The Annals of Applied Statistics
2024, Vol. 18, No. 1, 224–245
https://doi.org/10.1214/23-AOAS1787

NONNEGATIVE TENSOR COMPLETION FOR DYNAMIC COUNTERFACTUAL PREDICTION ON COVID-19 PANDEMIC

BY YAOMING ZHEN[1,a] AND JUNHUI WANG[2,b]

[1]*School of Data Science, City University of Hong Kong,* [a]*yzhen8-c@my.cityu.edu.hk*
[2]*Dapartment of Statistics, The Chinese University of Hong Kong,* [b]*junhuiwang@cuhk.edu.hk*

The COVID-19 pandemic has been a worldwide health crisis for the past three years, casting unprecedented challenges for policymakers in different countries and regions. While one country or region can only implement one social mobility restriction policy at a given time, it is of great interest for policy makers to decide whether to elevate or deelevate the restriction policy from time to time. This article proposes a novel nonnegative tensor completion method to predict the potential counterfactual outcomes of multifaceted social mobility restriction policies over time. The proposed method builds upon a low-rank tensor decomposition of the pandemic data, which also explicitly characterizes the ordinal nature of the mobility restriction strength and the smooth trend of the pandemic evolution over time. Its application to the COVID-19 pandemic data reveals some interesting facts regarding the impact of social mobility restriction policy on the spread of the virus. The effectiveness of the proposed method is also supported by its asymptotic estimation consistency and extensive numerical experiments on the synthetic datasets.

1. Introduction. The Coronavirus Disease 2019 (COVID-19) pandemic is caused by a novel coronavirus identified in 2019, called the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). According to the COVID-19 Dashboard at https://coronavirus.jhu.edu/map.html by the Center for Systems Science and Engineering at Johns Hopkins University, the pandemic has quickly spread over 200 countries and regions, with about 435 million confirmed cases and 5.9 million deaths as of March 1, 2022. Social distancing has been proved to be an effective measure to prevent the virus from spreading (Yan et al. (2021); Sinha and Chakraborty (2022); Quick, Dey and Lin (2021); Ye et al. (2021)), yet at the great cost of sacrificing the economy. Many governments have implemented various social mobility restriction policies to enforce social distancing, to a certain extent, so as to reduce the spread of the COVID-19 pandemic. It is thus of great significance for administration units or policymakers to understand the potential outcomes, or counterfactuals, of multifaceted public policies at a particular pandemic stage so that suitable policies can be adopted for the confronting situations.

In literature a great deal of existing works have been developed to investigate the counterfactual causal effects (Höfler (2005); Pearl (2009)) of different treatments or interventions. Imputation by regression is one of the most popular methods (Johansson, Shalit and sontaq (2016); Agarwal, Shah and Shen (2022); Fan, Masini and Medeiros (2022)), which requires a fully observed preintervention period and invariant relationships among the treated individuals before and after different interventions. More recently, Athey et al. (2021), Bai and Ng (2021), and Poulos et al. (2022) formulate the counterfactual prediction as a matrix completion problem, which requires that the missing entries must locate in a single block of the data matrix after appropriately permuting the rows and columns. This is equivalent to assuming

Received December 2022; revised April 2023.
Key words and phrases. Causal inference, imputation, informative missing, latent factor, tensor decomposition.

the existence of fully observed rows and columns. Another line of research related to counterfactual prediction is to estimate the average treatment effect (Athey, Imbens and Wager (2018); Fogarty (2020); Yadlowsky et al. (2021)) and derive the corresponding individualized treatment rules (Qi et al. (2020); Mo, Qi and Liu (2021); Mo and Liu (2022)).

However, the aforementioned methods are generally not suitable for the COVID-19 pandemic data due to a number of reasons. First, as one country or region only implements one social mobility restriction policy for a given time period, it leads to only one observation for each country or region at a given time. Second, the similarity of pandemic evolution patterns between two countries may vary over time due to different policy trajectories. Third, public policies are often multifaceted and multilevel, while existing imputation methods are restricted to binary intervention. For example, the number of quarantine days and the number of people allowed for social gathering are constantly adjusted under different pandemic situations.

In this paper we aim to conduct the counterfactual prediction of daily mortality rates of different countries or regions under various social mobility restriction policies. We formulate the task as a nonnegative tensor completion problem (Chen et al. (2019); Zhang and Ng (2022)) and propose to estimate the data tensor with a nonnegative tensor possessing multiple important features. Specifically, the estimated tensor admits a low-rank CANDECOMP/PARAFAC (CP) decomposition, where all the factor matrices of the expected data tensor to be nonnegative, the factor matrix in the policy mode is ordinal, and that in the time mode possesses a fusion structure. The nonnegativity requirement is imposed to ensure the expected data tensor to be nonnegative, corresponding to the nonnegative daily mortality rates. The ordinal structure is of great significance to characterize the ordinality of the mobility restriction strength of different public policies, and the fusion structure plays a crucial role in modeling the smooth trend of the pandemic evolution over time. The nonnegative tensor completion problem is then formulated in a least square format and subsequently optimized via an efficient alternative gradient descent algorithm.

The main contribution of this paper is three-fold. First, we develop a dynamic nonnegative tensor completion framework to predict the daily counterfactual mortality rates caused by the COVID-19 pandemic under various potential intervention policies in different countries or regions. The counterfactual prediction leads to a number of interesting results. For instance, it provides an accurate quantification of the effect of different social restriction policies as informative references for policymakers, and it reveals that countries or regions geographically close tend to share similar pandemic evolution patterns. Second, the developed method casts the tensor completion problem in a novel embedding framework, equipped with ordinal and fusion structures, which is able to simultaneously conduct tensor completion and numerical embeddings of the countries or regions, intervention policies, and time stamps. It provides a novel treatment to tackle the informative missing issue in the COVID-19 pandemic data. Third, we establish the asymptotic consistency of the estimated data tensor, which allows all the dimensions to diverge, including the number of treated individuals, intervention policies, and time stamps. Notably, the asymptotic theory is established under an informative missing assumption, and the effect of the missing pattern has also been precisely quantified and subsequently counted into the asymptotic convergence rate. This is in sharp contrast to the missing completely at random assumption in most existing tensor completion literature (Yuan and Zhang (2016, 2017); Xia and Yuan (2019)).

The rest of the paper is organized as follows. Section 2 provides a detailed description of the COVID-19 dataset and the data processing procedure. After introducing some necessary notations in Section 3, Section 4 presents the proposed dynamic counterfactual prediction framework and its associated nonnegative tensor completion problem and develops an alternative gradient descent algorithm to tackle the resultant optimization task. Section 5

establishes the asymptotic estimation consistency of the proposed method. Application of the proposed framework on the COVID-19 pandemic data is conducted in Section 6, followed by an extensive simulation study in Section 7. Technical proofs and necessary lemmas are relegated to the Supplementary Material.

2. COVID-19 data.

2.1. *Data description*. We integrate data from different sources to conduct counterfactual prediction of the social mobility restriction strength on the propagation severity of the COVID-19 pandemic. Particularly, the propagation severity is measured by the daily number of death cases caused by the pandemic, as suggested in Agarwal, Shah and Shen (2022), which points out that the number of death cases is a more faithful measurement than the number of infections in that the testing and reporting infection results may be inconsistent across different authorities. Also, the daily number of death cases of 193 countries and regions since January 22, 2020, are made publicly available by Dong, Du and Gardner (2020) at https://github.com/CSSEGISandData/COVID-19, including the numbers on the cruise ships Diamond Princess and Ms. Zaandam. More precisely, the daily number of death cases is used to measure the propagation intensity 23 days ago, due to the delay from infection of COVID-19 to actual death. This is because, as pointed out in literature (Lauer et al. (2020); Men et al. (2023); McAloon et al. (2020)), the median of the incubation period of COVID-19 is about five days while the duration from symptoms onset to death is about 18 days (Wilson et al. (2020)).

As it is generally believed, the social mobility restriction has an essential impact on the virus' diffusion among human beings. A relatively restricted social distancing policy will generally lead to a lower infection rate and hence fewer mortality cases. The strength of the social mobility restriction across 135 countries or regions is collected by Google's mobility report at https://www.google.com/covid19/mobility/, which measures the daily increment or decrement of mobility trends in percentages compared to basis across six different categories of places since February 15, 2020, including retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential.

We extract more than 7.5 million records from Google's mobility report, where each record consists of the daily increment or decrement of mobility strength by percentiles at various categories of places in a subregion of one of the 135 countries or regions on one of the 635 days range from February 15, 2020 to November 10, 2021. Note that not every record contains the mobility changing quantities of all six categories of places. For example, one record indicates that on June 1, 2021, the mobility strength increased 7% at groceries and pharmacies, increased 2% at residential, and decreased 5% at the workplace at Abitibi Regional County Municipality, Quebec, Canada, while the mobility changing quantities at retail and recreation parks as well as transit stations are missing. In fact, a total of 5,099,553 records contain missing values for one or more categories of places, but every record consists of the mobility changing quantity for at least one place category.

2.2. *Data processing*. We first extract the cumulative numbers of death cases for all of the 193 countries or regions from January 22, 2020 to November 10, 2021, from Dong, Du and Gardner (2020), resulting in a 193×659 data matrix, where its (i, j)th entry indicates the cumulative number of death cases of country or region i from January 22, 2020, to the subsequent jth day. A total of 23 countries, such as Solomon Islands, Holy See, Tonga, Palau, Kiribati, Marshall Islands, Samoa, and Micronesia, are removed from the analysis, as their cumulative numbers of death cases up to November 10, 2021, are too small for us to identify the pandemic "starting day," as in what follows. For the remaining 170 countries

or regions, we identify their first days with cumulative number of death cases more than 50 and treat those days as their corresponding starting days of the COVID-19 pandemic in the analysis. This step aligns with the intrinsic pandemic evolution of each country or region since the actual breakout time of the COVID-19 pandemic varies from country to country. Introducing a starting day could also avoid many noisy random patterns at the very beginning of the pandemic and thus enhance the signal of the data. A similar alignment of the pandemic starting day has also been employed in Agarwal, Shah and Shen (2022). We remark that the pandemic evolution may also be affected by various and possibly time-varying covariate information, such as seasonal effect, geographic location, medical treatment levels, and other social, economic, and political factors. A possible refinement of the alignment procedure shall take this covariate information into consideration, such as employing certain causal inference approaches to remove the confounding effects from the covariates.

We further delete those countries and regions with their starting day later than November 11, 2020, as we would like to study the COVID-19 pandemic evolution in a one-year period for all the studied countries, which further removes 24 countries or regions from the analysis. We then calculate the daily number of death cases for the remaining 146 countries or regions for a consecutive 365 days starting from their starting days. We finally obtain a 146×365 daily death case matrix, whose (i, j)th entry indicates the total number of death cases of country or region i at the jth day since its starting day.

Note that the COVID-19 pandemic generally broke out locally in some major cities with relatively dense populations of a country, where international passengers frequently traveled in and out, so it is reasonable to use the same number of cumulative death cases to identify the starting date. However, shortly after the break out, the pandemic would quickly spread over the whole country, and the population density can vary drastically from one country to another. Therefore, we adjust the daily death cases by dividing the population of the corresponding country in the analysis, leading to a much more stable measure. We thus obtain a 146×365 mortality rate matrix. The population of all the countries under investigation are available at https://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population.

Out of the 146 countries or regions, 110 of them are included in Google's mobility report, and thus we focus our analysis on these 110 countries or regions. Recall that the effects of mobility trends are reflected in the mortality rates 23 days later. Therefore, for each country or region, we extract all its records on Google's mobility report for a consecutive 365 days starting from 23 days prior to its starting day. For each record we first average the mobility changing quantities of all the available place categories in the particular day and subregion, and then for a given day in each country or region, we take the mean of the averaged mobility changing quantities of all the subregions within the country or region for that particular day. These processing steps yield a 110×365 mobility changing quantity matrix, where its (i, j)th entry indicates the mobility changing quantity by percentile of the ith country or region in the jth day. More precisely, the positive entries indicate the increments of the human mobility trend, while the negative entries indicate the decrements.

