

基于 vSEIdRm 模型的人口迁移以及 离汉交通管控对新冠肺炎疫情 发展的影响分析*

顾嘉 陈松蹊 董倩 邱宇谋

内容提要:不同于传统(Susceptible-Exposed-Infected-Removed) SEIR 流行病传播动力学模型,本文在近期研究的 Varying Coefficient Susceptible-Exposed-Infected-Diagnosed-Removed (vSEIdR) 模型基础上加上人口迁徙(Migration)模块,设计开发了 vSEIdRm 模型,该模型考虑了跨区域人口迁徙对疫情传播的影响,并允许流行病传播参数随时间变化。本文首先对人口迁移数据进行统计分析,建立其与各省新冠肺炎疫情发展的联系。之后,基于 vSEIdRm 模型估计了疫情初期各省来自武汉的输入病例数,并定量刻画了离汉交通管控的效果。研究结果显示,离汉交通管控措施有效地减少了各省的疫情规模。

关键词:流行病动态传播模型;人口迁徙;离汉交通管控影响

DOI: 10. 19343/j. cnki. 11-1302/c. 2021. 09. 009

中图分类号: F222. 3 **文献标识码:** A **文章编号:** 1002-4565(2021)09-0114-14

The Effect of Population Migration and Wuhan Lockdown on the Control of COVID-19 Based on vSEIdRm Model

Gu Jia Chen Songxi Dong Qian Qiu Yumou

Abstract: Different from the traditional SEIR (Susceptible-Exposed-Infected-Removed) epidemic model, based on the vSEIdR (Varying Coefficient Susceptible-Exposed-Infected-Diagnosed-Removed) model in our previous study, in this paper we add a population migration compartment and propose the vSEIdRm model, which takes the effect of cross-regional migration on the epidemic into consideration and allows the parameters to vary with time. We first conduct statistical analysis on the population migration data and connect the migration with the progression of the COVID-19 epidemic. Further, based on the new model, we estimate the would-be imported cases to the provinces from Wuhan in the absence of Wuhan lockdown which quantifies the effect of the Wuhan lockdown. Our results show that the Wuhan lockdown effectively reduces the scale of the epidemic in other provinces.

Key words: Epidemic Dynamic Transmission Model; Migration; Effect of Wuhan Lockdown

* 基金项目: 国家自然科学基金项目“面向管理决策大数据分析的理论与方法”(92046021)、“变系数流行病学模型的统计推断”(12071013)、“面向儿童脑发育障碍性疾病的神经机制建模与辅助诊疗算法”(12026607), 作者感谢北京大学统计科学中心和数量经济与数理金融教育部重点实验室(LMEQF)的帮助以及中国联通智慧足迹数据科技有限公司提供的人口迁移数据。



一、引言与文献综述

为应对新冠肺炎疫情蔓延,2020年1月23日10时起,武汉市采取了关闭离汉通道等措施,截止到1月24日12时,湖北共有13个城市区域公共交通停运^①。一个在控制病毒传播层面很自然的问题是如何量化评估武汉及湖北采取的上述防控措施对各省市疫情发展的影响?时值春运高峰,如果当时没有采取这些措施,巨大的人口流动又会对疫情起到什么作用?有学者将人口的迁徙与新冠肺炎疫情的时空分布建立联系,提出了一套疫情风险指标模型(Jia等,2020);Zhou等(2020)考虑了人口流动限制对于新冠肺炎疫情在深圳传播的影响;Zhang等(2020)探究了人口迁移对我国国内疫情传播的作用,并发现了两者之间非常强的正相关性;Cao等(2020)分析了导致我国疫情发展的系统性因素,其中人口迁徙是主要原因。还有学者探究了疫情在我国发生的最初50天内,管控措施起到了什么样的作用(Tian等,2020),其结论是离汉交通管控导致疫情到达全国其他城市的时间推迟了2.91天。北京大学陈松蹊课题组就新冠肺炎疫情提出了基于变系数SIR模型(Varying Coefficient Susceptible-Infected-Removed model),并且估计了各省的传染病再生系数 R_t ,及时评估疫情发展变化(Sun等,2020),之后又提出了vSEIdR模型并用于分析比较25国疫情控制效果,结果显示,疫情较晚发生的国家并没有吸取到较早发生国家的经验(Gu等,2020)。面对侵袭全球的第二波乃至第三波疫情,Yan等(2021)将无症状感染者纳入模型考虑,提出vSIADR模型,并结合线性混合效应模型分析各国管控措施对于疫情防控的效果。

目前关于人口迁徙对新冠肺炎疫情传播的相关工作,比如Kraemer等(2020)的人口迁徙数据来源是百度迁徙指数,但这只是反映人口多少的相对指标。本文使用来自国家统计局(中国联通智慧足迹提供)的2020年1月1日至2020年4月10日以及2019年1月20日至2019年4月10日的人口迁徙数据,其包含从湖北省和武汉市到全国其他30个省市自治区(不含港澳台,下同)的每日人口双向流动。本文利用提出的变系数SEIdRm模型(vSEIdRm),对传统SEIR流行病传播模型做出推广,将武汉市到其他各省的人口迁徙对各省市新冠肺炎疫情发展的影响考虑在内,并基于此模型进行情景分析,量化人口迁徙对各省市疫情发展的影响以及离汉交通管控的效果。若2020年1月23日武汉市未采取措施,则对于从1月24日起4周内(至2月20日)已经达到250例确诊病例的17个省市,其总计确诊病例将增加至27963例(95%置信区间:25127~31159例),确诊总人数增加比例达到151%(95%置信区间:126%~180%)。这意味着在一轮疫情开始的初期,及时切断病毒传播路径对有效控制疫情有重要作用,这也为有效防控冬季第二轮疫情反弹提供了经验。

二、数据汇总与描述性统计分析

(一)数据

本文使用的人口移动数据集包括2020年1月1日至2020年4月10日武汉市和湖北省向其其余30个省级行政单位的每日人口流动信息,武汉市向湖北省其余各市的流动信息,以及全国各省份之间的每日人口流动信息。数据还包括2019年1月20日至2019年4月10日的相应人口移动数据。按照农历计算,2020年1月23日(离汉交通管控日)是大年二十九。为评估离汉交通管控效果,本文使用2019年春节前两周数据(2019年1月20日至2019年2月3日)进行分析。本文使用的每日新增确诊病例、累计确诊、现存病例、治愈及死亡数据来自丁香园新型冠状病毒肺炎疫情

