DIDI三阶段模型处理结果

金鑫

2015年4月29日

# 1 前言

本文是didi数据的处理结果。

## 1.1 版本更新说明

1. 三阶段模型，大幅度提高订单预测精度
2. 去掉了部分数据的奇异值点，提高模型解释度

## 1.2 数据采集的一些注意事项：

1. 采集的数据在排名上尽量均衡，最好每500个排名至少保证有2个数据，均匀分布
2. 排名最高的2000名司机至少保证每1000名6个样本
3. 排名最高的100名司机至少保证4个样本
4. 排名最低的10000名司机每1000名保证4~6个样本
5. 时间上，保证同意一个城市的样本数据在同一天内采集
6. 数量上，中小城市至少100个样本，大城市至少200个样本
7. 避免排名和订单数都相同的重复选项
8. 按照模板保证每个城市数据在excel中sheet的固定顺序。

## 1.3 算法可能存在的问题：

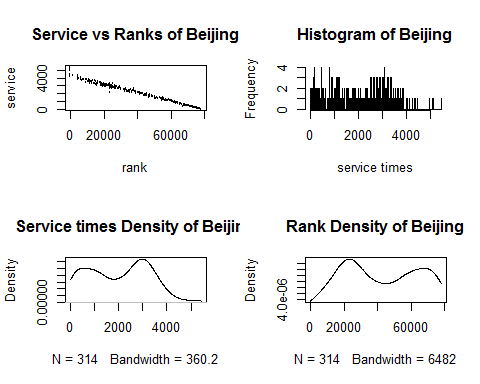
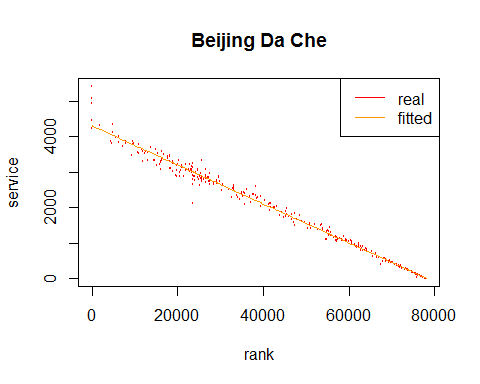
1. 南京的最高订单数可能不准，因为排名靠前的样本有点少。
2. 三阶段模型中的部分算法需要配合部分人工完成，比较费时，没有完全自动化
3. 对于司机数量，本算法只能覆盖到有订单司机，对于订单数=0的司机无法预测。
4. 建议结果要从其他角度交叉验证数据
5. 备注：本文稿处在自动处理试验阶段。

# 2 一阶段模型

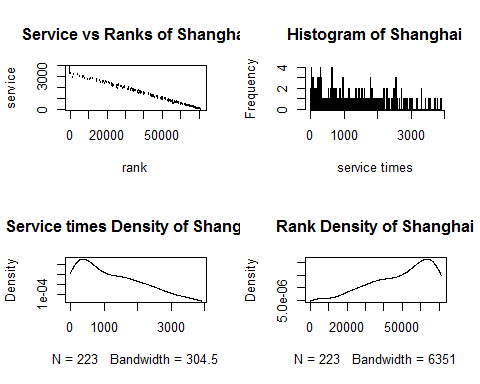
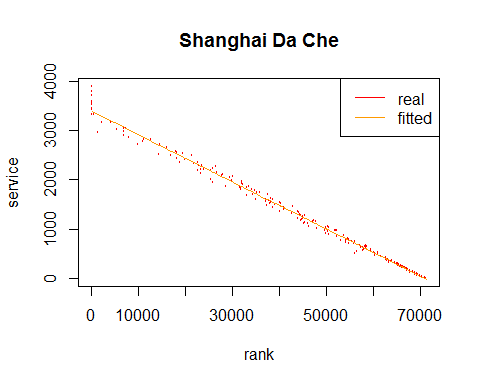
## 2.1 仿真结果图