Furthermore, a substantial chunk of the mobility data for Afghanistan and Serbia is missing in Google's mobility report, and thus we remove these two countries from the analysis. Besides, the starting day of Italy is March 2, 2020, while its mobility data are recorded only since February 15, 2020, and the starting day of Cabo Verde is September 18, 2020, while its mobility data on August 30, 2020, and September 6, 2020, are missing. To make full use of the data of these two countries, we impute the mobility flow from February 8 to February 14, 2020, in Italy, on August 30, 2020, in Coba Verde, and on September 6, 2020, in Coba Verde by using the data from February 15 to 28, 2020, in Italy, the data from August 25 to 29, 2020, and from August 31 to September 4, 2020, in Cabo Verde, and the data from September 1 to 5, 2020, and from September 7 to 11, 2020, in Coba Verde, respectively. These imputations are

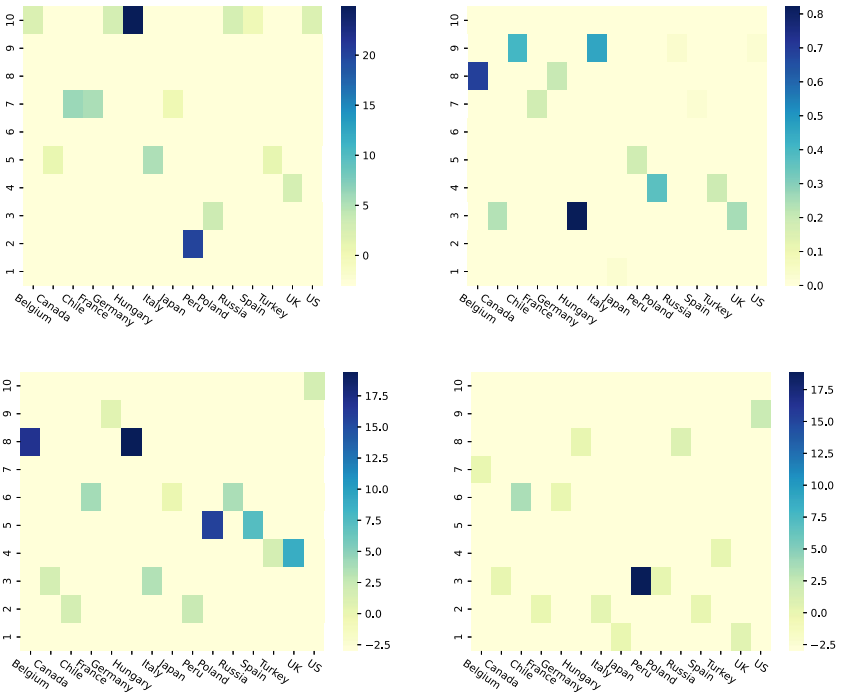

FIG. 1. *Mortality rates* (*multiplied by* 10^6) *of* 15 *selected countries on the first*, 120*th*, 240*th*, *and* 360*th days are displayed clockwise*. *The horizontal axis lists the* 15 *countries*, *the vertical axis indicates the activity restriction levels*, *the color intensity of each block indicates each country's mortality rate based on the corresponding public policies*, *and the light yellow color indicates missing mortality rates*.

based on the fact that the COVID-19 pandemic got more intense in the early breakout stage in Italy and Coba Verde, and their social mobility restriction got strengthened gradually.

After all the data processing steps, we obtain two 108×365 matrices, one about mortality rate and the other about mobility. Note that the mobility changing quantities are continuous and range from -46.62% to 0.93%. We then evenly and sequentially cut them into 10 parts, according to the corresponding percentiles, which thus categorizes the mobility changing quantities into 10 activity restriction public policies, with policy 1 being the strongest restriction one and policy 10 the weakest restriction one since they correspond to the parts with largest and smallest human mobility decrements, respectively. Clearly, this categorization procedure ensures the same strictness of policy restriction levels across different countries. Figure 1 illustrates the mortality rates of 15 selected countries or regions in four selected days.

It is clear that each country or region will only implement one public policy to control mobility for any single day, and different countries or regions may implement different policies. Consequently, each column of the figures in Figure 1 has only one observed mortality rate, and all other entries are missing. It is thus of great interest to conduct counterfactual prediction of the mortality rates under different public policies by leveraging information from other similar countries or regions. Such predicted counterfactual outcomes can provide critical guidelines for policymakers to determine whether to elevate or deelevate the public policy of social mobility control, in order to strike a better balance between COVID-19 combating and social economy development.

3. Notations. Throughout the paper we use boldface calligraphic Euler scripts ($\boldsymbol{\mathcal{A}}$) to denote tensors, boldface capital letters ($\boldsymbol{A}$) to denote matrices, boldface lowercase letters ($\boldsymbol{a}$) to denote vectors, and regular letters (a) to denote scalars. For an order 3 tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{N \times M \times T}$,

$\boldsymbol{\mathcal{A}}_{n,:,:} \in \mathbb{R}^{M\times T}$, $\boldsymbol{\mathcal{A}}_{:,m,:} \in \mathbb{R}^{N\times T}$, and $\boldsymbol{\mathcal{A}}_{:,:,t} \in \mathbb{R}^{N\times M}$ are the nth horizontal, mth lateral, and tth frontal slides of $\boldsymbol{\mathcal{A}}$, respectively, while $\boldsymbol{\mathcal{A}}_{:,m,t} \in \mathbb{R}^{N}$, $\boldsymbol{\mathcal{A}}_{n,:,t} \in \mathbb{R}^{M}$, and $\boldsymbol{\mathcal{A}}_{n,m,:} \in \mathbb{R}^{T}$ are the (m,t)th mode-1, (n,t)th mode-2, and (n,m)th mode-3 fibers of $\boldsymbol{\mathcal{A}}$, respectively. Similarly, for a matrix $\boldsymbol{A}$, $\boldsymbol{A}_{i,:}$ denotes its ith row and $\boldsymbol{A}_{:,j}$ denotes its jth column. For two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, the usual matrix product, Kronecker product, KhatriRao product (columnwise Kronecker product), and Hadamard product (entrywise product) between $\boldsymbol{A}$ and $\boldsymbol{B}$ are denoted as $\boldsymbol{AB}$, $\boldsymbol{A}\otimes\boldsymbol{B}$, $\boldsymbol{A}\odot\boldsymbol{B}$, and $\boldsymbol{A}*\boldsymbol{B}$, respectively. We also use $\boldsymbol{\mathcal{A}}*\boldsymbol{\mathcal{B}}$ to denote the Hadamard product between two tensors $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$. We use $\|\cdot\|$, $\|\cdot\|_0$, $\|\cdot\|_\infty$, $\|\cdot\|_{\infty,2}$, and $\|\cdot\|_F$ to denote the l_2-norm, l_0-norm, l_∞-norm of a vector, $l_{\infty,2}$-norm of a matrix, and the Frobenius norm of a matrix or tensor, respectively. For any integer n, denote $\mathbf{1}_n$ as the n-dimensional vector with all ones, and denote $[n]=\{1,2,\ldots,n\}$ to be the n-set.

The mode-1 product between a tensor $\boldsymbol{\mathcal{A}}\in\mathbb{R}^{N\times M\times T}$ and a matrix $\boldsymbol{A}\in\mathbb{R}^{N'\times N}$ is denoted as $\boldsymbol{\mathcal{A}}\times_1\boldsymbol{A}\in\mathbb{R}^{N'\times M\times T}$, whose entries are defined as

$$(\boldsymbol{\mathcal{A}}\times_1\boldsymbol{A})_{n',m,t}=\sum_{n=1}^{N}\boldsymbol{\mathcal{A}}_{n,m,t}\boldsymbol{A}_{n',n}\quad\text{for } n'\in[N],\, m\in[M],\text{ and } t\in[T].$$

The mode-2 or mode-3 product between a tensor and a matrix of appropriate dimension is defined in a similar fashion. Let $\mathcal{M}_1(\boldsymbol{\mathcal{A}})\in\mathbb{R}^{N\times MT}$, $\mathcal{M}_2(\boldsymbol{\mathcal{A}})\in\mathbb{R}^{M\times NT}$, and $\mathcal{M}_3(\boldsymbol{\mathcal{A}})\in\mathbb{R}^{T\times NM}$ be the mode-1, mode-2, and mode-3 matricization of $\boldsymbol{\mathcal{A}}$, respectively. Precisely, the (m,t)th mode-1 fiber $\boldsymbol{\mathcal{A}}_{:,m,t}$, (n,t)th mode-2 fiber $\boldsymbol{\mathcal{A}}_{n,:,t}$, and (n,m)th mode-3 fiber $\boldsymbol{\mathcal{A}}_{n,m,:}$ are placed at the $((t-1)M+m)$th column of $\mathcal{M}_1(\boldsymbol{\mathcal{A}})$, $((t-1)N+n)$th column of $\mathcal{M}_2(\boldsymbol{\mathcal{A}})$, and $((m-1)N+n)$th column of $\mathcal{M}_3(\boldsymbol{\mathcal{A}})$, respectively, for $n\in[N]$, $m\in[M]$, and $t\in[T]$.

The CANDECOMP/PARAFAC (CP) decomposition of $\boldsymbol{\mathcal{A}}$ has the form

$$\boldsymbol{\mathcal{A}}=\sum_{k=1}^{r}\boldsymbol{u}^{(k)}\circ\boldsymbol{v}^{(k)}\circ\boldsymbol{w}^{(k)},\tag{1}$$

where $\boldsymbol{u}^{(k)}\in\mathbb{R}^{N}$, $\boldsymbol{v}^{(k)}\in\mathbb{R}^{M}$, and $\boldsymbol{w}^{(k)}\in\mathbb{R}^{T}$ for $k\in[r]$ and $\circ$ stands for the vector outer product. The CP-rank of a tensor $\boldsymbol{u}^{(k)}\circ\boldsymbol{v}^{(k)}\circ\boldsymbol{w}^{(k)}$ is defined to be 1 (Kolda and Bader (2009)), for $k\in[r]$, and the CP decomposition in (1) decomposes $\boldsymbol{\mathcal{A}}$ into the summation of a number of rank-1 tensors. The minimal number of rank-1 tensors in the CP decomposition of $\boldsymbol{\mathcal{A}}$ is called the CP-rank of $\boldsymbol{\mathcal{A}}$. Let $\boldsymbol{\mathcal{I}}\in\{0,1\}^{r\times r\times r}$ be the order 3 r-dimensional identity tensor such that $\boldsymbol{\mathcal{I}}_{i_1,i_2,i_3}=1$ if $i_1=i_2=i_3$ and 0 otherwise, and let $\boldsymbol{U}\in\mathbb{R}^{N\times r}$, $\boldsymbol{V}\in\mathbb{R}^{M\times r}$, and $\boldsymbol{W}\in\mathbb{R}^{T\times r}$ such that $\boldsymbol{U}_{:,k}=\boldsymbol{u}^{(k)}$, $\boldsymbol{V}_{:,k}=\boldsymbol{v}^{(k)}$, and $\boldsymbol{W}_{:,k}=\boldsymbol{w}^{(k)}$. Then the CP decomposition in (1) can be equivalently rewritten as

$$\boldsymbol{\mathcal{A}}=\boldsymbol{\mathcal{I}}\times_1\boldsymbol{U}\times_2\boldsymbol{V}\times_3\boldsymbol{W}.$$

4. Dynamic counterfactual prediction. In this section we cast the COVID-19 pandemic data in a general dynamic counterfactual prediction framework in Section 4.1, then associate it with a tensor completion problem, and propose a nonnegative tensor decomposition model with various data-motivated structures in Section 4.2 to facilitate the tensor completion. An efficient alternative updated scheme is developed in Section 4.3. The key components of the proposed method in this section are illustrated in Figure 2.

4.1. *Dynamic counterfactual model.* Let $X_{n,t}\in[M]$ be the stochastic intervention (Papadogeorgou et al. (2022)) imposed to individual n at time t and $Y_{n,t}\in\mathbb{R}$ be the corresponding observed response affected by $X_{n,t}$ for $n\in[N]$ and $t\in[T]$, where N, M, and T are the number of treated individuals, possible intervention policies, and time stamps under investigation, respectively. For example, in the COVID-19 pandemic data, $X_{n,t}$ represents

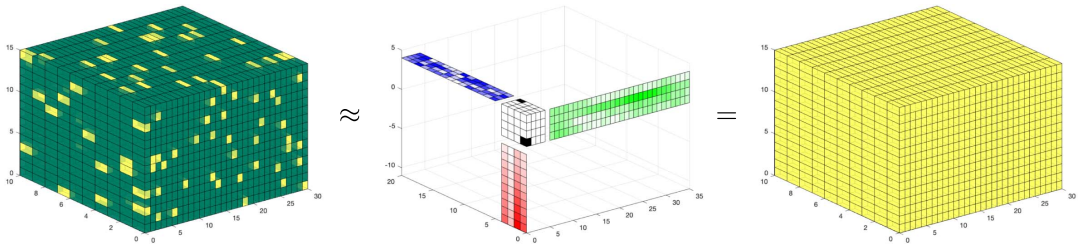

FIG. 2. *The observed COVID-19 pandemic data is illustrated in the left panel, where the yellow and green entries indicate observed and potential outcomes (missing entries), respectively. The counterfactual prediction problem is cast into a nonnegative tensor CP decomposition framework with various structures, as illustrated in the middle panel. Specifically, the core tensor is an identity tensor with black and white entries representing* 1 *and* 0*, while the blue, red, and light green factor matrices are the embeddings of countries, mobility restriction policies, and time stamps, respectively. The values of these three matrices are nonnegative, where the embeddings of mobility restriction policies are ordinal (the red color gets darker), and the embeddings of time stamps possess fusion structure (some adjacent entries over time share the same light green level). Finally, after estimating the embedding matrices, we can fully estimate the outcome tensor, as illustrated by the yellow tensor in the right panel, where the estimated yellow entries can be treated as counterfactual predictions.*

the human mobility restriction policy that country n adopted on the tth day, while $Y_{n,t}$ corresponds to the exact observed mortality rate of country n on the tth day. At time t only one intervention is imposed on individual n, and thus the outcomes of other $M-1$ interventions on individual n can not be observed. We term these $M-1$ interventions as counterfactual interventions and the corresponding outcomes as counterfactuals (Höfler (2005)).

