



## COVID-19 Multiple Quantile Regression Model and Application

---

Li Qian

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

August 1, 2021

投稿专题 (track) 序号: 7 投稿专题 (track) 名称: 数据科学与商务智能

# 新冠肺炎多元分位数回归模型及应用

李倩<sup>1</sup>

( 1. 中南大学 商学院, 长沙 410006 )

**摘要** 预测新冠肺炎是当前的研究热点之一, 传统线性回归模型采用最小二乘法对参数进行估计, 无法区分模型高估和低估的影响。因此, 本文采用分位数回归对新冠肺炎的累计确诊人数进行预测, 通过设定不同分位数系数区分高估和低估的损失值, 反映预测值的整体分布。最后通过对比实验说明了多元分位数回归模型在预测累计确诊人数方面的有效性, 并确定模型的最佳时间窗为 14 天。

**关键词** 新冠肺炎; 多元分位数回归; 预测

中图分类号 C931.6

## COVID-19 multiple quantile regression model and application

LI QIAN<sup>1</sup>

(1. Central South University, Changsha 410006, China)

**Abstract** Predicting COVID-19 is one of the current research hotspots. The traditional linear regression model uses the least square method to estimate the parameters, and it is impossible to distinguish between the effects of overestimation and underestimation of the model. Therefore, this article uses quantile regression to predict the cumulative number of confirmed cases of COVID-19, and distinguishes overestimated and underestimated loss values by setting different quantile coefficients to reflect the overall distribution of predicted values. Finally, a comparative experiment illustrates the effectiveness of the multiple quantile regression model in predicting the cumulative number of confirmed cases and it is found that the best time window is 14 days.

**Key words** COVID-19; multiple quantile regression; prediction

## 1 引言

新型冠状病毒肺炎(COVID-19), 简称“新冠肺炎”, 是由 2019 新型冠状病毒(SARS-CoV-2) 引起的肺炎。截止至 2021 年 4 月 20 日 14 时, 全球累计确诊人数超过 1 亿, 累计死亡 300 余万人, 其中, 印度、巴西、土耳其等多个国家新增确诊人数超过五万 (数据来自百度新型冠状病毒肺炎疫情实时大数据报告)。新冠肺炎传播如此迅速的主要原因包括: 第一, 新冠肺炎患者初期症状包括发热、乏力和干咳, 这些症状与普通感冒症状相似, 因此在感染初期难以察觉和区分。第二, 在经济全球化的驱动下, 大规模的人口流动直接或间接地导致了新冠病毒进一步传播。

---

通信作者: 李倩, 中南大学商学院, 研究生, e-mail: 18271835758@163.com

2019年发生的新冠肺炎不仅对个人生活产生了影响，并且对湖北省的农业、交通运输业和建筑业（刘世锦等，2020）、国内旅游业（夏杰长等，2020）甚至全球产业链（沈国兵，2020）产生了严重冲击。对新冠肺炎进行准确预测，可以提前采取有效措施减少新冠肺炎造成的人员伤亡与经济损失。

目前，新冠肺炎的预测模型研究主要包括两大类：基于动力学的传染病模型研究和基于机器学习的预测模型研究。其中，基于动力学的传染病模型将人群总数看作一个定值，并将其细分为易感人群、感染人群、康复人群、死亡人群、隔离人群以及暴露人群，常见模型有：SIR模型（张艳霞等，2020）、SIRD模型（Alòs等，2020）、SIQR模型（Crokidakis，2020）和SEIR模型（Wu等，2020；范如国等，2020）。利用传染病模型可以模拟预测某一个国家或地区的疫情爆发情况，但是无法考虑现实中存在的人口流动因素，因此，传染病模型在预测时会产生无法避免的偏差。基于机器学习的预测模型可以将气候因素（Zhu等，2020；Borah等，2020）、人口流动因素（Jia等，2020；Wu等，2020）和经济发展水平等因素考虑在内，常用的模型包括：随机森林（Chakraborti等，2020）、贝叶斯时间序列模型（Feroze，2020）、支持向量机（Ball，2021）和线性回归等，其中，Mojjada等（2020）通过对比LASSO回归、支持向量机和线性回归三种机器学习模型预测未来10天内新增感染COVID-19的人数、死亡率和恢复人数，验证了线性回归在预测新冠肺炎病例方面的有效性。

