

说明文档

（一）文件说明

本论文使用的数据代码压缩包命名为“2024-01185+数据”，解压之后里面有“2024-01185+程序代码”、“2024-01185+日志文件”和“2024-01185+说明文档”3个部分。

（1）“2024-01185+程序代码”里面包括26个文件，分别是“Master.do”“ChineseCustomData.do”“ForeignDemandShocks-HS96.do”“Data_CO2.do”“data.do”“Data_for_Table4.do”“Data_for_TableII4.do”“Figure1.do”“Figure2.do”“FigureII1.do”“FigureII2.do”“Table1.do”“Table2.do”“Table3.do”“Table4.do”“Table5.do”“Table6.do”“TableII1.do”“TableII2.do”“TableII3.do”“TableII4.do”“TableII5.do”“TableII6.do”“TableII7.do”“TableII8.do”“TableII9.do”文件：

“Master.do”是主程序运行文件，依次调用其他运行文件，完成数据处理和图表生成工作。“ChineseCustomData.do”“ForeignDemandShocks-HS96.do”“Data_CO2.do”“data.do”“Data_for_Table4.do”“Data_for_TableII4.do”是数据处理运行文件，依次用来处理中国企业层面贸易数据、计算企业层面外需增速、计算企业层面二氧化碳排放量和污染能源使用占比、生成实证部分使用的主要分析样本、生成正文表4机制分析使用的回归样本、生成在线附录稳健性检验部分表II4的回归样本。

“Table1.do”“Table2.do”“Table3.do”“Table4.do”“Table5.do”“Table6.do”“TableII1.do”“TableII2.do”“TableII3.do”“TableII4.do”“TableII5.do”“TableII6.do”“TableII7.do”“TableII8.do”“TableII9.do”是表格运行文件，主要生成表1关键变量的描述性统计、表2—表6以及在线附录II表II1—表II9的运行结果。

“Figure1.do”“Figure2.do”“FigureII1.do”“FigureII2.do”是图运行文件，主要生成图1、图2和在线附录II中的图II1和图II2证伪检验。

（2）“2024-01185+日志文件”中的“regression.log”是所有stata程序的运行日志文件，其他为图表运行结果文件。其中，“Table1.html”是表1关键变量的描述性统计；“Table3.csv”是表3总体二氧化碳排放增长分解；“Table2.html”、“Table4.html”——“Table6.html”分别是表2、表4—表6的运行结果；“Figure1.emf”和“Figure2.emf”是正文中图1和图2的运行结果；“TableII1.html”——“TableII9.html”是在线附录II中表II1—表II9的运行结果；“FigureII1.emf”是在线附录II中图II1的运行结果，“FigureII2_b_co2.emf”、“FigureII2_t_co2.emf”、“FigureII2_b_co2intensity.emf”、“FigureII2_t_co2intensity.emf”是在线附录II中图II2随机抽样运行200次证伪检验的运行结果。

需要说明的是，与正文中结果相比，表1—表6、表II1—表II9在文字排版、英文变量转变为对应的中文名称等方面做了细节调整。图1—图2及图II1—图II2在颜色和实虚线等方面做了细节调整。

（3）“2024-01185+说明文档”即是本说明文档。

（二）数据来源

本文使用的数据主要包括四部分：

（1）中国生态环境部提供的企业层面污染数据库。该数据被认为是目前国内最全面可靠的企业层面环境数据，为《中国环境统计年鉴》的数据来源。这些数据由占重点污染物排放总量85%以上的企业自行上报，并由地方环保部门进行抽查，确保数据质量。也有一些文

章验证了这套数据中企业能源投入和污染排放指标的可靠性(吕越和陈泳昌,2022;陈登科,2020)。该数据提供了2000—2014年442,546家企业的非平衡面板数据,包括能源使用、工业废气和废水排放、废气废水治理设施等企业环境行为指标和工业总产值、从业年限、行政区划等少量企业层面特征指标。其中,企业的能源使用又细分为原料煤、燃料煤、燃料油、柴油以及清洁燃气等。本文基于这些细分的能源使用情况,以及中国国家发展和改革委员会(NDRC)、政府间气候变化专门委员会(IPCC)提供的温室气体排放核算方法,测度了企业各年的二氧化碳排放量和二氧化碳排放强度(即单位产值的二氧化碳排放量)。值得说明的是,中国生态环境部的企业层面污染数据库的原始年份为2000—2014年,但其中燃料煤、燃料油以及洁净燃气等能源使用数据仅有2000—2010年数据,这导致基于能源投入计算的二氧化碳排放数据也仅到2000—2010年。