^① 采取区域公共交通停运措施的城市包括武汉、鄂州、仙桃、枝江、潜江、黄冈、赤壁、荆门、咸宁、黄石(含大冶市、阳新县)、当阳、恩施、孝感。



实时动态数据^①。

(二) 描述性统计分析

通过对离汉交通管控前后两周(以2020年1月23日为分隔点)从湖北以及武汉到全国其余30个省(自治区、直辖市)(不包含香港、澳门和台湾)平均人口流动和2019年农历同期的平均人口流动数据分析,可以发现,与2019年农历同期相比,2020年人口流动显著减少(见附图1^②和附图2^③)。通过测算离汉交通管控前后从湖北省向全国各省份的人口流动的改变率^④,可以发现,除安徽省外(68%)所有省份都达到了80%以上的下降率,其中四川、北京、福建、河北、贵州、广西、山西、辽宁、云南、甘肃、海南、吉林、内蒙古、宁夏、黑龙江和青海这16个省份下降率超过90%。结合离汉交通管控前后从湖北省向全国各省份人口流动的改变率测算结果来看,采取措施对限制人口流动确实起到了显著作用(见附图3^⑤)。同时,由于采取公共交通停运措施主要在武汉市实行,并逐步扩展到全省,可以发现如果以湖北省全体来看,采取措施前两周时间内湖北省往外省份流动总计超过1040万人,后两周内仍然有超过127万人流出湖北,下降约88%;相比较之下采取措施前两周武汉市往外省份流动总计超过185万人,后两周内武汉市往外省份流动总计约1.3万人,下降超过99%。另外,仅2020年1月22日零点至1月23日24点,武汉市总计流出人口达到119万,其中21万流向外省份,占前者的17.6%。

(三) 相关性分析

通过对采取公共交通停运措施前(2020年1月10日至2020年1月23日)从武汉市到各省份的人口流动(对数尺度下)与2020年1月23日起第一周各省份的确诊人数以及截至3月15日的各省份总确诊人数的分析(见附图4^⑥),结果显示人口流动和各省份确诊人数之间有较强的正相关性,Pearson相关系数(下同)分别为0.72和0.76(对应的单边t-检验p值分别为 4.3×10^{-6} 和 4.9×10^{-7})。进一步发现,离汉交通管控前不同时段下人口流动总数与各省份累计确诊人数的相关性是随时间而变化的,且可以观察得到两个结论:一是随时间推移,相关性逐渐上升,但是上升过程不单调,从1月27日至1月30日出现了约0.05的下降,之后基本单调上升;二是统计采取措施前不同时间段的人口流量总数对正相关的结论没有显著影响,且1月23日当天从武汉到达全国各省份的人数与最终的各省份确诊具有最高的正相关性,比考虑采取措施前更多天的人口流动总数计算的相关系数至少超出了0.02(见附图5^⑦)。

三、模型建立

(一) vSEIRm 模型

为考虑武汉市到各省份的人口迁移的作用并做定量分析,本文对传统的SEIR(Susceptible-Exposed-Infected-Removal)(Hethcote,2000)模型进行推广,提出vSEIRm模型。本文将非湖北省份的人群在同一时刻 t 分为5类状态:易感者 $S(t)$ 、感染但未出现症状或还未确诊者 $E(t)$ 、确诊者 $I(t)$ 和移出者 $R(t)$ (包含康复和死亡)。其中, E 和 I 状态均具有传染性,但是只有 I 状态的人群是可以被观测到的,即 E 状态的数据是缺失的。由于使用的是日数据,本文令

① <https://ncov.dxy.cn/ncovh5/view/pneumonia>。

② 因篇幅所限,采取措施前后人口流量对比以附图1展示,见《统计研究》网站所列附件。下同。

③ 因篇幅所限,采取措施后人口流量与去年同期流量对比以附图2展示。

④ 分别计算2020年1月23日24点前两周和后两周的平均流动,以此计算改变率。

⑤ 因篇幅所限,采取措施后人口流量下降比例以附图3展示。

⑥ 因篇幅所限,人口流量与确诊人数对比以附图4展示。

⑦ 因篇幅所限,相关性分析以附图5展示。



$$\frac{dS(t)}{dt} = S(t+1) - S(t), \frac{dE(t)}{dt} = E(t+1) - E(t);$$

$$\frac{dI(t)}{dt} = I(t+1) - I(t), \frac{dR(t)}{dt} = R(t+1) - R(t)。$$

vSEIdRm 的模型设定是, 给定 t 时刻已有的全状态变量信息 \mathcal{F}_t , 增量 $(\frac{dS(t)}{dt}, \frac{dE(t)}{dt}, \frac{dI(t)}{dt}, \frac{dR(t)}{dt})$ 的条件期望满足以下微分方程:

$$\begin{cases} \frac{dS(t)}{dt} = (1 - p_i^E)A(t) - (\beta_i^I I(t) + \beta_i^E E(t)) \frac{S(t)}{M} \\ \frac{dE(t)}{dt} = p_i^E A(t) + (\beta_i^I I(t) + \beta_i^E E(t)) \frac{S(t)}{M} - \alpha_i E(t) \\ \frac{dI(t)}{dt} = \alpha_i E(t) - \gamma_i I(t) \\ \frac{dR(t)}{dt} = \gamma_i I(t) \end{cases} \quad (1)$$

其中, $A(t)$ 代表 t 时刻从武汉到达该省份的人数, 对应的 p_i^E 代表 $A(t)$ 中感染者所占比例, β_i^I 和 β_i^E 分别代表 I 和 E 状态群体的感染力, α_i 和 γ_i 分别代表检出率和移除率, M 是该省份的人口总数。由于来自武汉市的人口迁移对各省份人口总数影响不大, 本文固定总人口数为常数 M 。定义 $N(t) = I(t) + R(t)$ 为 t 时刻累计确诊人数。 t 时刻有效再生数 R_t 可以定义为 (Gu 等, 2020):

$$R_t = \left(\frac{\beta_i^E}{\alpha_i} + \frac{\beta_i^I}{\gamma_i} \right) \frac{S(t)}{M} \quad (2)$$

有效传染再生数 R_t 代表的是一个染病者平均可以感染的易感者的人数, 因此 R_t 是度量一个地区的传染是在扩张 ($R_t > 1$) 还是在收缩 ($R_t < 1$) 的关键指标 (Nishiura 和 Chowell, 2009)。

进一步对模型式 (1) 指定的条件期望建模, 本文考虑条件独立 Poisson 分布。令 $\Delta S(t) = S(t+1) - S(t) - (1 - p_i^E)A(t)$, $\Delta E(t) = E(t+1) - E(t)$, $\Delta I(t) = I(t+1) - I(t)$, $\Delta R(t) = R(t+1) - R(t)$ 。于是有 $\Delta E(t) = p_i^E A(t) + \Delta S(t) - \Delta N(t)$, $\Delta I(t) = \Delta N(t) - \Delta R(t)$ 。由于实际情况中 $\Delta S(t)$ 、 $\Delta N(t)$ 、 $\Delta R(t)$ 分别代表染病、确诊和康复的过程, 生成机制相对独立, 因此可以假设它们在时刻 t 服从条件独立的泊松分布:

$$\begin{cases} -\Delta S(t) \sim \text{Poisson} \left\{ (\beta_i^I I(t) + \beta_i^E E(t)) \frac{S(t)}{M} \right\} \\ \Delta N(t) \sim \text{Poisson} \{ \alpha_i E(t) \} \\ \Delta R(t) \sim \text{Poisson} \{ \gamma_i I(t) \} \end{cases} \quad (3)$$

(二) 参数估计方法

模型中, 由于 $E(t)$ 不可观测, 出于参数可识别性 (Identification) 的考虑, 本文设定 $\beta_i^E = \frac{\beta_i^E}{r}$, 其中 $r > 1$ 是一个待定超参数。 $r > 1$ 是由于在此次新冠肺炎疫情中, 一旦某个个体被确诊, 他/她将居家隔离或是在医院隔离治疗, 这意味着确诊个体的传染力 β_i^I 会显著小于确诊前的传染力 β_i^E 。以下在给定 (r, α, p_i^E) 的前提下, 本文给出估计感染率 β_i^E 的方法, 之后将给出选择 (r, α, p_i^E) 的方案。参数估计中, 最主要的难点在于 E 状态的人群是不可观测的。根据本文模型假定, 有 $\Delta N(t) \approx$