值得注意的是，传统线性回归采用最小二乘法进行拟合，通过最小化预测值和实际值之间的差值的绝对值之和求解回归系数，相同单位的高估和低估的损失值是相同的。但是在现实生活中，经济发达、人口密度大的中心城市通过一定的资源浪费防止疫情大范围扩散是可行的，即模型高估的损失值小于模型低估的损失值。反之，对于经济欠发达、人口密度小的乡村城市来说，模型高估的损失值大于模型低估的损失值。

为了区分模型高估及低估的损失值，弥补线性回归无法估计自变量（解释变量）对因变量（被解释变量）的整体影响（Liu等，2021）的缺点，Koenker和Bassett(1978)提出了分位数回归方法，对解释变量和被解释变量在特定百分位数（即“分位数”）水平下的关系进行预测，从而更加全面的描述被解释变量的分布情况。另外，与最小二乘法相比，分位数回归不要求很强的正态分布，即在非正态情况下，分位数估计比最小二乘估计更加有效。

基于以上分析，本文采用多元分位数回归方法对新冠肺炎的累计确诊人数进行预测。首先，分别计算最高温度、最低温度、最佳空气质量、最差空气质量与武汉之间的物理距离、从武汉流入的人口数、人口总数、国内生产总值（GDP）等影响因素与新冠肺炎确诊人数之前的相关系数，提取出相关性较强的影响因素。然后，根据筛选出来的影响因素建立一个多元分位数回归模型，并对新冠肺炎累计确诊人数进行预测。

本文的结构如下：第二章介绍研究方法，第三章主要是实验结果的展示与分析，第四章对全文进行总结，并指出未来的研究方向。

## 2 方法

### 2.1 数据收集

本文收集了国内 282 个地级市在 2020 年 1 月最高温、最低温、最佳空气质量、最差空气质量、与武汉之间的物理距离 (KM)、从武汉流入的人口数 (2020 年 1 月 1 日至 2020 年 1 月 24 日)、人口总数 (万人)、GDP (万元) 和不同时期的新冠肺炎累计确诊人数数据。其中, 新冠肺炎数据来源于国家卫生健康委员会; 气候数据来源于 2345 天气网; 与武汉之间的距离、从武汉流入的人口数、人口总数、GDP 数据来源于参考文献 12 中的支撑数据。

### 2.2 相关系数

相关系数是反映两个变量之间相关程度的一个重要指标。对于两个随机变量  $X = (X_1, X_2, \dots, X_n)$  和  $Y = (Y_1, Y_2, \dots, Y_n)$  来说, 相关系数  $Cor(X, Y)$  的计算过程如下 (张应应, 2016): 首先计算随机变量的均值  $EX$  和  $EY$ , 然后根据公式 (1) 计算协方差  $Cov(X, Y)$ :

$$Cov(X, Y) = E[(X - EX)(Y - EY)] \quad (1)$$

接下来, 计算随机变量的方差  $Var(X)$  和  $Var(Y)$ , 最后根据公式 (2) 计算出变量之间的相关系数  $Cor(X, Y)$ :

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \quad (2)$$

### 2.3 分位数回归

本文根据 2020 年 1 月最低温、与武汉之间的距离、从武汉流入的人口数、人口总数、GDP 五个指标建立了一个多元分位数回归 (Multiple quantile regression) 模型。由于这五个指标的量纲单位存在较大差异, 为消除量纲影响需要先对数据进行归一化, 将数据映射到  $[0, 1]$  区间内。

