

基于 Bootstrap 算法的多国 Logistic 增长率 指标与变点检验

摘要:自 2020 年 1 月以来,新型冠状病毒疫情迅速在全球蔓延,各地确诊病例不断增加,社会经济生活受到了极大干扰,对此各国采取了多种指标来衡量本国的疫情状况。本文采用简单科学的 Logistic 函数的增长率指标衡量国家疫情发展情况,并同时估计增长率的变点参数,检测在不同时间上增长率的变化差异,分析各个国家疫情的防控情况,最后结合 Bootstrap 算法构造置信区间建立回归模型。本文对美国、意大利和日本三个国家从疫情初期到 2021 年 2 月 21 日的数据进行回归,发现各个地区疫情发展状况不同。此算法适用于多种情形下的疫情检测,对疫情当前状况和未来发展趋势能提供参考。

关键词: 新冠疫情; 增长率; 变点; Logistic 函数; Bootstrap 算法

1 引言

自新型冠状病毒肺炎(COVID-19)疫情爆发以来,全球累计一亿多人确诊,两百多个国家受到了不同程度的影响,而近来,部分国家又陆续出现第四波疫情,不同国家也出台了相应政策,欧盟每半个月调整一次准入国家和地区的名单,根据欧盟的标准,平均每 10 万人 7 天内新感染病例超过 200 例的国家和地区被视为高风险,新冠变种病毒传播现象严重的国家也将被列入联邦疾控机构的一份相关名单;美国一些州通过医院床位利用率,新冠病毒基本传染数和连续每日新增感染人数来判断该州是否能进行重启;意大利通过各地每日新增病例数量,将全国分为红色、橙色和绿色区域。红色区域的感染率最高,将关闭所有酒吧、餐馆和大多数商店,包括美发店和美容院,在橙色区域,餐馆和酒吧将关闭,但美发店和美容院可以继续营业。全球已接种疫苗超过 3 亿剂。发达国家和部分中等收入国家已开展大规模接种。美国以 8054 万剂领跑各国疫苗接种总量,中国和欧盟紧随其后,国际社会的支持一定程度上缓解了发展中国家的疫苗危机,但缺口仍然巨大¹。上述的这些疫情发展的指标略显刻板且缺乏科学性的解释,有的措施不力导致了疫情的反复爆发,如何来评估疫情发展情况、疫情什么时候出现拐点以及怎样才能预测未来疫情的发展等,已成为全球共同关注的重大问题。

本文通过对疫情期间 logistic 种群增长模型的增长率的变化来判断疫情的变化,通过寻找增长率的变化,来反映疫情的变化情况,为各国政府提供一个可以参考的直观的疫情数据。Logistic 方程是由比利时数学家兼生物学家皮埃尔·弗朗索瓦·韦吕勒(Pierre Francois Verhulst)在研究人口增长模型时提出,是对马尔萨斯人口模型(Malthus, 1798)的改进。Shen, C. Y.[1]通过将中国省级和国家级两个层面进行对比,通过采用 NLS 估计中国多个地区和其他选定国家的新的每日新冠肺炎病例的微分逻辑增长函数的参数,考察了逻辑增长模型的适用性,以及对新冠肺炎大流行和其他传染病研究的启示。结果表明 logistic 模型能够很好的拟合中国等 11 个国家的疫情数据。

¹ Data sources: <https://ourworldindata.org/covid-vaccinations>

随后对 Logistic 增长率建立带变点的面板数据模型，利用虚拟变量检验在间隔七天和间隔十四天的增长率中是否有变点出现，即检验增长率在间隔七天和间隔十四天间是否有显著差异，以上述数据来判断疫情变化程度及防控效果。变点理论基本定义是在一个序列或过程中，当某个统计特性在某时间点受系统性因素而非偶然性因素影响发生变化，我们就称该时间点为变点。Dehning, J, et al.[2]通过将已建立的流行病学模型与贝叶斯推论相结合，分析了新感染有效增长率的时间依赖性。着眼于在德国传播的 COVID-19，发现有效增长率的变化点与公开宣布的干预措施的时间密切相关。因此，可以量化干预措施的效果，并且可以将相应的变更点合并到对未来方案和案例数量的预测中。

同时由于样本量比较少，增长率参数以及变点参数并不一定符合标准的 t 分布，所以采用基于残差抽样的 bootstrap[3]进行置信区间的计算。Bootstrap 是一类非参数 Monte Carlo 方法,其实质是对观测信息进行再抽样，进而对总体的分布特性进行统计推断。该方法充分利用了给定的观测信息，不需要模型其他的假设和增加新的观测，并且具有稳健性和效率高的特点。Zhang, Sheng, et al.[4]用伽玛分布拟合报告的序列区间来估计 COVID-19 爆发早期的 R_0 ，模拟可能的累积流行轨迹和未来的日发病率传染性，随后用 bootstrap 重采样法估计 R_0 值 95%可信区间 (CI) 的中位数为 2.28 (2.06-2.52)，得出了除非采取严格的感染管理和控制措施，否则 COVID-19 可能会在造成更大的爆发的结论。

研究结果表明，不同地区的国家疫情爆发的先后时间不同，亚洲是最先爆发疫情的，而不同地区的国家疫情发展也不同步。所以对于不同地区的不同国家应该采取不同的防疫措施。增长率的置信区间大部分都不包含 x 轴，拒绝原假设，说明模型的参数为显著，具有统计学意义。同时增长率与变点参数几乎同步变化，证明了模型的准确性和可靠性。

在其它文章中，Cooper[6]研究了由于新型 COVID-19 疾病的传播而对大流行采取建模方法的有效性，并开发了一个易受感染的去除 (SIR) 模型，为调查其在社区内的传播提供了理论框架。Fna, B, et al.[7]提出了一个用于 COVID-19 疾病传播的分门别类的数学模型，他们计算基本繁殖数阈值，研究疾病自由平衡的局部稳定性，研究基本繁殖数，并调查模型对每一个参数变化的敏感性。Aza, B, et al.[8]已应用简单的经常性神经网络 (RNN)、长期短期内存 (LSTM)、双向 LSTM (BILSTM)、门控常发单元 (GRUS) 和变异自动编码器 (VAE) 算法进行 COVID-19 病例的全球预测。这项研究基于从意大利、西班牙、法国、中国、美国和澳大利亚等六个国家收集的每日确诊和复发病例。Wang, P., et al.[9]将 2020 年 6 月 16 日之前更新的 COVID-19 流行病学数据集集成到符合流行趋势上限的物流模型中，然后将上限值输入 FbProphet 模型，这是一个基于机器学习的时间系列预测模型，以推导流行病曲线并预测流行趋势。Annas, S., et al.[10]建立了 COVID-19 的 SEIR 模型，并利用印度尼西亚地区的数据对 SEIR 模型对 COVID-19 扩散的稳定性进行分析和数值模拟。