## Warning: package 'iplots' was built under R version 3.1.3

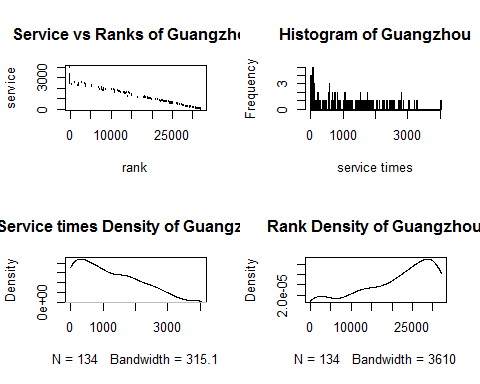
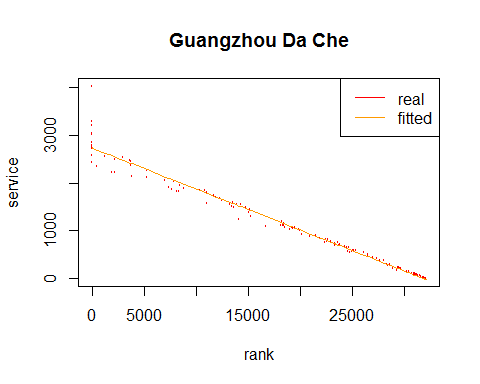
## Beijing

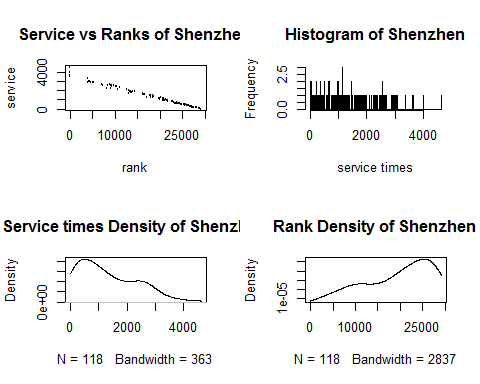
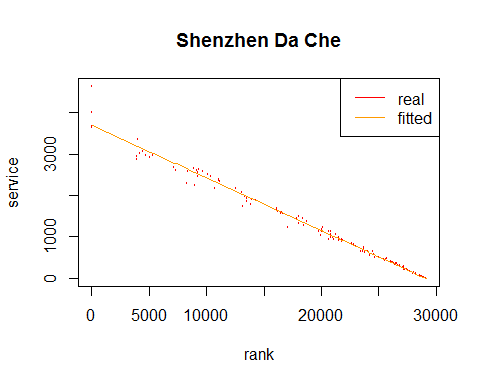
## Shanghai

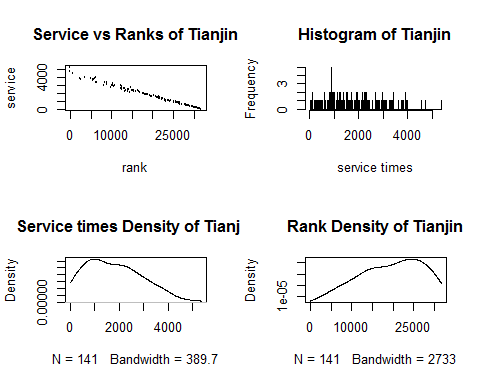
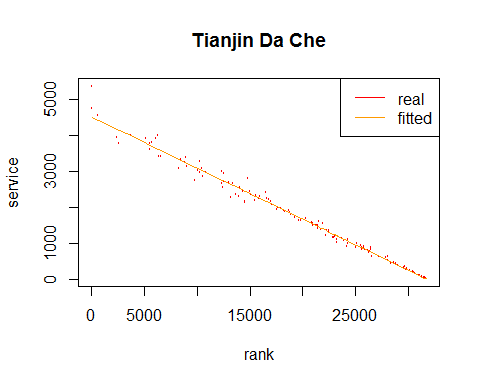
## Guangzhou

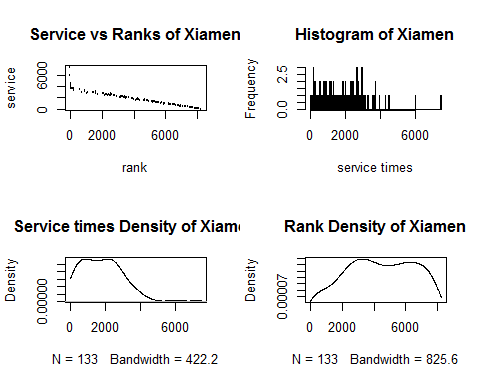
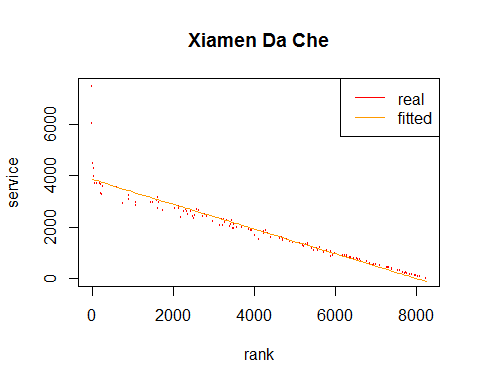
## Shenzhen