#### 2.3.1 模型表示

假设  $i$  表示自变量的序数 ( $i = 1, 2, \dots, 5$ ), 则新冠肺炎的多元分位数回归模型如下:

$$Y_q = \sum_{i=1}^5 X_i * \beta_i + b, \quad 0 < q < 1 \quad (3)$$

其中,  $X_1$  表示 2020 年 1 月最低温,  $X_2$  表示该城市与武汉之间的物理距离,  $X_3$  表示从武汉流入该城市的人口总数,  $X_4$  表示该城市的人口总数,  $X_5$  表示该城市的 GDP,  $q$  是分位数系数,  $Y_q$  代表不同分位数水平下的累计确诊人数,  $b$  为截距,  $\beta_i$  是待估计的参数。

#### 2.3.2 损失函数

多元分位数回归模型的优化目标是 minimized 加权绝对误差之和, 计算公式如公式 (4):

$$L_q = \min \left( \sum_{Y_q \geq \hat{Y}} q * |Y_q - \hat{Y}| + \sum_{Y_q < \hat{Y}} (1 - q) * |Y_q - \hat{Y}| \right) \quad (4)$$

式中,  $Y_q$  代表模型预测的累计确诊人数,  $\hat{Y}$  代表实际的累计确诊人数,  $L_q$  代表不同分位数  $q$  水平下的损失函数, 当  $q < 0.5$  时, 模型低估的损失大于高估的损失; 当  $q > 0.5$  时, 模型高

估的损失大于低估的损失；当  $q = 0.5$  时，模型高估的损失与低估的损失相等，模型简化成传统多元线性回归模型。

### 2.3.3 评估指标

传统线性回归模型往往采用  $R^2$  评估模型拟合效果，分位数回归中采用的是 Laitila(1993)提出的一种伪拟合测度指标 Pseudo  $R^2$ 。Pseudo  $R^2$  是样本数量较大时保留了传统  $R^2$  (简称  $R^2$ ) 的相似性的泛化形式。Pseudo  $R^2$  的计算方式如下：首先从定性因变量模型中获取估计系数和回归样本协方差矩阵，回归样本协方差矩阵的计算公式如下：

$$\hat{\Sigma}_X = m^{-1} \sum_{i=1}^m (X_i - \bar{X})'(X_i - \bar{X}) \quad (5)$$

其中， $X_i$  是一个  $1 \times k$  的回归向量， $\bar{X}$  代表样本均值。将计算所得协方差矩阵代入如下公式计算出  $R^2$ ：

$$R^2 = \frac{\hat{\beta}' \hat{\Sigma}_X \hat{\beta}}{\hat{\beta}' \hat{\Sigma}_X \hat{\beta} + \hat{\sigma}^2} \quad (6)$$

其中， $\sigma^2$  代表方差， $\hat{\sigma}^2$  代表残差平方和的均值， $\tilde{\beta}$  和  $\tilde{\sigma}^2$  是  $\beta$  和  $\sigma^2$  的一致估计量，如果估计量一致，Pseudo  $R^2$  在概率上收敛于  $R^2$  的极限。即 Pseudo  $R^2$  的计算公式如下：

$$\text{Pseudo } R^2 = \frac{\tilde{\beta}' \tilde{\Sigma}_X \tilde{\beta}}{\tilde{\beta}' \tilde{\Sigma}_X \tilde{\beta} + \tilde{\sigma}^2} \quad (7)$$

在普通最小二乘估计的线性回归中，Pseudo  $R^2$  简化为  $R^2$ 。

## 3 实验

本文实验主要基于 python 3.7 和 IBM SPSS Statistics 26 实现。

### 3.1 相关系数

在研究气候因素对新冠肺炎的影响时，Zhu 和 Xie(2020)采用广义加性模型 (GAM) 研究了印度尼西亚雅加达的平均气温与新冠肺炎感染病例数量的相关性，发现在气温低于  $3^\circ\text{C}$  时，平均气温与新冠肺炎病例数线性正相关。但是，Borah 等 (2020) 从数学变量的角度，通过分析全球的气候数据 (主要是温度、湿度) 和新冠肺炎感染病例数据发现，爆发初期温度较高的地区，病毒的传播速度较慢。这与 Zhu 和 Xie(2020) 的研究结果存在差异。因此，为了检验中国国内城市的气温对新冠肺炎传播的影响，本文分别计算了湖北省内城市 (包括恩施、黄石、黄冈、荆州、荆门、宜昌、十堰、襄阳、孝感、咸宁、随州) 的变量之间的相关系数和除湖北省之外的其它省份城市变量之间的相关系数。相关系数矩阵图如图 1 和图 2 所示。