Denote $Y_{n,t}^{(m)}$ as the potential outcome if $X_{n,t}=m$. We follow the standard treatment in Rubin's potential outcome framework (Rosenbaum and Rubin (1983)) to assume the following identifiability conditions: *(consistency)* $Y_{n,t}=Y_{n,t}^{(X_{n,t})}$, *(unconfoundness)* $X_{n,t}$ is independent with $\{Y_{n,t}^{(m)}\}_{m=1}^{M}$, and *(strict overlap)* $P(X_{n,t}=m)\geq p_{\min}$ for some $p_{\min}>0$, for $m\in[M]$. Clearly, the "unconfoundness" condition states that the intervention $X_{n,t}$ is independent of the set of all potential outcomes $\{Y_{n,t}^{(m)}\}_{m=1}^{M}$, whereas the "consistent" condition states that the observed outcome $Y_{n,t}$ strictly depends on the intervention $X_{n,t}$. Denote $\theta_{n,m,t}=\mathbb{E}Y_{n,t}^{(m)}$, the identifiability conditions allow us to write

$$\mathbb{E}Y_{n,t}=\sum_{m=1}^{M}\mathbb{E}\big[Y_{n,t}^{(m)}|X_{n,t}=m\big]P(X_{n,t}=m)=\sum_{m=1}^{M}\theta_{n,m,t}P(X_{n,t}=m) \tag{2}$$

for $n\in[N]$, $m\in[M]$ and $t\in[T]$.

A few remarks are in order. First, when $X_{n,t}=m$, the only observation is $Y_{n,t}=Y_{n,t}^{(m)}$, while $Y_{n,t}^{(1)},\ldots,Y_{n,t}^{(m-1)},Y_{n,t}^{(m+1)},\ldots Y_{n,t}^{(M)}$ are the counterfactuals. Second, model (2) allows multiple potential interventions (Qi et al. (2020); Mo and Liu (2022); Newey and Stouli (2022)), which is a natural generalization of the conventional potential outcome framework, where only two interventions are considered (Höfler (2005); Mo, Qi and Liu (2021)), namely, control and treatment. It also allows the intervention, outcome, and its expectation to be time-varying, as in the causal framework of spatial-temporal point process (Papadogeorgou et al. (2022)). Third, the major difference distinguishes the proposed model from conventional counterfactual framework is that the expected value of the potential outcome $\theta_{n,m,t}$ can vary from individual to individual, while existing potential outcome frameworks usually assume that individuals are randomly sampled from a source population, and thus $Y_{1,t}^{(m)},\ldots Y_{N,t}^{(m)}$ are independently and identically distributed for any $m\in[M]$ and $t\in[T]$. This coincides with the recent set up of individualized treatment rules (Fan et al. (2022); Mo and Liu

(2022)), where individualized covariate information will provide treatment-related and possibly treatment-free effects on the potential outcomes, leading to different expected potential outcomes for different individuals, even under the same treatment at the same time.

Instead of estimating the average treatment effect of one intervention across all individuals, it is of great interest to estimate the expected responds $\theta_{n,m,t}$'s for different individuals. However, the heterogeneous distributions of $Y_{n,t}^{(m)}$ cast great challenges to the estimation, since there are only NT observations, while the number of parameters to estimate is NMT, which is impossible to obtain consistent estimation unless reasonable structural assumption is imposed on $\theta_{n,m,t}$'s. To this end, we propose to formulate the counterfactual prediction problem in a tensor completion framework so as to reduce the intrinsic number of parameters and facilitate counterfactual prediction.

4.2. *Nonnegative tensor decomposition.* Let $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{N\times M\times T}$ be the data tensor such that $\boldsymbol{\mathcal{A}}_{n,m,t} = Y_{n,t}^{(m)}$ for $n \in [N]$, $m \in [M]$, and $t \in [T]$. Note that if $X_{n,t} = m$, then $\boldsymbol{\mathcal{A}}_{n,m,t}$ is observed while $\boldsymbol{\mathcal{A}}_{n,m',t}$ is missing for all other $m' \neq m$. Consequently, the problem of predicting the potential outcomes affected by the counterfactual interventions is now converted to impute the missing entries of the data tensor $\boldsymbol{\mathcal{A}}$. To fully characterize the dynamic intervention process, we consider a nonnegative CP decomposition model,

$$\mathbb{E}\boldsymbol{\mathcal{A}} = \boldsymbol{\Theta} = \boldsymbol{\mathcal{I}} \times_1 \boldsymbol{U} \times_2 \boldsymbol{V} \times_3 \boldsymbol{W}, \tag{3}$$

where $\boldsymbol{\Theta}_{n,m,t} = \theta_{n,m,t}$ by the definition of $\boldsymbol{\mathcal{A}}$, $\boldsymbol{\mathcal{I}}$ is an order 3 r-dimensional identity tensor and $\boldsymbol{U}_{n,k} \geq 0$, $\boldsymbol{V}_{m,k} \geq 0$, and $\boldsymbol{W}_{t,k} \geq 0$ for $n \in [N]$, $m \in [M]$, $t \in [T]$, and $k \in [r]$. For more details about the tensor CP decomposition, one can refer to Kolda and Bader (2009) for a comprehensive review.

Clearly, (3) assures the nonnegative elements in the mean tensor $\boldsymbol{\Theta}$, corresponding to the nonnegative expected daily mortality rates in the COVID-19 dataset. In addition, the rows of $\boldsymbol{U}$, $\boldsymbol{V}$, and $\boldsymbol{W}$ serve as the numeric embeddings of the corresponding entities, interventions, and time stamps in the latent spaces. Therefore, the proposed nonnegative CP decomposition model allows one to develop efficient algorithms to conduct counterfactual prediction and numerical embedding simultaneously. Moreover, (3) reduces the inherent number of parameters in $\boldsymbol{\Theta}$ from NMT to $(N + M + T)r$ so that consistently estimating the tensor parameter $\boldsymbol{\Theta}$ is possible based on the NT observed entries in $\boldsymbol{\mathcal{A}}$ if $r \ll \min\{N, T\}$. As in the tensor recommender system literature (Bi, Qu and Shen (2018); Zhang et al. (2021)) that some users may share similar preferences on items over time, the low-rank structure on $\boldsymbol{\Theta}$ essentially imposed similar evolution patterns of the causal effects with certain subgroups of treated individuals. In the COVID-19 dataset, the difference of the counterfactual pandemic evolution patterns between counties i and j is fully controlled by the difference of their embeddings $\boldsymbol{U}_{i,.}$ and $\boldsymbol{U}_{j,.}$, while this heterogeneous embeddings still allow that every treated individual is sampled from a different population. Finally, we also remark that the CP decomposition (3) may not be identifiable. For instance, (3) will be still valid if one permutes the columns of $\boldsymbol{U}$, $\boldsymbol{V}$, and $\boldsymbol{W}$ in the same fashion simultaneously or multiples the kth column of $\boldsymbol{U}$, $\boldsymbol{V}$, $\boldsymbol{W}$ by positive constants $c_{\boldsymbol{U}}^{(k)}$, $c_{\boldsymbol{V}}^{(k)}$, and $c_{\boldsymbol{W}}^{(k)}$ with $c_{\boldsymbol{U}}^{(k)} c_{\boldsymbol{V}}^{(k)} c_{\boldsymbol{W}}^{(k)} = 1$. Fortunately, the primary goal of this paper is to estimate the expected counterfactual effect $\boldsymbol{\Theta}$, instead of its CP decomposition factors, and hence it does not suffer from the identifiability issue induced by the tensor decomposition.

In what follows, we introduce some data-motivated properties on the CP factors to better facilitate tensor completion. Note that, in the COVID-19 dataset, the human mobility restriction policies are ordered by 10 different monotonic levels of strength. This leads to a

monotonic effect of intervention levels on the potential outcomes, following the general belief that mortality rate decreases with the strength of mobility flow restriction policies. This motivates us to introduce an ordinal structure on $\boldsymbol{V}$. Particularly, we assume that

$$\boldsymbol{V}_{1,:} = \boldsymbol{\Delta}_1, \quad \text{and} \quad \boldsymbol{V}_{m,:} = \boldsymbol{V}_{m-1,:} + \boldsymbol{\Delta}_m \quad \text{for } m = 2, \ldots, M,$$

where $\boldsymbol{\Delta}_1$, $\boldsymbol{\Delta}_2$,..., $\boldsymbol{\Delta}_M$ are some r-dimensional nonnegative vectors. The expected gap between policies $m+1$ and m for country n on the tth day is

$$\boldsymbol{\Theta}_{n,m+1,t} - \boldsymbol{\Theta}_{n,m,t} = \boldsymbol{\mathcal{I}} \times_1 \boldsymbol{U}_{n,:}^\top \times_2 \Delta_{m+1}^\top \times_3 \boldsymbol{W}_{t,:}^\top,$$

which may vary from country to country and from day to day. Also, Δ_{m+1} is allowed to vary with $m \in [M-1]$ as well. It is clear that country heterogeneity and time-varying trends have been incorporated in the proposed method.

Let $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1, \ldots, \boldsymbol{\Delta}_M)^\top$ and $\boldsymbol{L} \in \{0,1\}^{M \times M}$ be a lower triangular matrix such that $\boldsymbol{L}_{m,m'} = 1$ if $m \geq m'$ and zero otherwise for $m, m' \in [M]$. It follows that $\boldsymbol{V}$ admits the following decomposition:

$$\boldsymbol{V} = \boldsymbol{L}\boldsymbol{\Delta}. \tag{4}$$

We remark that $\boldsymbol{\Delta}$ is the intrinsic free parameters of $\boldsymbol{V}$, and the nonnegativity constraint on $\boldsymbol{\Delta}$ is sufficient to assure both the ordinal structure and nonnegativity condition on $\boldsymbol{V}$. Therefore, the embedding $\boldsymbol{V}$ is able to reflect the ordinal nature of the mobility restriction policies.

Furthermore, to incorporate the dynamic flow and smooth trend of the COVID-19 pandemic evolution, as illustrated by three selected countries in Figure 4 in Section 6, we introduce a fusion structure on $\boldsymbol{\Theta}$ through the temporal-mode factor $\boldsymbol{W}$. Specifically, let $D : \mathbb{R}^T \to \mathbb{R}^{T-1}$ be the first-order difference operator such that $D(\boldsymbol{a}) = (a_2 - a_1, \ldots, a_T - a_{T-1})$ for any $\boldsymbol{a} = (a_1, \ldots, a_T) \in \mathbb{R}^T$. We then introduce a fusion structure on $\boldsymbol{W}$ such that

$$\|D(\boldsymbol{W}_{:,k})\|_0 \leq \kappa - 1 \quad \text{for } k \in [r], \tag{5}$$

where κ is a positive integer fusion parameter. It is clear that the constraint (5) implies that $\boldsymbol{W}_{:,k}$ is successively cut into at most κ segments and the coordinates of $\boldsymbol{W}_{:,k}$ within each segment are identical. We remark that (5) can be extended to allow different fusion parameters for different columns of $\boldsymbol{W}$ at the cost of increased computational burden on model tuning. The fusion structure imposed in the temporal-mode factor matrix has been popularly employed in the dynamic tensor decomposition literature, such as dynamic tensor clustering (Sun and Li (2019)), dynamic network analysis (Zhang, Sun and Li (2020)), and dynamic tensor regression (Zhou et al. (2023)), among many others.

Let $\Gamma_{\boldsymbol{U}} \subset \mathbb{R}^{N \times r}$, $\Gamma_{\boldsymbol{V}} \subset R^{M \times r}$, and $\Gamma_{\boldsymbol{W}}(\kappa) \in \mathbb{R}^{T \times r}$ be the domains of $\boldsymbol{U}$, $\boldsymbol{V}$, and $\boldsymbol{W}$, respectively. Specifically, $\Gamma_{\boldsymbol{U}}$ is the set of all $N \times r$ nonnegative matrices. For any $\boldsymbol{V} \in \Gamma_{\boldsymbol{V}}$, $\boldsymbol{V}$ possesses the ordinal decomposition (4). The set $\Gamma_{\boldsymbol{W}}(\kappa)$ consists of all $T \times r$ nonnegative matrices $\boldsymbol{W}$ that satisfy $\|D(\boldsymbol{W}_{:,k})\|_0 \leq \kappa - 1$ for $k \in [r]$. Denote

$$\Gamma = \left\{\boldsymbol{\Theta} = \boldsymbol{\mathcal{I}} \times_1 \boldsymbol{U} \times_2 \boldsymbol{V} \times_3 \boldsymbol{W} : (\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W}) \in \Gamma_{\boldsymbol{U}} \times \Gamma_{\boldsymbol{V}} \times \Gamma_{\boldsymbol{W}}(\kappa)\right\}$$

as the parameter space of $\boldsymbol{\Theta}$ and Ω as the index set of the observed entries in $\boldsymbol{\mathcal{A}}$. The proposed nonnegative tensor completion formulation is organized as

$$\min_{\boldsymbol{\Theta} \in \Gamma} \frac{1}{2} \|P_\Omega(\boldsymbol{\Theta} - \boldsymbol{\mathcal{A}})\|_F^2, \tag{6}$$

where $P_\Omega(\cdot)$ is the entrywise projection that zeros the entries with indexes out of Ω. Formally, $P_\Omega(\boldsymbol{\Theta} - \boldsymbol{\mathcal{A}})$ is an $N \times M \times T$ tensor with

$$\left(P_\Omega(\boldsymbol{\Theta} - \boldsymbol{\mathcal{A}})\right)_{n,m,t} = \begin{cases} \boldsymbol{\Theta}_{n,m,t} - \boldsymbol{\mathcal{A}}_{n,m,t} & \text{if } (n,m,t) \in \Omega; \\ 0 & \text{otherwise,} \end{cases}$$

for $n \in [N]$, $m \in [M]$ and $t \in [T]$. The least square formulation is computationally efficient, does not require any distributional assumption of the data tensor, as in likelihood-based formulations, and hence lets flexibility for theoretical development. The constraint set Γ narrows down the support of $\boldsymbol{\Theta}$ and helps avoid overfitting in terms of counterfactual prediction on the complement of Ω.

