

政府干预下离汉人员对疫情的影响

虞睿灵

董洁

常佳琪

摘要

COVID-19 疫情于 2020 年年初在中国大陆地区快速蔓延，为有效防控疫情，中央切断源头，对湖北武汉市进行全面封城。本文通过相关分析、互相关分析分析离汉人员的分布差异和离汉人员对其他省份疫情的影响，通过构建 Arimax 模型来评估武汉封城令这一政策对疫情防控的效果。结果表明，离汉人员在一定程度上对其他省份的疫情存在影响，武汉封城令这一政策干预确实有效的减少了疫情传播。

1. 引言

1.1 背景及研究意义

COVID-19 于 2019 年 12 月在湖北省武汉市出现，其显著的人传人特性伴随着高密度的人口流动导致病毒迅速蔓延，成为新中国成立以来在我国发生的传播速度最快、感染范围最广、防控难度最大的一次重大突发公共卫生事件。

在正值春运时期的疫情爆发初期，大量武汉群众在紧张氛围下纷纷快速逃离武汉，在一定程度上加大了疫情防控的难度。因此，为有效防控疫情，中央于 2020 年 1 月 23 日对武汉进行全面封城。本文旨在通过分析疫情初期离汉人员对其他 30 个省市自治区新冠病例的影响，来评估武汉封城这一政策干预对疫情防控的效果。

1.2 文献综述

在新冠疫情爆发初期，我国各省市自治区在中央领导下多次通过政府干预对疫情进行防控。随着我国新增确诊病例数量的快速减少，各项政府干预政策受到国内外各界高度关注，部分学者也对这些政策效果进行了分析与评价。

其中，林华珍教授团队（2020）主要通过建立时间序列模型得到病例的实时潜在传输率，评估各省公共干预对新冠病例传播率的影响；方亚、石再兴教授（2020）运用 arimax 模型结合互相关分析，研究了交通量和疫情发病率的时间滞后关系，评估武汉封城令对新冠疫情累计发病率的影响；杨华磊等人（2020）运用 SIR 模型分析人口迁徙对新冠病毒感

染人数的影响；廖可等人（2020）运用格兰杰因果检验分析迁徙指数与新增病例数之间的因果关系；陈田木等人（2020）通过构建 SEIAR 模型拟合厦门市新冠肺炎在政府干预前后的有效再生数，评估厦门市政府综合干预的防控效果。

1.3 研究思路

本文主要通过研究疫情初期离汉人员与其他 30 个省市自治区疫情的相关关系，评估武汉封城这一政策干预对疫情发展的影响效果。其具体研究思路如下：1. 通过相关分析分析武汉迁出到各省人数差异的原因；2. 运用格兰杰因果检验，分析累计确诊病例与累计迁出人数的因果关系；3. 通过预白化互相关分析，研究确诊病例与迁出人数之间的滞后相关关系；4. 根据互相关分析结果分类构建 Arimax 模型，评估武汉封城政策对疫情防控的效果。

2. 理论分析

2.1 格兰杰因果检验

Granger 因果关系检验一般用于检验两个变量之间的变动是否存在因果关系，根据变量间因果关系的影响方向不同，其结果可分为单向关系和双向关系。如果一个变量过去发生的变动影响另一个变量当前发生变化，则为单向影响关系；如果两个变量过去发生的变化都影响另外的变量现在发生变化，则二者存在双向 Granger 因果检验关系。Granger 因果检验的完成，需要借助受约束的 F 检验来完成，其相关公式如下：

$$Y_t = \sum_{i=1}^m \alpha_i X_{t-i} + \sum_{i=1}^m \beta_i Y_{t-i} + \mu_i$$

$$F = \frac{(RSS_R - RSS_U) / m}{RSS_U / (n - k)}$$

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \dots = \alpha_m = 0$$

其中，当 F 统计量满足 $F < F_{\alpha}(m, n - k)$ 时，变量 X 是变量 Y 的 Granger 原因。

2.2 互相关分析

互相关函数是描述随机信号 x_t 和 y_t 在任意两个不同时刻 s,t 的取值之间的相关程度，是一种非常有用的测度两个变量之间相关强度和方向的函数。

由于总体的互相关函数是未知的，为了讨论两个时间序列的互相关函数，通常用一个跨度为 n 的样本来估计总体互相关函数，假设这个跨度为 n 的样本为 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 。