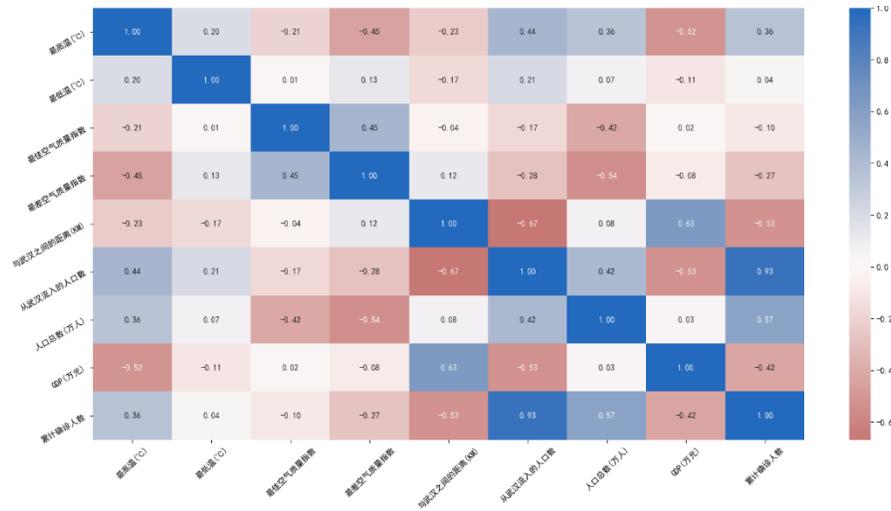


图 1 湖北省内城市变量之间的相关系数矩阵图



图 2 其它省份城市变量之间的相关系数矩阵图

由图 1 和图 2 可知，中国 296 个城市的气温与新冠肺炎确诊人数是正相关的，但是不同城市的气候与新冠肺炎累计确诊人数之间的相关系数不同，即气温对新冠肺炎确诊人数的影响具有地区的差异性。

另外，通过对比两个相关系数矩阵图可以发现：（1）对于湖北省内城市来说，最高温和累计确诊人数之间的相关性强于最低温和累计确诊人数之间的相关性，与其他省份城市的情况相反。考虑到湖北省内城市和其他省份城市数量上的差异，最终选取最低温作为多元分位数回归模型的自变量之一。（2）湖北省内城市的空气质量指数与累计确诊人数之间的相关性是负向的，而其他省份城市是正向的，因此在建模过程中不考虑空气质量这一指标。

### 3.2 多元分位数回归

根据相关系数矩阵选择 2020 年 1 月最低温、与武汉之间的距离、从武汉流入的人口数、人口总数和 GDP 作为分位数回归模型的输入变量，2020 年 2 月 14 日的累计确诊人数为输出变量。

不同分位数水平下模型的参数估计结果如表 1 所示。

表 1 不同分位数水平下模型的参数估计

参数	q=0.05	q=0.25	q=0.5	q=0.75	q=0.95
截距	-10.951	-5.114	1.164	1.064	-9.084
1 月最低温	8.716	4.298	4.251	4.568	23.832
与武汉之间的距离	4.970	3.085	-5.887	-1.515	38.049
从武汉流入人口数	2336.546	2817.745	3211.763	4507.553	6919.745
人口总数	27.563	40.223	59.222	91.649	151.296
GDP	121.411	158.778	193.857	262.619	312.306

随着分位数系数从 0.05 增加到 0.95，人口总数的系数从 27.563 上升到 151.296，从武汉流入人口数和 GDP 的回归系数也不断增大，自变量对被解释变量的回归系数不断增大。由此可见，不同分位数水平下，自变量对因变量的影响程度是不同的。对于北京、上海、广州、深圳等发达城市来说，模型低估的损失值大于高估的损失值，取分位数系数  $q = 0.95$ ，参数估计结果如表 2 所示。