4.3. *Alternative gradient descent algorithm.* Let $\boldsymbol{\delta} \in \{0,1\}^{N\times M\times T}$ be the indicator tensor such that $\boldsymbol{\delta}_{n,m,t} = 1$ if $(n,m,t) \in \Omega$ and 0 otherwise. Clearly, $\delta_{n,m,t} = 1$ if and only if the mobility restriction policy imposed to country n on the tth day $X_{n,t}$ equals m for $n \in [N]$, $m \in [M]$, and $t \in [T]$. As a result, $\{X_{n,,m,t} : n \in [N], m \in [M], t \in [T]\}$, Ω, and $\boldsymbol{\delta}$ essentially refer to the same thing but have different representation forms in different contexts. Denote the objective function as

$$f(\boldsymbol{U},\boldsymbol{V},\boldsymbol{W}) = \frac{1}{2}\big\|P_\Omega(\boldsymbol{\mathcal{I}}\times_1\boldsymbol{U}\times_2\boldsymbol{V}\times_3\boldsymbol{W}-\boldsymbol{\mathcal{A}})\big\|_F^2 = \frac{1}{2}\big\|(\boldsymbol{\mathcal{I}}\times_1\boldsymbol{U}\times_2\boldsymbol{V}\times_3\boldsymbol{W}-\boldsymbol{\mathcal{A}})*\boldsymbol{\delta}\big\|_F^2.$$

We next develop an alternative gradient descent (AGD) algorithm to minimize $f(\boldsymbol{U},\boldsymbol{V},\boldsymbol{W})$, subject to $\boldsymbol{U}\in\Gamma_{\boldsymbol{U}}$, $\boldsymbol{V}\in\Gamma_{\boldsymbol{V}}$, and $\boldsymbol{W}\in\Gamma_{\boldsymbol{W}}$. Clearly, $f(\boldsymbol{U},\boldsymbol{V},\boldsymbol{W})$ is blockwise multiconvex (Xu and Yin (2013)) on $(\boldsymbol{U},\boldsymbol{V},\boldsymbol{W})$, yet the feasible region is nonconvex, due to the l_0-norm constraint in $\Gamma_{\boldsymbol{W}}$, leading to that the optimization problem (6) can be solved only locally.

Given $\boldsymbol{U}^{(s)}$, $\boldsymbol{V}^{(s)} = \boldsymbol{L}\boldsymbol{\Delta}^{(s)}$, and $\boldsymbol{W}^{(s)}$ at iteration s, it follows from the mode-1 matricization operator that

$$\begin{aligned} f\big(\boldsymbol{U}^{(s)},\boldsymbol{V}^{(s)},\boldsymbol{W}^{(s)}\big) &= \frac{1}{2}\big\|\big(\boldsymbol{U}^{(s)}\mathcal{M}_1(\boldsymbol{\mathcal{I}})\big(\boldsymbol{W}^{(s)}\otimes\boldsymbol{V}^{(s)}\big)^\top-\mathcal{M}_1(\boldsymbol{\mathcal{A}})\big)*\mathcal{M}_1(\boldsymbol{\delta})\big\|_F^2 \\ &= \frac{1}{2}\big\|\big(\boldsymbol{U}^{(s)}\big(\boldsymbol{W}^{(s)}\odot\boldsymbol{V}^{(s)}\big)^\top\big)*\mathcal{M}_1(\boldsymbol{\delta})-\mathcal{M}_1(\boldsymbol{\mathcal{A}}*\boldsymbol{\delta})\big\|_F^2. \end{aligned}$$

Therefore, the gradient of $f(\boldsymbol{U},\boldsymbol{V},\boldsymbol{W})$, with respect to $\boldsymbol{U}$ evaluated at $(\boldsymbol{U}^{(s)},\boldsymbol{V}^{(s)},\boldsymbol{W}^{(s)})$, is

$$\nabla_{\boldsymbol{U}} f\big(\boldsymbol{U}^{(s)},\boldsymbol{V}^{(s)},\boldsymbol{W}^{(s)}\big) = \mathcal{M}_1\big(\big(\boldsymbol{\mathcal{I}}\times_1\boldsymbol{U}^{(s)}\times_2\boldsymbol{V}^{(s)}\times_3\boldsymbol{W}^{(s)}-\boldsymbol{\mathcal{A}}\big)*\boldsymbol{\delta}\big)\big(\boldsymbol{W}^{(s)}\odot\boldsymbol{V}^{(s)}\big).$$

With a step size $\eta > 0$, we update $\widetilde{\boldsymbol{U}}^{(s+1)}$ as

$$\widetilde{\boldsymbol{U}}^{(s+1)} = \boldsymbol{U}^{(s)} - \eta\nabla_{\boldsymbol{U}} f\big(\boldsymbol{U}^{(s)},\boldsymbol{V}^{(s)},\boldsymbol{W}^{(s)}\big) \tag{7}$$

and then obtain $\boldsymbol{U}^{(s+1)}$ by thresholding the negative entries of $\widetilde{\boldsymbol{U}}^{(s+1)}$ with zeros.

Next, given $\boldsymbol{U}^{(s+1)}$ and $\boldsymbol{W}^{(s)}$, we update $\widetilde{\boldsymbol{\Delta}}^{(s+1)}$ along the gradient of $f(\boldsymbol{U},\boldsymbol{L}\boldsymbol{\Delta},\boldsymbol{W})$, with respect to $\boldsymbol{\Delta}$ evaluated at $(\boldsymbol{U}^{(s+1)},\boldsymbol{L}\boldsymbol{\Delta}^{(s)},\boldsymbol{W}^{(s)})$, as

$$\begin{aligned} \widetilde{\boldsymbol{\Delta}}^{(s+1)} &= \boldsymbol{\Delta}^{(s)} - \eta\nabla_{\boldsymbol{\Delta}} f\big(\boldsymbol{U}^{(s+1)},\boldsymbol{L}\boldsymbol{\Delta}^{(s)},\boldsymbol{W}^{(s)}\big) \quad \text{where} \\ &\nabla_{\boldsymbol{\Delta}} f\big(\boldsymbol{U}^{(\boldsymbol{s+1})},\boldsymbol{L}\boldsymbol{\Delta}^{(s)},\boldsymbol{W}^{(s)}\big) \\ &\quad = \boldsymbol{L}^\top\mathcal{M}_2\big(\big(\boldsymbol{\mathcal{I}}\times_1\boldsymbol{U}^{(s+1)}\times_2\boldsymbol{V}^{(s)}\times_3\boldsymbol{W}^{(s)}-\boldsymbol{\mathcal{A}}\big)*\boldsymbol{\delta}\big)\big(\boldsymbol{W}^{(s)}\odot\boldsymbol{U}^{(s+1)}\big). \end{aligned} \tag{8}$$

Further, $\boldsymbol{\Delta}^{(s+1)}$ is obtained by thresholding the negative entries of $\widetilde{\boldsymbol{\Delta}}^{(s+1)}$ with zeros.

Finally, given $\boldsymbol{U}^{(s+1)}$, $\boldsymbol{V}^{(s+1)} = \boldsymbol{L}\boldsymbol{\Delta}^{(s+1)}$, and $\boldsymbol{W}^{(s)}$, we update $\widetilde{\boldsymbol{W}}$ along the direction of the gradient of $f(\boldsymbol{U},\boldsymbol{V},\boldsymbol{W})$, with respect to $\boldsymbol{W}$ evaluated at $(\boldsymbol{U}^{(s+1)},\boldsymbol{V}^{(s+1)},\boldsymbol{W}^{(s)})$,

$$\begin{aligned} \widetilde{\boldsymbol{W}}^{(s+1)} &= \boldsymbol{W}^{(s)} - \eta\nabla_{\boldsymbol{W}} f\big(\boldsymbol{U}^{(s+1)},\boldsymbol{V}^{(s+1)},\boldsymbol{W}^{(s)}\big) \quad \text{where} \\ &\nabla_{\boldsymbol{W}} f\big(\boldsymbol{U}^{(s+1)},\boldsymbol{V}^{(s+1)},\boldsymbol{W}^{(s)}\big) \\ &\quad = \mathcal{M}_3\big(\big(\boldsymbol{\mathcal{I}}\times_1\boldsymbol{U}^{(s+1)}\times_2\boldsymbol{V}^{(s+1)}\times_3\boldsymbol{W}^{(s)}-\boldsymbol{\mathcal{A}}\big)*\boldsymbol{\delta}\big)\big(\boldsymbol{V}^{(s+1)}\odot\boldsymbol{U}^{(s+1)}\big). \end{aligned} \tag{9}$$

Algorithm 1: Alternative gradient descent (AGD)

Input : Observed data tensor $\mathcal{A} * \boldsymbol{\delta}$, indicator tensor $\boldsymbol{\delta}$, CP rank r, step size $\eta = 10^{-3}$, and fusion parameter κ.

Output: Estimators of $\boldsymbol{U}$, $\boldsymbol{V}$ and $\boldsymbol{W}$.

1 Initialization. The entries of $\boldsymbol{U}^{(0)}$, $\boldsymbol{\Delta}^{(0)}$, and $\boldsymbol{W}^{(0)}$ are independently drawn from uniform(0, 2). Set $s = 0$.

2 **repeat**

3 Step 1: Compute $\widetilde{\boldsymbol{U}}^{(s+1)}$ according to (7), and project it to the nonnegative orthant to obtain $\boldsymbol{U}^{(s+1)}$.

4 Step 2: Compute $\widetilde{\boldsymbol{\Delta}}^{(s+1)}$ according to (8), and project it to the nonnegative orthant to obtain $\boldsymbol{\Delta}^{(s+1)}$.

5 Step 3: Compute $\widetilde{\boldsymbol{W}}^{(s+1)}$ according to (9) , and project it to the nonnegative orthant to obtain $\overline{\boldsymbol{W}}^{(s+1)}$. Employ the fusion operator on the columns of $\overline{\boldsymbol{W}}^{(s+1)}$ to obtain $\boldsymbol{W}^{(s+1)}$ as in (10).

6 **if** $f(\boldsymbol{U}^{(s+1)}, \boldsymbol{V}^{(s+1)}, \boldsymbol{W}^{(s+1)}) \geq f(\boldsymbol{U}^{(s)}, \boldsymbol{V}^{(s)}, \boldsymbol{W}^{(s)})$, **then**

7 reduce step size $\eta \leftarrow \eta/2$, and back to line 3.

8 **end**

9 $s \leftarrow s + 1$.

10 **until** $\eta < 10^{-8}$;

Let $\overline{\boldsymbol{W}}^{(s+1)}$ be the nonnegative version of $\widetilde{\boldsymbol{W}}^{(s+1)}$ such that the negative entries of $\widetilde{\boldsymbol{W}}^{(s+1)}$ are truncated by zeros. Clearly, $\|D\overline{\boldsymbol{W}}^{(s+1)}\|_0 \leq \|D\widetilde{\boldsymbol{W}}^{(s+1)}\|_0$, implying that the fusion structure in $\widetilde{\boldsymbol{W}}^{(s+1)}$ will never be destroyed by the nonnegative thresholding. We then introduce a fusion operator $F(\cdot, \cdot)$ to force the columns of $\overline{\boldsymbol{W}}^{(s+1)}$ to satisfy the l_0-norm constraints (5). Particularly, for any positive integer d, positive integer τ, and vector $\boldsymbol{a} \in \mathbb{R}^d$, $F(\boldsymbol{a}, \tau)$ first truncates $D(\boldsymbol{a})$ by keeping the $\tau - 1$ coordinates of $D(\boldsymbol{a})$ with largest absolute values and zeroing all other coordinates. This partitions the coordinates of $\boldsymbol{a}$ into τ groups such that $\boldsymbol{a}_{j+1}$ and $\boldsymbol{a}_j$ are in the same group if the jth coordinate of the truncated version of $D(\boldsymbol{a})$ is zero, for $j \in [d-1]$. The ith coordinate of $F(\boldsymbol{a}, \tau)$ is then defined to be the average value of all the coordinates of $\boldsymbol{a}$, inside the group that a_i belongs to, for $i \in [d]$. For the fusion parameter κ, we then obtain $\boldsymbol{W}^{(s+1)}$ as

$$\boldsymbol{W}_{:,k}^{(s+1)} = F\big(\overline{\boldsymbol{W}}_{:,k}^{(s+1)}, \kappa\big) \quad \text{for } k \in [r]. \tag{10}$$

The developed AGD algorithm is summarized in Algorithm 1. It employs an initial step size 10^{-3} and reduces it to half only when overshooting happens. The algorithm stops when $\eta < 10^{-8}$, which shares the same spirit of controlling estimation error, and the estimate $\widehat{\boldsymbol{\Theta}}$ is guaranteed to be close to a stationary point.