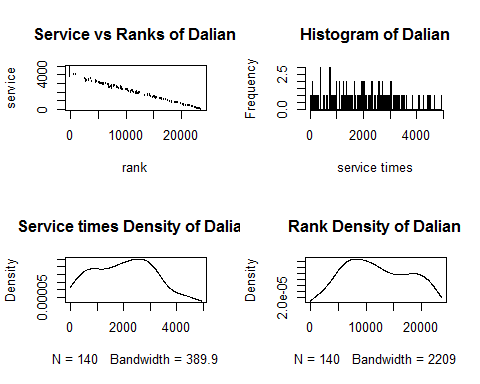
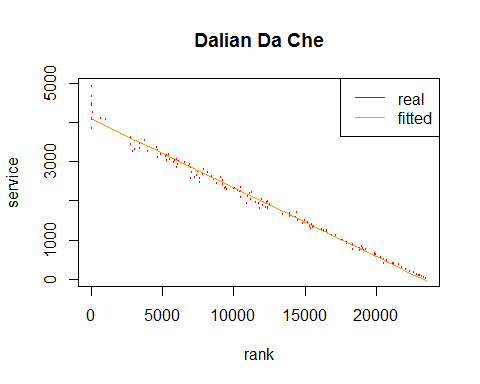
## Tianjin

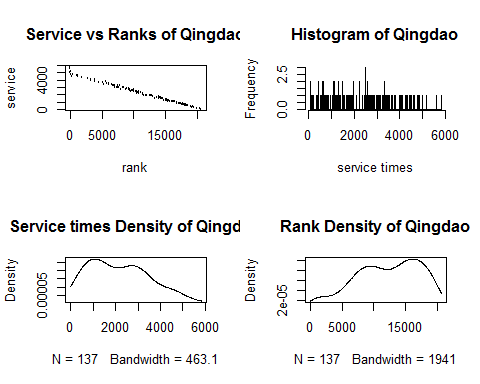
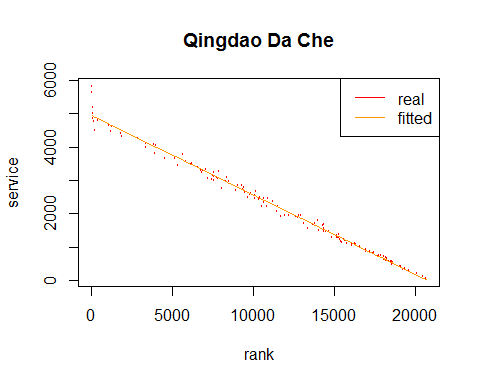
## Xiamen

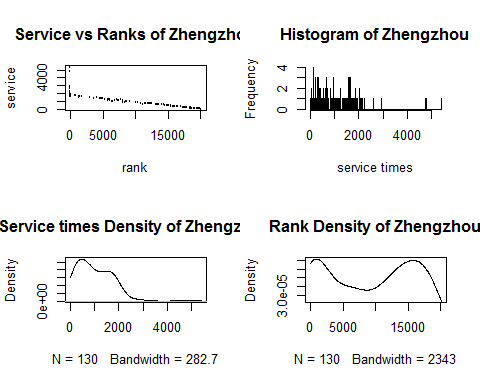
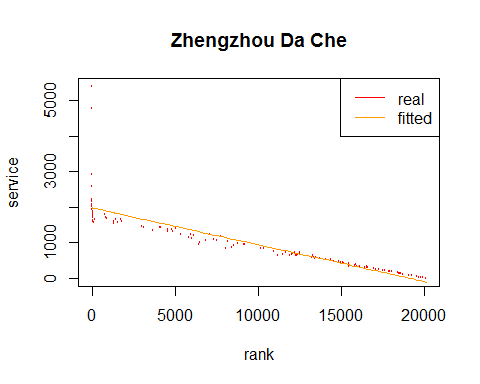
## Dalian

