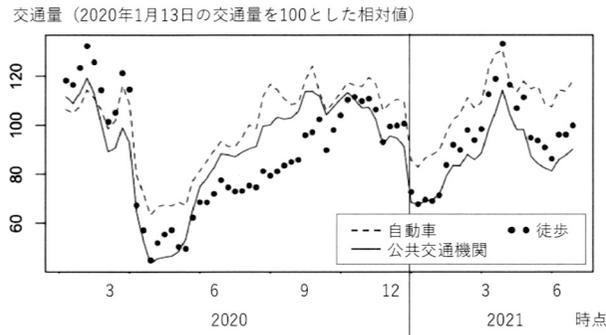


図 8. 東京都内の週別の移動量. Apple mobility trend で提供されている日別データを集計して作成した 2020 年 1 月 13 日を 100 とした相対値.

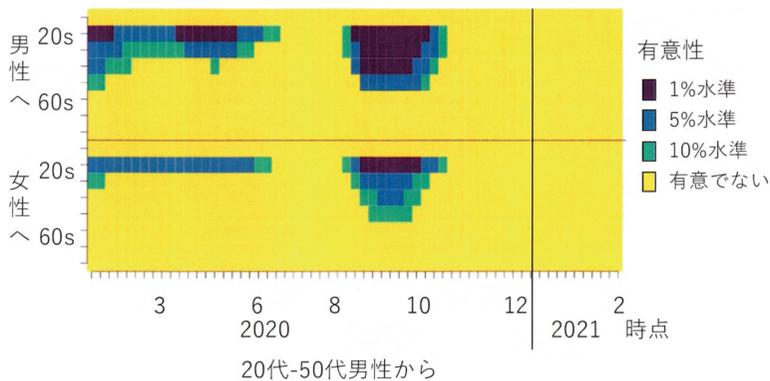


図 9. 男女別・世代別のインパルス応答(12 週後まで)の和の推定値(20-50 代男性からの影響)の統計的有意性.

うことで統計的な有意性を評価した. 図 9 によれば 2020 年 10 月頃までは「20-50 代男性の感染者が増えた後に 20 代で感染が増える」という相関傾向が有意であった. また人流が回復した 8-10 月には「20-50 代男性の感染者が増えた後に同世代男女の感染が増える」という相関傾向が有意となった. ここでの分析は相関に着目したものであり, 背後にある因果関係を明らかにするためには更なる解析が不可欠なものの, この結果は 20-50 代男性の陽性者をできる限り抑えることが必要であることを示唆するものである.

参考までに非負制約を用いない局所 VAR モデルのインパルス応答の 3 ヶ月先までの和を同様に推定して図 10 に視覚化した. この図から, 例えば 60 代以上女性からの影響が大きい点や, 負の相互相関が散見される点, 推定値の分散が大きい点など, 全体として解釈困難な結果となっていることがわかる. 本研究で導入した非負制約が解釈しやすい結果を得る上で重要であることを再確認した.

4. 首都圏を対象とした分析

4.1 データ

前章に引き続き, JX Press 社提供の週別陽性者数データ(2020 年 1 月 18 日~2021 年 6 月 11 日の 73 週)を使用する. 対象地域は東京都, 神奈川県, 埼玉県, 千葉県であり都道府県間の相