As for the CP-rank r of the approximating data tensor and the fusion parameter κ, we treat them as tuning parameters and select these tuning parameters through minimizing the following BIC criterion:

$$\mathrm{BIC} = \big\| P_\Omega(\mathcal{I} \times_1 \boldsymbol{U} \times_2 \boldsymbol{V} \times_3 \boldsymbol{W} - \mathcal{A}) \big\|_F^2 + (Nr + Mr + \kappa r) \log(NMT).$$

Herein, the factor $(Nr + Mr + \kappa r)$ is the number of free parameters, which serves as the degree of freedom in the nonnegative tensor decomposition model. Similar BIC criteria have been widely employed to select tensor rank and other tuning parameters in literature

(Goutte and Amini (2010); Sun and Li (2019); Hu, Lee and Wang (2022)). We further remark that the computational complexity of the proposed method primarily lies in the computation of the gradients, while the complexities of the thresholding as well as fusion operations are negligible. Since the computation of $\nabla_{\boldsymbol{U}} f(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W})$, $\nabla_{\boldsymbol{\Delta}} f(\boldsymbol{U}, \boldsymbol{L}\boldsymbol{\Delta}, \boldsymbol{W})$, and $\nabla_{\boldsymbol{W}} f(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{W})$ are all of order $O(rNMT)$, the computational complexity of Algorithm 1 is of order $O(rNMTS)$ with S iterations, which is quite efficient and scales linearly with the size of the data tensor. Finally, we note that the scales and frequency of the negative entries in $\widetilde{\boldsymbol{U}}^{(s+1)}$, $\widetilde{\boldsymbol{\Delta}}^{(\boldsymbol{s+1})}$, and $\widetilde{\boldsymbol{W}}^{(s+1)}$ depend on the step size η of the algorithm. The thresholding steps for $\widetilde{\boldsymbol{U}}^{(s+1)}$, $\widetilde{\boldsymbol{\Delta}}^{(\boldsymbol{s+1})}$, and $\widetilde{\boldsymbol{W}}^{(s+1)}$ may introduce some small estimation bias. This issue can be addressed by further reducing the step size to assure positiveness of the estimated CP factors at the expense of increased computational cost.

5. Asymptotic properties. In this section we establish the asymptotic consistency of the estimated parameter tensor $\hat{\boldsymbol{\Theta}}$ obtained by the proposed dynamic nonnegative tensor completion method. The estimation consistency of $\hat{\boldsymbol{\Theta}}$ is evaluated via the $\boldsymbol{\alpha}$-weighted F-norm (Mandal and Parkes (2022)) of $\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*$, where $\boldsymbol{\Theta}^*$ is the underlying true parameter tensor. Specifically, let $\boldsymbol{\alpha} \in \mathbb{R}^M$ be a weighting vector with $\boldsymbol{\alpha}_m > 0$ for $m \in [M]$, and $\sum_{m=1}^M \boldsymbol{\alpha}_m = 1$. For any tensor $\boldsymbol{\mathcal{T}}^{(1)} \in \mathbb{R}^{N\times M\times T}$, its $\boldsymbol{\alpha}$-weighted F-norm is defined as

$$\|\boldsymbol{\mathcal{T}}^{(1)}\|_{F(\boldsymbol{\alpha})} = \sqrt{\langle \boldsymbol{\mathcal{T}}^{(1)}, \boldsymbol{\mathcal{T}}^{(1)} \rangle_{\boldsymbol{\alpha}}},$$

where $\langle \boldsymbol{\mathcal{T}}^{(\mathbf{1})}, \boldsymbol{\mathcal{T}}^{(\mathbf{2})} \rangle_{\boldsymbol{\alpha}}$ is the $\boldsymbol{\alpha}$-weighted inner product between two tensors $\boldsymbol{\mathcal{T}}^{(1)}$, $\boldsymbol{\mathcal{T}}^{(2)} \in \mathbb{R}^{N\times M\times T}$, defined as

$$\langle \boldsymbol{\mathcal{T}}^{(1)}, \boldsymbol{\mathcal{T}}^{(2)} \rangle_{\boldsymbol{\alpha}} = \sum_{m=1}^{M} \boldsymbol{\alpha}_m \sum_{n=1}^{N} \sum_{t=1}^{T} \boldsymbol{\mathcal{T}}^{(1)}_{n,m,t} \boldsymbol{\mathcal{T}}^{(2)}_{n,m,t}.$$

As $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_M$ are strictly positive, the defined $\boldsymbol{\alpha}$-weighted inner product is positive definite, symmetric, and bilinear, and hence it is indeed an inner product, inducing an valid $\boldsymbol{\alpha}$-weighted F-norm.

The following technical assumptions are made.

ASSUMPTION A. Suppose that the interventions $X_{n,t}$, $n \in [N]$, $t \in [T]$ are independently sampled from a multinomial distribution on $[M]$ with parameter $\boldsymbol{p} = (p_1, \ldots, p_M)$. Equivalently, $\boldsymbol{\delta}_{n,:,t}, n \in [N], t \in [T]$ are independently sampled from a multinomial distribution on $\{\boldsymbol{e_m}\}_{m=1}^M$ with parameter $\boldsymbol{p}$, where $\boldsymbol{e_m}$ is the mth canonical basis in $\mathbb{R}^M$.

Assumption A specifies the stochastic mechanism on the interventions with a multinomial distribution, leading to the missing mechanism in $\boldsymbol{\mathcal{A}}$. Specifically, for each (n, t) pair, only one $\boldsymbol{\mathcal{A}}_{n,m,t}$ is observed, and all other $\boldsymbol{\mathcal{A}}_{n,m',t}$'s with $m' \neq m$ are missing. Clearly, the entries of $\boldsymbol{\mathcal{A}}$ are missing not at random, which differs from most existing matrix or tensor completion methods, assuming the data are missing completely at random (Yuan and Zhang (2016, 2017); Xia and Yuan (2019); Liu and Moitra (2020)) or missing at random (Xue and Qu (2021)). Moreover, the matrix or tensor regression imputation methods also fail to tackle this missing mechanism, as they often require the existence of a block of complete observations (Agarwal, Shah and Shen (2022); Bai and Ng (2021)).

ASSUMPTION B. Suppose that $\boldsymbol{\mathcal{A}}_{n,m,t}$'s are independent nonnegative sub-Gaussian random variables with mean $\boldsymbol{\Theta}^*_{n,m,t}$ and a universal sub-Gaussian parameter $\sigma > 0$. Specifically, for any $\lambda > 0$,

$$\mathbb{E}e^{\lambda(\boldsymbol{\mathcal{A}}_{n,m,t} - \boldsymbol{\Theta}^*_{n,m,t})} \leq e^{\lambda^2\sigma^2/2} \quad \text{for } i \in [N], m \in [M] \text{ and } t \in [T].$$

Assumption B characterizes the tail distribution behavior of the entries of $\mathcal{A}$, which is satisfied by many popular distributions, such as a single-side truncated normal distribution or the distribution of any nonnegative bounded random variable.

Note that a fundamental condition making tensor completion possible is the so-called incoherence condition, which essentially ensures that the rows of the factor matrices of a tensor decomposition to have comparable magnitude. Definitions of the incoherence measures vary from the tensor decomposition formats, such as the incoherence measures for orthogonal CP decomposition in Jain and Oh (2014); Cai et al. (2019), Tucker decomposition in Cai, Li and Xia (2022a), and tensor train decomposition in Cai, Li and Xia (2022b). Since, in this paper, we consider a general CP decomposition that does not require the factor matrices to be orthogonal, we introduce the following nonorthogonal CP incoherence measure for a tensor $\boldsymbol{\Theta}$ with CP-rank r:

$$\mu(\boldsymbol{\Theta}) = \min_{(\boldsymbol{U},\boldsymbol{V},\boldsymbol{W}):\boldsymbol{\Theta}=\mathcal{I}\times_1\boldsymbol{U}\times_2\boldsymbol{V}\times_3\boldsymbol{W}} \max_{n,m,t}\{\|\boldsymbol{U}_{n,:}\|, \|\boldsymbol{V}_{m,:}\|, \|\boldsymbol{W}_{t,:}\|\}. \tag{11}$$

Clearly, though the CP decomposition of $\boldsymbol{\Theta}$ is not identifiable, the nonorthogonal CP incoherence measure is well defined, as it is defined on a best behaved decomposition in the sense that the maximum l_2 norm of the rows of the factor matrices is minimized. In the following theoretical analysis, we consider those parameter $\boldsymbol{\Theta}$ has CP rank r and its nonorthogonal CP incoherence measure satisfies

$$\mu(\boldsymbol{\Theta}) \le \xi \tag{12}$$

for some $\xi > 0$. Clearly, $\mu(\boldsymbol{\Theta}) < \xi$ implies that $\boldsymbol{\Theta}_{n,m,t} \le \xi^3$ for any n, m, and t. Further, for any $\boldsymbol{W} \in \Gamma_{\boldsymbol{W}}(\kappa)$, its kth column $\boldsymbol{W}_{:,k}$ can be split into at most κ segments, each consisting of identical entries. Therefore, there exists a matrix $\tilde{\boldsymbol{w}} \in [-\xi,\xi]^{\kappa\times r}$ such that the common entry value for the lth segment of $\boldsymbol{W}_{:,k}$ is $\tilde{\boldsymbol{w}}_{l,k}$. We then call $\tilde{\boldsymbol{w}}$ a generator of $\boldsymbol{W}$. Note that different columns may have different segment splittings.

ASSUMPTION C. For any $\boldsymbol{W}^{(1)}, \boldsymbol{W}^{(2)} \in \Gamma_{\boldsymbol{W}}(\kappa)$ and their corresponding generators $\tilde{\boldsymbol{w}}^{(1)}$ and $\tilde{\boldsymbol{w}}^{(2)}$, assumes that there exists an absolute constant C such that

$$\|\boldsymbol{W}^{(1)} - \boldsymbol{W}^{(2)}\|_F \le C\sqrt{T}\|\tilde{\boldsymbol{w}}^{(1)} - \tilde{\boldsymbol{w}}^{(2)}\|_{\infty,2}. \tag{13}$$

Assumption C is imposed to reduce the complexity of $\Gamma_{\boldsymbol{W}}$, according to the fusion structure encoded in $\boldsymbol{W}$, such that the perturbation of $\boldsymbol{W} \in \Gamma_{\boldsymbol{W}}$ can be well controlled by that of its generator $\tilde{\boldsymbol{w}} \in [-\xi,\xi]^{\kappa\times r}$. It is clear that Assumption C holds with $C = 1$ when $\kappa = 1$ or $\kappa = T$. Further, as long as $\boldsymbol{W}^{(1)}$ and $\boldsymbol{W}^{(2)}$ share the same segment splittings within each column, it follows from the Hölder inequality that $\|\boldsymbol{W}^{(1)} - \boldsymbol{W}^{(2)}\|_F^2 \le T\|\tilde{\boldsymbol{w}}^{(1)} - \tilde{\boldsymbol{w}}^{(2)}\|_{\infty,2}^2$, and thus Assumption C is also satisfied with $C = 1$.

THEOREM 5.1. *Under Assumptions A–C, for any estimator* $\hat{\boldsymbol{\Theta}}$ *satisfying*

$$\frac{1}{2NT}\|(\hat{\boldsymbol{\Theta}} - \mathcal{A}) * \boldsymbol{\delta}\|_F^2 \le \frac{1}{2NT}\|(\boldsymbol{\Theta}^* - \mathcal{A}) * \boldsymbol{\delta}\|_F^2 + \epsilon$$

with probability at least $1 - \frac{2}{(NT)^2} - \frac{2}{(NT)^{(N+M+\kappa)r}(C\sqrt{r})^{\kappa r}}$, *it holds true that*

$$\frac{1}{NT}\|\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_{F(\boldsymbol{p})}^2 \le 4\epsilon,$$

where $\epsilon = \max\{\frac{2^{11}\sigma^2((N+M+\kappa)r\log(NT)+\kappa r\log(C\sqrt{r}))}{NT}, \xi^6\sqrt{\frac{\log(NT)}{NT}}\}$.

The proof of Theorem 5.1 is provided in the Supplementary Material (Zhen and Wang (2024)). Essentially, Theorem 5.1 assures that $\frac{1}{NT}\|\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|^2_{F(\boldsymbol{p})} = O_p(\epsilon)$ as long as the objective value of $\hat{\boldsymbol{\Theta}}$ is sufficiently close to that of $\boldsymbol{\Theta}^*$. Herein, the convergence rate ϵ is governed by two competing terms, where the first term depends on the number of parameters $(N + M + \kappa)r$, the number of observed entries NT, and the sub-Gaussian parameter σ, while the second term comes from the randomness of the missing pattern. We further remark that ξ^6 in the second term of ϵ is essentially an upper bound of all the squared entries in $\boldsymbol{\Theta}$, and thus one may replace it back to have a slightly tighter upper bound. As a theoretical example with $M, \kappa, r, \xi, \sigma$ fixed as constants, we have $\epsilon = O(\frac{1}{T})$ if $T = O(N)$, and $\epsilon = O(\frac{1}{\sqrt{NT}})$ if $N = O(T)$, up to some logarithm factors.

Furthermore, it follows from the definition of $\boldsymbol{\alpha}$-weighted F-norm and Theorem 5.1 that

$$\frac{1}{NMT}\big\|\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\big\|_F^2 \le \frac{4\epsilon}{Mp_{\min}}$$

with probability tending to 1, where $p_{\min} = \min_{m\in M} p_m$, suggesting that the convergence property of $\hat{\boldsymbol{\Theta}}$ in terms of F-norm is indeed also affected by the missing mechanism. Specifically, $\frac{1}{NMT}\|\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_F^2$ can achieve the same convergence rate as $\frac{1}{NT}\|\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|^2_{F(\boldsymbol{p})}$ if the number of observed entries under each interventions are relatively balanced, whereas the convergence rate can be much slower if the missing pattern is unbalanced since $p_{\min}$ can be rather small.

