

学术论文规范与论文撰写——2022年秋季学期博士生课程

第1课

数据与代码管理

陈方豪 助理教授

经济学院 特区港澳经济研究所

2022年9月14日



暨南大学
JINAN UNIVERSITY

自我介绍

◆ 教育背景

- 2017/09-2022/06 北京大学 国家发展研究院 经济学博士
- 2019/12-2020/04 多伦多大学 经济系 访问学者
- 2013/09-2017/06 西南财经大学 经济与管理研究院 经济学学士
- 2016/08-2016/12 德州农工大学 经济系 交换生



暨南大学
JINAN UNIVERSITY

研究领域

- ◆ 产业经济学、城市与区域经济学、国际经济学
- ◆ 开放经济下的产业变迁, 特别是中国的产业发展与中国不断深化参与的国际贸易与跨国投资之间的关系。
- ◆ “研究外企入华、理解本土产业、服务中企出海”
- ◆ 关键词: 华人华侨、出海 (国际贸易、FDI)、产业



发表与审稿

◆ 发表论文

- [个人所得税减免会增加劳动供给吗?——来自准自然实验的证据](#)
叶菁菁, 吴燕, 陈方豪*, 王宇晴
《管理世界》, 2017, 12: 20-32
- [以侨为侨: 侨资企业与中国的外向型发展](#)
陈方豪*, 熊瑞驰
《中国经济学》(创刊号), 2022, 1: 118-158
- 突破城市发展的边界: 撤县设区与城市内部协同发展
陈方豪*, 叶菁菁
《经济学(季刊)》, Forthcoming
- 禀赋结构一致的技术调整: 技术可转换性、人力资本与结构转型
陈方豪*, 樊仲琛
《经济学(季刊)》, Forthcoming

◆ 审稿经历

- China Economic Review
- China and World Economy
- 经济学(季刊)



两次课程的安排

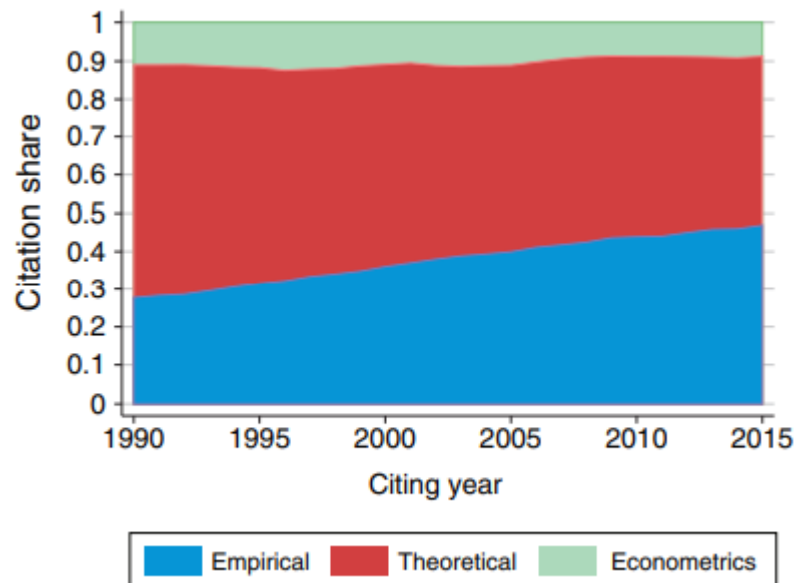
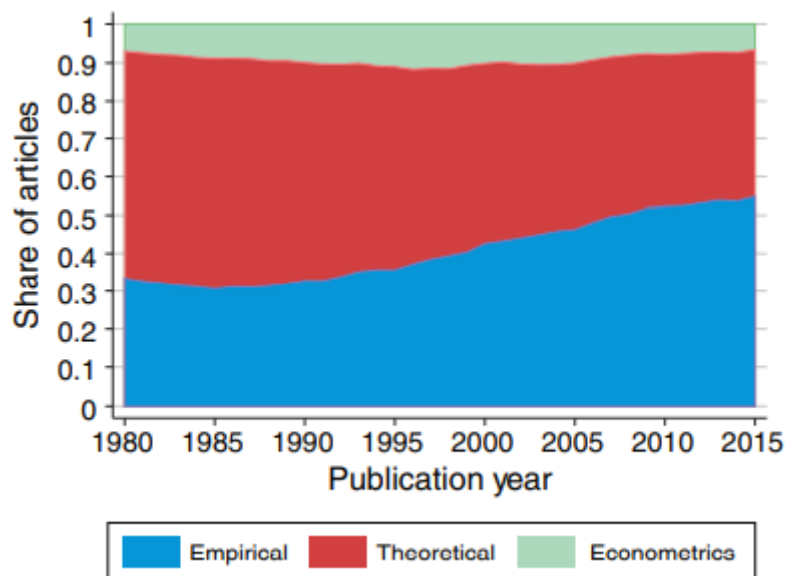
◆ 9月14日：务实