表 2  $q = 0.95$  分位数下模型参数估计

参数	系数	标准误差	t	自由度	显著性	下限	上限
截距	-9.084	56.3286	-0.161	276	0.872	-119.973	101.804
1 月最低温	23.832	55.5353	0.429	276	0.668	-85.494	133.159
与武汉之间的 距离	38.049	67.4391	0.564	276	0.573	-94.712	170.809
从武汉流入人 口数	6919.745	117.3622	58.961	276	0.000	6688.707	7150.784
人口总数	151.296	129.8009	1.166	276	0.245	-104.230	406.821
GDP	312.306	98.3265	3.176	276	0.002	118.741	505.871

通过实验结果发现：武汉流入人口和 GDP 对累计确诊人数的影响具有显著性。不同分位数水平下武汉流入人口和 GDP 的参数估计结果如图 3 所示：

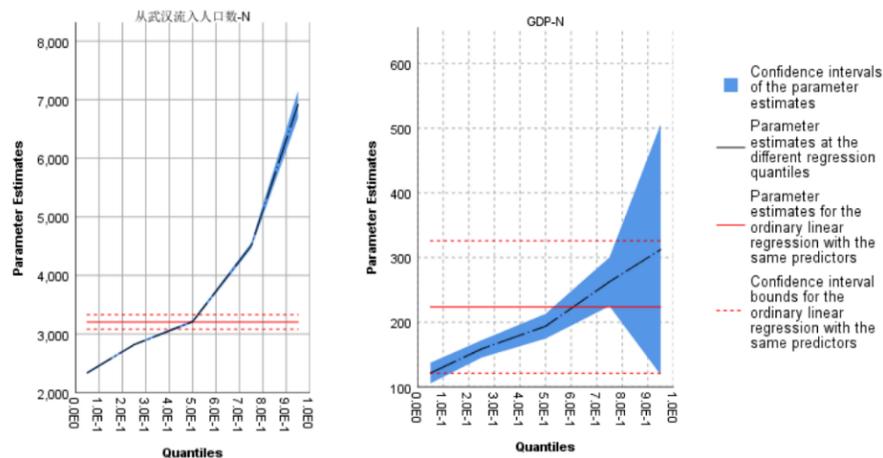


图 3 单因素参数估计图

其中，蓝色区域表示参数估计的置信区间，黑线表示不同分位数水平下的参数估计，红色实线表示传统线性回归的参数估计，红色虚线表示传统线性回归 95%置信区间边界。不同分位数水平下根据从武汉流入人口数和 GDP 预测累计确诊人数的结果如图 4 所示：