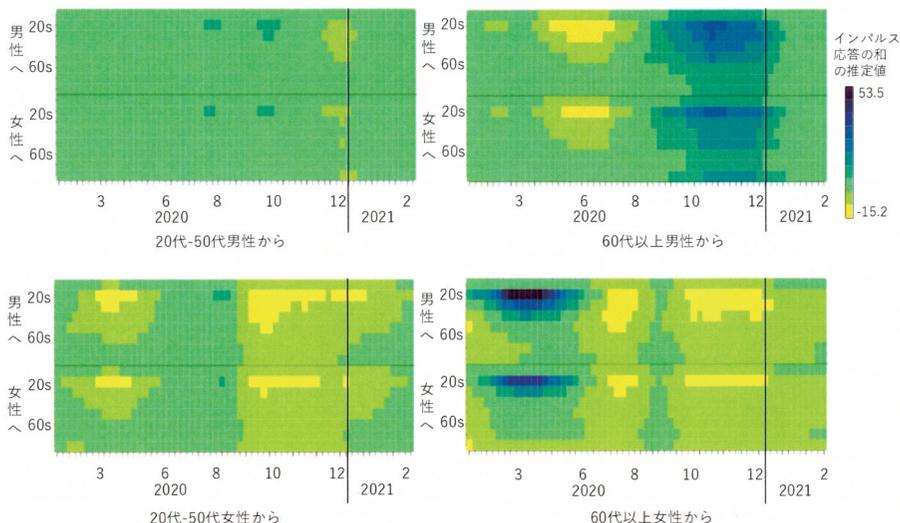


図 10. 非負制約なしの VAR モデルから得られたインパルス応答(12 週後まで)の和の推定値 (左上：20-50 代男性からの影響, 右上：60 代以上男性からの影響, 左下：20-50 代女性からの影響, 右下：60 代以上女性からの影響)。

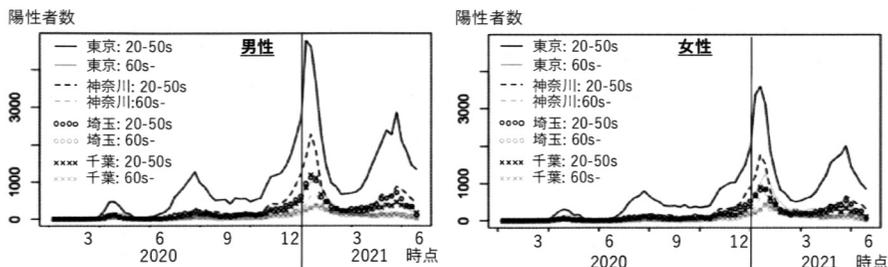


図 11. 1 都 3 県における男女別・世代別の陽性者数(左：男性, 右：女性)。

相互相関も考慮した分析を行う。都道府県別・世代別の陽性者数は図 11 に示す通りである。この図から、都内在住の 20-50 代男女の感染が特に多いことが確認できる。

予備的検討の結果、前章と同様の世代に 8 区分するとパラメータ数が膨大となり ((世代数 $8 \times$ 男女別 $2 \times$ 地区数 4)² = 4,096 個), パラメータが安定的に推定できないと判断した。そこで本章では、1 都 3 県毎に 20-50 代男性, 20-50 代女性, 60 代以上男性, 60 代以上女性の 4 グループに分けることにした。これは次の理由による：労働世代(20-50 代)の影響を把握することは在宅勤務などの労働世代への対策を検討する上で重要なため；前章で示した通り男性からの影響は 20-50 代男性かそれ以外かで大きく異なっていたため。16 グループ(4 グループ \times 4 都県) \times 73 週からなる多変量時系列データに前章の VAR モデルを適用した。

4.2 結果

相互相関で説明されない時間トレンドを捉える係数である $a_{g,t}$ の推定結果を図 12 にプロットした。この図から、都内在住の 20-50 代男女は第 2~4 波で係数値が上下しているのに対し、他県ではそのような傾向が弱いことがわかる。東京都の特異性が、陽性者数の時間変化パター

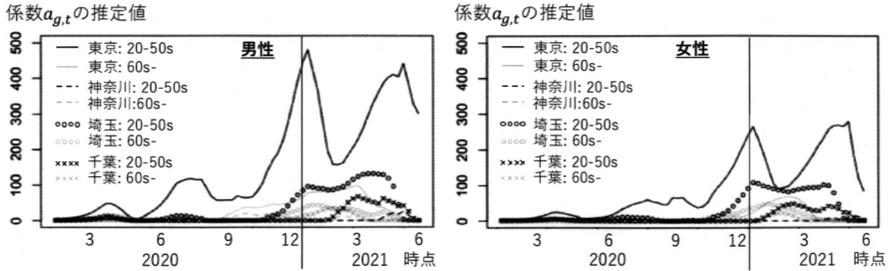


図 12. 係数 $a_{g,t}$ の都道府県別・世代別の推定結果(左:男性, 右:女性).