样本的互协方差函数为:

$$\gamma_{xy}(k) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(y_{t+k} - \bar{y}) & k = 0, 1, 2, \dots \\ \frac{1}{n} \sum_{t=1}^{n+k} (y_t - \bar{y})(x_{t-k} - \bar{x}) & k = 0, -1, -2, \dots \end{cases}$$

样本的互相关系数为:

$$\hat{\rho}_{xy}(k) = \frac{\gamma_{xy}(k)}{S_x S_y}$$

然而, 由于 CCF 受 x 变量的时间序列结构以及 x 和 y 序列随时间的任何“共同”趋势的影响, 因此在对两时间序列做互相关分析前通常要进行预白化处理, 及利用 X 变量过滤 Y 变量, 消除两者的共同趋势。

2.3 Arimax 分析

Arimax 模型, 是在对变量协整关系研究的基础上将时间序列分析与多元回归分析结合提出得一种模型。它是在 Arima 模型的基础上引入解释变量, 建立了附带解释变量的 Arima 模型, 也称扩展的 Arima 模型, 从而提高时间序列的预测精度。Arimax 模型可表示为:

$$y_t = \beta x_t + n_t$$

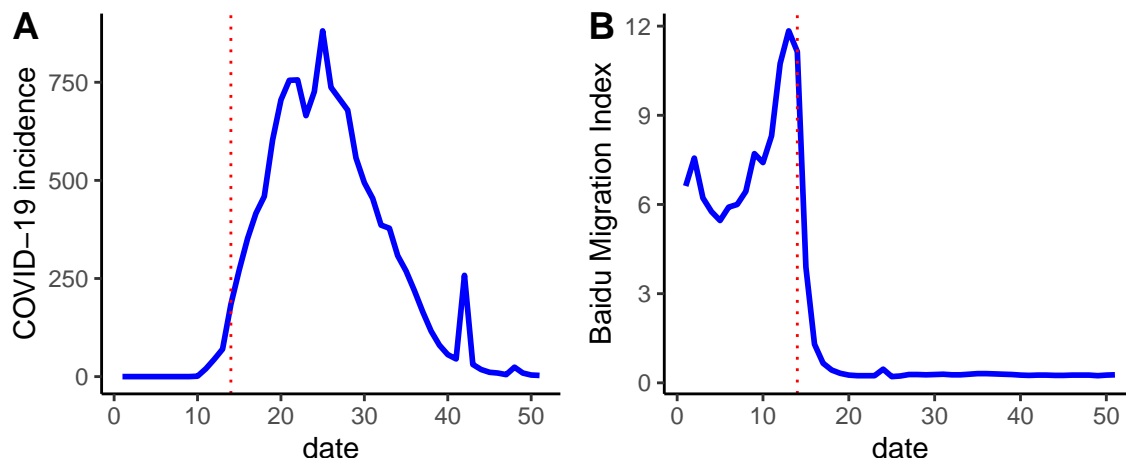
$$n_t = \mu + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_p \varepsilon_{t-p}$$

其中, μ 是常数项, $\alpha_i(1, 2, \dots, p)$ 和 $\varepsilon_i(1, 2, \dots, p)$ 为待估的参数, ε_p 为方差为 σ^2 的白噪声过程, X 是引入的解释变量, β 为其待估的参数。