本文结构安排如下。第二节从 logistic 函数导出增长率参数的回归模型，并给

出增长率的估计方法以及变点检验方法。第三节介绍了研究数据。第四节给出增长率参数的估计结果和变点的检验结果。第五节是结论。

2 Logistic 增长率的回归模型和变点检验

2.1 Logistic 增长率回归模型

在进行流行病学研究的时候,我们通常会使用 Logistic 模型进行建模和研究,传染病力学模型的微分方程为:

$$\frac{dN_t}{dt} = \beta \cdot \left(1 - \frac{N_t}{K}\right) N_t \quad (1)$$

$$N_{t_0} = N_0 \quad (2)$$

其中, N_0 为疫情初期 t_0 时刻感染人数; N_t 为 t 时刻感染人数; β 为感染人数的增长率; K 为最终感染人数(可变), 假设疫情持续蔓延, 各地区最终感染数为该地区人口总数。

解微分方程可得:

$$N_t = \frac{K}{1 + C \cdot e^{-\beta(t-t_0)}}, \quad C = \frac{K - N_0}{N_0} \quad (3)$$

将微分方程(1)离散化为

$$\frac{\Delta N_t}{\Delta t} = \beta \cdot \left(1 - \frac{N_t}{K}\right) N_t \quad (5)$$

令 $\Delta t = 1$, 则

$$N_{t+1} - N_t = \beta \cdot \left(1 - \frac{N_t}{K}\right) N_t \quad (6)$$

得到预测方程:

$$\frac{N_{t+1} - N_t}{K} = (1 + \beta) \cdot N_t - \beta \cdot \frac{N_t^2}{K} \quad (7)$$

方程(6)两边同时除以 K , 得到

$$\frac{N_{t+1} - N_t}{K} = \beta \cdot \left(\frac{K - N_t}{K}\right) \frac{N_t}{K} \quad (8)$$

当 $N_{t+1} - N_t = 0$ 时, 可以得出第 $t+1$ 天无新增确诊人数, 疫情情况明显好转的结论, 即 $\beta = 0$ 。为了寻找这个疫情的转折点, 本文通过 OLS 回归方程来估计 β 并检验 $\beta = 0$ 的显著性来进行研究。

令 $y_{t+1} = \frac{N_{t+1} - N_t}{K}$, $x_t = \left(\frac{K - N_t}{K}\right) \frac{N_t}{K}$, 代入(8)式中得到数学方程:

$$y_{t+1} = \beta \cdot x_t \quad (9)$$

其中, y_{t+1} 和 x_t 的含义分别为第 $t+1$ 天增加的确诊(感染)人数占总人数的比率和第 t 天已确诊人数比率和未确诊人数比率的乘积。

2.2 截面数据估计

采用截面数据估计的方式计算动态增长率 β , 截面数据是指由同一时期、不同个体的一个或多个统计指标所组成的数据集。以美国的数据为例, 采用美国 56 个州的截面数据, 选择 1 天作为时间点, 利用截面数据回归估计这个国家增长率, 移动时间窗口向前, 得到一系列随时间(窗口)变化的动态增长率。

$$y_{t+1,i} = \alpha_i + \beta_i x_{t,i} + \varepsilon_{t+1,i} \quad (10)$$

其中 α_i 是截距项, $\varepsilon_{t+1,i}$ 是第 i 州第 $t+1$ 天的误差项, β_i 是增长率

我们利用截面数据, 结合方程 (10) 估计每一天的增长率 $\hat{\beta}_t$, 移动时间窗口估计得到增长率的一个时间序列。这里仅估计增长率的截面数据分析疫情发展情况, 不考虑感染人数的预测问题。注意到回归模型 (10) 式中没有常数项。当自变量 $x_{t,i} = 0$ 时, 如果存在非零的常数项使得因变量 $y_{t+1,i} \neq 0$, 这与实际情况不符。

方程 (10) 也是一个预测回归模型, 在估计参数的基础上, 我们可以利用第 t 天的 x_t (第 t 天已确诊人数比率和未确诊人数比率的乘积) 预测第 $t+1$ 的 y_{t+1} (第 $t+1$ 天增加的确诊(感染)人数占总人数的比率)。

利用 OLS 估计方程 (10) 中增长率 β_i , 再用 t 统计量方法检验原假设 $H_0: \beta_i = 0$ 和对立假设 $H_1: \beta_i \neq 0$ 。如果给定的显著性水平 α , 关于估计参数 $\hat{\beta}_i$ 的 t 统计量不大于其临界值, 则不拒绝 H_0 ; 否则拒绝 H_0 。

2.3 变点检测

接下来引入时间序列中变点来分析疫情的发展变化。本文利用截面数据, 采用 Logistic 回归模型, 引入时间序列中变点这一概念(change points or breaks)来分析疫情的发展变化。为了检验第 s 天和第 $s+k$ 天的增长率是否有显著差异, 设定虚拟变量 d_t 。当 $t=s$ 时, $d_t=0$, 当 $t=s+k$ 时, $d_t=1$ 。带变点的面板回归模型为

$$y_{t+1,i} = \alpha + \beta x_{t,i} + \beta d_t + \varepsilon_{t+1,i}, \quad (11)$$

其中 $i=1, \dots, m$, $t=s, s+k$ 再次利用 OLS 估计回归模型 (11) 式, 并应用 t 检验方法检验变点参数 $\hat{\beta}$ 的显著性。

对参数原假设 $H_0: \beta=0$ 的检验结果有以下三种情况:

(i) 不拒绝 H_0 , 表示第 $s+k$ 天的增长率和第 s 天的增长率相等, 疫情情况没有发生显著变化。

(ii) 拒绝 H_0 且 $\hat{\beta}$ 的符号为正, 表明第 $s+k$ 天的增长率显著大于第 s 天的增长率, 疫情向更坏的方向发展。

(iii) 拒绝 H_0 且 $\hat{\beta}$ 的符号为负, 表明第 $s+k$ 天的增长率显著小于第 s 天的增长率, 疫情向好的方向发展。

方程（8）中 $k=1$ 时，检验 β 的显著性就是检验相邻两天的增长率是否发生显著性变化；当 $k=7$ 时检验相隔一周的增长率是否发生了显著变化。不同的 k 值可以比较不同时间上的增长率情况。