- 第1课：数据与代码管理
- 第2课：作图的原则与技巧

◆ 9月21日：务虚

- 第3课：论文写作的心法与技法
- 第4课：以学术为业

实证研究的黄金时代



Joshua Angrist, Pierre Azoulay, Glenn Ellison, Ryan Hill, and Susan Feng Lu, “Economic Research Evolves: Fields and Style”, American Economic Review: Papers & Proceedings 2017, 107(5): 293–297

- ◆ 实证（**empirical**）论文在数量和引用量上相对于其他类型都呈上升趋势
- ◆ 即使是理论（**theoretical**）论文也越来越经常被要求提供实证证据
- ◆ 计量（**econometrics**）论文根本上也是为了实证研究而服务



数据分析是实证论文的核心部分

- ◆ 提出一个好的问题→收集清理数据→统计分析→最后才以论文形式呈现
- ◆ 根据参会人、审稿人的反馈，重新调整数据和代码，修改内容，循环往复
- ◆ 70%以上的时间花费在撰写代码、清理数据、整理更新图表上面
- ◆ 提高数据分析的工作效率≈撰写整篇论文的工作效率

可是，你是否经常遇到以下情况？

- ◆ 打开很久没有动过的研究项目文件夹，面对一堆标题如“data.xlsx”、“结果10.doc”、“results_最终版.dta”等意义不明的文件们不知从何下手.....
- ◆ 好不容易找到了想要的文件，但已经彻底忘记自己是如何得到这个结果的；导师希望你更改之前的一个技术参数重新跑一遍结果，你不得不从头操作一遍数据.....
- ◆ 在开展一个项目的过程中产生了多个版本的结果，但你已经无法回忆起每个版本的结果的差异是在哪个环节产生的（样本、变量定义、模型设定.....）
- ◆ 导师希望检查一遍你的代码，确保结果的准确性，但你分享给导师的文件导师却表示看不懂，完全无从下手.....
- ◆

Workflow的重要性：华为与IBM

◆ 华为引进IBM的管理流程：痛苦但有效

另一项重大的ISC（集成供应链）变革，经过IBM顾问为期5年的指导，华为核心竞争力的全面提升更是非常明显：

1998年12月IBM顾问在对华为供应链进行变革之前，曾对华为的运行现状做过一次详细的摸底调查，那时候华为的订单及时交货率为30%，而世界级企业平均为90%；

华为的库存周转率为3.6次/年，而世界级企业平均为9.4次/年；华为的订单履行周期为20—25天，而世界级企业平均为10天左右……

2003年12月，IBM顾问再次给华为做出的考核数据显示：订单及时交货率已达到65%，库存周转率则上升到5.7次/年，而订单的履行周期也缩短到17天。

Workflow的重要性：华为与IBM

◆ 华为引进IBM的管理流程：痛苦但有效

U4

“美式”诊断触痛任正非

1995年，华为斥资1000万元从美国和德国引进二套先进管理系统后，顾问们基本上都听从华为人的“指挥”，甚至华为人还能对他们的系统提出各种各样的“改进方法”和“优化意见”，当然最后以流程“七疮八孔”为代价而告终。

但此次华为人发现：IBM顾问与以前的顾问截然不同，他们不仅趾高气扬，而且态度也非常强硬。

比如在进行访谈期间，顾问们一律要求华为干部讲英语，而且有时提问相关问题时，不让他们过多地问“为什么”，只允许他们回答“是”或“不是”。



本节课你将会学习到：

- ◆ 基于STATA的高效代码与数据管理workflow
 - [为什么选择STATA?](#)
 - 虽以STATA为例，但原则通用于所有数据处理分析软件
- ◆ 进一步提高
 - 学习资料
 - 计量方法：《高级计量经济学及Stata运用》（陈强）
 - 规范与技巧：[Replication File](#)
 - 公众号：连玉君、爬虫俱乐部
 - 日常积累：百度+好记性不如烂笔头+建立模板
 - 探索（exploratory）阶段的小技巧
 - `if 1==0{}` `ctrl+/` `/**//` `global` / 批量回归