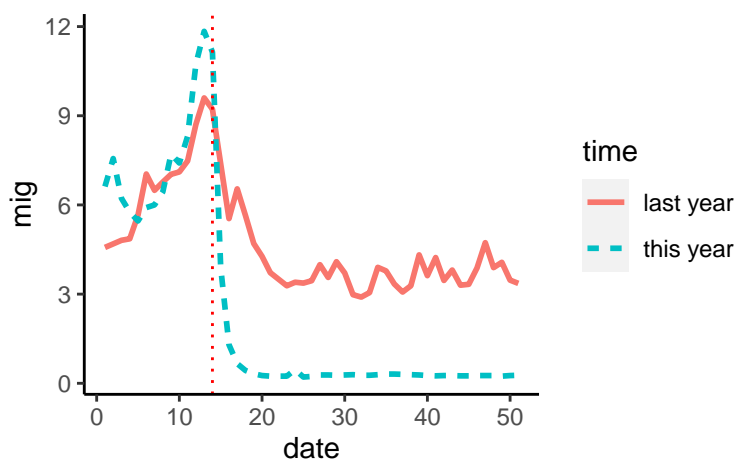
3 描述性统计分析

3.1 全国确诊数和迁徙人数趋势分析

本文主要探究的是武汉迁出人员和确诊病例之间的关系, 下图是 2020 年 1 月 10 日-2020 年 2 月 29 日的武汉是百度迁徙指数和全国除湖北地区以外的新增病例数的时间趋势图, 图中红色的虚线代表的是武汉封城政策。由图 A 可以看出, 全国新增病例数在武汉封城之后呈现出快速增长的趋势并很快达到了最高点。而从图 B 可以看出, 武汉封城政策从很大程度上减少了离汉人数, 迁徙指数逼近 0, 从而探讨这一政策的实施是否对新冠疫情的控制有显著作用。



下图是 2020 年 1 月 10 日-2020 年 2 月 2 武汉市迁徙指数与去年农历同期武汉市迁徙指数的对比图。由图可以看出，随着春运的到来，武汉迁出人数逐渐增加，在封城前今年的迁出人数甚至比去年还要多，而在采取了封城政策以后，迁徙指数大幅度下降。



3.2 各省确诊数和迁徙人数对比分析

下图中图 A 是全国累计确诊情况的分布，所用到的数据是截止 2020 年 2 月 29 日的各省累计确诊数据。由图中可以看出，除湖北外累计确诊最多的是广东省，其次是河南省、浙江省、湖南省、安徽省，由图中颜色分布可以看出，湖北附近的省份颜色较深。