6. Counterfactual prediction on COVID-19. We now apply the proposed dynamic nonnegative tensor completion method to conduct counterfactual prediction on the COVID-19 pandemic data $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{108\times 10\times 365}$, where 90% of the mortality rates are missing. Before carrying out Algorithm 1, we first employ the BIC criterion to determine the CP-rank r of $\mathbb{E}\boldsymbol{\mathcal{A}}$ and the fusion parameter κ. Minimizing the BIC criterion over the searching space $\{1, 2, \ldots, 40\} \times \{1, 2, \ldots, 40\}$ yields that $(r, \kappa) = (18, 23)$. Clearly, the suggested κ is substantially smaller than 365, suggesting that the COVID-19 pandemic indeed evolves smoothly over time.

With these tuned parameters, Algorithm 1 is carried out to produce the estimated numeric embeddings of the countries, activity restriction policies, and time stamps, denoted as $\hat{\boldsymbol{U}}$, $\hat{\boldsymbol{V}}$, and $\hat{\boldsymbol{W}}$, respectively. The counterfactual prediction on the mortality rates is then given by $\hat{\boldsymbol{\Theta}} = \boldsymbol{\mathcal{I}} \times_1 \hat{\boldsymbol{U}} \times_2 \hat{\boldsymbol{V}} \times_3 \hat{\boldsymbol{W}}$. The submatrices corresponding to the 15 selected countries in Figure 1 of the first, 120th, 240th, and 360th mode-3 slides of $\hat{\boldsymbol{\Theta}}$ are displayed in Figure 3, which can be regarded as the estimation of the completed version of Figure 1.

It is clear from Figure 3 that the estimated mortality rate decreases with the mobility flow restriction level as the result of the ordinal structure imposed on $\boldsymbol{V}$, which is motivated by the general belief that more strict activity restriction policy prevents large gatherings and close contacts, and is crucial for slowing down the spreads of the COVID-19 virus. More importantly, the completed tensor $\hat{\boldsymbol{\Theta}}$ also provides trustworthy prediction on the potential mortality rates for each country with different activity restriction policies. In addition, the fluctuation range of the estimation mortality rates in those countries with higher true mortality rates also tend to be larger. This is reasonable, as the pandemic evolution in those countries that have already suffered from severe pandemic situation are more sensitive to the public activity restriction policies.

To further investigate the counterfactual prediction accuracy over the temporal mode, we turn to study the mode-1 slides of $\hat{\boldsymbol{\Theta}}$. For illustration, we focus our analysis on Brazil, Japan, and the United Kingdom. For each country the top row of Figure 4 displays its evolution curve by connecting the true mortality rates over time and a synthetic evolution curve by

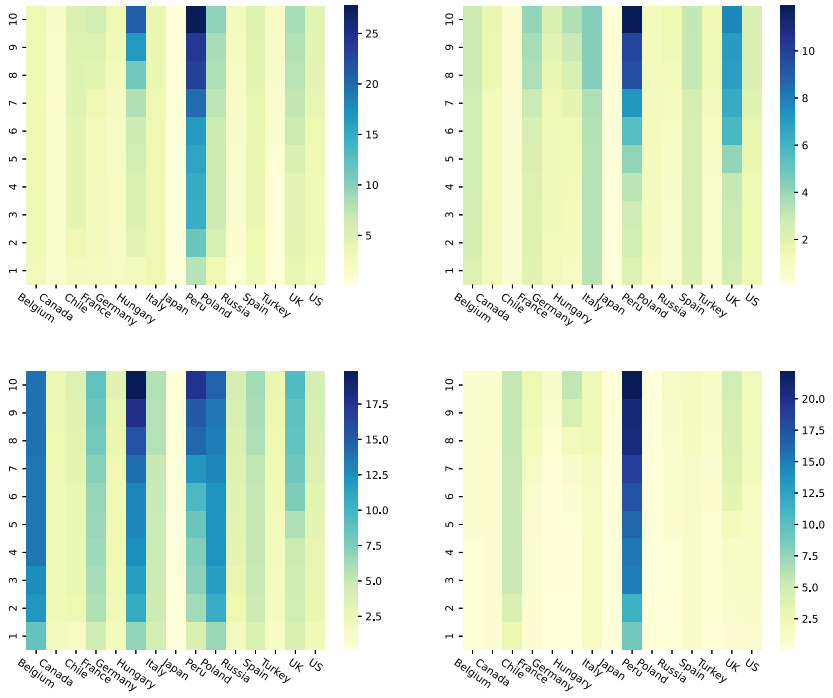

FIG. 3. *The estimated mortality rates* (*multiplied by* 10^6) *of the same* 15 *selected countries in Figure* 1 *on the first*, 120*th*, 240*th*, *and* 360*th days are displayed clockwise*.

connecting the estimated mortality rates in $\hat{\boldsymbol{\Theta}}$, given the actual mobility flow level every day. It is evident that the synthetic evolution trends closely coincide with the true curves in all three countries, while they are capable of avoiding overfitting by possessing smoother evolution patterns than the real curves. The bottom row of Figure 4 displays the cumulative predicted mortality rates of Brazil, Japan, and the United Kingdom over time under different

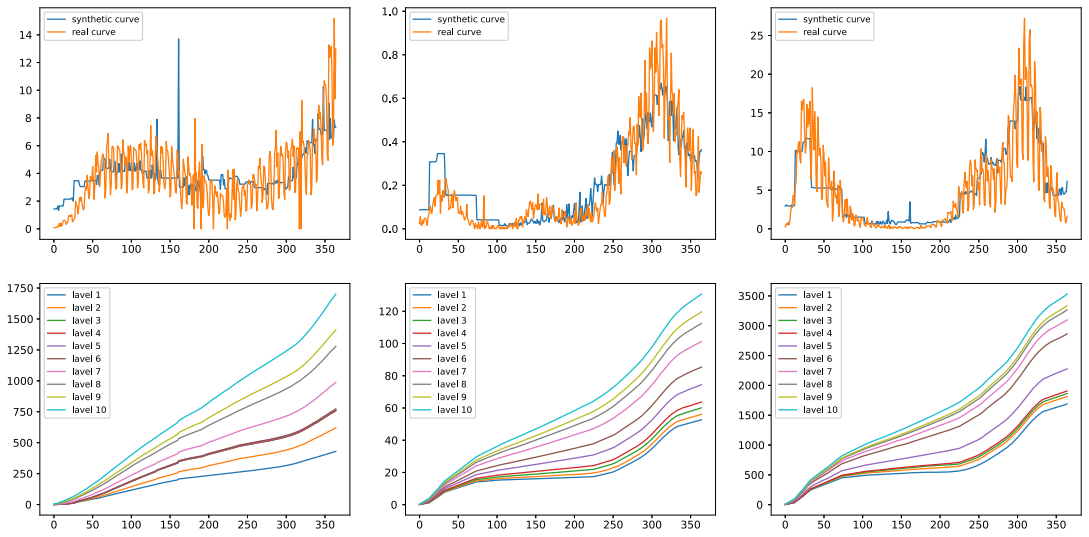

FIG. 4. *The top row displays the synthetic and true evolution curves of mortality rates in Brazil* (*Left*), *Japan* (*Middle*), *and the United Kingdom* (*Right*). *The bottom row displays the predicted cumulative mortality rates by the proposed method under different social mobility restriction policies in Brazil* (*Left*), *Japan* (*Middle*), *and the United Kingdom* (*Right*).

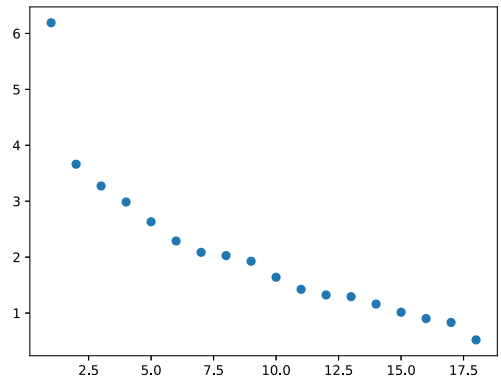

FIG. 5. *A clear elbow point at the seventh leading singular value of* $\hat{\boldsymbol{U}}$.

mobility flow levels. Clearly, the predicted mortality rates under more restrictive policies are much smaller than other loose ones, and the difference among the predicted mortality rates under different polices become larger as time goes by.

As a by-product we next turn to investigate the topological structure of the estimated latent positions of the countries. The singular values of $\hat{\boldsymbol{U}}$ are displayed in Figure 5, where the seventh leading singular value is a clear elbow point (Ji and Jin (2016); Rohe, Qin and Yu (2016)), suggesting that the countries can be roughly clustered into six communities.

We hence employ a K-means algorithm with $K = 6$ to detect communities in $\hat{\boldsymbol{U}}$, and the results are summarized in Figure 6. It is interesting to note that community 1 consists of the United States and many Western Europe countries, such as Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Portugal, Spain, United Kingdom, and Sweden. Gathering around the Caribbean Sea, Colombia, Venezuela, Panama, Nicaragua have been clustered into community 2. The majority of the countries in community 3 locate in Latin America, including Mexico, Belize, Honduras, Haiti, Ecuador, Peru, Chile, Bolivia, and Brazil. Communities 4 and 6 mainly contain countries in Eastern Europe and Africa. For example, a clique in Balkan Peninsula consists of Greece, Macedonia, Bulgaria, Romania, Hungary, Bosnia and Herzegovina, and Croatia, and a clique in southwest Sahara together with Gulf of Guinea consists of Mali, Niger, Senegal, Burkina Faso, Ghana, Cameroon, and Gabon. Community 5 only contains a few countries and is not as geographically clear as other communities, while we still can see that Argentina and Uruguay, Iraq, and Jordan are two adjacency country pairs in community 5. This community structure also suggests that the evolution of the COVID-19

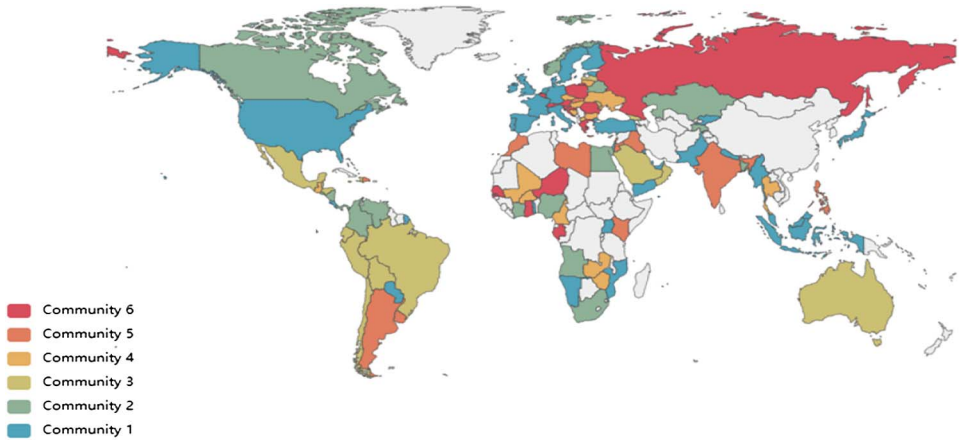

FIG. 6. *Communities estimated from the COVID-19 dataset.*

pandemic share great similarities among countries geographically close to each other, which is sensible, as it is easy for the virus and its variants to spread from one country to its neighboring countries.

7. Simulation study. To further scrutinize its empirical performance, we apply the proposed dynamic nonnegative tensor completion model, denoted as NTC, to various synthetic data and compares it against the standard tensor completion algorithm without considering the structural information (Bi, Qu and Shen (2018)), denoted as TC. We evaluate their numerical performance by the relative parameter estimation error,

$$\text{Err} = \frac{\|\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}^*\|_F}{\|\boldsymbol{\Theta}^*\|_F}. \tag{14}$$

7.1. *Synthetic data.* The synthetic data is generated as follows. First, we generate $\boldsymbol{U}^*_{n,k} \sim \text{Unif}(0, 2)$, $\boldsymbol{V}'_{m,k} \sim \text{Unif}(0, 2)$, and $\boldsymbol{W}'_{t,k} \sim \text{Unir}(0, 2)$ for $n \in [N]$, $m \in [M]$, $t \in [T]$, and $k \in [r]$, where Unif$(0, 2)$ stands for the uniform distribution in the interval $[0, 2]$. The kth column of $\boldsymbol{V}^*$ is then obtained by sorting the kth column of $\boldsymbol{V}'$ in ascending order for $k \in [r]$, and thus we have $\boldsymbol{\Delta} = \boldsymbol{L}^{-1}\boldsymbol{V}^*$. For each $k \in [r]$, we randomly and uniformly select $\kappa - 1$ elements from $\{2, 3, \ldots, T-1\}$ without replacement and denote the selected values as $\tau_1 < \cdots < \tau_{\kappa-1}$. $\boldsymbol{W}^*_{t,k}$ is then computed as the average value of $\boldsymbol{W}'_{\tau_{i-1},k}, \boldsymbol{W}'_{\tau_{i-1}+1,k}, \ldots, \boldsymbol{W}'_{\tau_i-1,k}$ if $\tau_{i-1} \le t < \tau_i$, for $i \in [\kappa]$, where τ_0 and τ_κ are defined as 1 and $T+1$, respectively. Second, we calculate $\boldsymbol{\Theta}^* = \boldsymbol{\mathcal{I}} \times_1 \boldsymbol{U}^* \times_2 \boldsymbol{V}^* \times_3 \boldsymbol{W}^*$, generate $\boldsymbol{\mathcal{A}} = \boldsymbol{\Theta}^* + \boldsymbol{\mathcal{E}}$, and threshold the negative entries of $\boldsymbol{\mathcal{A}}$ with zeros, where the entries of $\boldsymbol{\mathcal{E}}$ are independently generated from the truncated normal distribution $N_{[-2,2]}(0, 1)$. Third, given $\boldsymbol{p} = (p_1, \ldots, p_M)^\top$, the observation indicator tensor $\boldsymbol{\delta}$ is generated according to Assumption A:

SCENARIO 1. In this scenario we study the effect of the tensor size on the estimation accuracy. Specifically, we vary $N \in \{50, 100, 150, 200, 250\}$, $M \in \{10, 20, 30\}$, and $T \in \{100, 200, 300, 400, 500\}$ and fix $r = 10$, $\kappa = 5$, and $p_m = \frac{1}{M}$ for $m \in [M]$. For each triplet (N, M, T), we replicate the experiment 50 times, and the averaged relative estimation errors of NTC and TC and their standard errors are reported in Tables 1–3.