以美国疫情为例，根据每一天，美国各州的疫情数据进行 logistic 回归，对假设 $\tau=0$ 进行显著性检验，数据分别间隔 1 天，7 天，14 天取得，最终绘制出 τ 和 τ 的 95% 的置信区间曲线。

2.4 基于残差的 Bootstrap 方法

该方法通过对回归拟合后的残差进行重抽样得到 Bootstrap 复制。

第一步，首先通过最小二乘方法对参数 β_0, β_1 进行估计，从而得到残差：

$$\hat{\varepsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 \cdot X_i;$$

第二步，对中心化的残差 $\{\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}, i = 1, \dots, n\}$ 进行又放回抽样得到 Bootstrap 的误差 ε_i^* 。令 $Y_i^* = \hat{\beta}_0 + \hat{\beta}_1 \cdot X_i + \varepsilon_i^*$ ；

第三步，基于 (X_i, Y_i^*) 重新拟合回归模型，得到 $\hat{\beta}_0^*, \hat{\beta}_1^*$ ；

第四步，重复步骤 2-3 B 次，得到相应 Bootstrap 复制 $\hat{\beta}_{01}^*, \hat{\beta}_{11}^*, \dots, \hat{\beta}_{0B}^*, \hat{\beta}_{1B}^*$ ；

第五步，对所有生成的 β 值进行从小到大排序，将其作为 β_1 的分布，选取显著性水平对应的百分位数作为回归参数 β_1 的置信区间。

3 数据

截止到 2021 年 3 月 15 日，全球新冠肺炎确诊病例数为 119875420，疫情仍在某些国家肆虐。其中美国累计确诊病例为 29438775，为全球最高，意大利累计确诊病例为 3223142，日本累计确诊病例 447326²。

本文选择的样本数据包含了自疫情初期 2020 年 1 月 21 日至 2021 年 2 月 21 日期间的疫情发展信息，包括每日新冠肺炎累计确诊人数、新增确诊人数及各州（或各县）的人口总数，数据来源于 Wind 数据库及 Harvard Dataverse[5]，具体的样本数据来源选取如表 1 所示，美国整体以 56 个州为样本，意大利整体以 20 个区为样本，日本整体以 47 个行政区（都、道、府、县）为样本。

表 1 样本数据来源

国家	日期	包含地区	数据来源
美国	2020/1/21-2021/2/21	56 个州	Wind 数据库
意大利	2020/1/31-2021/2/21	20 个区	Wind 数据库
日本	2020/3/7-2020/2/21	47 个行政区	Wind 数据库

4 结果

基于选择的样本数据，利用回归方程（10）求出七日面板数据估计增长率 $\hat{\beta}_7$

² Data source: <https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>

和十四日面板数据 $\hat{\beta}_{14}$ ，再用 bootstrap 方法随机抽样生成增长率的置信区间，根据参数 $\hat{\beta}_7$ 和 $\hat{\beta}_{14}$ 的数据量级及显著性结果来判断疫情发展情况和防控效果。利用回归方程（11）分别估计相隔一周变点参数 τ_7 和相隔两周变点参数 τ_{14} ，也采用 bootstrap 算法生成变点参数的置信区间，并检验参数显著性，判断不同时间上增长率的变动情况。由于使用一天的截面数据大部分情况下都没有较为显著的统计学意义，在此不讨论使用一天截面数据的结果。

参数估计及检验结果如图 1 至图 6 所示。图 1 至图 6 分别美国、意大利和日本参数 $\hat{\beta}_7$ 和 $\hat{\beta}_{14}$ ， τ_7 和 τ_{14} 的时序图，且给出 95% 的置信区间。如果 0 包含在置信区间内，则原假设成立，参数不显著；否则估计的参数显著。其中每个国家均分解为疫情前期和后期，前期指的是疫情刚爆发，对疫情的治理比较欠缺的时期；后期指的是在经过政府干预治理后不断好转的时期。

需要说明的是，采用 bootstrap 算法生成的置信区间具有一定随机性，每次结果可能略微不同，同时随着抽样次数的增加，置信区间分布会越来越接近真实区间分布。

4.1 美国

2 月末到 3 月末：

增长率突然上升并剧烈波动， τ 值来回穿过横坐标，且大部分时候为正。说明新冠疫情在此时爆发，政府采取了相应的措施但效果不佳，部分人民并不听取政府建议，政府内部官员也意见不一。疫情在此时很不稳定，并向着更坏的方向发展。

4 月初到 6 月中旬：

增长率逐渐下降， τ 值大部分情况为负，在横坐标轴上波动，表明增长率下降明显。这是因为随着新冠疫情愈发严重，政府加大了管控力度，自 3 月中旬启动全国封锁措施阻止新冠病毒传播，扩大社交距离、居家令、停课令等措施取得了成效，疫情上升曲线渐趋平缓，人们也逐渐意识到病毒的危害性，疫情得到了逐步缓解。

6 月中旬到 7 月中旬：

增长率显著上升， τ 值显著为正。代表此阶段增长率不断增大，疫情在不断恶化。主要是因为之前一段时间种族歧视问题导致的多地暴乱集会的“后遗症”，同时美国政府为了选举需要重启经济而轻视疫情，以及美国部分民众受政府影响不重视保持社交距离和戴口罩，导致疫情形势恶化。