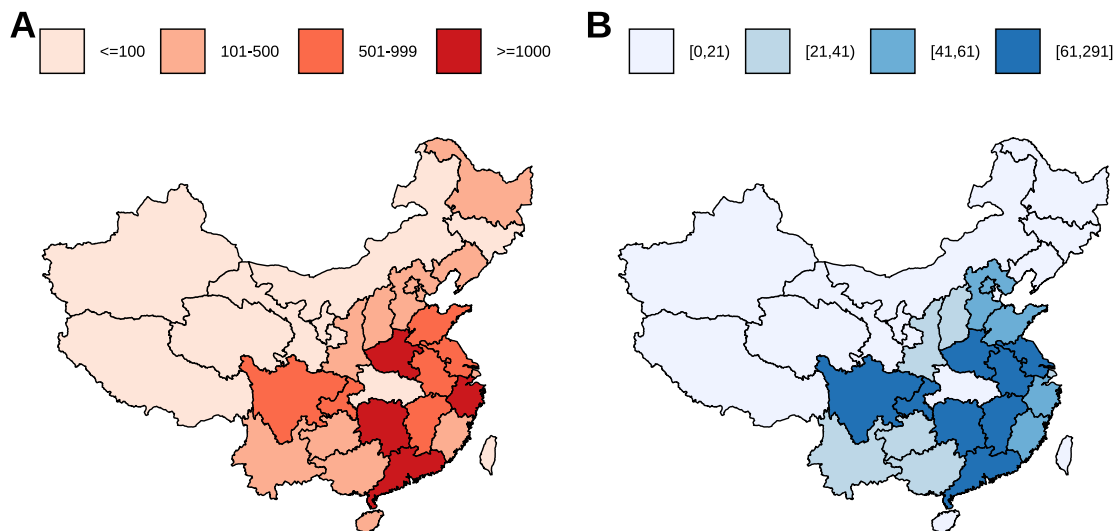
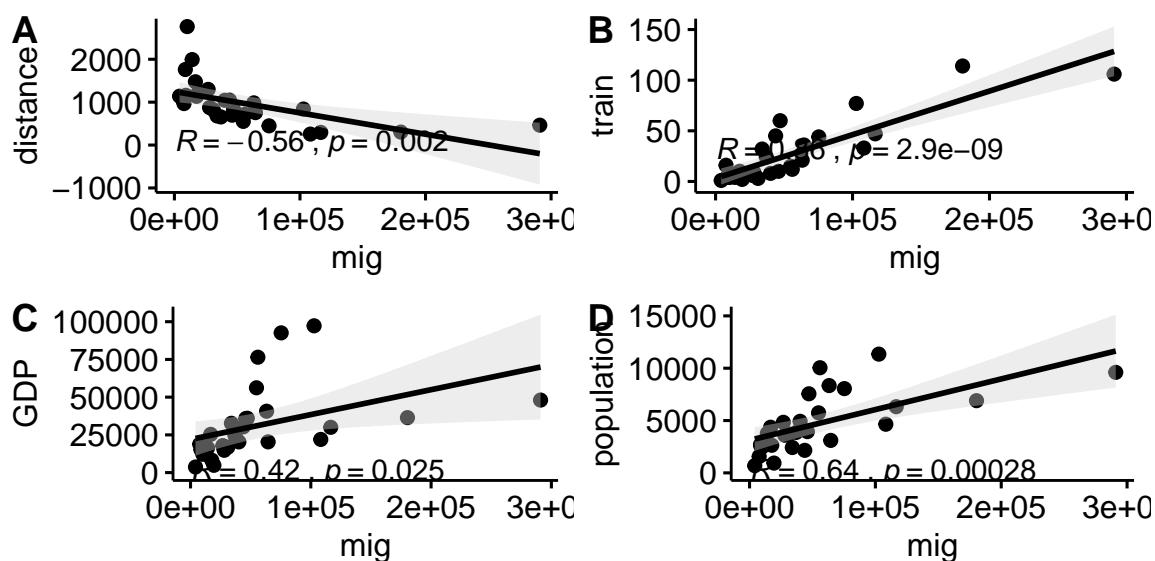


图 B 是对离汉人员的去向统计，所统计的是 2020 年 1 月 10 日——2020 年 2 月 29 日武汉去往各个省的累计迁出人口 (单位：千人)，由图可以看出，武汉的迁出人员去往河南的最多，其次是湖南、安徽、江西等，人数排名前十的省份大多在湖北周边。

4. 实证分析

4.1 相关分析



以上是针对可能影响各省从武汉迁入人口数的指标做的相关分析的结果，图 A 是迁出人口和距离的相关关系图，这里的距离值得是各省省会城市距武汉的距离。从图中可以看

出，武汉迁到各省的累计人口与武汉离其省会城市的距离成反比，相关系数为-0.56。图 B 为迁出人口和火车趟数的相关关系图，火车趟数在一定程度上可以代表该省的交通便利程度，由图可以看出，二者之间呈现出正相关关系，相关系数为 0.86。图 C 为迁出人口和各省的 GDP 的相关关系图，地区生产总值反映该地区的经济实力和市场规模，从图中可以看出，武汉迁到各省的人口与其地区生产总值呈现出正相关关系，相关系数为 0.42。图 D 是迁出人口与各省年末总人口的相关关系图，二者之间存在正相关关系，相关系数为 0.66。以上所研究的四个指标均与武汉迁到各省的累计人数有相关关系，且其 P 值均显著。

4.2 格兰杰因果关系检验

本文为了分析各省确诊病例与其从武汉迁入的人口总数之间的关系，对二者进行了格兰杰因果关系检验，在检验之前首先对两列时间序列数据的平稳性进行检验，检验得到的结果如下表所示：