SCENARIO 2. In this scenario we study the effect of the balance structure of the missing pattern on the estimation accuracy. Specifically, we fix $N = 108$, $M = 10$, and $T = 365$ to be the same of the COVID-19 dataset. We further fix $r = 10$, $\kappa = 5$. For the weight vector $\boldsymbol{p}$, we first generate $\boldsymbol{p}' \sim N_M(\boldsymbol{0}_M, \rho^2 \boldsymbol{I}_M)$ for $\rho \in \{1, 1.5, 2, 2.5, 3\}$, where $\boldsymbol{0}_M$ is the M-dimensional vector with all zeros, $\boldsymbol{I}_M$ is the Mth order identity matrix, and $N_M(\boldsymbol{0}_M, \rho^2 \boldsymbol{I}_M)$ stands for the M-variate normal distribution with mean $\boldsymbol{0}_M$ and covariance matrix $\rho^2 \boldsymbol{I}_M$. We then normalize $\boldsymbol{p}'$ by the soft-max operator to obtain $\boldsymbol{p}$; that is, $\boldsymbol{p}_m = \frac{\exp(\boldsymbol{p}'_m)}{\sum_{j=1}^M \exp(\boldsymbol{p}'_j)}$ for $m \in [M]$. For each realization of ρ or equivalently $\boldsymbol{p}$, the averaged relative estimation errors of NTC and TC and their corresponding 95% confidence intervals over 50 independent replications are displayed in the left penal of Figure 7.

It is evident from Tables 1–3 and Figure 7 that NTC consistently outperforms TC in terms of averaged relative estimation errors and its standard errors in both scenarios. In addition, given N and M, the estimation error by NTC decays as T increases, whereas that of TC does not. This demonstrates the advantage of introducing a fusion structure into the time-mode factor matrix to reduce the effective degrees of free parameters, as otherwise the number of free parameters and observations will both scale linearly with T. Besides that, the estimation errors of both NTC and TC increase as M increases. This is because of the fact that, with

TABLE 1
The averaged relative estimation errors of NTC and TC and their standard errors over 50 *independent replications with* $M = 10$ *in Scenario* 1

N		$T = 100$	$T = 200$	$T = 300$	$T = 400$	$T = 500$
50	NTC	0.0344 (0.0007)	0.0320 (0.0006)	0.0315 (0.0007)	0.0301 (0.0006)	0.0303 (0.0007)
	TC	0.0529 (0.0007)	0.0500 (0.0009)	0.0490 (0.0009)	0.0491 (0.0007)	0.0518 (0.0014)
100	NTC	0.0408 (0.0007)	0.0353 (0.0007)	0.0327 (0.0008)	0.0323 (0.0007)	0.0325 (0.0006)
	TC	0.0579 (0.0008)	0.0581 (0.0010)	0.0587 (0.0015)	0.0550 (0.0014)	0.0651 (0.0015)
150	NTC	0.0410 (0.0007)	0.0366 (0.0005)	0.0345 (0.0005)	0.0330 (0.0005)	0.0312 (0.0005)
	TC	0.0594 (0.0009)	0.0631 (0.0010)	0.0685 (0.0014)	0.0667 (0.0012)	0.0617 (0.0015)
200	NTC	0.0434 (0.0007)	0.0373 (0.0006)	0.0356 (0.0005)	0.0334 (0.0005)	0.0332 (0.0006)
	TC	0.0614 (0.0010)	0.0667 (0.0010)	0.0729 (0.0008)	0.0695 (0.0012)	0.0743 (0.0018)
250	NTC	0.0440 (0.0007)	0.0379 (0.0006)	0.0356 (0.0004)	0.0345 (0.0006)	0.0333 (0.0004)
	TC	0.0635 (0.0013)	0.0676 (0.0015)	0.0709 (0.0008)	0.0771 (0.0017)	0.0768 (0.0013)

TABLE 2
The averaged relative estimation errors of NTC and TC and their standard errors over 50 *independent replications with* $M = 20$ *in Scenario* 1

N		$T = 100$	$T = 200$	$T = 300$	$T = 400$	$T = 500$
50	NTC	0.0388 (0.0006)	0.0334 (0.0005)	0.0318 (0.0005)	0.0298 (0.0004)	0.0289 (0.0004)
	TC	0.0590 (0.0007)	0.0558 (0.0010)	0.0623 (0.0014)	0.0610 (0.0009)	0.0761 (0.0024)
100	NTC	0.0401 (0.0008)	0.0335 (0.0005)	0.0321 (0.0005)	0.0300 (0.0003)	0.0301 (0.0004)
	TC	0.0636 (0.0009)	0.0663 (0.0007)	0.0738 (0.0016)	0.0725 (0.0005)	0.0894 (0.0024)
150	NTC	0.0420 (0.0007)	0.0354 (0.0005)	0.0319 (0.0004)	0.0305 (0.0005)	0.0286 (0.0004)
	TC	0.0700 (0.0011)	0.0742 (0.0013)	0.0791 (0.0018)	0.0853 (0.0011)	0.0822 (0.0011)
200	NTC	0.0413 (0.0007)	0.0352 (0.0006)	0.0340 (0.0005)	0.0300 (0.0005)	0.0308 (0.0006)
	TC	0.0687 (0.0010)	0.0740 (0.0009)	0.0838 (0.0013)	0.0828 (0.0004)	0.1033 (0.0019)
250	NTC	0.0443 (0.0011)	0.0407 (0.0010)	0.0326 (0.0006)	0.0375 (0.0009)	0.0317 (0.0007)
	TC	0.0796 (0.0012)	0.0853 (0.0012)	0.0807 (0.0009)	0.1009 (0.0012)	0.0908 (0.0007)

TABLE 3
The averaged relative estimation errors of NTC and TC and their standard errors over 50 *independent replications with* $M = 30$ *in Scenario* 1

N		$T = 100$	$T = 200$	$T = 300$	$T = 400$	$T = 500$
50	NTC	0.0401 (0.0006)	0.0352 (0.0005)	0.0304 (0.0005)	0.0290 (0.0005)	0.0275 (0.0004)
	TC	0.0656 (0.0006)	0.0711 (0.0008)	0.0743 (0.0022)	0.0892 (0.0009)	0.0859 (0.0012)
100	NTC	0.0413 (0.0007)	0.0349 (0.0005)	0.0304 (0.0006)	0.0309 (0.0004)	0.0288 (0.0004)
	TC	0.0704 (0.0012)	0.0789 (0.0009)	0.0861 (0.0016)	0.0976 (0.0009)	0.0943 (0.0014)
150	NTC	0.0418 (0.0007)	0.0339 (0.0006)	0.0319 (0.0005)	0.0294 (0.0005)	0.0309 (0.0006)
	TC	0.0760 (0.0015)	0.0839 (0.0009)	0.0884 (0.0009)	0.0985 (0.0020)	0.1108 (0.0023)
200	NTC	0.0448 (0.0011)	0.0393 (0.0009)	0.0319 (0.0005)	0.0371 (0.0009)	0.0306 (0.0006)
	TC	0.0815 (0.0014)	0.0884 (0.0015)	0.0969 (0.0018)	0.1078 (0.0012)	0.1026 (0.0017)
250	NTC	0.0455 (0.0011)	0.0412 (0.0006)	0.0499 (0.0015)	0.0372 (0.0007)	0.0340 (0.0016)
	TC	0.0812 (0.0008)	0.0880 (0.0005)	0.1090 (0.0009)	0.1001 (0.0010)	0.1357 (0.0022)

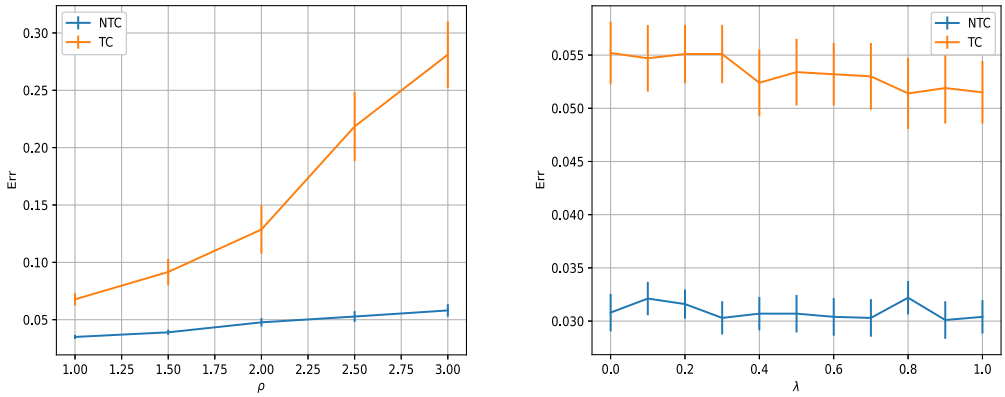

FIG. 7. *Left*: *The averaged relative estimation errors, given by NTC and TC, and their corresponding* 95% *confidence intervals for various values of* ρ. *Right*: *The averaged relative estimation errors, given by NTC and TC, and their corresponding* 95% *confidence intervals for various values of* λ.

fixing N and T, the number of parameters increases as M increases, while the number of observations remains invariant. Moreover, it can be seen from Tables 1–3 that the performance of NTC is relatively more robust than that of TC to the increasing M, and from Figure 7 that the performance of NTC is substantially more robust than that of TC to unbalance missing patterns. This is possibly due to the fact that NTC takes the ordinal structure of the intervention policies into consideration and allows for informative missing patterns, which alleviates the increment of the relative estimation error caused by the increasing M or ρ.

7.2. *Sensitivity analysis*. The proposed NTC method assumes the intervention patterns or, equivalently, the missing mechanisms to be independent among different countries and different time stamps. Yet policymakers might intend to adjust social mobility restriction policy if the mortality rates of their country or their neighboring countries significantly changed in previous days. This subsection carries out a sensitivity analysis of the proposed method to the independent intervention assumption.

Particularly, let the propensity score vector for country n on the tth day be

$$\boldsymbol{p}_{n,t} = (1-\lambda)\boldsymbol{p}_0 + \lambda \boldsymbol{p}_{n,t}^{(x)},$$

where $\boldsymbol{p}_0 \in \mathbb{R}^M$ is the fixed propensity score vector, $\boldsymbol{p}_{n,t}^{(x)}$ is the individualized and instant propensity score vector depending on the history of the mortality rates up to the $(t-1)$-day of country n and its neighboring countries, and $\lambda \in [0,1]$ balances the weights of $\boldsymbol{p}_0$ and $\boldsymbol{p}_{n,t}^{(x)}$. We conduct the sensitivity analysis by varying $\lambda \in \{0, 0.1, 0.2, \ldots, 1.0\}$.

We set $N = 108$, $M = 10$, and $T = 365$ to mimic the COVID-19 dataset. The data-generating scheme is the same as in Section 7.1, while the missing mechanism is different. Specifically, we set $p_0 = 0.1 \times \mathbf{1}_{10}$ and generate $\boldsymbol{p}_{n,t}^{(x)}$ based on additional spatial-temporal information. In the first $T_0 = 15$ days, we set $\lambda = 0$ for all countries so that we can compute the median of the first NT_0 observed outcomes, denoted as γ. It serves as a threshold to determine whether a mortality rate is high. For each country the closest four neighboring countries are considered. We further assume Markov property in $\boldsymbol{p}_{n,t}^{(x)}$ so that $\boldsymbol{p}_{n,t}^{(x)}$ depends on the mortality rate on the $(t-1)$th day only. We then set $\boldsymbol{p}_{n,t}^{(x)} = (\frac{1}{5}\mathbf{1}_5^\top, \mathbf{0}_5^\top)^\top$ if either the observed mortality rate of country n or the averaged observed mortality rate of its neighboring countries excesses γ and set $\boldsymbol{p}_{n,t}^{(x)} = (\mathbf{0}_5^\top, \frac{1}{5}\mathbf{1}_5^\top)^\top$ otherwise. For each λ the averaged relative estimation errors of NTC and TC and their corresponding 95% confidence intervals over 50 independent replications are displayed in the right penal of Figure 7. It is clear that both

NTC and TC are relatively robust against the variation of λ, suggesting that both methods do not heavily rely on the independent intervention assumption. Moreover, NTC consistently delivers smaller relative estimation error than TC, confirming the adequacy of the structural constraints.