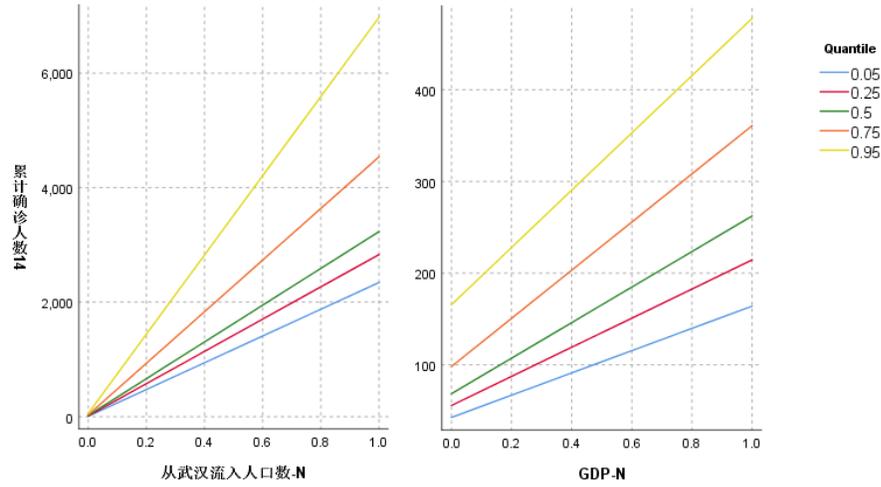


图 4 单因素累计确诊人数预测图

根据从武汉流入人口数-累计确诊人数的预测图中可知：（1）从武汉流入的人口数越多，累计确诊人数越多。因此，合理控制疫情爆发地区的人口流动是控制疫情的必要手段之一。（2）在相同的从武汉流入人口水平下，0.05 分位数的预测值与最小二乘估计值的差值远远小于 0.95 分位数的预测值与最小二乘估计值的差值，而且流入人口数越多，差值越大。这种非对称分布意味着从武汉流入人口基数越大，大范围爆发疫情的可能性越大。GDP 和累计确诊人数的预测关系与武汉流入人口数类似，GDP 越高，国民经济活跃度越高，人群的接触也更加频繁（袁宏永等，2020）。

另外，多元分位数回归模型预测值与实际观察值在不同分位数水平下的分布情况如图 5 所示：

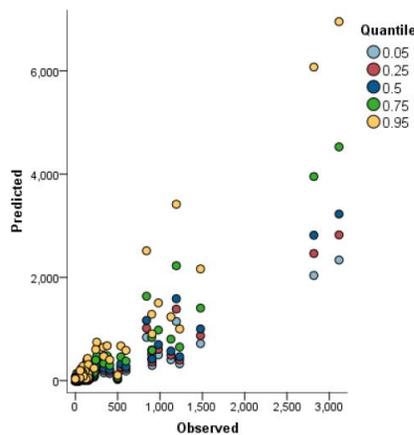


图 5 多元分位数回归预测值-观察值对比图

图 5 中横坐标代表实际观察值，纵坐标代表模型预测值。浅蓝色圆圈、红色圆圈、深蓝色圆圈、绿色圆圈、黄色圆圈分别表示分位数系数为 0.05、0.25、0.5、0.75、0.95 时的多元

分位数回归预测结果。

### 3.3 对比实验

为了验证多元分位数回归模型的有效性，本文对比了该模型与一元分位数回归模型的实验结果。另外，为了使多元分位数回归模型的预测更加准确，本文对比了模型在不同时间窗（即距离 2020 年 1 月 31 日的天数）的累计确诊人数的预测效果。

#### 3.3.1 一元分位数回归与多元分位数回归的对比实验

表 3 一元分位数回归与多元分位数回归的对比结果

自变量	模型 指标	分位数水平				
		q = 0.05	q = 0.25	q = 0.5	q = 0.75	q = 0.95
1 月最低温	Pseudo R2	0.008	0.009	0.015	0.018	<b>0.075</b>
	MAE	92.8128	87.7247	84.5496	92.5957	411.4238
与武汉之间的 距离	Pseudo R2	0.017	0.025	0.033	0.035	<b>0.162</b>
	MAE	91.7188	86.1116	83.0038	91.7009	491.9962
从武汉流入人 口数	Pseudo R2	0.434	0.491	0.562	0.649	<b>0.821</b>
	MAE	51.9977	439.3277	37.5638	49.7576	119.0335
人口总数	Pseudo R2	0.040	0.057	0.073	<b>0.080</b>	0.019
	MAE	89.3217	82.4637	79.5495	88.2758	215.6013
GDP	Pseudo R2	0.066	0.092	<b>0.105</b>	0.097	0.019
	MAE	84.9200	79.6424	76.8594	82.2154	187.7533
多元分位数回 归	Pseudo R2	0.479	0.540	0.632	0.706	<b>0.854</b>
	MAE	45.6662	35.9487	31.5511	43.2463	95.5478

表中粗体表示该行的最大值，可以发现：（1）除了人口总数和 GDP 之外，其它模型均在分位数系数为 0.95 时表现最佳。（2）在 Pseudo R2 方面，多元分位数回归模型和基于武汉流入人口数的一元分位数回归模型的表现较为优异，都超过了 0.8，但是多元分位数回归模型在不同分位数水平下的表现均优于一元分位数回归模型。（3）当分位数系数从 0.75 增加到 0.95 时，基于武汉流入人口数的一元分位数回归模型的绝对误差猛增，不如多元分位数回归模型发挥稳定。