原假设	P 值	检验结果
case does not Granger Cause mig	0.9807	不拒绝
mig does not Granger Cause case	0.0184	拒绝

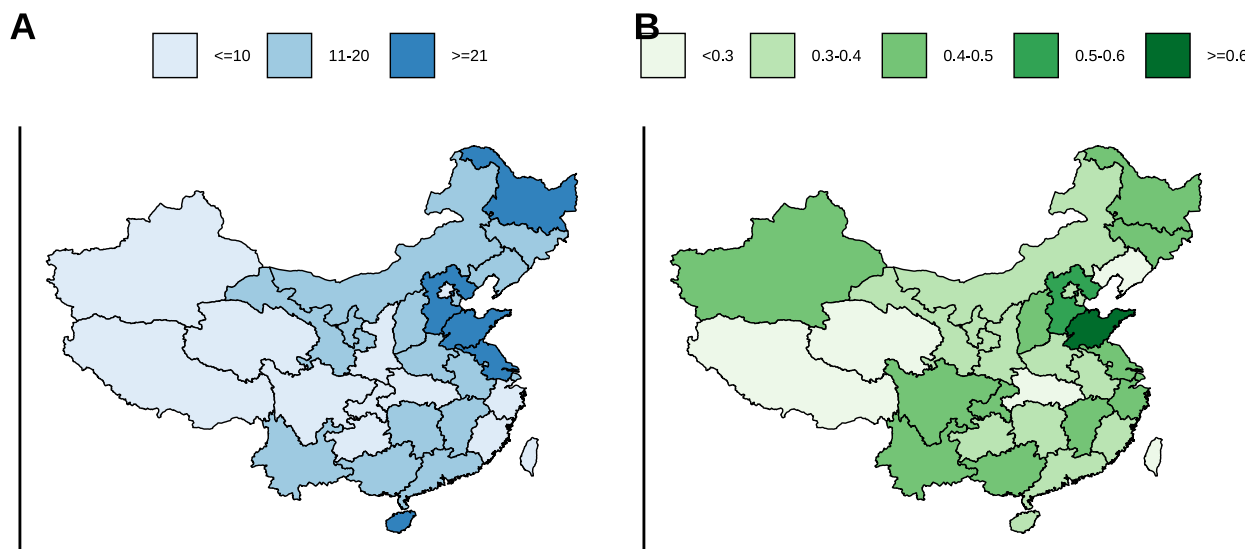
上表中 mig 代表的是从武汉迁入的人口总数，case 代表的是各省的确诊病例，原假设是前者不是后者的因，检验得到的 P 值为 0.0184，故拒绝原假设，即前者是后者的因。因此，接下来通过互相关分析来进一步探究二者之间的关系。

4.3 互相关分析

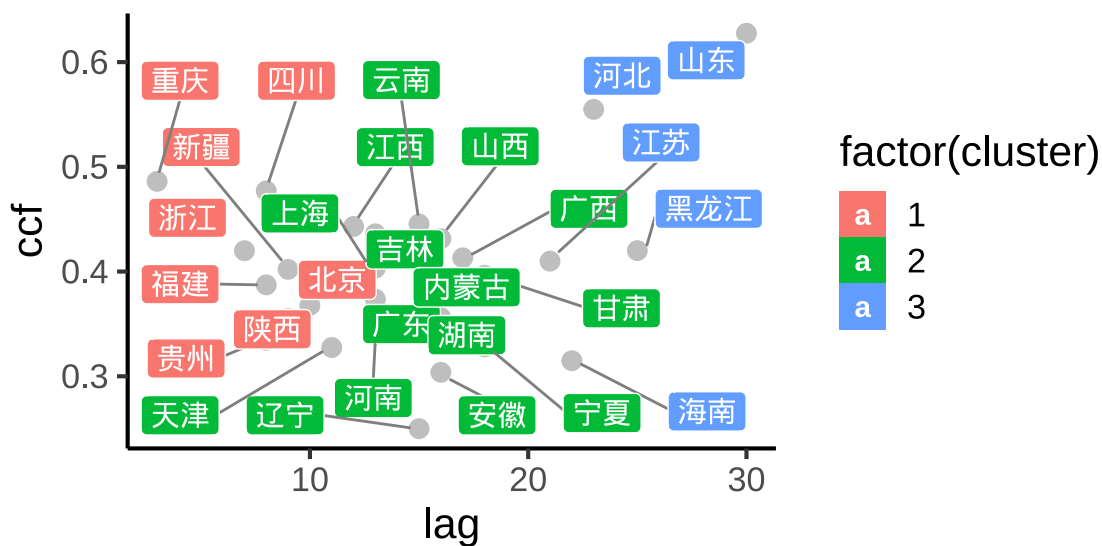
由前面的理论可知，如果直接用两列数据做互相关分析，会因为存在公共趋势而导致结果有偏，故需要先对数据进行预白化。预白化的第一步是用 X 变量拟合一个 ARIMA 模型，并计算模型残差；第二步是用 X 的 ARIMA 模型对 y 变量进行过滤。由于本次研究中西藏和青海的疫情数量较小，故仅研究除去西藏和青海的省份的迁徙与疫情的互相关分析。结果见下图：