10 月份到 1 月份：

增长率显著为正且迅速上升， τ 值也显著为正，代表疫情显著恶化，秋冬疫情的第三次袭击开始。这是由于美国部分州眼见疫情好转，为了恢复经济解除部分防疫措施所致。

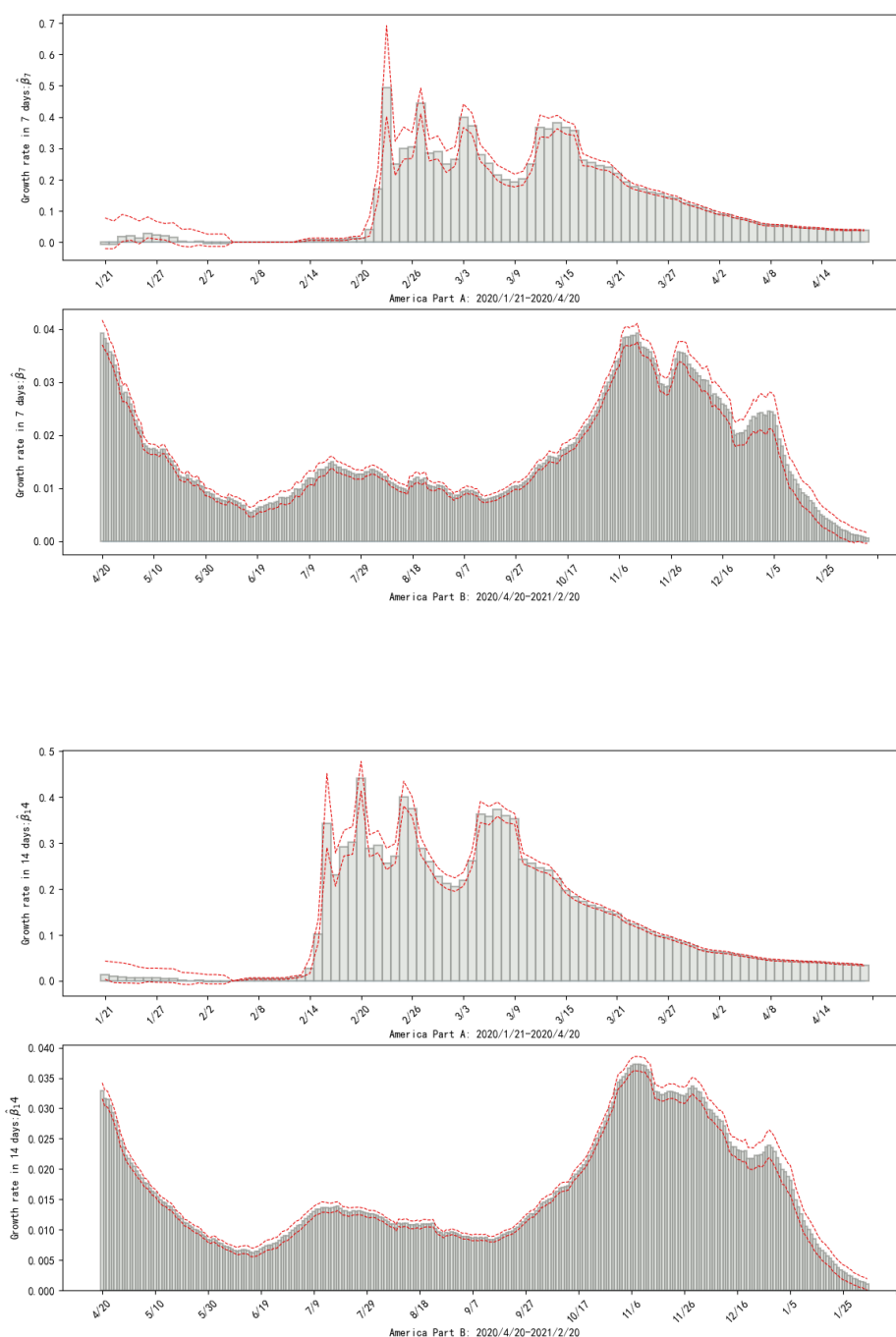


图 1 美国增长率: 黑色条形表示参数估计值, 灰色阴影表示 95%置信区间. 横坐标是时间 (单位: 日), 纵坐标

标 $\hat{\beta}_7$ 是每周增长率, $\hat{\beta}_{14}$ 是每两周增长率.

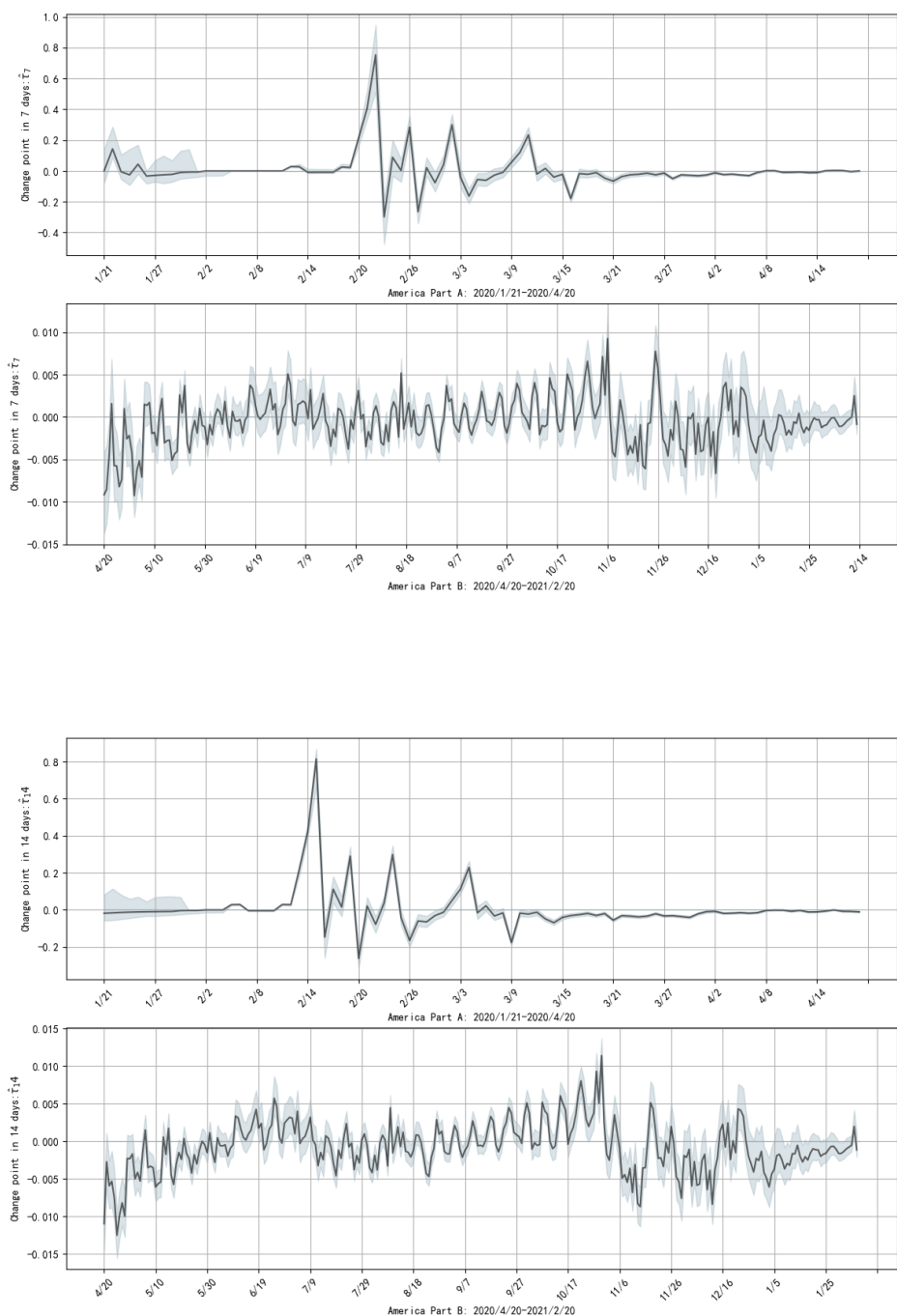


图2 美国变点参数: 黑色条形表示估计值, 灰色阴影表示 95%置信区间.横坐标是时间 (单位:日), 纵坐标 τ_7 是间隔一周估计值, τ_{14} 是间隔两周估计值.