$\alpha_i E(t)$, 于是可以得到 $E(t)$ 的一个“估计量”: $\hat{E}(t) = \frac{\Delta N(t)}{\alpha_i}$ 。根据式(1)中第二条方程, 可以建立以下关系式:

$$\hat{E}(t+1) - \hat{E}(t) = p_i^E A(t) + \widetilde{\beta}_i^E \left(\frac{I(t)}{r} + \hat{E}(t) \right) - \alpha_i \hat{E}(t) \quad (4)$$

其中, $\widetilde{\beta}_i^E = \beta_i^E \frac{S(t)}{M}$ 为根据人口中易感染群体所占比例做出调整的传染力参数。定义:

$$Y_i = \hat{E}(t+1) + (\alpha_i - 1) \hat{E}(t) - p_i^E A(t), X_i = \frac{I(t)}{r} + \hat{E}(t) \quad (5)$$

在时刻 t , 可以用 Y_i 对 X_i 做局部核回归来获得参数估计 $\widetilde{\beta}_i^E$, 即关于 β 最小化以下目标函数:

$$\sum_{i=1}^T (Y - X_i \beta)^2 B\left(\frac{t-i}{h}\right) \quad (6)$$

其中, B 是一种从通常核函数调整过的边界核函数 (Boundary Kernel) (Jones, 1993), h 是平滑带宽 (bandwidth), 下文统一使用带宽 $h=7$ 。使用边界核函数的目的是为了消除在 $t=T$ 附近, 即接近边界时的估计偏差 (bias)。于是可以获得以下核估计量 (Kernel Estimator):

$$\widetilde{\beta}_i^E = \frac{\sum_{i=1}^T X_i Y_i B\left(\frac{t-i}{h}\right)}{\sum_{i=1}^T X_i^2 B\left(\frac{t-i}{h}\right)} \quad (7)$$

令 $\hat{S}(t) = N - N(t) - \hat{E}(t)$ 和 $\widehat{\beta}_i^E = \widetilde{\beta}_i^E \frac{M}{\hat{S}(t)}$, 从而 $\widehat{\beta}_i^E = \frac{\widetilde{\beta}_i^E}{r}$ 。由式(3)的最后一条方程可以发现

$E[R(t) | \mathcal{F}_t] = \gamma_t I(t)$, 于是可以用 $\Delta R(t)$ 对 $I(t)$ 做不带截距项的局部核回归来获得移出率 γ_t 的估计:

$$\hat{\gamma}_t = \frac{\sum_{i=1}^T I(i) \Delta R(i) B\left(\frac{t-i}{h}\right)}{\sum_{i=1}^T I(i)^2 B\left(\frac{t-i}{h}\right)} \quad (8)$$

对于比例 r 的选取, 由于估计方法对参数的选取不敏感 (Gu 等, 2020), 本文以下计算分析中, 统一采用 $r=5$ 。定义有效传染再生系数 R_t 的估计量 \hat{R}_t 为:

$$\hat{R}_t = \left(\frac{1}{\alpha_i} + \frac{1}{r \hat{\gamma}_t} \right) \widehat{\beta}_i^E \quad (9)$$

(三) 利用参数自助法 (Parametric Bootstrap) 构造参数置信区间

为了对上述参数估计进行统计推断, 本文在条件 Poisson 增量下的 vSEIIdRm 模型式(3)下, 给定利用原始数据得出的“估计量” ($\hat{S}(t), \hat{E}(t), \widehat{\beta}_i^E, \hat{\gamma}_t$), 使用自助法构造估计的置信区间。在每一时刻 t , 依据如下条件 Poisson 模型产生 $t+1$ 天的增量:

$$\begin{cases} -\Delta^* S(t) \sim \text{Poisson} \left\{ (\widehat{\beta}_i^E I(t) + \hat{\beta}_i^E \hat{E}(t)) \frac{\hat{S}(t)}{M} \right\} \\ \Delta^* N(t) \sim \text{Poisson} \{ \alpha_i \hat{E}(t) \} \\ \Delta^* R(t) \sim \text{Poisson} \{ \hat{\gamma}_t I(t) \} \end{cases} \quad (10)$$



根据式(10)的条件 Poisson 分布设定以及给定的初始值 $(S(0), E(0), I(0), R(0))$, 可以反复生成传染病传播过程曲线 $(S^*(t), E^*(t), I^*(t), R^*(t))$ 。重复 $B(B=500)$ 次上述抽样, 将再抽样得到的数据记为 $D^b = \{S^{*b}(t), E^{*b}(t), I^{*b}(t), R^{*b}(t)\}_{t=1}^T, b=1, 2, \dots, B$ 。对每一组数据, 使用前文给出的参数估计方法得到 $(\hat{\beta}_i^{l,b}, \hat{\beta}_i^{E,b}, \hat{\gamma}_i^b)$ 。在 t 时刻对再抽样得到的估计关于利用原始数据得到的估计 $(\hat{\beta}_i^l, \hat{\beta}_i^E, \hat{\gamma}_i)$ 做中心化(Centering), 使得再抽样估计的均值等于 $(\hat{\beta}_i^l, \hat{\beta}_i^E, \hat{\gamma}_i)$ 。定义 t 时刻各参数的 95% 置信区间端点分别为中心化再抽样估计的 2.5% 和 97.5% 分位数。

(四) 模拟试验

如果没有把从武汉到各省的人口迁移项 $p_i^E A(t)$ 考虑进来情况会如何? 即模型的错误设定(Mis-specification)会对参数估计造成怎样的影响? 本文通过两组模拟试验进行分析。简单起见, 本文将所有的参数都被设定为常数, 并按照式(3)所述的方式生成数据, 其中: $M=5 \times 10^8$ 代表总人口, $p_i^E A(t) = 20, 50, 100$ 代表每日输入病例数; $E(0) = 5, I(0) = 5$ 代表初始感染与确诊人数; $\beta_i^E = 0.15, \alpha = 0.25, \gamma = \frac{1}{21}$ 分别代表感染率(E 状态时)、确诊率和移出率; $r=5$, 这意味着 $\beta_i^E = 5\beta_i^I$; 在此设定下基本再生数 R_0 (即 $\frac{S(t)}{M} = 1$ 时)为 1.23。

对于每个给定的 $p_i^E A(t)$, 基于带条件 Poisson 增量的 vSEIdRm 模型式(3)生成数据, 进而估计参数 β_i^E 。重复实验 $B=500$ 次, 将估计的均值曲线绘制成图 1。由图 1 左图可以发现, 忽略人口迁移(即在回归中忽略 Y_i 中的 $p_i^E A(t)$ 项)会导致早期对参数 β_i^E 的显著高估, 在模拟情形下疫情发生一周后仍然有超过 100% 的高估。随着疫情的发展, 当总体确诊人数不断增加, 相比之下输入病例作用不太重要时, 估计又逐渐收敛到真值 0.15, 但是收敛过程非常缓慢, 与真值差距缩小到 10% 以内需要大约 2 个月(57 天)的时间。