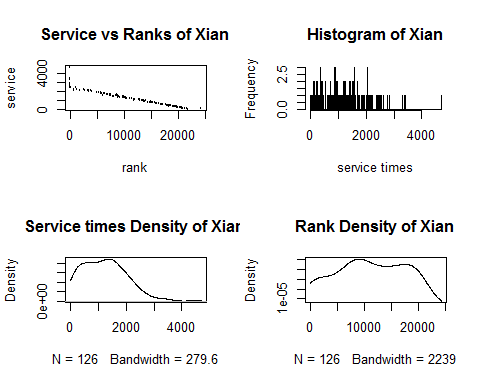
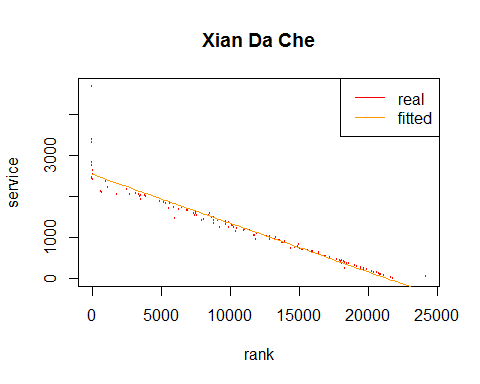
## Qingdao

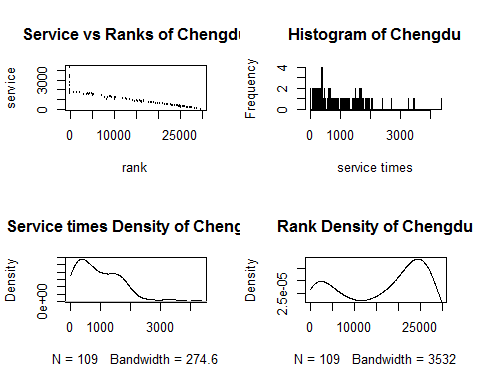
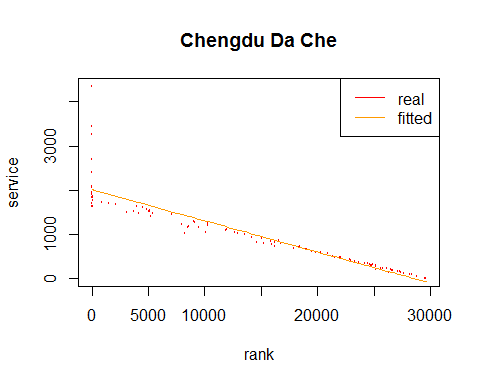
## Zhengzhou

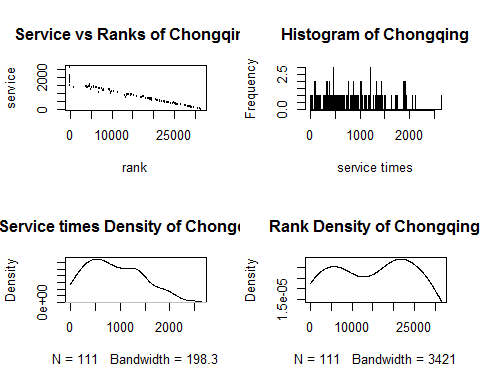
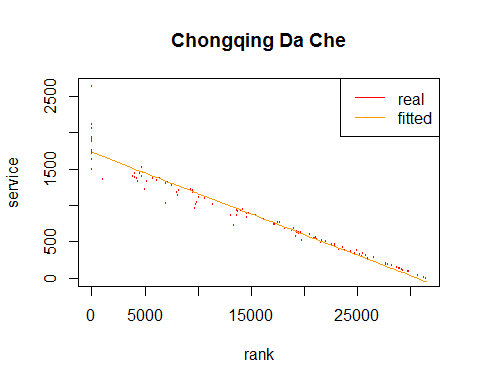
## Xian