4.2 意大利

2 月到 9 月:

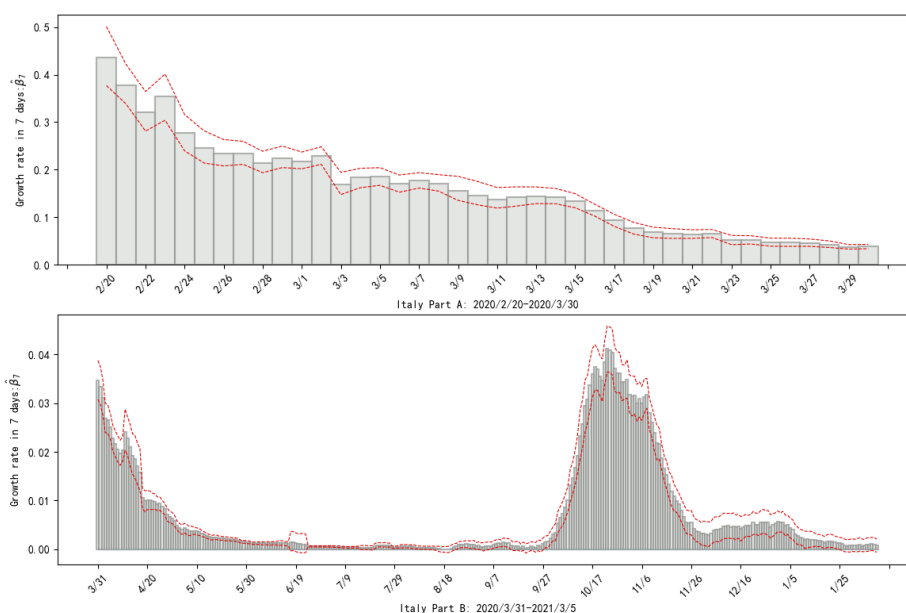
增长率接近于最高峰, 随后持续下降; τ 值来回穿过横坐标, 大部分时候为负.说明新冠疫情在 2 月份爆发, 但是随后政府干预措施得当, 成效显著, 疫情

逐渐得到控制。这是由于意大利部长会议在 2 月 23 日通过法令对病例集中的北部 11 个市镇交通采取进出限制，并暂停生产、公众集会及教学活动，全境加强宣传和消毒，对疫情发展有一定程度的抑制。而且 3 月 1 日意大利隔离法案升级到意大利全境，意大利政府接连发布关闭学校、企业、公园等封城令，暂停航班，虽然没能完全控制住疫情的发展，但是从感染率增长结果来看，也在一定程度上降低了疫情的增长率的快速变化。

另外，6 月 15 日附近出现增长率不显著的情况，可以解读为意大利防疫措施成效显著，此时根据政府规划，政府也将从 6 月 15 日进一步放松管控，在保持社交距离并佩戴口罩的前提下重新开放儿童夏令营、舞厅、影院等场所。

10 月到 11 月：

增长率突然升高，随后下降，增长率的置信区间大部分没有包含 0，在 95% 水平上具有显著性。 τ 值迅速上升为正，随后在横坐标轴上波动，表明从 10 月到 1 月增长率先上升后下降。说明此时疫情反弹，秋冬季第二波疫情爆发，这与政府放宽防疫措施有关。



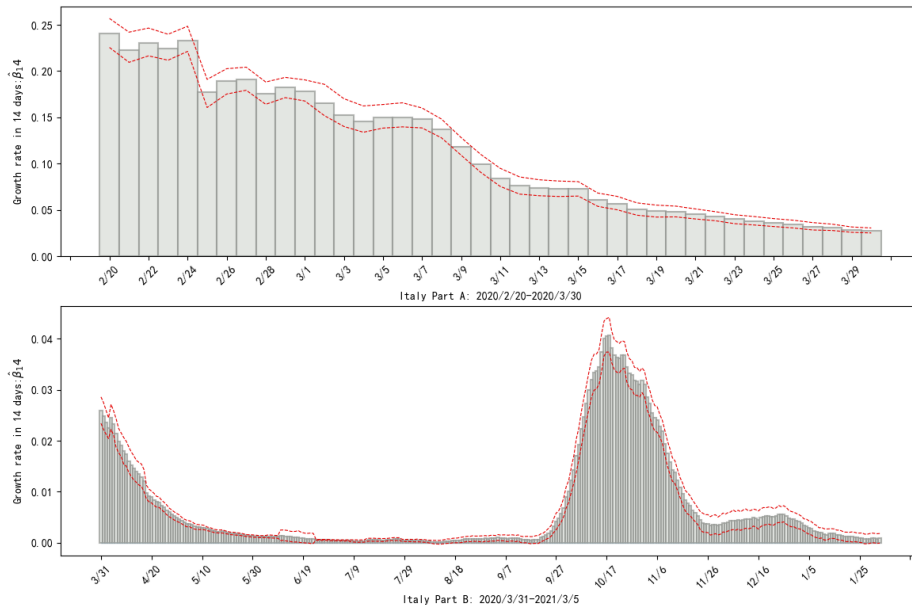
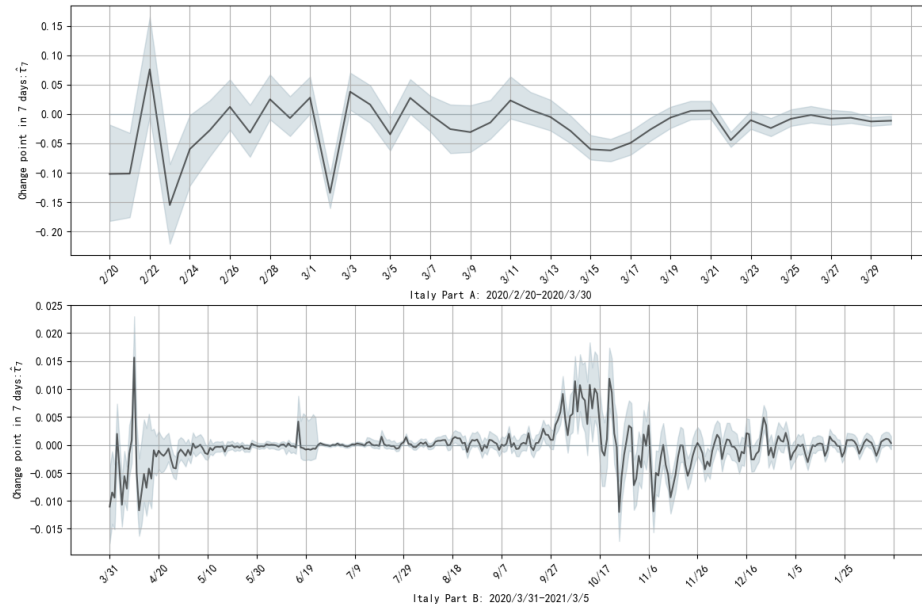