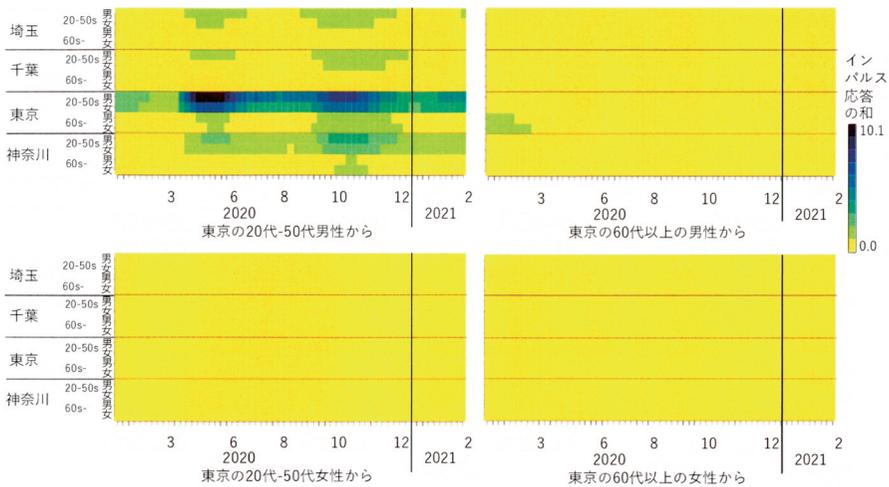


図 13. 東京都からのインパルス応答(12 週後まで)の和の推定値. 色が濃いほど, その後 12 週先までに及ぼす影響が大きいことを意味する(左上:20-50 代男性からの影響, 右上:60 代以上男性からの影響, 左下:20-50 代女性からの影響, 右下:60 代以上女性からの影響).

ンの観点で確認された。

図 13 は東京都内からのインパルス応答の大きさを図 7 と同様に図示したものである。この図から、「期間を通して、都内 20-50 代男性の陽性者増加が都内同世代の陽性者数増加に寄与する傾向」が確認できる。特に 5 月 9 日から 15 日の週はインパルス応答が最大であり、この週に都内 20-50 代男性の陽性者が 1 人増えると、その後 3 ヶ月間で都内陽性者が 10.54 人増えると推定された(週あたり 0.88 人)。また 10 月 17 日から 23 日の週もまたインパルス応答のピークとなっており、陽性者が 1 人増加することがその後 3 ヶ月間の 8.03 人の陽性者数増加につながると推定された。また「第 3 波の最中では、都内 20-50 代男性の陽性者の増加が、他県の陽性者の増加にも寄与すること」が確認された。一方、60 代以上男性や女性からのインパルス応答は極めて小さくなった。図 7 の結果も踏まえると、都内では 20-50 代男性の感染をできるだけ減らすことが重要といえよう。

最後に、埼玉県、千葉県、神奈川県 の 20-50 代男性からのインパルス応答の大きさを図 14 に視覚化した。この図から、東京都以外の県からのインパルス応答は比較的小さいことが確認できる。なお、千葉県の 2020 年初頭に値が大きくなっているが、同時期は陽性者数が極めて少