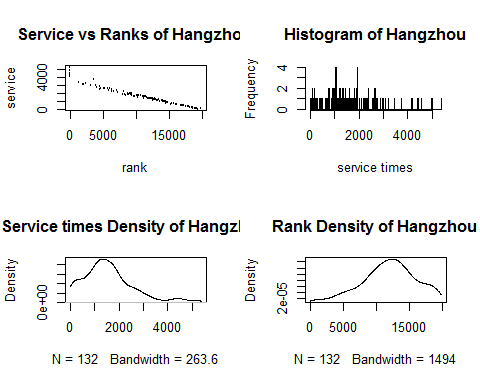
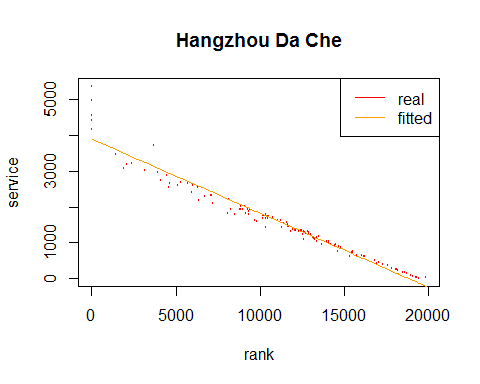
## Chengdu

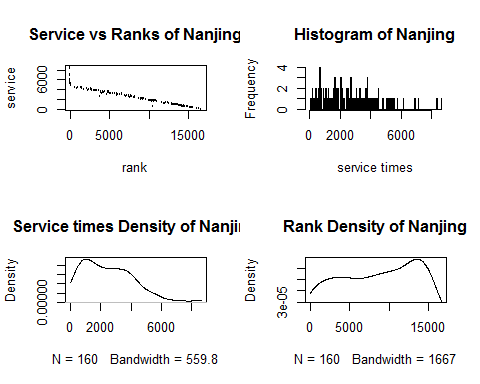
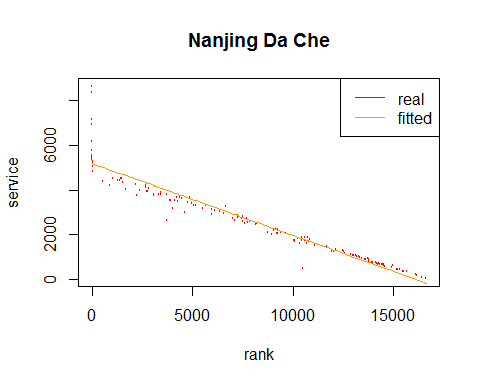
## Chongqing

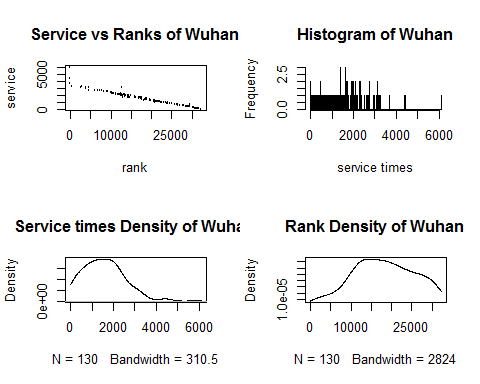
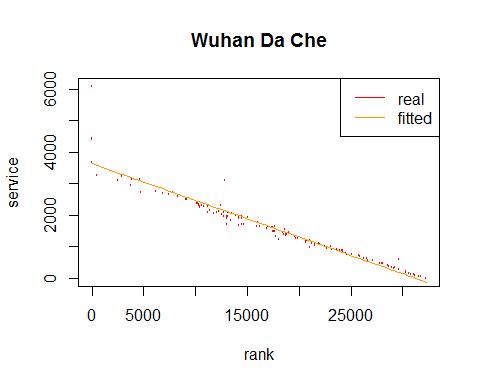
## Hangzhou

