数据说明

1. 论文的原始数据主要为以下几个方面：包括首台套保险数据集、上市公司数据集、上市公司子公司数据集、中国铁路货运量数据、地区宏观经济数据集以及中国工商注册信息大数据，样本区间为2010-2020年。

1.微观数据集

（1）首台套保险的数据来源于中华人民共和国工业和信息化部官方网站，课题组连续多年对该数据进行跟踪，并通过手工整理构建了本文的数据基础。

（2）子公司存量信息以及撤销信息数据分别来源于CSMAR数据库的“上市公司子公司联营合营情况表”和“子公司退出明细表”。但由于数据库中子公司地区数据存在大量缺失，因此本文在CSMAR数据的基础上，通过企查查、Python程序、百度地图以及搜索引擎等方式，对缺失的子公司信息进行了补全，整理获得了详细的子公司地区分布数据，并将其与母公司地区数据匹配，从而构建了母公司-子公司-年度数据集。

（3）其他财务数据以及公司治理数据来自于CSMAR数据库。

2.宏观数据集

（1）中国铁路货运量数据包括省际间的铁路货运量数据和各地区的铁路货运量。其中，省际间的铁路货运量数据来源于《中国交通年鉴》。各地区的铁路货运量数据来源于《中国统计年鉴》。

（2）地区宏观经济数据集来源于《中国统计年鉴》。

（3）中国工商注册信息大数据来源于锐思数据库（RESSET）“中国企业大数据平台”。

二、论文代码中（正文图表代码和附录图表代码）使用的衍生数据主要为以下四个方面：包括“ALL\_Data.dta”、“PSM.dta”、“PSM\_DID.dta”以及“Area\_Data.dta”。

1.“ALL\_Data.dta”为经过原始数据合并以及初始筛选后得到的2010-2020年沪深A股上市公司的面板数据集。

2.“PSM.dta”为通过选取企业参保前一年一系列特征变量作为协变量，并采用一比一、无放回倾向得分匹配法筛选得到的处理组和对照组数据集。

3.“PSM\_DID.dta”为处理组和对照组数据集与原始面板数据匹配合并后本文回归所使用的面板数据集。

4、“Area\_Data.dta”为所有宏观数据匹配合并后得到的地区宏观面板数据集。