图 1 右图展示的是另一组对比, 其中 vSEIdRm 和 vSEIdR 分别代表考虑人口迁移和不考虑人口迁移的参数估计方法。可以发现, 虽然两种方法都逐渐向真值收敛, vSEIdRm 关于 β_i^E 的估计方法仅需大约两周时间就能达到 10% 以内误差。而此时忽略人口迁移的 vSEIdR 模型仍然有超过 50% 的高估。若换算作 R_t 的值, 此时 vSEIdRm 方法的估计值为 1.13, 而 vSEIdR 方法估计值为 1.85, 与真值(1.23)差异较大。准确掌握真实的有效传染再生数 R_t 对实际防控疫情过程中及时掌握疫情严重程度有重要意义。

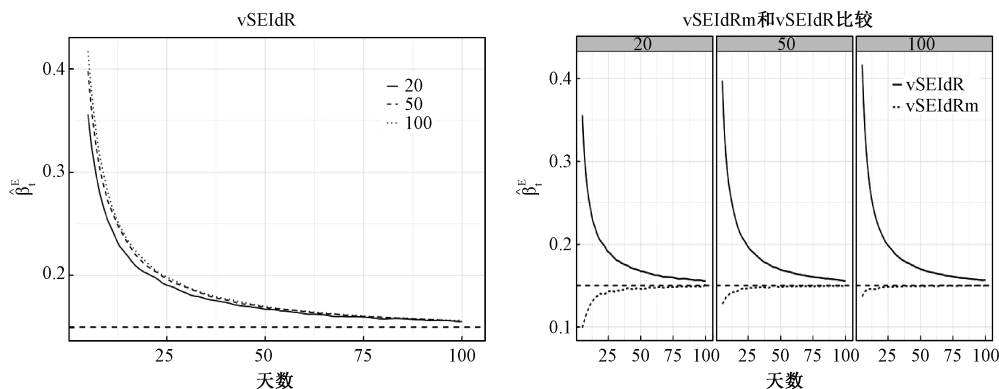


图 1 忽略人口迁移项对传染率参数估计的影响

(五) 基于观测数据的 p_i^E 估计

由于早期检测效率等原因, 从武汉到各省份的人口流动数据 $A(t)$ 中的新冠肺炎病例所占比例



是未知的,需要基于模型对该参数做出估计,为此本文做出以下假定。令2020年1月10日为 $t=1$,1月10日至1月23日是离汉交通管控前两周,考虑到病毒的潜伏期及14天隔离期(Lauer等,2020),本文以此两周时间段作为病毒大范围扩散到各省份的时间窗口。

1. 对某一选定省份 P ,以及给定的参数 η ,令 p_t^E 服从以下变化模式:

$$p_t^E(\eta) = \begin{cases} \frac{\eta}{100} e^{\frac{R_0}{14}t}, & t \leq 14 \\ 0, & 15 \leq t \end{cases} \quad (11)$$

$p_t^E(\eta)$ 的设定理由如下: $t=14$ 对应1月23日,当天上午十时起,离汉交通管控正式实施,可以近似认为1月24日之后流出人群都经过了严格的检查,本文令此后的参数 $p_t^E=0$ 。1月10日至1月23日是病例输出的主要时间段,在这一时间段内武汉经历了一段病毒快速传播的过程,检测率 α 和移出率 γ 都可以近似认为是0。对于武汉市内这一时间段的新冠肺炎传播,由式(1)的第二条方程有 $\frac{S(t)}{M} \approx 1$,近似地:

$$\frac{dE(t)}{dt} \approx \beta_t^E E(t) \frac{S(t)}{M} \approx \beta_t^E E(t)$$

可以得出结论,疫情初期 E 状态人群在武汉市内经历了一段指数增长的过程,这导致武汉市初期向各省份的病例输出率 p_t^E 的指数增长。求解以上微分方程,有

$$E(t) = E(0) e^{\beta_t^E t} = E(0) e^{\frac{R_0}{14}t}$$

其中, $R_0 = \beta^E D$ 是武汉市此次疫情的基本再生数,设定为5.7(Sanche等,2020), $D=14$ 是从感染到确诊的平均时间长度。由此可以得出前述采取措施前两周内 p_t^E 的参数化形式(11)。

2. 诊断率 α 的选取。

令 α 服从以下变化模式:

$$\alpha_t(p^E) = \begin{cases} 0, & 1 \leq t \leq 10 \\ \frac{1}{3.5} - \xi_1(p^E) \\ \frac{1}{14}(t-11) + \xi_1(p^E), & 11 \leq t < 25 \\ \frac{1}{3.5}, & t > 25 \end{cases} \quad (12)$$

大多数省份的病例从2020年1月21日开始出现比较明显的增长,并且随着时间推移,检测能力有一个快速上升的过程,因此本文将 α_t 设定为一个分段线性函数,并且以1月21日作为非零 α_t 起始点。根据钟南山团队发表的文章(Guan等,2020),本文将平稳期的检出率 α 给定在1/3.5(即大约3.5日从染病到被确诊,这反映了随着疫情发展,检测能力的快速上升)。而关于左端点(ξ_1),可以进行如下考虑:对某一省份 P ,定义 $E_0 = \sum_{t=0}^{13} p_t^E A^P(t)$,其中 $A^P(t)$ 代表第 t 天武汉向该省份的人口流动, E_0 可以代表离汉交通管控前向该省份输入的病例总数。根据式(1)的第2,3式,近似地有:

$$\alpha_{13} \approx \frac{N(14) - N(13)}{E_0}$$

可以假定采取措施前1月21日和1月23日这两天的检出率差别不大,于是有 $\xi_1 = \alpha_{11} \approx \alpha_{13}$ 。由此本文将检出率 α_t 定义为如式(12)所给定的 p_t^E 的函数。



3. 基于确诊曲线 $N(t)$ 拟合效果的参数 $(p_i^E(\eta), \beta_i^E, \beta_i^I, \gamma_i)$ 估计。

给定 $p_i^E(\eta)$, 本文首先估计检出率 α_i , 然后进一步利用第三部分介绍的参数估计方法, 得出其余参数估计。从 $t=14$ 起, 以 $(S(13), E(13), I(13), R(13)) = (M - E_0, E_0 - N(13), N(13), 0)$ 作为初始值, 在省份 P 可以根据式 (1) 所确定的增量模型生成一条基于模型的确诊曲线, 记为 $N^H(t)$, 其中 M 代表该省份人口总数。考虑 1 月 24 日 ($t=15$) 起 4 周的拟合效果, 定义如下距离函数:

$$D(\eta) = \sqrt{\frac{1}{28} \sum_{t=15}^{42} (N(t) - N^H(t))^2} \quad (13)$$

在参数值范围 $\eta \geq 0$ 内极小化 $D(\eta)$, 可以得到相应的参数估计。

四、疫情初期输出病例估计及情景分析

(一) 输入病例估计

本文选取了 17 个截至 2020 年 2 月 20 日确诊病例达到 250 人的省份^①, 根据前文基于距离的估计方法, 对相应参数进行估计, 其中确诊曲线的拟合效果见图 2, 参数估计展示在表 1。由图 2 来看, 该方法的拟合效果比较理想, 为本文的估计方法提供了有力支撑。表 1 首先给出了各省份 η 取最优值时目标函数式 (13) 的值, 可以发现最优函数值的取值范围是 $[0.25, 2.21]$, 平均值是 1.16 (标准误 SE=0.16), 这意味着平均意义下拟合值与实际确诊数字每天的误差不到 2 例。