## Nanjing

## Wuhan

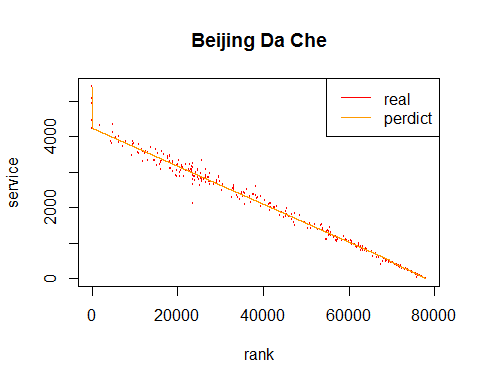
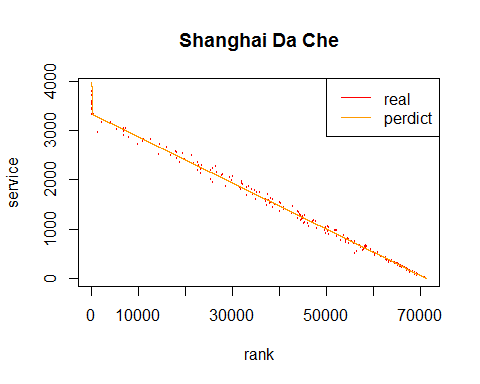
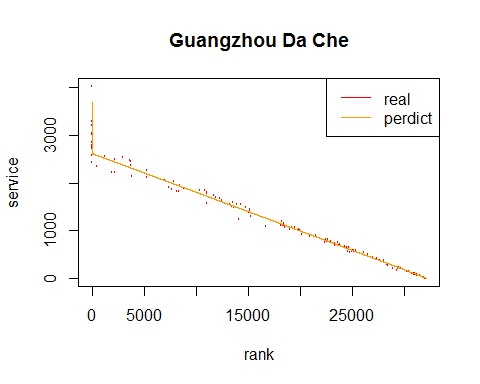
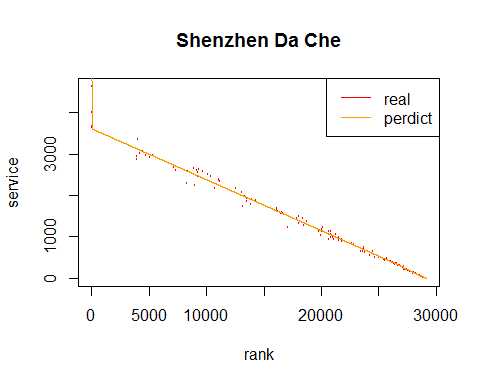
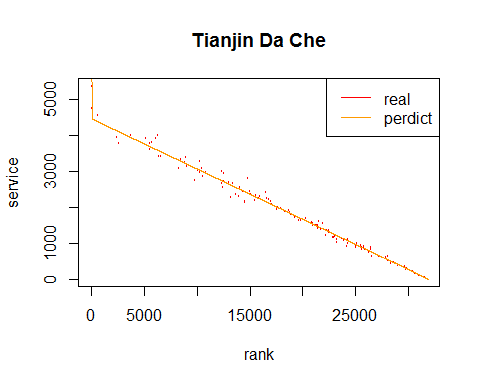
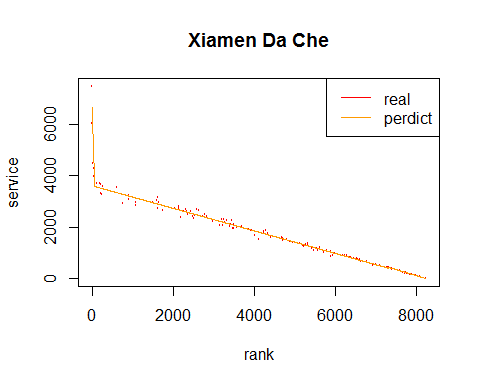
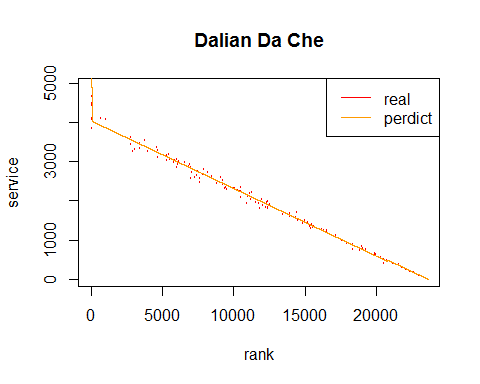
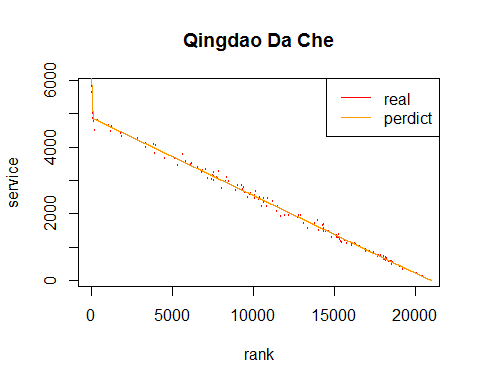
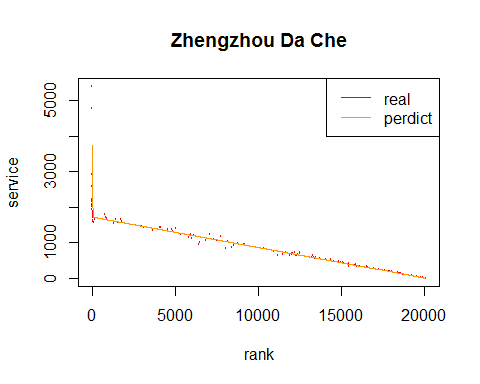
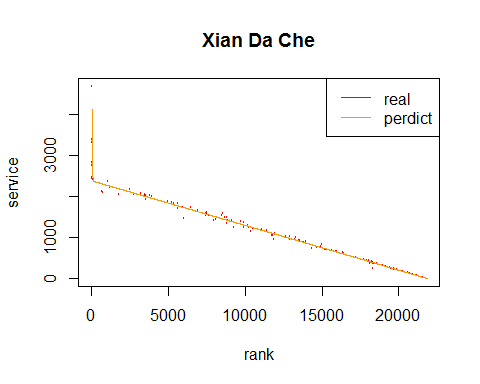
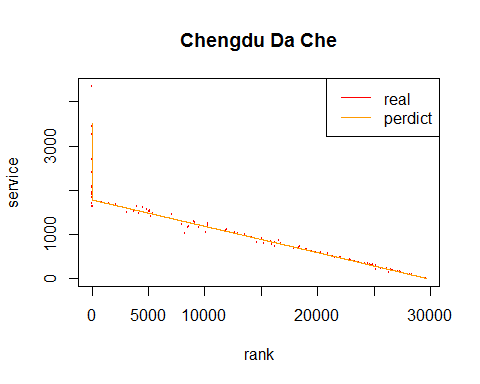
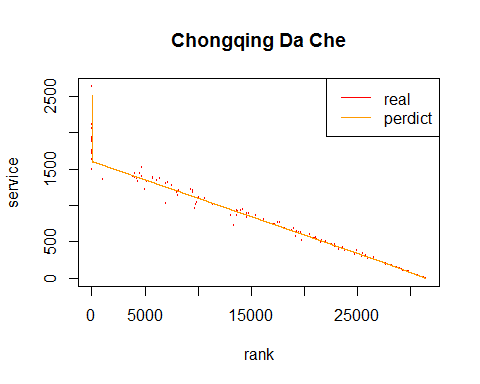
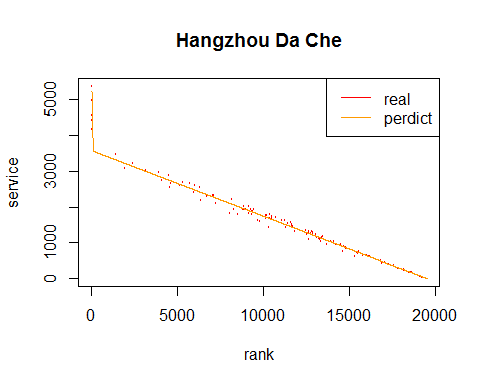
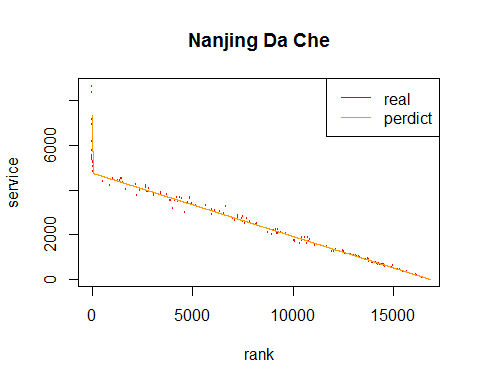
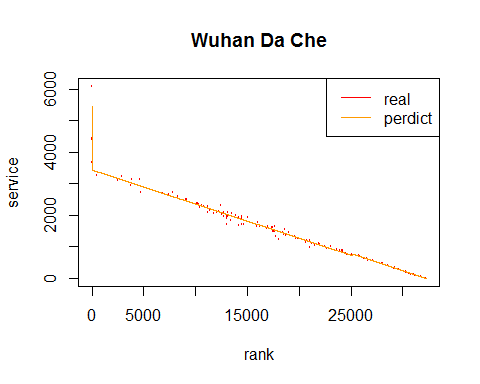
 