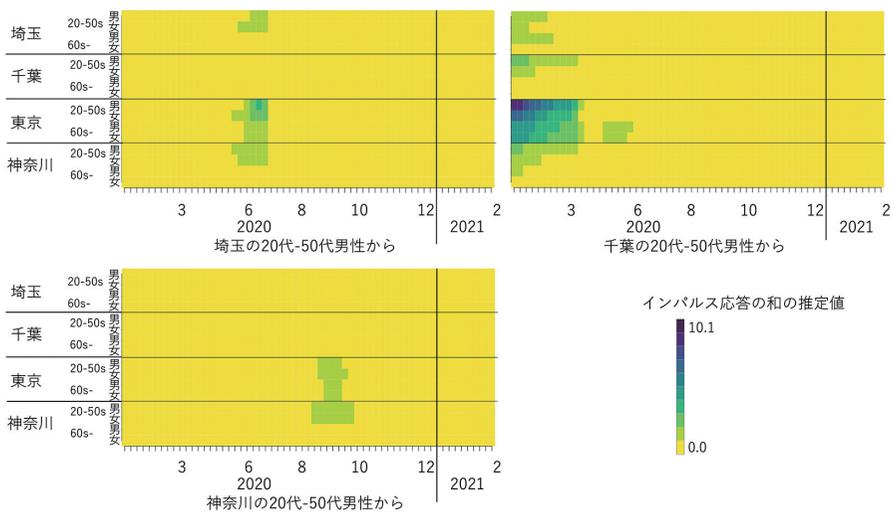


図 14. その他県に居住する 20-50 代男性からのインパルス応答(12 週後まで)の和の推定値。色が濃いほど、その後 12 週先までに及ぼす影響が大きいことを意味する(左上：埼玉県からの影響，右上：千葉県からの影響，左下：神奈川県からの影響)。

ないため推定値の信頼性は低い。

5. まとめ・考察

本研究では世代・男女別間の相互相関に着目した COVID-19 陽性者数の分析を、非負制約付きの局所 VAR モデルを用いて行った。それにより、例えば「都内では 20-50 代男性が感染した後、同世代・他世代での陽性者数が増加する傾向」、「同傾向が、人流が増加した 2020 年 8~10 月頃に顕著である点」、「20-50 代男性と 20 代女性の陽性者数(の時間トレンド $a_{g,t}$) が波毎に大きく増加する点」、「東京都で感染者が増えた後に他県の陽性者が増える傾向」など、一般に予想されていた事柄を、データ解析を通してある程度確認することができた。2 点目に関しては就業状況に関係している可能性がある。例えば女性の 20 代と 30 代以上で大きな傾向の差があるのは、30 歳以上では子供のいる世帯数が 20 代より多くなることから、在宅勤務を行う女性の数がより多く、ウイルスへの曝露が 20 代より少なかったためと考えられる。また、男性 20-50 代とそれ以外で傾向が異なったのは、労働者の大半が同世代であり曝露が多かったためと考えられる。得られた結果は、世代・男女別なども踏まえた対策を検討するための材料として役立つことが期待できる。

一方で、本研究で用いた統計モデルは感染・潜伏・流行といったメカニズムを考慮しない記述統計的なものである。SIR モデルのような感染メカニズムを考慮したモデルと組み合わせた解析を進めることは、世代間・男女間の相互相関関係を考慮しながら、より解釈性・説得性のある分析を行うための第一歩として重要である。また、本研究では人流や政策の影響がより直接的に影響する陽性者数を分析対象としたが、医療リソースへの影響や社会的関心などを踏まえると、重症者数や死者数も考慮して緊急事態宣言や在宅勤務などの各種政策の効果をより詳細に分析する必要がある。以上に加え、曜日の効果や休日・平日の違いを考慮するためには日次での分析が必要である。また、感染状況に加え、人々の心理状態や経済状況などの、より多様な要因間の相互影響関係を統合的に分析することもまた、COVID-19 の影響をより網羅的に