表 1 17 省份的拟合效果 $D(\eta)$ 以及相应的参数估计值

	省份	$D(\eta)$	η	E_0	N	E_0/N 比例	$\overline{R_1}$	$\overline{R_2}$
1	福建	0.43	0.0075	169	287	0.59	7.8	7.5
2	陕西*	0.25	0.0149	98	240	0.41	12.3	11.6
3	浙江	1.55	0.0630	395	1193	0.33	12.9	12.5
4	重庆*	1.08	0.0067	253	563	0.45	12.8	12.6
5	河南*	1.91	0.0015	511	1233	0.41	13.9	13.4
6	湖南*	2.17	0.0020	332	986	0.34	14.2	13.4
7	四川	1.06	0.0292	208	523	0.40	14.8	14.2
8	河北	0.95	0.0053	115	308	0.37	15.5	14.7
9	北京	0.59	0.0064	129	448	0.29	15.9	15.4
10	江苏	1.20	0.0054	203	614	0.33	16.9	15.9
11	上海	0.41	0.0138	68	352	0.19	16.4	16.0
12	广东	2.21	0.0061	295	1339	0.22	18.1	17.5
13	安徽*	1.50	0.0025	228	969	0.24	18.6	17.9
14	江西*	1.69	0.0027	205	914	0.22	19.3	18.6
15	山东	1.72	0.0259	251	738	0.34	20.4	20.0
16	广西	0.46	0.0027	35	251	0.14	20.7	20.4
17	黑龙江	0.61	0.0298	60	477	0.13	27.4	26.8
总数				3550	11435			
均值 (标准误)		1.16 (0.16)	0.0075 (0.0020)	209 (31)	673 (89)	0.32 (0.03)	16.3 (1.1)	15.8 (1.0)

注: E_0 代表封城前武汉市向各省份输出的病例数, N 代表截至 3 月 15 日各省份确诊人数, 比例计算的是两者比值, 即输入型病例占各省份总体病例的比值, $\overline{R_1}$ 和 $\overline{R_2}$ 分别代表在 vSEIdR 和 vSEIdRm 模型下计算的累计传染力指标式 (14), 与湖北相邻省份用 * 号标示, 17 省份排名按照累积传染力指标 $\overline{R_2}$ 由小到大排列。

比例 (E_0/N) 越小, 代表一个输入病例带来的二次乃至三次传播越严重, 即疫情控制效果较差。

① 本文选取的省份包括安徽、北京、重庆、福建、广东、广西、河北、黑龙江、河南、湖南、江苏、江西、陕西、山东、上海、四川和浙江。

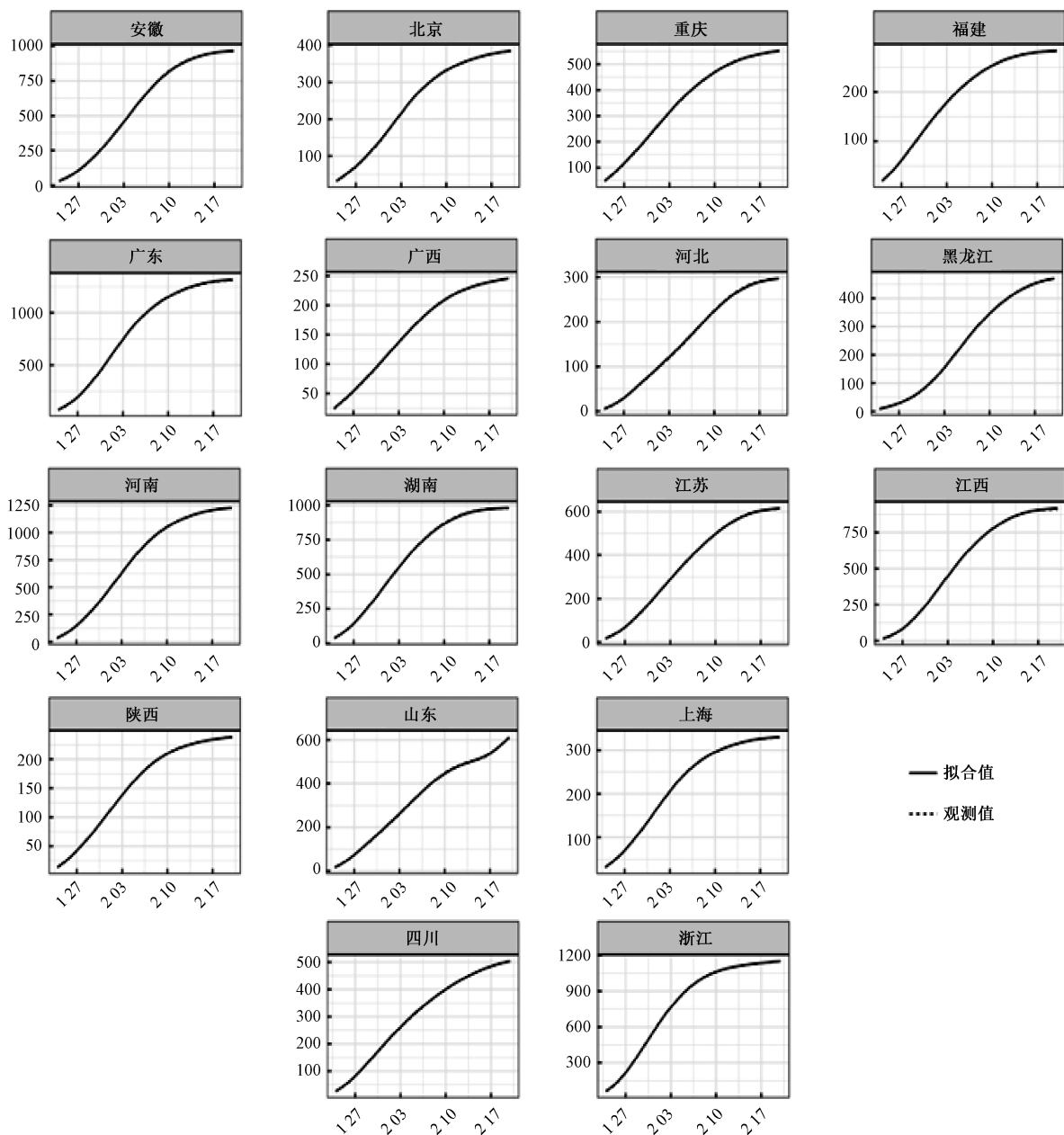


图2 2020年1月24日至2月20日17省份确诊人数及相应拟合值的曲线

这一点可以通过和有效传染再生数 R_t 的估计值建立联系而得到验证。令2020年1月10日为 $t=1$, 考虑离汉交通管控后4周内(1月24日至2月20日) R_t 曲线下面积, 定义累积传染力指标 \bar{R} :

$$\bar{R} = \frac{1}{2} (R_{15} + R_{28} + 2 \sum_{t=16}^{27} R_t) \quad (14)$$

图3显示输入病例比例与累积传染力指标的 Pearson 相关系数达到了-0.85(对应单边 t-检验的 p 值为 8.05×10^{-6})。根据有效传染再生数的定义(2), 可以认为比例 E_0/N 是一个直观反映疫情控制效果的度量。本文将17省份分为两组, 分别是与湖北相邻6个地区; 其余11个地区, 以考察两组间比例的差异。相邻组的平均比例是34.5%(SE:4%), 其余地区组的平均比例是30.3%(SE:4%), 这意味着对于相邻地区来说, 即便面临着更大的输入压力, 其管控仍然做得很不错, 甚至好于