## 2.2 结果统计表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| city | ndriver | maxorder | meanorder | totorder |
| Beijing | 78134 | 5425 | 2154.1363 | 168311287 |
| Shanghai | 70961 | 3893 | 1695.2174 | 120294324 |
| Guangzhou | 31838 | 4029 | 1369.4865 | 43601710 |
| Shenzhen | 29082 | 4631 | 1846.5742 | 53702071 |
| Tianjin | 31833 | 5354 | 2253.4482 | 71734015 |
| Xiamen | 7990 | 7494 | 1932.9031 | 15443896 |
| Dalian | 23307 | 4919 | 2043.7654 | 47634040 |
| Qingdao | 20805 | 5816 | 2470.1799 | 51392092 |
| Zhengzhou | 19122 | 5405 | 991.3378 | 18956362 |
| Xian | 21329 | 4693 | 1272.5661 | 27142562 |
| Chengdu | 28474 | 4352 | 1005.1826 | 28621568 |
| Chongqing | 30683 | 2634 | 864.6275 | 26529365 |
| Hangzhou | 18841 | 5378 | 1951.5021 | 36768252 |
| Nanjing | 16152 | 8623 | 2590.4091 | 41840287 |
| Wuhan | 31288 | 6089 | 1820.4756 | 56959041 |