三原则

- ◆ 可理解性（understandable）
 - 分门别类建立文件夹
 - 所有文件名都应是self-explanatory
 - 代码分区、缩进、分行
 - 重要的步骤、意义不明的位置有注释
- ◆ 可追溯性（trackable）
 - 以代码记录下所有操作，避免人工操作记录丢失
 - 记录下所有的input/output flow，分门别类保存
 - 版本管理（version control）
- ◆ 可分享性（replicable）
 - 只需更改根目录即可让他人复制整个项目
 - README
 - 变量标签（label）



六个步骤

- ◆ 1. 建立文件夹
- ◆ 2. 规范命名文件
- ◆ 3. 建立分工的do文件群
- ◆ 4. 使用相对路径控制input/output flow
- ◆ 5. 建立良好的代码风格
- ◆ 6. 在 do文件之间建立连接

1. 建立文件夹：个人的模板

文档 > Research > PlaceBasedPolicy				在 PlaceBasedPolicy 中搜索	
<input type="checkbox"/>	名称	修改日期	类型	大小	
	Crude_Data	2022/4/15 15:10	文件夹		
	Documentation	2022/9/11 10:35	文件夹		
	dofiles	2022/3/17 15:58	文件夹		
	Draft	2022/3/17 16:00	文件夹		
	figures	2022/3/17 15:58	文件夹		
	GIS_Data	2022/3/17 15:59	文件夹		
	Literature	2022/6/6 11:27	文件夹		
	logfiles	2022/3/17 15:58	文件夹		
	Presentations	2022/3/17 16:00	文件夹		
	raw_data	2022/4/15 15:19	文件夹		
	tables	2022/3/17 15:58	文件夹		
	temp_data	2022/3/17 15:58	文件夹		
	working_data	2022/3/17 15:58	文件夹		

- ◆ 文件命名：尽量使用英文；不使用空格和特殊符号；用“_”不用“-”
- ◆ 大写字母开头的文件夹用来放置需要手动操作的文件
- ◆ 小写字母开头的文件夹用来放置代码自动化操作产生的文件



1. 建立文件夹：个人的模板

文档 > Research > PlaceBasedPolicy				在 PlaceBasedPolicy 中搜索	
<input type="checkbox"/>	名称	修改日期	类型	大小	
	Crude_Data	2022/4/15 15:10	文件夹		
	Documentation	2022/9/11 10:35	文件夹		
	dofiles	2022/3/17 15:58	文件夹		
	Draft	2022/3/17 16:00	文件夹		
	figures	2022/3/17 15:58	文件夹		
	GIS_Data	2022/3/17 15:59	文件夹		
	Literature	2022/6/6 11:27	文件夹		
	logfiles	2022/3/17 15:58	文件夹		
	Presentations	2022/3/17 16:00	文件夹		
	raw_data	2022/4/15 15:19	文件夹		
	tables	2022/3/17 15:58	文件夹		
	temp_data	2022/3/17 15:58	文件夹		
	working_data	2022/3/17 15:58	文件夹		

- ◆ Crude_Data: 一些需要手动处理或尚未电子化的数据，例如网上下载的原始档案
- ◆ Documentation: 背景文件、阶段性结果整理、技术性参考资料、和合作者的讨论
- ◆ Draft: 论文的文稿（初始稿，修改稿，定稿）



1. 建立文件夹：个人的模板

文档 > Research > PlaceBasedPolicy				在 PlaceBasedPolicy 中搜索	
<input type="checkbox"/>	名称	修改日期	类型	大小	
	Crude_Data	2022/4/15 15:10	文件夹		
	Documentation	2022/9/11 10:35	文件夹		
	dofiles	2022/3/17 15:58	文件夹		
	Draft	2022/3/17 16:00	文件夹		
	figures	2022/3/17 15:58	文件夹		
	GIS_Data	2022/3/17 15:59	文件夹		
	Literature	2022/6/6 11:27	文件夹		
	logfiles	2022/3/17 15:58	文件夹		
	Presentations	2022/3/17 16:00	文件夹		
	raw_data	2022/4/15 15:19	文件夹		
	tables	2022/3/17 15:58	文件夹		
	temp_data	2022/3/17 15:58	文件夹		
	working_data	2022/3/17 15:58	文件夹		