图3 意大利增长率: 黑色条形表示参数估计值, 灰色阴影表示 95%置信区间.横坐标是时间 (单位:日), 纵

坐标 $\hat{\beta}_7$ 是每周增长率, $\hat{\beta}_{14}$ 是每两周增长率.



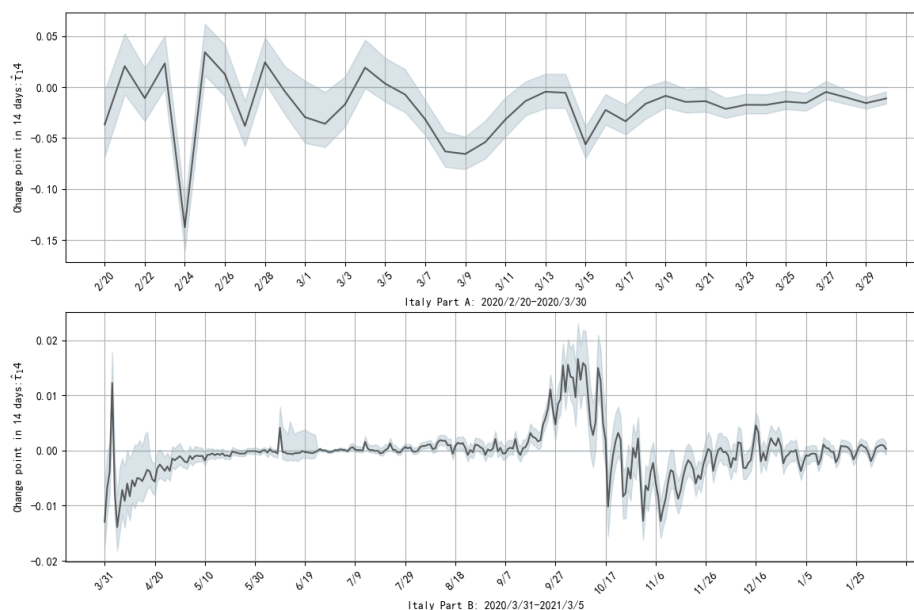


图 4 意大利变点参数: 黑色条形表示估计值, 灰色阴影表示 95%置信区间.横坐标是时间 (单位:日), 纵坐标 τ_7 是间隔一周估计值, τ_{14} 是间隔两周估计值.

4.3 日本

3 月中旬到 4 月中旬第一波疫情:

3 月份增长率显著为正, τ 值显著为正, 说明第一波疫情在 3 月份席卷而来。主要因为 2 月, 日本发现首个新冠病例, 安倍政府为保奥运会, 一直淡化疫情, 导致疫情逐步加重, 酿成第一波大爆发。

6 月末到 8 月初第二波疫情

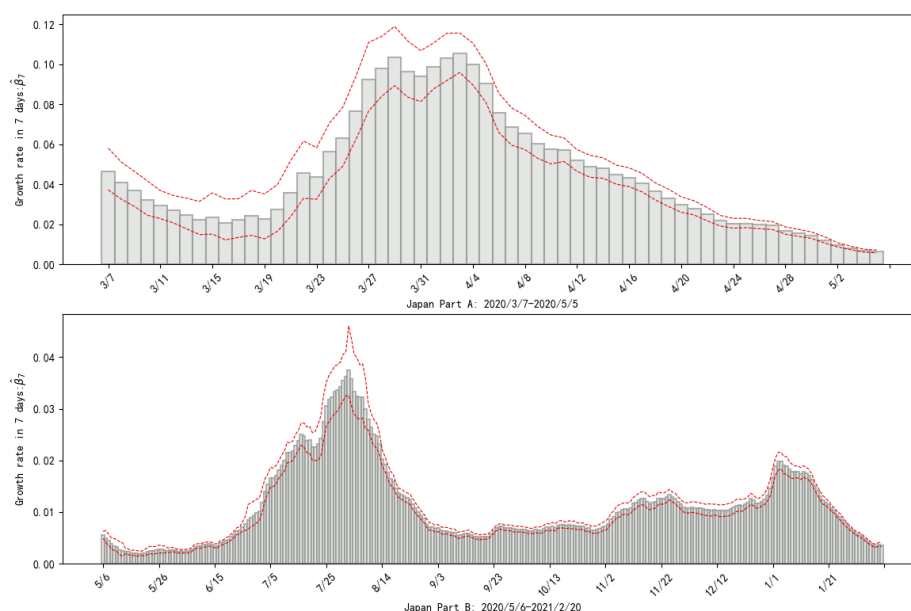
7 月份增长率显著为正, τ 值显著为正, 说明第一波疫情在 7 月份席卷而来。主要因为东京在中央政府解除紧急状态后, 在 6 月 2 日, 发布了都政府制定的“东京警报”, 此后于 6 月 11 日解除警报, 到 6 月 19 日, 又解除了夜间娱乐场所的营业限制, 全面解除停业限制, 而这为疫情反弹埋下了伏笔。