#### 3.3.2 不同时间窗的对比实验

在时间窗对比实验中，本文根据 2020 年 1 月 1 日至 1 月 24 日期间从武汉流入人口数据预测 2020 年 2 月的累计确诊人数，原因有两点：（1）武汉市政府于 2020 年 1 月 23 日采取“封城”隔离措施后，导致 1 月 24 日之后武汉流出人口锐减。（2）由于数据上报机制不完善和检测手段的缺失，2 月 1 日以前的疫情数据存在大量缺失和失真。因此，本文设置起始日期为 2 月 1 日，并根据新冠肺炎潜伏期平均为 3~5 天，最短 1 天，最长 29 天的特点设置了时间间隔。

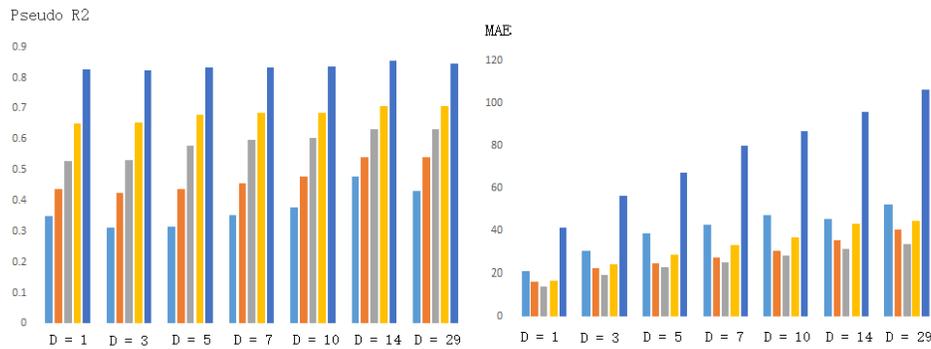


图 6 不同时间窗的对比图

通过对比可以发现，多元分位数回归模型的 Pseudo R2 最大值均出现在 D = 14 天（即 2 月 14 日累计确诊人数）和 D = 29 天（即 2 月 29 日累计确诊人数），且 D = 14 天的绝对误差始终小于 D = 29 天的绝对误差。因此，多元分位数回归模型对 14 天之后的累计确诊人数的预测最为准确。

## 4 总结与讨论

综上所述，得到以下结论：第一，气温因素对新冠肺炎确诊人数的影响存在地区差异性，根据气温预测新冠肺炎的传播需要进一步细分地区。第二，采用多元分位数回归区分模型高估和低估的损失值，反映新冠肺炎累计确诊人数的整体分布是可行的。并且，综合考虑气候、人口流动、经济因素能够提高模型预测的准确性。第三，多元分位数回归模型预测累计确诊人数的最佳预测时间窗是 14 天，这个数值与新冠肺炎平均最长潜伏期接近。

在未来的研究中，可以从以下两个方面展开。首先，本文只对某一时间间隔内累积的新冠肺炎确诊人数进行预测，不能表示出每天确诊人数的变化，因此，在未来的研究中可以进一步引入时间序列。其次，目前中国国内的新冠肺炎基本得到控制，未来需要将此方法应用于国外疫情爆发初期的城市或地区。

## 参考文献

- [1] 百度新型冠状病毒肺炎疫情实时大数据报告 [EB/OL]. [https://voice.baidu.com/act/newpneumonia/newpneumonia/?from=osari\\_aladin\\_banner#tab4](https://voice.baidu.com/act/newpneumonia/newpneumonia/?from=osari_aladin_banner#tab4).
- [2] 刘世锦, 韩阳, 王大伟. 基于投入产出架构的新冠肺炎疫情冲击路径分析与应对政策 [J]. 管理世界, 2020, 36(05): 1-12+51+263.
- [3] 夏杰长, 丰晓旭. 新冠肺炎疫情对旅游业的冲击与对策 [J]. 中国流通经济, 2020, 34(03): 3-10.
- [4] 沈国兵. “新冠肺炎”疫情对我国外贸和就业的冲击及纾困举措 [J]. 上海对外经贸大学学报, 2020, 27(02): 16-25.
- [5] 张艳霞, 李进. 基于 SIR 模型的新冠肺炎疫情传播预测分析 [J]. 安徽工业大学学报(自然科学版), 2020, 37(01): 94-101.
- [6] Alòs E, Mancino M E, Merino R. A fractional model for the COVID-19 pandemic: Application to Italian data [J].