- ◆ GIS_Data: 一些地理信息的数据，可能会涉及到其他鼠标流软件如ArcGIS
- ◆ Literature: 文献
- ◆ Presentations: 开会时用于展示的PPT



1. 建立文件夹：个人的模板

文档 > Research > PlaceBasedPolicy				在 PlaceBasedPolicy 中搜索	
<input type="checkbox"/>	名称	修改日期	类型	大小	
	Crude_Data	2022/4/15 15:10	文件夹		
	Documentation	2022/9/11 10:35	文件夹		
	dofiles	2022/3/17 15:58	文件夹		
	Draft	2022/3/17 16:00	文件夹		
	figures	2022/3/17 15:58	文件夹		
	GIS_Data	2022/3/17 15:59	文件夹		
	Literature	2022/6/6 11:27	文件夹		
	logfiles	2022/3/17 15:58	文件夹		
	Presentations	2022/3/17 16:00	文件夹		
	raw_data	2022/4/15 15:19	文件夹		
	tables	2022/3/17 15:58	文件夹		
	temp_data	2022/3/17 15:58	文件夹		
	working_data	2022/3/17 15:58	文件夹		

- ◆ dofiles: 存放do文件
- ◆ figures: 存放图片
- ◆ tables: 存放表格
- ◆ logfiles: 存放日志文件



1. 建立文件夹：个人的模板

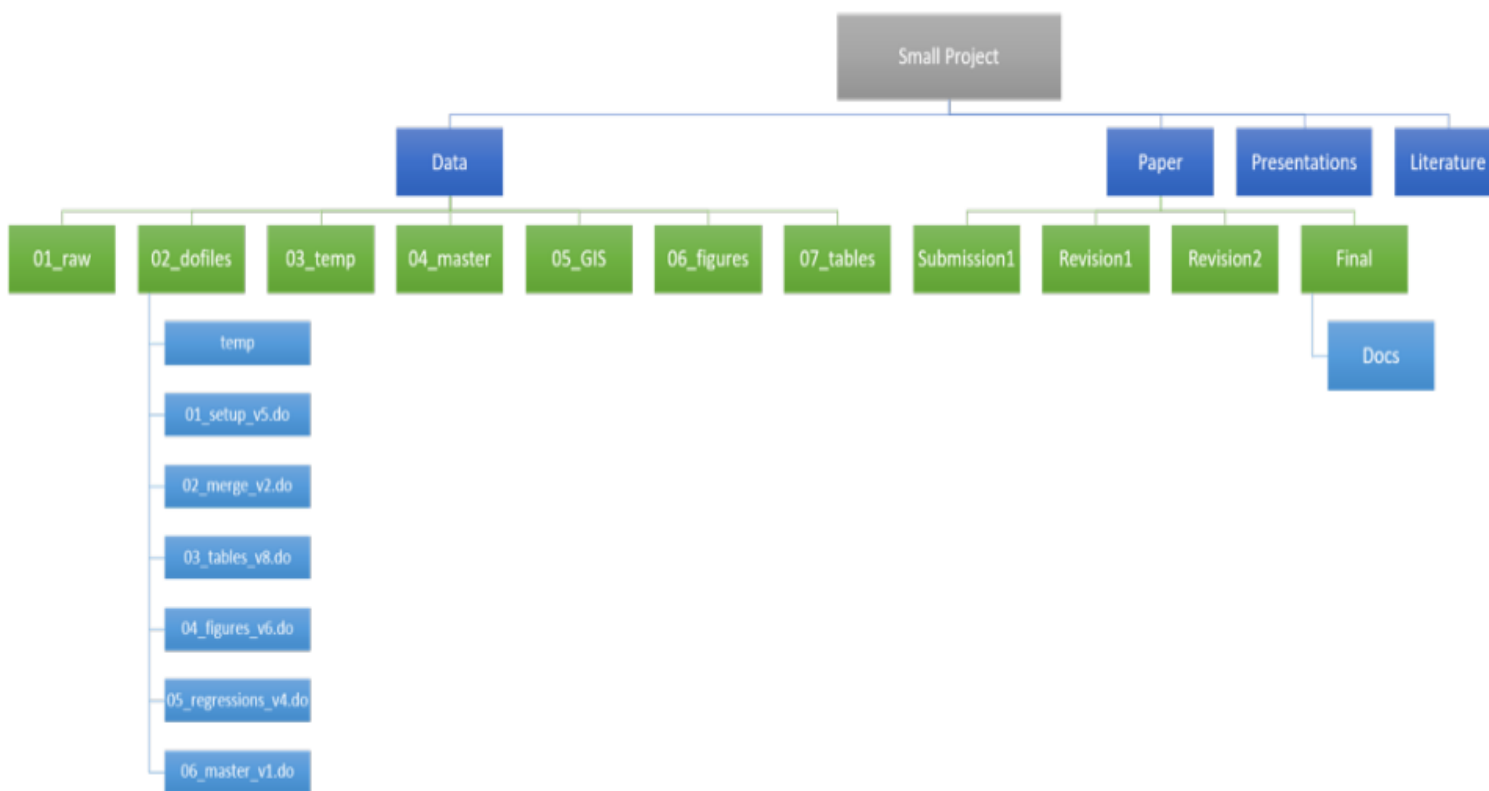
文档 > Research > PlaceBasedPolicy				在 PlaceBasedPolicy 中搜索	
<input type="checkbox"/>	名称	修改日期	类型	大小	
	Crude_Data	2022/4/15 15:10	文件夹		
	Documentation	2022/9/11 10:35	文件夹		
	dofiles	2022/3/17 15:58	文件夹		
	Draft	2022/3/17 16:00	文件夹		
	figures	2022/3/17 15:58	文件夹		
	GIS_Data	2022/3/17 15:59	文件夹		
	Literature	2022/6/6 11:27	文件夹		
	logfiles	2022/3/17 15:58	文件夹		
	Presentations	2022/3/17 16:00	文件夹		
	raw_data	2022/4/15 15:19	文件夹		
	tables	2022/3/17 15:58	文件夹		
	temp_data	2022/3/17 15:58	文件夹		
	working_data	2022/3/17 15:58	文件夹		