下图左图展示了人口迁移与新冠肺炎新增病例之间积极的时间关联。在靠近湖北武汉的省市，比如陕西、重庆、湖南，时间滞后期都在 10 天之内，所以它们对离汉人员是比较敏感的；从武汉流入人口较多的省份，比如浙江、广东和四川，时间滞后期也比较短，尤其是浙江省，是除了湖北以外第一个进入疫情一级响应的省份；从武汉迁往安徽，江苏，进而北上迁往山东河北的人，其时滞天数比较长，但时间较长并不意味着对输入病例的反应较弱。相反，这反映了这些省份累积病例的总体变化不仅是由输入病例驱动，还通过本地传播感染。在远离武汉的地区，如果滞后时间较长，比如黑龙江、海南，通过观察截截止

2月29日各省累计病例数，发现累积发病人数较高，这可能是由于起初一代病人确诊时间长，扩大了向二代、三代病例传播的几率。右图显示，除西藏、新疆和湖北以外，各省份的离汉人员与新冠肺炎新增病例有着积极的正相关关系。



从下图中可以看出，根据每个省份最显著的滞后期数，有 29% 的省份的滞后期在 10 天以内，58% 的省份滞后期在 11-20 天；18% 的省份的滞后期为 20 以上。下文将对每个集群建立 ARIMAX 模型来对疫情数据进行拟合和预测。