其余地区,不过这一差异在统计意义下并不显著,两样本单侧 t -检验的 p 值是 0.23。所有地区中,福建省的估计比例值最高(58.9%),这意味着该省在这 17 省份的疫情防控中表现最为出色。而在相邻地区组中,可以看到江西和安徽省的比例显著低于其他相邻地区(江西、安徽分别为 22.4% 和 23.5%,相邻地区除去江西、安徽平均比例 40.2%(SE:2%)),说明这两个省的表现有所不足。这一点也可以从图 4 的传染再生系数估计曲线中看出:江西的基本再生数 R_0 估计达到了 2.4;而安徽和湖南虽然 R_0 值非常接近(分别是 1.4 和 1.1),但湖南 R_t 降低到 1 以下仅仅用了 5 天(1 月 27 日),而安徽则花了 12 天(2 月 3 日)。17 个省份中共有 5 个省份有效再生数 R_t 始终没有超过 1,分别为重庆(R_t 最高值为 0.79,下同)、福建(0.56)、河南(0.96)、陕西(0.83)和四川(0.83),这意味着疫情在这 5 个省份始终处于有效管控之下。

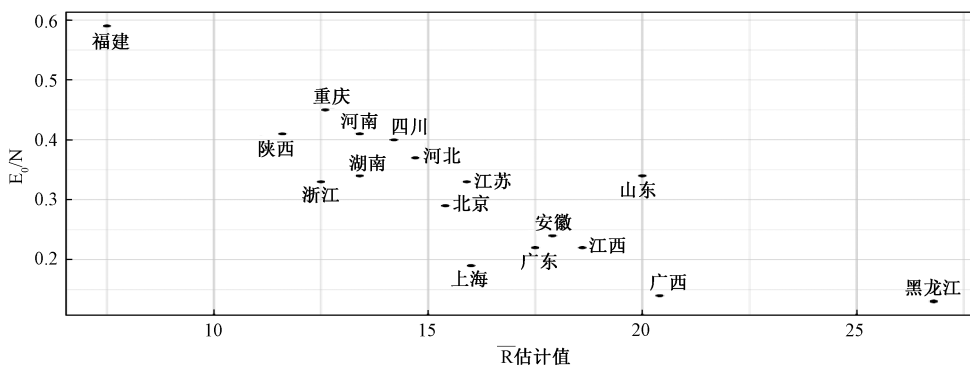


图 3 17 省份累积传染力指标 \bar{R} 与对应输入病例占总确诊数比例的散点图

注:数据截至 3 月 15 日,并在 vSEIdRm 模型下计算。

总量上,从武汉往 17 个省份的输出病例总数估计值是 3555 例,占这些地区截至 3 月 15 日总确诊病例数(11435 例)的 32%,这说明近三分之一的病例是输入型病例,从侧面验证了离汉交通管控的重要意义,在当代快捷交通运输的背景之下,若没有及时果断的旅行限制,病毒的传播严重程度将不可设想。

本文也通过累积传染力指标(14)来对各地的疫情防控效果做出评估,表 1 分别展示了在 vSEIdR 和 vSEIdRm 模型下计算的累积传染力指标 \bar{R}_1 和 \bar{R}_2 。结果显示,没有考虑人口移动的累计传染再生数 \bar{R}_1 比 \bar{R}_2 平均小 0.56,主要是未将输入病例从总确诊数中扣除所产生的估计偏差。在 vSEIdRm 模型下,累积传染力指标的最小值为 7.5,最大值为 26.8,后者是前者的 3.6 倍。

(二) 情景分析

在 2020 年 1 月 23 日离汉交通管控之后,各地也立刻采取了管控措施。为了能够更好地评价管控措施效果,一个自然的问题是:如果当时没有实施管控措施,疫情发展会如何变化?情况会有多严重?本文试图通过情景试验模拟来回答这个问题。

1. 模拟试验设计。

对于一个给定的省份,为了评价离汉交通管控对该省份自采取管控措施起 4 周(2020 年 1 月 24 日至 2 月 20 日)的疫情发展影响,本文设计如下模拟情景,(1)令该省份从武汉流入的流动人群中的患病比例保持在 1 月 23 日的估计值(即采取管控措施前的最高值)水平上,保持两周,再用两周时间逐渐下降到 0,数学表达式为:

