**说明文档**

1. **数据来源**

**（1）出租车数据集**

本文出租车数据来自上海市出租车市场，由两期组成：2015年-2017年数据集，和2019年数据集1和数据集2。每笔行程数据为驾驶员的一次载客记录，信息见附录表A1，包含驾驶员编号、车牌编号、上车时间、下车时间、行驶里程、营运收入等信息。

2015年-2017年数据集：data2015、data2016、data2017。为上海某大型巡游车公司的随机抽样，包含3,285名出租车驾驶员三年内的完整工作记录，共计34,900,942笔行程。

2019年数据集1：网约\_trip。为全市网约车驾驶员的随机抽样，包含4,994名网约车驾驶员一周内的完整工作记录，共计390,551笔行程

2019年数据集2：巡游\_trip。为全市巡游车驾驶员的随机抽样，以及4,230名驾巡游车驶员在同一周内的完整工作记录，共计356,553笔行程。

**（2）天气数据集**

本文的天气数据集weather来自上海市气象局，为以小时为单位的上海平均温度和降雨量。

1. **数据与代码之间的关系**

**（1）数据结构化处理程序**

以下程序在实证研究中被调用：

data\_process\_2015\_2017.do：将2015年-2017年数据集中的原始行程数据转换为小时数据所需程序。

data\_process\_2019.do：将2019年数据集中的原始行程数据转换为小时数据所需程序。

data\_process\_shift.do：将2019年数据集中的原始行程数据转换为班次数据所需程序。

**（2）正文中所使用的实证分析程序**

empirical\_analysis\_2015\_2017.do：2015年-2017年小时数据进行实证分析的程序。生成表2 PanelA、Panel C，表3。

empirical\_analysis\_2019.do：2019年小时数据进行模型一实证分析的程序。生成表2 PanelA，表4 Panel A，表4 Panel B。

empirical\_shift\_2019.do：2019年班次数据进行模型二实证分析的程序。生成表2 Panel B，表5。

hinton\_plot.py：画系数图的程序，生成图1。

empirical\_analysis\_2019\_robust.do：2019年小时数据进行稳健性分析的程序。生成表6 Panel A、Panel B。

1. **附录中所使用的实证分析程序**

empirical\_shift\_2019.do：2019年班次数据进行模型二实证分析的程序。生成图III1。

empirical\_analysis\_2019\_robust.do：2019年小时数据进行稳健性分析的程序。生成表V1、VI1，图V1。

data\_process\_shift\_VII.do：2019年班次数据进行驾驶员临时休息决策分析的程序。生成表VII1。