2020.

- [7] Crokidakis N. COVID-19 spreading in Rio de Janeiro, Brazil: Do the policies of social isolation really work?[J]. *Chaos, Solitons & Fractals*, 2020, 136.
- [8] Wu J T, Kathy L, Leung G M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study[J]. *Lancet (London, England)*, 2020, 2020年 395 卷 10225 期:689-697 页.
- [9] 范如国, 王奕博, 罗明, 张应青, 朱超平. 基于 SEIR 的新冠肺炎传播模型及拐点预测分析[J]. *电子科技大学学报*, 2020, 49(03):369-374.
- [10] Zhu Y, Xie J. Association between ambient temperature and COVID-19 infection in 122 cities from China[J]. *Science of The Total Environment*, 2020, 724:138201.
- [11] Borah M J, Hazarika B, Panda S K. Examining the correlation between the weather conditions and COVID pandemic in India: A mathematical evidence[J]. *Results in Physics*, 2020, 19.
- [12] Jia J S, Lu X, Yuan Y, et al. Population flow drives spatio-temporal distribution of COVID-19 in China[J]. *Nature*, 2020, 582(7812):1-11.
- [13] Wu J T, Kathy L, Leung G M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study[J]. *Lancet (London, England)*, 2020, 2020年 395 卷 10225 期:689-697 页.
- [14] Chakraborti S, Maiti A, Pramanik S, et al. Evaluating the plausible application of advanced machine learnings in exploring determinant factors of present pandemic: A case for continent specific COVID 19 analysis[J]. *Science of The Total Environment*, 2020, 765(2).
- [15] Feroze N. Forecasting the patterns of COVID-19 and causal impacts of lockdown in top five affected countries using Bayesian Structural Time Series Models[J]. *Chaos, Solitons & Fractals*, 2020, 140.
- [16] Ball S. Data Analysis of Covid-19 Pandemic and Short-Term Cumulative Case Forecasting Using Machine Learning Time Series Models[J]. *Chaos Solitons & Fractals*, 2021, 142:110512.
- [17] Mojada R K, Yadav A, Prabhu A V, et al. Journal Pre-proofs Machine Learning Models for covid-19 future forecasting[J]. *Materials Today: Proceedings*, 2020.
- [18] Liu Y, Lin B, Xu B. Modeling the impact of energy abundance on economic growth and CO2 emissions by quantile regression: Evidence from China[J]. *Energy*, 2021(3):120416.
- [19] Koenker R, Bassett G W. Regression quantiles[J]. *Econometrica*, 1978, 46(1):211-244.
- [20] 张应应. 总体或样本的协方差(矩阵)和相关系数(矩阵)的系统定义[J]. *统计与决策*, 2016(08):20-24.
- [21] Laitila T. A pseudo- R<sup>2</sup> measure for limited and qualitative dependent variable models[J]. *Journal of Econometrics*, 1993, 56( 3):341-355.
- [22] Xie J, Zhu Y. Association between ambient temperature and COVID-19 infection in 122 cities from China. *Science of the Total Environment*, 724 (2020).
- [23] Borah M J, Hazarika B, Panda S K, Nieto J J. Examining the correlation between the weather conditions and COVID pandemic in India: A mathematical evidence. *Results in Physics*, 19 (2020).
- [24] 袁宏永, 陈涛, 黄丽达. 全国及湖北省新冠肺炎综合风险指数分析及比较[J]. *科技导报*, 2020, 38(4):21-28.

#### 作者简介:

李倩 (1997-), 女, 中南大学商学院, 研究生, 研究方向为传染病模型、风险评估。