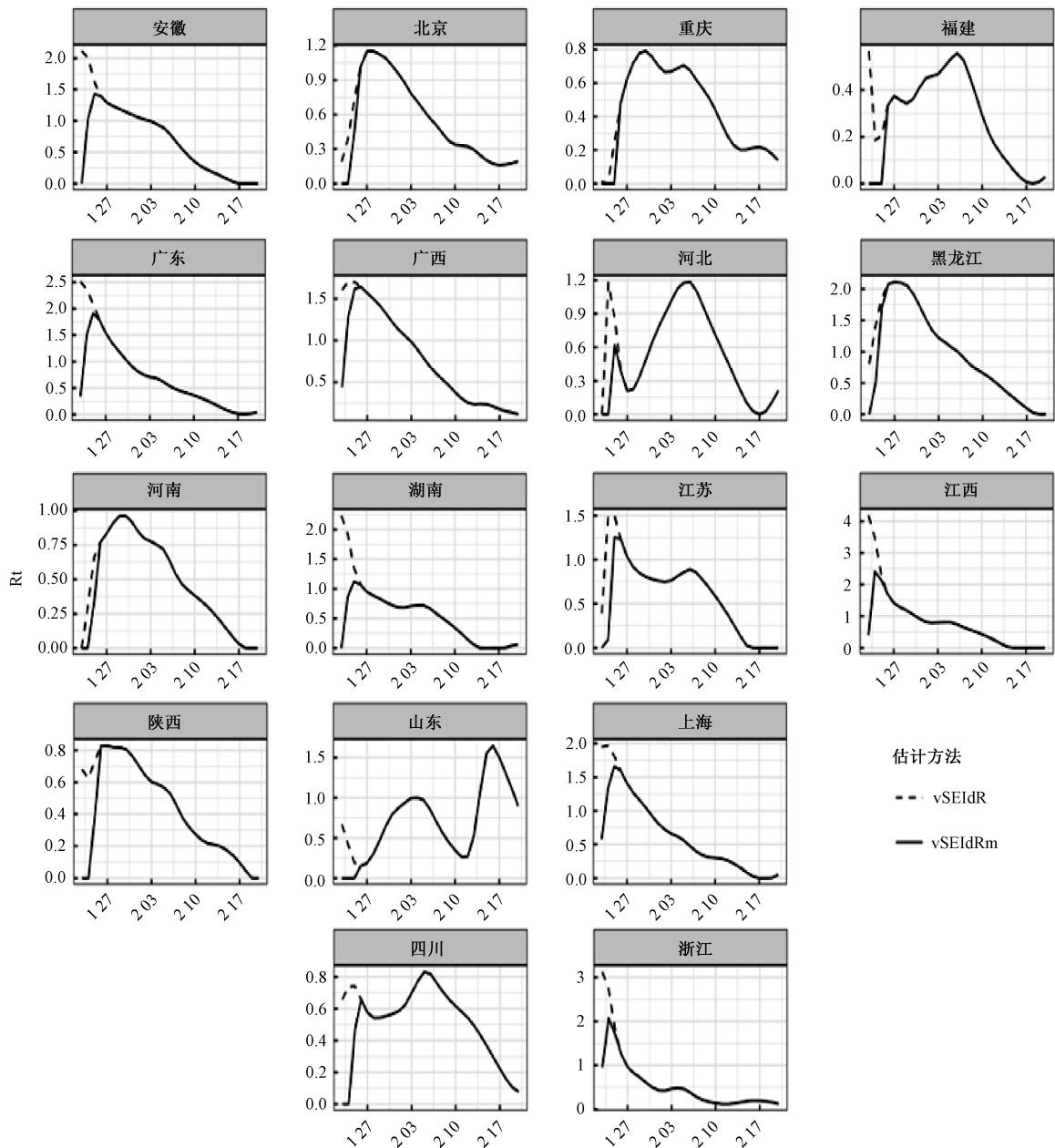


图4 传染再生系数估计曲线图

$$p_i^E = \begin{cases} \frac{\hat{\eta}}{100} e^{R_0}, & 14 \leq t \leq 27 \\ \frac{\hat{\eta}}{100} e^{R_0} \left(1 - \frac{t-27}{14}\right), & 28 \leq t \leq 41 \\ 0, & t > 41 \end{cases} \quad (15)$$

以上仍以2020年1月10日为第一天 $t=1$ 。

(2) 从2020年1月24日起的4周(至2020年2月20日),使用2019年农历同期(2019年2月4日至2019年3月3日)的人口流动数据替换2020年的人口流动数据。

(3) 由于本文主要分析的是离汉交通管控效果的影响,检出率、传染率以及移出率估计值 $\hat{\alpha}$,



$\hat{\beta}_i, \hat{\gamma}_i$ 均使用该省份已经估计的值并保持不变。

2. 结果分析。

截至 2020 年 2 月 20 日(离汉交通管控后 4 周), 17 个省份确诊总人数为 11121 人, 死亡 81 人。在前述的不采取管控措施情形下(见图 5, 表 2), 确诊总人数将增加至 27963 人(95% 置信区间: 25127~31159 人), 增加 151%, 其中来自武汉输入病例为 12508 例, 占总确诊人数 43%, 死亡人数将增加至 155 人, 增加 98%。各地确诊人数增加比例平均值为 154%(标准误: 22%), 死亡人数增加比例平均值 162%(标准误: 61%, 由于实际情形下江苏零死亡, 故计算死亡增加比例均值及相应标准误时不考虑在内, 下同)。从数值上可以直接得出结论: 如果当时没有果断采取隔离措施, 疫

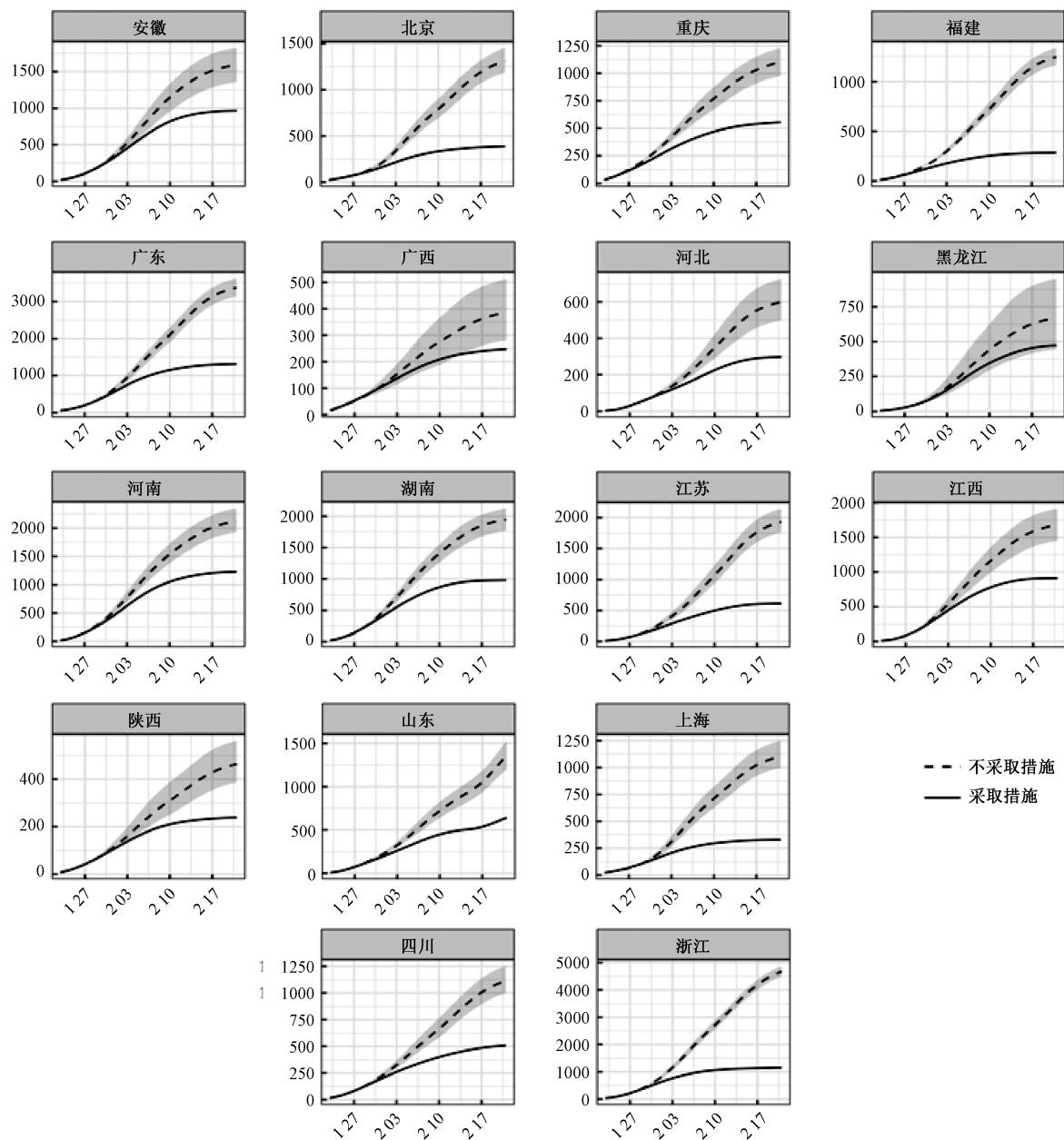


图 5 17 省份确诊人数曲线对比

注: 灰色区域代表不采取管控措施情形模拟确诊曲线的 95% 置信域。



情将比后来实际发生的情况严重的多。分地区来看,如果当时没有采取管控措施,北京、福建、浙江、上海、江苏和广东等东部经济大省以及重要城市将受到最为严重的影响,确诊人数的平均增长比例将达到 259%(标准误:25%),死亡人数平均增加比例为 294%(标准误:80%)。与湖北相邻的 6 个地区管控措施总体表现良好,因此武汉不采取管控措施带来的影响相比之下会小一些,但平均来看确诊人数增幅仍然达到了 94%(标准误:10%),死亡人数增幅达到 119%(标准误:26%)。而黑龙江省受到相对较小影响的主要原因是人流量绝对值较小:2019 年 2 月 3 日至 2019 年 3 月 2 日(对应武汉 2020 年 1 月 23 日采取管控措施起一个月)从武汉到黑龙江省的日均人流量是 76 人。但即使是这较小的人流量绝对值,在武汉不采取管控措施的情形下,依旧会给黑龙江带来超过 50%的确诊人数增幅(死亡人数增加 26%)。

