**《得失之祸：股市下跌对交通事故的溢出影响》代码说明**

贺佳　罗淑婧　李冰清　任远

**一、本文使用的数据信息**

1. 机动车保险出险理赔数据

机动车保险出险理赔数据来自某总部位于上海的财产保险公司2012年-2016年上海市机动车保险出险理赔数据，其中详细记载了出险时间、出险地址、被保险人性别、年龄、驾驶车辆车型、保险保费和理赔金额。出险数据中并未记载车辆价格信息，本文从汽车之家网站（https://www.autohome.com.cn/）爬取了车辆价格信息，并通过车型与出险理赔数据匹配。

2.上海市每日交通事故数据

上海市每日交通事故数据为机动车保险出险理赔数据的衍生数据。本文根据出险理赔数据，以当日出险数量衡量上海市日度层面交通事故数量，包括全天和不同时段交通事故数（开盘前、开盘后、交易时间、收盘后、开盘后各细分时段）、金融区和非金融区交通事故、高股市关联和低股市关联车辆交通事故、分赔率交通事故、分性别人群交通事故、分年龄交通事故、分车价交通事故。其中，金融区和非金融区交通事故根据出险理赔数据中关于出险地址的信息进行划分，高股市关联车辆和低股市关联车辆以2017年西南财经大学中国家庭金融调查（CHFS2017）关于被调查家庭拥有的汽车品牌及其是否拥有股票账户股票的回答为依据进行划分。

3.股市收益率数据

本文所使用的股市收益率数据来源于CSMAR数据库，包括上证指数日度收益率、沪深300指数日度收益率、经过去一年标准差调整的上证指数日度收益率。回归中使用上证指数日度收益率作为核心解释变量，其余收益率指标用于稳健性检验（详见附录Ⅱ表A3）。此外，我们还从CSMAR数据库获取了融资融券标的证券清单和上市公司日度收益率，分别计算了融券标的证券收益率与非融券标的证券收益率，用于检验做空机制的影响（详见附录Ⅲ）。

4.气象数据

本文所使用的气象数据主要来自两个数据库。其中，气温、露点温度、风速和降雨量数据来自美国国家海洋和大气管理局（NOAA）的上海市主要气象站每三小时报告，空气质量数据来自美国国务院中国空气质量监测项目的上海市逐时PM2.5浓度（ug/m3）。上述气象原始数据经城市-日度层面平均处理后得到上海市当日平均气象数据作为控制变量。

5. 彩票奖池数据

彩票奖池数据包括超级大乐透和双色球历史奖池，用于检验注意力机制。超级大乐透历史奖池金额爬取自中国体彩网历史开奖公告（https://www.lottery.gov.cn/kj/kjlb.html?dlt），双色球历史奖池数据爬取自中国福彩网历史开奖公告（https://www.cwl.gov.cn/ygkj/wqkjgg/）。

6．股吧数据

股吧数据来自于CNRDS数据库股吧评论数据库，该数据库是基于我国网络股吧论坛中不同类型的帖子评论进行统计研究而建立的专业性财经文本数据库。本文使用的相关数据包括：上市公司股吧帖子总量、负面帖子量和阅读数，经日度层面汇总计算得到股吧负面帖子占比和阅读量，分别用于衡量投资者负面情绪（详见附录Ⅱ表A2）和股市注意力（详见正文表4）。

7. 出行数据

出行数据来自于2016年8月上海市滴滴出行数据集，其中记录了每笔出行订单起终点时间、起终点经纬度的行程记录。我们结合高德地图兴趣点数据（Point of interest, POI）识别金融出行订单和非金融出行订单，用于检验出行渠道的影响（详见附录Ⅳ）。

**二、代码介绍及相关数据集**

“2024-00143\_程序代码.do”文件为本文用于生成文章图表的程序代码。“2024-00143\_日志文件”为运行该程序的日志文件。其中涉及的数据集如下：

表 1 数据文件介绍

|  |  |  |  |
| --- | --- | --- | --- |
| **数据文件** | **介绍** | **变量内容** | **相关表格** |
| claims.dta | 机动车保险出险理赔数据 | 出险人性别*male*、年龄*age*、车辆价格*price* | 表1 |
| reg.dta | 回归分析主要数据集 | 上海市交通事故数据（*acc\_\**）  收益率变量(*ret\**)  气象控制变量 | 文章主要表格 |
| 彩票奖池金额*jackpost* | 表4 |
| 股吧阅读量*readnum*  股吧负面帖子占比*negpost* | 表4  附录ⅡA2 |
| reg\_fin.dta | 金融区与非金融区因果推断检验数据集 | 上海市开盘后交通事故（分金融/非金融区）*acc\_afteropen\_fin*  金融区虚拟变量*FIN*  上证指数收益率*retindex1*  气象控制变量 | 表3 |
| reg\_stockcar.dta | 车辆股市参与倾向因果推断 | 上海市开盘后交通事故（分高/低车辆股市参与倾向）*acc\_afteropen\_stockcar*  高股市参与倾向虚拟变量*stock\_car*  上证指数收益率*retindex1*  气象控制变量 | 附录Ⅴ |
| reg\_taxi.dta | 出行倾向检验 | 上海市金融出行与非金融出行量*order*  金融出行虚拟变量*fin\_taxi*  上证指数收益率*retindex1*  气象控制变量 | 附录Ⅳ |