(2) 企业层面和国家—产品层面贸易数据。本文利用中国海关总署提供的2000—2014年中国企业—贸易伙伴国—HS8位产品层面的贸易数据和CEPII BACI数据库提供的1996-2014年出口国—进口国—HS6位产品层面的贸易数据,构建了企业层面按照初年出口份额加权的外需增速指标。对于HS编码跨年匹配的问题,本文采用1996版HS产品分类,并基于联合国统计司提供的对应码将不同年份的HS产品分类统一至HS1996层面。值得说明的是,上述企业层面和国家—产品层面贸易数据可更新至2014年之后的年份,但受制于企业层面碳排放数据的可得性,文章实证分析部分基准回归及相关稳健性检验主要使用的数据截至2010年、部分排放强度下降的机制检验分析使用的数据截至2014年。

此外,在构建实证分析样本时,还需将企业层面污染数据库与海关企业交易层面贸易数据匹配。考虑到两套数据库的企业识别码并不一致,因此本文使用了去除干扰词^①后的企业名称和年份进行精确匹配。在将贸易数据与污染数据匹配时,本文还将污染数据库中变更企业名称或组织机构代码的情况进行了调整。在匹配数据的基础上,本文进一步构建了企业的出口增速、总产值增速以及出口强度等指标。其中,出口强度等于以人民币计出口金额与工业总产值现价的比值,并利用美元兑人民币平均汇率将海关贸易数据中以美元计出口金额转化为以人民币计出口金额。此外,本文还利用中国企业交易层面贸易数据中的所有制类型变量,构建了不同所有制类型企业的虚拟变量^②。

(3) 规模以上工业企业数据。中国国家统计局提供的规模以上工业企业数据提供了企业基本信息数据以及资产负债表、现金流量表、利润表中的多个财务变量。本文使用注册类型、研发投入、开业年份以及固定资本合计四个指标。其中,注册类型是企业层面所有制变量的补充信息来源,开业年份和固定资产合计则用来构建企业资本设备年限指标。本文首先使用组织机构代码和年份信息,进一步匹配规模以上工业企业数据库和基准回归样本(即工业企业污染数据与海关贸易数据的匹配样本)。然后,使用企业名称和年份信息来匹配作为

^① 干扰词包括名称中的标点符号、特殊字母、英文括号等。在去除干扰词后,将两套数据库中的企业名称进一步统一(例如将有限公司相关的表达统一为有限公司),而后进行精确匹配。

^② 私营企业包括个体工商户、私营企业、股份有限公司、股份合作企业、私营股份有限公司、私营独资企业、私营有限责任公司、私营合伙企业。集体企业和国有企业包括国有企业、集体企业、集体联营企业、国有联营企业、国有独资公司、国有与集体联营企业、其他联营企业、其他有限责任公司、其他企业。外资企业包括登记注册类型为中外合作企业、中外合资企业、外商独资企业、港澳台独资企业、港澳台商投资股份有限公司、外资(独资)企业、外商独资企业、外商投资股份有限公司、合资经营企业(港或澳台资)、合作经营企业(港或澳台资)的企业。

补充。

(4) 专利数据。在机制检验部分, 本文还使用了中国国家知识产权局提供的 1998-2014 年申请并公开的专利申请数据, 以识别并测度企业申请的绿色专利情况。具体而言, 在进行数据清洗后, 使用去除干扰词后的企业名称, 将基准回归样本与企业专利申请数据进行匹配, 并通过分类号和主分类号筛选绿色专利。

(三) Dofile 里面使用的指标说明

co2_ndrc_win: 按照中国国家发展和改革委员会 (NDRC) 提供的温室气体排放核算方法, 测度的企业各年的二氧化碳排放量, 并剔除左右两端 1% 分位数异常值。

co2_IPCC_win: 按照政府间气候变化专门委员会 (IPCC) 提供的温室气体排放核算方法, 测度的企业各年的二氧化碳排放量, 并剔除左右两端 1% 分位数异常值。

排放强度_ndrc_win: 按照 NDRC 提供的温室气体排放核算方法, 测度的企业各年的二氧化碳排放强度 (即单位产值的二氧化碳排放量), 并剔除左右两端 1% 分位数异常值。