4.4 ARIMAX 模型

建立 ARIMAX 模型通常分为四步：

1、对数据进行处理，并进行平稳化检验；

- 2、考察观测变量和输入变量的互相关关系以确定滞后回归项；
- 3、建立 ARIMAX 模型，确定 p, d, q 的值；
- 4、进行模型检验及预测。

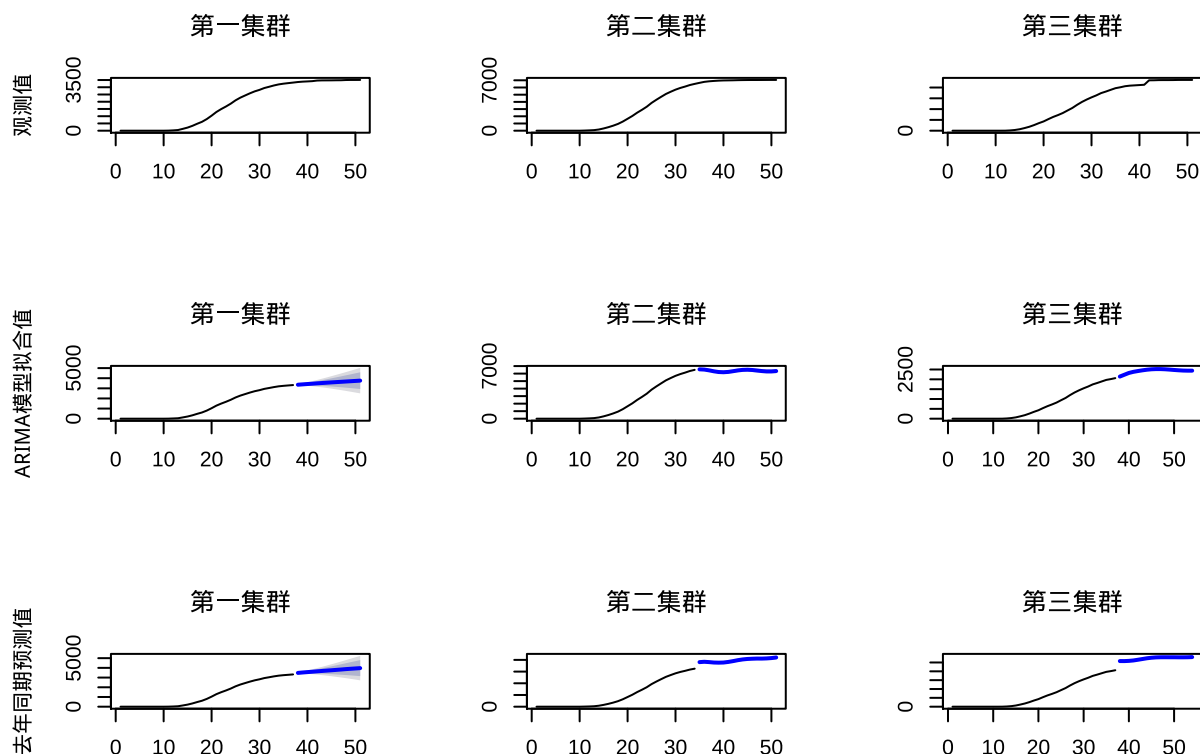
对三大集群的疫情数据进行 ADF 平稳性检验，发现在 5% 显著性水平下，全部通过检验。

类别	Dickey-Fuller	p-value	结果
第一集群	-3.8212	0.0298	平稳
第二集群	-4.6805	<0.01	平稳
第三集群	-3.6175	0.0453	平稳

考察三大集群经过平稳化后的迁移数据和疫情数据的互相关关系，发现分别滞后 8 阶、滞后 16 阶与滞后 26 阶最显著，这也与三大集群是按不同滞后阶数的汇总数据情况吻合。分别利用滞后 8 阶、滞后 16 阶与滞后 26 阶建立三个集群的 ARIMAX 模型，三个模型系数的 P 值在 10% 的显著性水平下通过检验。进行 LB 检验的结果发现，三个模型的残差均是独立和随机的序列，最终得到三个模型的信息见下表：

滞后阶数	Model	RMSE	MAE	AIC	BIC	百分比
1-10	(1,1,0)	25.66	20.26	282.03	287.5	29%
11-20	(2,0,0)	51.18	41.87	207.09	211.54	53%
20 以上	(2,0,0)	12.85	10.37	102	103.99	18%

运用上述建立的 ARIMAX 模型对疫情数据进行拟合和预测。选取 2 月 15 日为疫情转折点，1 月 10 日至转折点前的数据用来建模，预测转折点后的疫情发展情况并与真实观测值比较，判断模型准确性。再用去年同期迁徙数据进行预测，预测值与真实值之间的差距即为武汉封城令的政策效果。下图第一行展示了三大集群疫情观测值趋势，由于滞后阶数在 11-20 阶的省份占比最多，故可明显看出第二集群疫情人数远高于第一和第三集群。第二行展示了利用 2020 年的迁徙数据建立的 ARIMAX 模型及其预测值，与各集群真实值相比相差较小，说明模型建立准确。第三行展示了利用 2019 年农历同期迁徙数据对疫情发展做出的预测，可以看出三大集群去年同期数据预测均有明显的“跃升”趋势，即由于没有武汉封城令的干扰，去年迁移人数必定大于今年迁徙人数，进而预测的疫情感染人数也会有大幅攀升。最终本次研究得到，截止 2 月 29 日，第一集群用 ARIMAX 模型预测出疫情人数是 3759(95%；2501, 5018) 人，真实值是 3507 人，用去年同期数据预测为 3972(95%；2712, 5230) 人；第二集群预测的疫情人数是 6335(95%；6064；6607) 人，真实值是 7053 人，去年同期数据预测为 8409 人 (95%，8138，8680)。第三类预测出疫情人数是 2441(95%；2375，2407) 人，真实值是 2353 人，用去年同期数据预测为 2805 人 (95%；2739，2872) 人。