夜生活场所成为了疫情反弹的“温床”, 尤其是东京娱乐场所为中心形成了首都圈的传播链条。比如, 6 月 26 日东京确诊的 54 例中 31 人为牛郎酒吧等夜晚闹市区相关人员; 6 月 30 日确诊的 54 例新增病例中, 15 例是夜间营业娱乐场所的工作人员和顾客。传染链条此后随着人的移动进一步延展至临近地区。和东京相邻的横滨市在 6 月 30 日通报了 28 例新增, 其中 26 例来自同一家牛郎夜店。据报道, 其中有人之前在东京某发生过感染的夜店工作, 这家横滨的夜店目前已

累计确诊 32 例。另有和东京相邻的埼玉县也发生了类似事件，当地一家夜总会的工作人员在东京感染后，进而形成了夜总会内的传播。

11 月初到 1 月末第三波疫情

在 12 月末的时候增长率又一次升高降低，可以看出疫情有轻微反弹，增长率的置信区间大部分没有包含 0，在 95%水平上具有显著性。 τ 值显著为正，代表增长显著增加。主要因为日本没有正视和重视抗疫，在疫情防控措施上，日本没有实施全国范围的封锁，在疫情尚未遏制的情况下急于开放，甚至鼓励外出旅行激活经济；日本企业、餐饮业和商业基本没有关闭。不受限的社会流动，为病毒传播和交叉感染提供了基础。



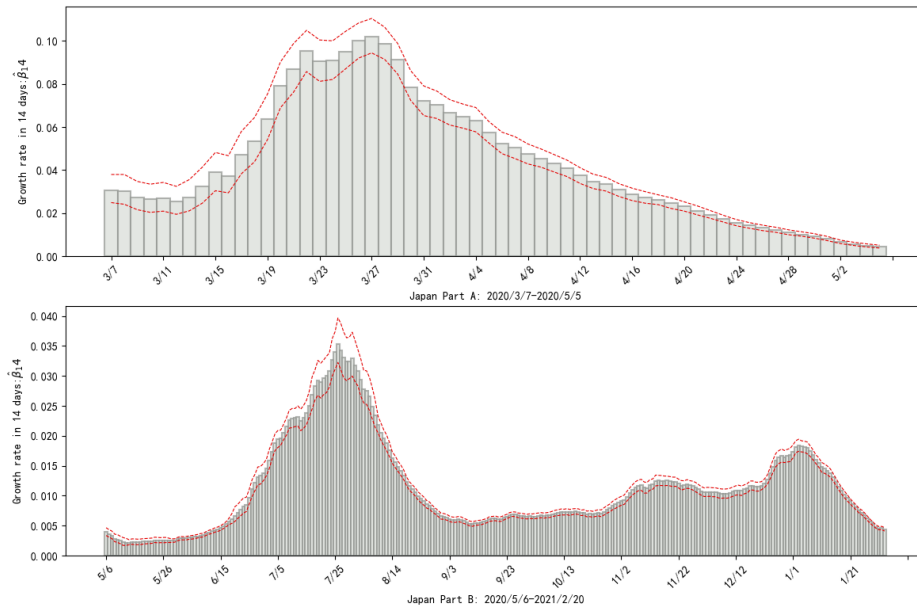
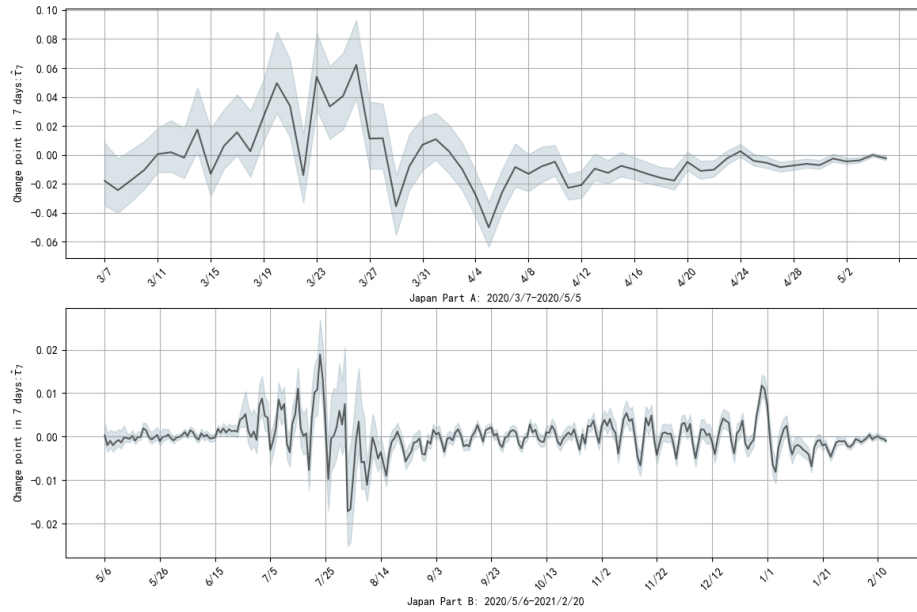


图 5 日本增长率: 黑色条形表示参数估计值, 灰色阴影表示 95%置信区间.横坐标是时间 (单位:日), 纵坐

标 $\hat{\beta}_7$ 是每周增长率, $\hat{\beta}_{14}$ 是每两周增长率.



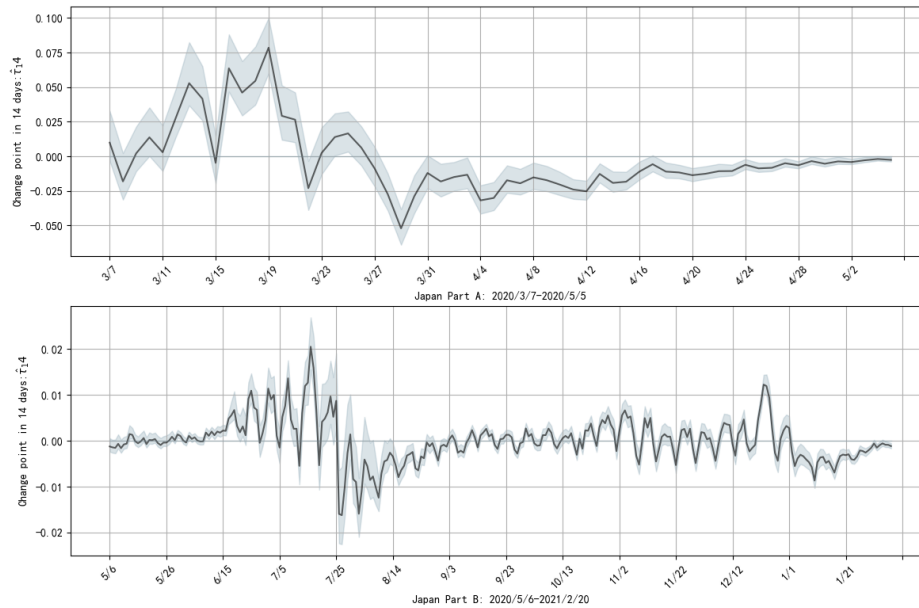


图 6 日本变点参数: 黑色条形表示估计值, 灰色阴影表示 95%置信区间.横坐标是时间 (单位:日), 纵坐标 τ_7 是间隔一周估计值, τ_{14} 是间隔两周估计值.