排放强度_IPCC_win: 按照 IPCC 提供的温室气体排放核算方法, 测度的企业各年的二氧化碳排放强度 (即单位产值的二氧化碳排放量), 并剔除左右两端 1% 分位数异常值。

工业总产值现价万元_win: 以现价计算的企业各年的工业总产值, 剔除左右两端 1% 分位数异常值, 单位为万元。

出口_win: 以人民币计企业出口金额, 利用美元兑人民币平均汇率将海关贸易数据中以美元计出口金额 (记为 **export** 变量) 转化为以人民币计出口金额, 剔除左右两端 1% 分位数异常值, 单位为万元。

exportintensity_win: 企业—年份层面的出口强度, 等于以人民币计出口金额与工业总产值现价的比值, 并剔除左右两端 1% 分位数异常值。

delta_co2_ndrc: 按照 NDRC 方法计算的二氧化碳排放增速, 其中增速采用 Davis-Haltiwanger 增长率刻画。假定变量 x 在 t 期和 $t-1$ 期取值分别为 x_t 和 x_{t-1} , 则其 Davis-Haltiwanger 增长率为 $(x_t - x_{t-1}) / [0.5 \times ((x_t + x_{t-1}))]$ 。

delta_co2_IPCC: 按照 IPCC 方法计算的二氧化碳排放增速, 其中增速采用 Davis-Haltiwanger 增长率刻画。

delta_排放强度_ndrc: 按照 NDRC 方法计算的二氧化碳排放强度增速, 其中增速采用 Davis-Haltiwanger 增长率刻画。

delta_排放强度_IPCC: 按照 IPCC 方法计算的二氧化碳排放强度增速, 其中增速采用 Davis-Haltiwanger 增长率刻画。

delta_output: 以现价计算的企业各年的工业总产值增速, 其中增速采用 Davis-Haltiwanger 增长率刻画。

delta_export: 以美元计企业出口额增速, 增速采用 Davis-Haltiwanger 增长率刻画。

delta_ZFD: 参考 Barrows and Ollivier (2021), 在企业层面按照初年贸易份额加权、构建的外部需求增速指标, 其中增速采用 Davis-Haltiwanger 增长率刻画。

delta_ZFD_ein_win_L0L1: **delta_ZFD** 变量与 **exportintensity_win** 变量 1 期滞后项的交乘项。

delta_ZFD_ein_win_L1L1: **delta_ZFD** 变量 1 期滞后项与 **exportintensity_win** 变量 1 期滞

后项的交乘项。

delta_ZFD_ein_win_L2L1: delta_ZFD 变量 2 期滞后项与 exportintensity_win 变量 1 期滞后项的交乘项。

delta_ZFD_ein_win_L3L1: delta_ZFD 变量 3 期滞后项与 exportintensity_win 变量 1 期滞后项的交乘项。

delta_ZFD_ein_win_F1L1: delta_ZFD 变量 1 期提前项与 exportintensity_win 变量 1 期滞后项的交乘项。

delta_ZFD_ein_win_F2L1: delta_ZFD 变量 2 期提前项与 exportintensity_win 变量 1 期滞后项的交乘项。

exportintensity_t0_win: 企业在样本初年的出口强度, 并剔除左右两端 1%分位数异常值。

lnco2_ndrc_t0_win: 企业在样本初年的按照 NDRC 方法计算的二氧化碳排放量对数, 并剔除左右两端 1%分位数异常值。

lnco2_IPCC_t0_win: 企业在样本初年的按照 IPCC 方法计算的二氧化碳排放量对数, 并剔除左右两端 1%分位数异常值。

ln 工业总产值现价万元_t0_win: 企业在样本初年的以现价计算的工业总产值对数, 并剔除左右两端 1%分位数异常值。

ln 排放强度_ndrc_t0_win: 企业在样本初年的按照 NDRC 方法计算的二氧化碳排放强度对数, 并剔除左右两端 1%分位数异常值。

ln 排放强度_IPCC_t0_win: 企业在样本初年的按照 IPCC 方法计算的二氧化碳排放强度对数, 并剔除左右两端 1%分位数异常值。

g_E_ndrc: 按照 NDRC 方法计算的所有样本企业加总的二氧化碳排放量增速, 其中增速基于以下公式计算, $g_x = \frac{x_{t_2} - x_{t_1}}{x_{t_1}}$, 表示变量 x 在 t_1 到 t_2 期的增速。