8. Conclusion and discussions. This article proposes a novel nonnegative tensor completion method to conduct counterfactual prediction of mortality rates of the COVID-19 pandemic in different countries under various social mobility restriction policies. The proposed method is built upon a CP decomposition of the observed pandemic data, under the constraints that all the factor matrices are nonnegative, the factor matrix in the mobility mode is ordinal, and the factor matrix in the time mode possesses the fusion structure. The ordinal structure on the mobility mode is motivated from the general belief that restrictive social mobility policies reduce down the mortality rates. We also find out that neighboring countries that are geographically close tend to share similar pandemic evolution patterns. The effectiveness of the proposed method is also supported by numerical experiments on various synthetic datasets and asymptotic consistency in terms of parameter estimation.

We would like to remark that the independent intervention assumption is one of the key assumptions to facilitate the theoretical development, while real-life dynamic intervention mechanism can be more complicated. For example, countries may decide their intervention policy by optimizing their expected potential outcomes or taking the possibly time-varying covariate information into consideration. Consequently, it can be reasonable to allow the intervention $X_{n,t}$ depend on the historical observations and the current available covariate information, leading to possible extension of $\{X_{n,t}\}_{t=1}^{T}$ to a martingale or other discrete-time stochastic processes. Another possible extension of the current missing mechanism is to allow a small fixed part of the data tensor to be missing, such as periodic missing for one country. The current theoretic framework can still accommodate such a fixed and stochastic mixture missing pattern as long as the number of observations is substantially larger than the inherent number of parameters to be estimated. Without diluting the focus of the article, we leave there two lines of research as future works.

Acknowledgments. We thank the Associate Editor and two anonymous referees, whose constructive comments and suggestions have led to significant improvements of the article. We also thank Miss Ruixuan Zhao for her helpful discussion on counterfactual prediction.

Funding. This work is supported in part by HK RGC Grants GRF-11304520, GRF-11301521, GRF-11311022, CUHK Startup Grant 4937091, and CUHK Direct Grant 4053588.

SUPPLEMENTARY MATERIAL

Supplementary Materials to "Nonnegative tensor completion for dynamic counterfactual prediction on COVID-19 pandemic" (DOI: 10.1214/23-AOAS1787SUPP; .zip). The Supplementary Material contain necessary lemmas, technique proofs, processed data and relevant codes for all numerical experiments.

REFERENCES

AGARWAL, A., SHAH, D. and SHEN, D. (2022). Synthetic interventions. ArXiv preprint. Available at arXiv:2006.07691.

ATHEY, S., BAYATI, M., DOUDCHENKO, N., IMBENS, G. and KHOSRAVI, K. (2021). Matrix completion methods for causal panel data models. *J. Amer. Statist. Assoc.* **116** 1716–1730. MR4353709 https://doi.org/10.1080/01621459.2021.1891924

ATHEY, S., IMBENS, G. W. and WAGER, S. (2018). Approximate residual balancing: Debiased inference of average treatment effects in high dimensions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 597–623. MR3849336 https://doi.org/10.1111/rssb.12268

BAI, J. and NG, S. (2021). Matrix completion, counterfactuals, and factor analysis of missing data. *J. Amer. Statist. Assoc.* **116** 1746–1763. MR4353711 https://doi.org/10.1080/01621459.2021.1967163

BI, X., QU, A. and SHEN, X. (2018). Multilayer tensor factorization with applications to recommender systems. *Ann. Statist.* **46** 3308–3333. MR3852653 https://doi.org/10.1214/17-AOS1659

CAI, C., LI, G., POOR, H. V. and CHEN, Y. (2019). Nonconvex low-rank tensor completion from noisy data. *Adv. Neural Inf. Process. Syst.* **32**.

CAI, J.-F., LI, J. and XIA, D. (2022a). Generalized low-rank plus sparse tensor estimation by fast Riemannian optimization. *J. Amer. Statist. Assoc.* 1–17.

CAI, J.-F., LI, J. and XIA, D. (2022b). Provable tensor-train format tensor completion by Riemannian optimization. *J. Mach. Learn. Res.* **23** Paper No. [123], 77. MR4577075

CHEN, B., SUN, T., ZHOU, Z. and ZENG, Y. (2019). Nonnegative tensor completion via low-rank Tucker decomposition: Model and algorithm. *IEEE Access* 95903–95914.

DONG, E., DU, H. and GARDNER, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20** 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1

FAN, J., MASINI, R. and MEDEIROS, M. C. (2022). Do we exploit all information for counterfactual analysis? Benefits of factor models and idiosyncratic correction. *J. Amer. Statist. Assoc.* **117** 574–590. MR4436297 https://doi.org/10.1080/01621459.2021.2004895

FAN, Y., LU, X., ZHAO, J., FU, H. and LIU, Y. (2022). Estimating individualized treatment rules for treatments with hierarchical structure. *Electron. J. Stat.* **16** 737–784. MR4366820 https://doi.org/10.1214/21-ejs1948

FOGARTY, C. B. (2020). Studentized sensitivity analysis for the sample average treatment effect in paired observational studies. *J. Amer. Statist. Assoc.* **115** 1518–1530. MR4143482 https://doi.org/10.1080/01621459.2019.1632072

GOUTTE, C. and AMINI, M.-R. (2010). Probabilistic tensor factorization and model selection. In *Tensors, Kernels, and Machine Learning* (*TKLM* 2010) 1–4.

HÖFLER, M. (2005). Causal inference based on counterfactuals. *BMC Med. Res. Methodol.* **5** 1–12.

HU, J., LEE, C. and WANG, M. (2022). Generalized tensor decomposition with features on multiple modes. *J. Comput. Graph. Statist.* **31** 204–218. MR4387221 https://doi.org/10.1080/10618600.2021.1978471

JAIN, P. and OH, S. (2014). Provable tensor factorization with missing data. *Adv. Neural Inf. Process. Syst.* **27**.

JI, P. and JIN, J. (2016). Coauthorship and citation networks for statisticians. *Ann. Appl. Stat.* **10** 1779–1812. MR3592033 https://doi.org/10.1214/15-AOAS896

JOHANSSON, F., SHALIT, U. and SONTAG, D. (2016). Learning representation for counterfactual inference. In *International Conference on Machine Learning* 3020–3029.

KOLDA, T. G. and BADER, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51** 455–500. MR2535056 https://doi.org/10.1137/07070111X

LAUER, S. A., GRANTZ, K. H., BI, Q., JONES, F. K., ZHENG, Q., MEREDITH, H. R., AZMAN, A. S., REICH, N. G. and LESSLER, J. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Ann. Intern. Med.* **172** 577–582. https://doi.org/10.7326/M20-0504

LIU, A. and MOITRA, A. (2020). Tensor completion made practical. In *Advances in Neural Information Precessing Systems* **33** 18905–8916.

MANDAL, D. and PARKES, D. (2022). Weighted tensor completion for time-series causal inference. ArXiv preprint. Available at arXiv:1902.04646.

MCALOON, C., COLLINS, Á., HUNT, K., BARBER, A., BYRNE, A. W., CASEY, M., GRIFFIN, J., LANE, E., MCEVOY, D. et al. (2020). Incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research. *BMJ Open* **10** e039652.

MEN, K., LI, Y., WANG, X., ZHANG, G., HU, J., GAO, Y., HAN, A. and LIU, W. (2023). Estimate the incubation period of coronavirus 2019 (COVID-19). *Comput. Biol. Med.* **158** 106794.

MO, W. and LIU, Y. (2022). Efficient learning of optimal individualized treatment rules for heteroscedastic or misspecified treatment-free effect models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 440–472. MR4412993 https://doi.org/10.1111/rssb.12474

MO, W., QI, Z. and LIU, Y. (2021). Learning optimal distributionally robust individualized treatment rules. *J. Amer. Statist. Assoc.* **116** 659–674. MR4270012 https://doi.org/10.1080/01621459.2020.1796359

NEWEY, W. K. and STOULI, S. (2022). Heterogeneous coefficients, control variables and identification of multiple treatment effects. *Biometrika* **109** 865–872. MR4472853 https://doi.org/10.1093/biomet/asab060

PAPADOGEORGOU, G., IMAI, K., LYALL, J. and LI, F. (2022). Causal inference with spatio-temporal data: Estimating the effects of airstrikes on insurgent violence in Iraq. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 1969–1999. MR4515563 https://doi.org/10.1111/rssb.12548

PEARL, J. (2009). Causal inference in statistics: An overview. *Stat. Surv.* **3** 96–146. MR2545291 https://doi.org/10.1214/09-SS057

POULOS, J., ALBANESE, A., MERCATANTI, A. and LI, F. (2022). Retrospective causal inference via matrix completion, with an evaluation of the effect of European integration on cross-border employment. ArXiv preprint. Available at arXiv:2106.00788.

QI, Z., LIU, D., FU, H. and LIU, Y. (2020). Multi-armed angle-based direct learning for estimating optimal individualized treatment rules with various outcomes. *J. Amer. Statist. Assoc.* **115** 678–691. MR4107672 https://doi.org/10.1080/01621459.2018.1529597

QUICK, C., DEY, R. and LIN, X. (2021). Regression models for understanding COVID-19 epidemic dynamics with incomplete data. *J. Amer. Statist. Assoc.* **116** 1561–1577. MR4353694 https://doi.org/10.1080/01621459.2021.2001339

ROHE, K., QIN, T. and YU, B. (2016). Co-clustering directed graphs to discover asymmetries and directional communities. *Proc. Natl. Acad. Sci. USA* **113** 12679–12684. MR3576189 https://doi.org/10.1073/pnas.1525793113

ROSENBAUM, P. R. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41–55. MR0742974 https://doi.org/10.1093/biomet/70.1.41

SINHA, S. and CHAKRABORTY, M. (2022). Causal analysis and prediction of human mobility in the US during the COVID-19 pandemic. ArXiv preprint. Available at arXiv:2111.12272.

SUN, W. W. and LI, L. (2019). Dynamic tensor clustering. *J. Amer. Statist. Assoc.* **114** 1894–1907. MR4047308 https://doi.org/10.1080/01621459.2018.1527701

WILSON, N., KVALSVIG, A., BARNARD, L. T. and BAKER, M. G. (2020). Case-fatality risk estimates for COVID-19 calculated by using a lag time for fatality. *Emerg. Infec. Dis.* **26** 1339–1441. https://doi.org/10.3201/eid2606.200320

XIA, D. and YUAN, M. (2019). On polynomial time methods for exact low-rank tensor completion. *Found. Comput. Math.* **19** 1265–1313. MR4029842 https://doi.org/10.1007/s10208-018-09408-6

XU, Y. and YIN, W. (2013). A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM J. Imaging Sci.* **6** 1758–1789. MR3105787 https://doi.org/10.1137/120887795

XUE, F. and QU, A. (2021). Integrating multisource block-wise missing data in model selection. *J. Amer. Statist. Assoc.* **116** 1914–1927. MR4353722 https://doi.org/10.1080/01621459.2020.1751176

YADLOWSKY, S., PELLEGRINI, F., LIONETTO, F., BRAUNE, S. and TIAN, L. (2021). Estimation and validation of ratio-based conditional average treatment effects using observational data. *J. Amer. Statist. Assoc.* **116** 335–352. MR4227698 https://doi.org/10.1080/01621459.2020.1772080

YAN, H., ZHU, Y., GU, J., HUANG, Y., SUN, H., ZHANG, X., WANG, Y., QIU, Y. and CHEN, S. X. (2021). Better strategies for containing COVID-19 pandemic: A study of 25 countries via a vSIADR model. *Proc. R. Soc. A* **477** Paper No. 20200440, 25. MR4258333

YE, Y., ZHANG, Q., CAO, Z., CHEN, F. Y. YAN, H., STANLEY, H. E. and ZENG, D. D. (2021). Impact of export restrictions on the global personal protective equipment trade network during COVID-19. *Adv. Theory Simul.* 2100352.

YUAN, M. and ZHANG, C.-H. (2016). On tensor completion via nuclear norm minimization. *Found. Comput. Math.* **16** 1031–1068. MR3529132 https://doi.org/10.1007/s10208-015-9269-5

YUAN, M. and ZHANG, C.-H. (2017). Incoherent tensor norms and their applications in higher order tensor completion. *IEEE Trans. Inf. Theory* **63** 6753–6766. MR3707566 https://doi.org/10.1109/TIT.2017.2724549

ZHANG, J., SUN, W. W. and LI, L. (2020). Mixed-effect time-varying network model and application in brain connectivity analysis. *J. Amer. Statist. Assoc.* **115** 2022–2036. MR4189774 https://doi.org/10.1080/01621459.2019.1677242

ZHANG, X. and NG, M. K. (2022). Sparse nonnegative tensor factorization and completion with noisy observations. *IEEE Trans. Inf. Theory* **68** 2551–2572. MR4413569

ZHANG, Y., BI, X., TANG, N. and QU, A. (2021). Dynamic tensor recommender systems. *J. Mach. Learn. Res.* **22** Paper No. 65, 35. MR4253758

ZHEN, Y. and WANG, J. (2024). Supplement to "Nonnegative tensor completion for dynamic counterfactual prediction on COVID-19 pandemic." https://doi.org/10.1214/23-AOAS1787SUPP

ZHOU, J., SUN, W. W., ZHANG, J. and LI, L. (2023). Partially observed dynamic tensor response regression. *J. Amer. Statist. Assoc.* **118** 424–439. MR4571132 https://doi.org/10.1080/01621459.2021.1938082