5 结论与不足

5.1 结论

自 2020 年 1 月 10 日起，武汉迁出客流量开始快速增长，而在 1 月 23 日实施封城令后，出境客流量急剧下降。截至 1 月 23 日，武汉迁出人员中有 71% 前往湖北其他城市，29% 迁移到其他省份。湖北省迁移数据与各省新冠疫情确诊人数呈现相似的波动格局，格兰杰因果检验发现人口迁移数据是各省疫情增长的原因之一。

利用拟合的 ARIMA 模型对各省疫情数据进行预白化，计算出各省二者之间的相关系数。按照滞后期数的不同划分为三大集群，29% 的省份的滞后期在 10 天以内，58% 的省份滞后期在 11-20 天；18% 的省份的滞后期为 20 以上。

我们研究了武汉迁入人口数的潜在影响因素。分析表明，武汉迁入人口数与距离武汉的距离呈负相关，相关系数-0.56($P=0.002$)，与火车趟数、GDP 和各省人口数呈现正相关，相关系数分别为 0.86($P=29e-09$)、0.42($P=0.025$)、0.64($P=0.0003$)。

每个省份集群的最优 ARIMAX 模型如上表所示。滞后 1-10 天的省份 ARIMA(1,1,0) 最优，滞后 11-20 和滞后 20 天以上的省份均为 ARIMA(2,0,0) 最优。运用去年农历同期数

据估计, 三集群省份武汉封城令可能分别减少 465、1356 和 452 人, 一共会减少 2273(95%: 679, 3869) 例一代病例, 在对新出现的传染病缺乏治疗和免疫的情况下, 交通管控这种应急法被证明是一种有效的公共卫生干预措施。我们的研究结果为政府机构确定最佳的疾病控制策略和改善对未来突发公共卫生事件的准备提供了重要依据。

5.2 不足

本文研究也存在一定不足, 首先, 新冠疫情诊断能力和标准的变化可能导致改变了确诊人数的长期趋势。例如, 在疫情爆发初期, 新冠肺炎曾被误诊, 后来由于诊断标准扩大, 在 2 月 13 日左右病例激增, 这可能在某种程度上扭曲了预测的准确性。其次, 迁徙数据与疫情数据之间的时间相关性可能会被其他预防措施混淆, 如加强自我隔离、广泛使用口罩和手部卫生处理。最后, 本文分析的省份不包括西藏、青海和湖北, 有一定局限性。

参考文献

1. Huazhen Lin, et al. *Comparative Analysis of Early Dynamic Trends in Novel Coronavirus Outbreak: A Modeling Framework*. medRxiv 2020.02.21.20026468.
2. Zaixing Shi,Ya Fang. *Temporal relationship between outbound traffic from Wuhan and the 2019 coronavirus disease (COVID-19) incidence in China*.medRxiv 2020.03.15.20034199.
3. 杨华磊, 吴远洋, 蔺雪钰. 新型冠状病毒肺炎、人口迁移与疫情扩散防控. 中国管理科学 [J], 2020(03): 1-10.
4. 廖可等. 人口迁入与新增确诊数的趋势关系及因果量化分析. 数学建模及其应用 [J], 2020(01): 23-28.
5. 陈田木等. 厦门市新型冠状病毒肺炎人群传播能力计算与防控措施效果的模拟评估. 厦门大学学报 (自然科学版), 2020(03): 1-8.