g_V: 所有样本企业加总的总产出增速, 其中增速基于以下公式计算, $g_x = \frac{x_{t_2} - x_{t_1}}{x_{t_1}}$, 表示变量 x 在 t_1 到 t_2 期的增速。

g_ita: 按照产出份额加权平均的二氧化碳排放强度增速, 其中增速基于以下公式计算, $g_x = \frac{x_{t_2} - x_{t_1}}{x_{t_1}}$, 表示变量 x 在 t_1 到 t_2 期的增速。

g_V_g_ita: g_V 变量与 g_ita 变量的交乘项。

within: 参考 Brandt et al. (2012), 将加权平均的二氧化碳排放强度增速分解得到的企业内效应 (within firm effect)。

across: 参考 Brandt et al. (2012), 将加权平均的二氧化碳排放强度增速分解得到的构成效应。

entryeffect: 参考 Brandt et al. (2012), 将加权平均的二氧化碳排放强度增速分解得到的进入效应 (entrants effect)。

exiteffect: 参考 Brandt et al. (2012), 将加权平均的二氧化碳排放强度增速分解得到的退出效应 (exiting effect)。

delta_NRG_win: 企业层面污染能源使用占比 NRG_{it} 增速, 其中增速采用 Davis-Haltiwanger

增长率刻画，并剔除左右两端 1%分位数异常值。

delta_废水治理设施数套_win: 企业层面废水治理设施数（套）增速，其中增速采用 Davis-Haltiwanger 增长率刻画，并剔除左右两端 1%分位数异常值。

delta_废气治理设施数套_win: 企业层面废气治理设施数（套）增速，其中增速采用 Davis-Haltiwanger 增长率刻画，并剔除左右两端 1%分位数异常值。

delta_VIN_win: 企业层面资本设备年限指标增速，其中增速采用 Davis-Haltiwanger 增长率刻画，并剔除左右两端 1%分位数异常值。

delta_Mix_ndrc_win: 按照 NDRC 方法计算的企业定基二氧化碳排放强度增速，其中增速采用 Davis-Haltiwanger 增长率刻画，并剔除左右两端 1%分位数异常值。

tfp_acf2_ty: 参考 Akerberg et al. (2015) 方法度量的企业的全要素生产率。

ln_GreenIPC 总数_分类: 基于分类号识别的企业绿色专利总数对数。

ln_GreenIPC 发明_分类: 基于分类号识别的企业绿色发明专利数对数。

ln_GreenIPC 实用_分类: 基于分类号识别的企业绿色实用新型专利数对数。

ln_GreenIPC 总数_主分类: 基于主分类号识别的企业绿色专利总数对数。

ln_GreenIPC 发明_主分类: 基于主分类号识别的企业绿色发明专利数对数。

ln_GreenIPC 实用_主分类: 基于主分类号识别的企业绿色实用新型专利数对数。

NewEntry: 进入企业虚拟变量，当企业 t 期出口为正、 $t-1$ 期出口为零定义为新进入企业，此时进入企业虚拟变量取 1、反之取 0。

Continuing: 在位企业虚拟变量，当企业 $t-1$ 期和 t 期出口均为正时，在位企业虚拟变量取 1、反之取 0。

Exiting: 退出企业虚拟变量，当 $t-1$ 期出口为正、 t 期出口为零时，退出企业虚拟变量取 1、反之取 0。

ln 排放强度_ndrc_dum75_100: 按照 NDRC 方法计算的二氧化碳排放强度对数位于 75 分位数以上的企业虚拟变量。

year: 年份。

参考文献

- [1] Akerberg, D.A., K. Caves, and G. Frazer, "Identification properties of recent production function estimators", *Econometrica*, 2015, 83(6), 2411-2451.
- [2] Barrows, G., and H. Ollivier, "Foreign demand, developing country exports, and CO2 emissions: Firm-level evidence from India", *Journal of Development Economics*, 2021, 149, 102587.
- [3] Brandt, L., J. Van Biesebroeck, and Y. Zhang, "Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing", *Journal of Development Economics*, 2012, 97(2), 339-351.
- [4] 陈登科, "贸易壁垒下降与环境污染改善——来自中国企业污染数据的新证据", 《经济研究》, 2020 年第 12 期, 第 98-114 页。
- [5] 吕越、陈泳昌, "上游服务业开放与制造业企业的环境污染", 《财贸经济》, 2022 年第 6 期, 第 114-127 页。