5 结论与讨论

本研究根据生态学种群增长的 Logistic 函数, 提出增长率指标, 并建立增长率估计回归模型和变点回归模型, 随后采用基于残差抽样的 bootstrap 算法得出置信区间。基于美国、意大利和日本三个国家的行政区数据, 利用滚动窗口估计 Logistic 增长率并进行显著性检验, 为评估疫情发展和重启经济提供参考。根据增长率估计值的数量级发现各个国家及地区的疫情发展阶段不同, 由变点检验的显著性结果, 发现各个国家及地区的防控效果不同。

对三个国家每天的疫情发展情况进行比较、归纳和总结, 可以发现: 第一, 不同地区的国家疫情爆发的先后时间不同, 在本文中, 日本是最先爆发的, 其次是意大利, 随后是美国, 截止到 2021 年 2 月 21 日, 它们增长率最终都接近 0, 说明最后疫情终将被人类控制住。第二, 在实时计算增长率估计值和显著性检验过程中, 即使增长率不显著的情况持续了很长一段时间, 本模型也能快速识别出疫情再次出现明显反弹的时点, 同时增长率变化和变点参数变化几乎同步, 说明了模型的准确性和可靠性。第三, 不同地区的国家, 疫情发展不同步。日本先爆发疫情, 随后管控得当, 快速控制住疫情, 虽然后面有两次轻微的疫情反弹, 但总体上是能够看出增长率呈一个下降的趋势。意大利疫情爆发后增长率在快速上升后开始剧烈波动, 说明政府防疫措施一般, 人民执行力不够高。在持续下降一段时间后, 疫情突然又开始反弹, 这和政府放松防疫措施的政策是相关的。美国在这方面则表现得最为糟糕, 在增长率快速地上升缓慢地下降后又是快速地上升再下降, 这说明美国疫情十分严峻, 政府监管不力, 必须要加强防疫措施。由此可见, 对不同地区应该采用不同的防控管理措施。同时, 本文大部分时段增长率的置信区间都不包含 0, 说明模型得到了数据的支持, 是具有统计学上的显著性意义的。

本文考虑的增长率估计和实时的变点检验适用性(目标对象)广, 灵活度高。在地区上,

除了对国家的行政区进行增长率的估计外，也可以推广到微观上的某个具体的地方，如美国纽约、中国武汉等。利用更小行政单位统计的感染人数数据，采用相同的方法即可以对该地区的增长率进行估计和显著性检验。在时间上，在进行增长率的变点检验时，除了以天、周作为比较的时间间隔、还可以以月为间距进行检验，能够自由地选择时间，从而更明显地判断疫情变化程度及防控效果。同时本文考虑到小样本问题，采用基于残差的 bootstrap 算法得出置信区间，同时还解决了变量不服从 t 分布的问题，使结果更加精确。

Logistic 算法科学权威，适合作为疫情判断的补充标准，从增长率的显著性和数级大小可以判断国家疫情状况，作为国家间统一的参考指标，同时也能用于预测感染人数，预测未来的疫情发展状况。

由于研究结果是基于数据驱动计算得到，依赖于数据的真实性，所以详尽准确的数据资料十分重要。各个疫情国家需要获取准确的疫情数据，利用估计的增长率为评估疫情和重启经济提供科学建议。

参考文献

- [1]. Shen C Y . Logistic growth modelling of COVID-19 proliferation in China and its international implications[J]. International Journal of Infectious Diseases, 2020, 96: 582-589.
- [2]. Dehning J, Zierenberg J, F Paul Spitzner F P, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions[J/OL]. Science, 2020, 368. DOI: 10.1126/science.abb9789.
- [3]. Chatterjee, A., & Lahiri, S. (2010). Asymptotic properties of the residual bootstrap for lasso estimators. Proceedings of the American Mathematical Society, 138(12), 4497-4509.
- [4]. Zhang, Sheng, et al. "Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis." International journal of infectious diseases 93 (2020): 201-204.
- [5]. Hu T, Guan W, Zhu X, et al. Building an open resources repository for COVID-19 research[J], Data and Information Management, published online ahead of print, 2020.000010247820200012. <https://doi.org/10.2478/dim-2020-0012>.
- [6]. Cooper, I. , A. Mondal , and C. G. Antonopoulos . "A SIR model assumption for the spread of COVID-19 in different communities." Chaos Solitons & Fractals 139(2020):110057.
- [7]. Fna, B , et al. "Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan - ScienceDirect." Chaos, Solitons & Fractals 135.
- [8]. Aza, B , et al. "Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study - ScienceDirect." Chaos, Solitons & Fractals 140.
- [9]. Wang, P. , et al. "Prediction of Epidemic Trends in COVID-19 with Logistic Model and Machine Learning Technics." Chaos Solitons & Fractals 139(2020):110058.
- [10]. Annas, S. , et al. "Stability Analysis and Numerical Simulation of SEIR Model for pandemic COVID-19 spread in Indonesia." Chaos Solitons & Fractals 139(2020).

Multi-Country Logistic Growth Rate Index and Change Point Test Based on Bootstrap Algorithm

Abstract: Since January 2020, the new coronavirus epidemic has spread rapidly around the world, and the number of confirmed cases in various places has been increasing, and social and economic life has been greatly disrupted. In this regard, various countries have adopted a variety of indicators

to measure the epidemic situation in their countries. This article uses a simple and scientific logistic function growth rate indicator to measure the development of the national epidemic, and at the same time estimates the change point parameters of the growth rate, detects the difference in the growth rate at different times, analyzes the prevention and control of the epidemic in each country, and finally combines Bootstrap The algorithm constructs the confidence interval and establishes the regression model. This article regressed the data of the United States, Italy, and Japan from the early stage of the epidemic to February 21, 2021, and found that the development of the epidemic situation in each region is different. This algorithm is suitable for epidemic detection in a variety of situations, and can provide a reference for the current situation and future development trends of the epidemic.

Keywords: New crown epidemic, growth rate, change point, Logistic function, Bootstrap algorithm