以dta格式存放数据于如下文件夹：

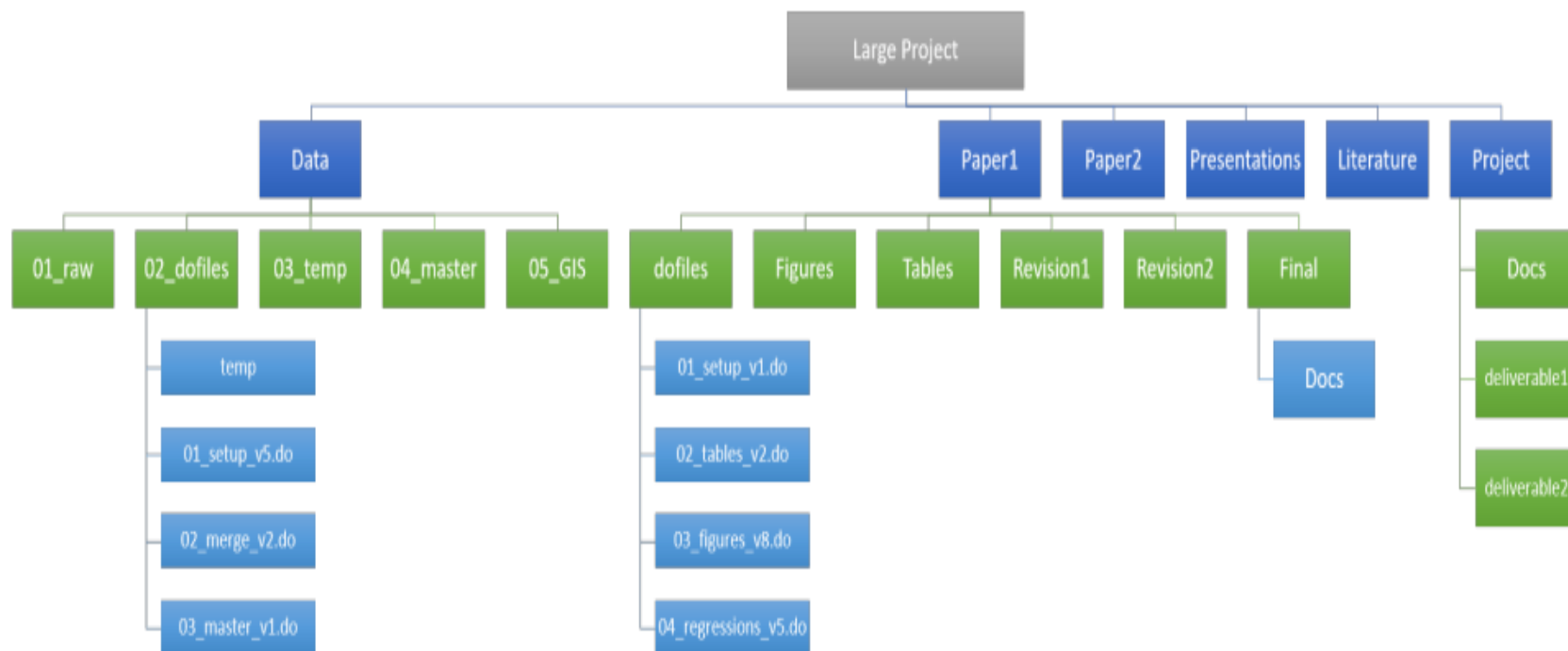
- ◆ raw_data: 原始数据文件
- ◆ temp_data: 所有中间步骤产生的数据文件
- ◆ working_data: 最后用于进行回归分析的数据文件