分析する上で重要な今後の課題である。

謝 辞

本研究は、JST 未来社会創造事業、JPMJMI20B2、ならびに情報・システム研究機構 COVID-19 対応プロジェクトの支援を受けたものである。

参 考 文 献

- Bringmann, L.F., Ferrer, E., Hamaker, E.L., Borsboom, D. and Tuerlinckx, F. (2018). Modeling nonstationary emotion dynamics in dyads using a time-varying vector-autoregressive model, *Multivariate Behavioral Research*, **53**(3), 293–314.
- Casas, I. and Fernandez-Casal, R. (2019). tvReg: Time-varying coefficient linear regression for single and multi-equations in R, *SSRN*, DOI: 10.2139/ssrn.3363526.
- Chen, Y.C., Lu, P.E., Chang, C.S. and Liu, T.H. (2020). A time-dependent SIR model for COVID-19 with undetectable infected persons, *IEEE Transactions on Network Science and Engineering*, **7**(4), 3279–3294.
- Christopoulos, D.K. and Len-Ledesma, M.A. (2008). Testing for Granger (non-) causality in a time-varying coefficient VAR model, *Journal of Forecasting*, **27**(4), 293–303.
- Cooper, I., Mondal, A. and Antonopoulos, C.G. (2020). A SIR model assumption for the spread of COVID-19 in different communities, *Chaos, Solitons & Fractals*, **139**, 110057.
- Gefang, D. (2014). Bayesian doubly adaptive elastic-net Lasso for VAR shrinkage, *International Journal of Forecasting*, **30**(1), 1–11.
- Griette, Q., Magal, P. and Seydi, O. (2020). Unreported cases for age dependent COVID-19 outbreak in Japan, *Biology*, **9**(6), 132.
- Haslbeck, J.M., Bringmann, L.F. and Waldorp, L.J. (2021). A tutorial on estimating time-varying vector autoregressive models, *Multivariate Behavioral Research*, **56**(1), 120–149.
- Kim, H. and Park, H. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis, *Bioinformatics*, **23**(12), 1495–1502.
- Kobayashi, G., Sugawara, S., Tamae, H. and Ozu, T. (2020). Predicting intervention effect for COVID-19 in Japan: State space modeling approach, *BioScience Trends*, **14**(3), 174–181.
- Koyama, S., Horie, T. and Shinomoto, S. (2021). Estimating the time-varying reproduction number of COVID-19 with a state-space method, *PLoS Computational Biology*, **17**(1), e1008679.
- Mizumoto, K., Omori, R. and Nishiura, H. (2020). Age specificity of cases and attack rate of novel coronavirus disease (COVID-19), *MedRxiv*, DOI: 10.1101/2020.03.09.20033142.
- Murakami, D. and Matsui, T. (2021). Improved log-Gaussian approximation for over-dispersed Poisson regression: Application to spatial analysis of COVID-19, *ArXiv*, 2104.13588.
- Wood, S.N. (2021). Inferring UK COVID-19 fatal infection trajectories from daily mortality data: Were infections already in decline before the UK lockdowns?, *Biometrics*, DOI: 10.1111/biom.13462.

Time Series Analysis of Inter-generational and Inter-regional Dependence of COVID-19 Cases

Daisuke Murakami and Tomoko Matsui

The Institute of Statistical Mathematics

The growing severity of the novel COVID-19 coronavirus (SARS-CoV-2) is pushing national and local governments to launch effective countermeasures. To aid in this effort, we performed an analysis of COVID-19 data focusing on the cross-correlation of the number of COVID-19 cases between generations, genders, and prefectures. Although we used a vector autoregressive (VAR) model, the basic VAR model has two disadvantages: (i) the parameters are time-invariant, making it difficult to capture the COVID-19 situation, which is changing rapidly; (ii) many of the parameters in the model can be spuriously correlated. Therefore, we used a local regression modeling approach for (i) and non-negative constraints for (ii) in our VAR model. We used the model to analyze COVID-19 cases in Tokyo. There are three main findings. (1) The temporal patterns of cases differed between the working and not-working generations for men and between people in their 20s and people not in their 20s for women. (2) The spread of the virus among working people may have triggered a rapid increase in the number of infected people in Tokyo during the winter of 2020. (3) The influence of men in their 20s-50s should be particularly pronounced around August-October 2020, when the human flow recovers. Analysis of the number of COVID-19 cases in the Tokyo metropolitan area confirmed that controlling the spread of the infection among the working generation is important.