# 3 三阶段模型

## 3.1 仿真结果图

## 3.2 结果统计表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| city | ndriver | maxorder | meanorder | totorder |
| Beijing | 77890 | 5425 | 2147.6459 | 167280143 |
| Shanghai | 71283 | 3962 | 1667.6045 | 118871852 |
| Guangzhou | 32147 | 4029 | 1317.7198 | 42360737 |
| Shenzhen | 29136 | 5000 | 1817.5501 | 52956140 |
| Tianjin | 31909 | 5558 | 2233.4041 | 71265690 |
| Xiamen | 8226 | 7494 | 1821.9151 | 14987074 |
| Dalian | 23644 | 5182 | 2004.8731 | 47403220 |
| Qingdao | 20994 | 6072 | 2432.6333 | 51070703 |
| Zhengzhou | 20092 | 5405 | 868.2775 | 17445431 |
| Xian | 21853 | 4693 | 1198.0386 | 26180738 |
| Chengdu | 29615 | 4352 | 893.9261 | 26473621 |
| Chongqing | 31465 | 2634 | 810.5958 | 25505398 |
| Hangzhou | 19529 | 5378 | 1789.7422 | 34951876 |
| Nanjing | 16856 | 8623 | 2381.2531 | 40138402 |
| Wuhan | 32251 | 6089 | 1707.5801 | 55071165 |

# 4 算法对比

三阶段结果减去-一阶段结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| city | ndriver | maxorder | meanorder | totorder |
| Beijing | -244 | 0 | -6.490368 | -1031144 |
| Shanghai | 322 | 69 | -27.612923 | -1422472 |
| Guangzhou | 309 | 0 | -51.766707 | -1240973 |
| Shenzhen | 54 | 369 | -29.024094 | -745931 |
| Tianjin | 76 | 204 | -20.044096 | -468325 |
| Xiamen | 236 | 0 | -110.987982 | -456822 |
| Dalian | 337 | 263 | -38.892274 | -230820 |
| Qingdao | 189 | 256 | -37.546584 | -321389 |
| Zhengzhou | 970 | 0 | -123.060357 | -1510931 |
| Xian | 524 | 0 | -74.527462 | -961824 |
| Chengdu | 1141 | 0 | -111.256468 | -2147947 |
| Chongqing | 782 | 0 | -54.031644 | -1023967 |
| Hangzhou | 688 | 0 | -161.759920 | -1816376 |
| Nanjing | 704 | 0 | -209.155967 | -1701885 |
| Wuhan | 963 | 0 | -112.895539 | -1887876 |