表 2 17 个省份截至 2020 年 2 月 20 日实际确诊人数(即采取管控措施情形)和

武汉不采取管控措施情形的确诊人数比较

省份	确诊人数(采取管控措施)	确诊人数(不采取管控措施)	增加人数	增加比例(%)
福建	283	1259	975	343
陕西*	238	485	246	103
浙江	1154	4842	3688	319
重庆*	554	1109	555	100
河南*	1227	2165	937	76
湖南*	980	2103	1122	114
四川	507	1156	649	128
河北	298	615	317	106
北京	387	1334	947	245
江苏	613	1988	1374	224
上海	329	1151	821	249
广东	1314	3626	2311	176
安徽*	966	1748	781	81
江西*	912	1842	930	102
山东	641	1394	753	117
广西	246	432	185	75
黑龙江	472	714	242	51
总数	11121	27963	16842	151

注:与湖北相邻的 6 个地区用*号标出。

五、结论

本文分析了人口流动大数据,建立了人口流动与疫情严重程度的相关性,为以后应对此类公共卫生事件提供了参考。进一步,本文拓展了传统的流行病传播模型——SEIR 模型,将确诊前感染和人口迁移带来的跨区域疫情传播纳入考虑,并允许参数随时间变化,构建了 vSEIRm 模型,使得微分动力学方程对疫情发展的建模更加符合实际情况。本文依据此模型分析了离汉交通管控的效果,结果表明如果不采取管控措施将造成巨大损失。同时,由于现代交通技术的发达,产生传染病疫情时,受到最严重影响的不一定是相邻地区,对于经贸联系紧密、人口流动频繁的经济发达地区,必须及早做好响应,执行严格的防疫政策,以免遭受更大的损失。对于经贸联系并不那么紧密或是人口流动相对较少的地区也不能掉以轻心,正如 Ruan 等(2020)指出,国家之间(在本次疫情初期即省际)的疫情扩散仅需要少数(5~10)个感染者入境,事实上黑龙江省本轮疫情就经历了一段本地聚集性传播的过程。一次成功的疫情防控需要外防输入和内防本地传播的有机结合。



参考文献

- [1] Cao Z C, Tang F, Chen C, et al. Impact of Systematic Factors on the Outbreak Outcomes of the Novel COVID-19 Disease in China: Factor Analysis Study[J]. Journal of Medical Internet Research, 2020, 22(11):e23853.
- [2] Gu J, Yan H, Huang Y X, et al. Comparing Containment Measures among Nations by Epidemiological Effects of COVID-19[J]. National Science Review, 2020, 7(12):1847-1851.
- [3] Gu J, Yan H, Huang Y X, et al. Better Strategies for Containing COVID-19 Epidemics: A Study of 25 Countries via an Extended Varying Coefficient SEIR Model[J]. Medrxiv, 2020.
- [4] Guan W J, Ni Z Y, Zhong N S, et al. Clinical Characteristics of Coronavirus Disease 2019 in China[J]. The New England Journal of Medicine, 2020, 382:1708-1720.
- [5] Hethcote H W. The Mathematics of Infectious Diseases [J]. SIAM Review, 2000, 42: 599-653.
- [6] Jones M. Simple Boundary Correction for Kernel Density Estimation[J]. Statistics and Computing, 1993, 3:135-146.
- [7] Jia J, Xin L, Yun Y, et al. Population Flow Drives Spatio-Temporal Distribution of COVID-19 in China[J]. Nature, 2020(582): 1-11.
- [8] Kraemer M, Yang C H, Gutierrez B, et al. The Effect of Human Mobility and Control Measures on the COVID-19 Epidemic in China [J]. Science, 2020, 368(6490): 493-497.
- [9] Lauer S A, Grantz K H, Bi Q F, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application [J]. Annals of Internal Medicine, 2020, 172(9): 577-582.
- [10] Nishiura H, Chowell G. Mathematical and Statistical Estimation Approaches in Epidemiology[M]. Springer, 2009.
- [11] Ruan Y S, Luo Z D, Wu C I, et al. On the Founder Effect in COVID-19 Outbreaks-How Many Infected Travelers May Have Started Them All? [J] National Science Review, 2020, 8(1).
- [12] Sun H X, Qiu Y M, Yan H, et al. Tracking Reproductivity of COVID-19 Epidemic in China with Varying Coefficient SIR Model (with discussion)[J]. Journal of Data Science, 2020, 18(3): 455-472.
- [13] Sanche S, Lin Y T, Xu C G, et al. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2[J]. Emerging Infectious Diseases, 2020, 26(7):1470-1477.
- [14] Tian H Y, Liu Y H, Li Y D, et al. An Investigation of Transmission Control Measures During the First 50 Days of the COVID-19 Epidemic in China[J]. Science, 2020, 368: 638-642.
- [15] Yan H, Zhu Y R, Gu J, et al. Better Strategies for Containing COVID-19 Pandemic: A Study of 25 Countries via a vSIADR Model [J]. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2021, 477(2248).
- [16] Zhang C, Chen C, Shen W, et al. Impact of Population Movement on the Spread of 2019-nCoV in China[J]. Emerging Microbes & Infections, 2020, 9(1):988-990.
- [17] Zhou Y, Xu R Z, Hu D S, et al. Effects of Human Mobility Restrictions on the Spread of COVID-19 in Shenzhen, China: A Modelling Study Using Mobile Phone Data[J]. The Lancet, 2020, 2(8):714-424.

作者简介

顾嘉, 北京大学数学科学学院统计科学中心博士研究生。研究方向为流行病数据建模与分布式统计推断。

陈松蹊(通讯作者), 北京大学讲席教授, 北京大学光华管理学院、数学科学学院教授。研究方向为数理统计、超高维统计推断、大数据算法、大气环境统计、计量经济。电子邮箱: songxichen@pku.edu.cn。

董倩, 国家统计局统计科学研究所副研究员。研究方向为贝叶斯分析及预测、机器学习、数据挖掘、价格指数编制及数字经济测算。

邱宇谋, Iowa State University 助理教授。研究方向为大维协方差矩阵统计推断, 高维统计推断及其在基因组学中的应用, 脑成像统计分析以及因果推断。

(责任编辑: 张晓梅)