1. 建立文件夹：管理小项目



1. 建立文件夹：管理大项目



2. 规范命名文件

- ◆ 命名格式：（序号）_功能_（作者）_（版本）_yymmdd
 - “序号”：[对文件按一定顺序排列](#)（不必需）
 - “功能”：应是self-explanatory，不会产生歧义
 - ~~ESIEC_datacleaning_partA_FanghaoChen_190917.do~~→
clean_partA_cfh_190917.do
 - “作者”：用于记录合作项目中不同作者的操作痕迹
 - “版本”：补充的说明（例如投稿时有时要求删去作者信息）
 - “yymmdd”：版本管理（version control）
 - ~~Draft_v5.docx~~→ Draft_220914.docx
 - 不同类型的文件通过相同的日期后缀进行对应，保证可追溯
 - Draft_220914.docx, main_220914.do
- ◆ 上述原则对所有文件（文稿、图片、表格……）都适用
- ◆ 同样适用于变量命名：
 - rename age35 ShareDoctor_age35



3. 建立分工的do文件群

- ◆ 把所有代码写在一个do文件里
 - 内容庞杂
 - debug 费时费力
- ◆ 按照**功能**将代码拆分成一个文件群，存放于dofiles文件夹下：
 - 00_main_220914.do → **220914**这个版本所有结果可由运行该文件一键生成
 - 01_setup_220914.do → **数据清理**
 - 011_setup_custom_220913.do
 - 012_setup_citystats_220910.do
 - 02_merge_220914.do → **整合清理后的数据组成用于分析的数据**
 - 03_tables_220914.do → **输出表格**
 - 031_tables_sumstat_220911.do
 - 032_tables_cityreg_220911.do
 - 032_tables_firmreg_220912.do
 - 04_figures_220914.do → **输出图片**
- ◆ 严格遵照功能切分代码。例如，生成变量/打标签的工作统一放在setup与merge中完成



4.使用相对路径控制input/output flow

- ◆ 固定每一个do文件的开头，以宏（global）来控制输入输出

```
/******  
Industrial Development in China  
  
Merging Datasets:  
year-prefecture-cic2d level  
  
This version: 2022/8/31 Fanghao Chen  
  
*****/  
  
clear  
set more off  
capture log close  
  
global root="C:\Users\Chinh\Documents\Research\ChinaSpatial\DataCode_Package_22Jun2022" //Chen  
  
global dofiles= "$root/dofiles"  
global logfiles= "$root/logfiles"  
global raw_data= "$root/Data/raw_data"  
global temp_data= "$root/Data/temp_data"  
global working_data= "$root/Data/working_data"  
global tables= "$root/tables"  
global figures= "$root/figures"  
  
log using $logfiles/merge_year_region4d_cic2d_220831.log, replace  
  
/******
```



4.使用相对路径控制input/output flow

- ◆ 固定每一个do文件的开头，以宏（global）来控制输入输出

```
/******
```

```
Industrial Development in China
```

```
Merging Datasets:
```

```
year-prefecture-cic2d level
```

```
This version: 2022/8/31 Fanghao Chen
```

```
***** /
```

```
clear  
set more off  
capture log close
```

```
global root="C:\Users\Chinh\Documents\Research\ChinaSpatial\DataCode_Package_22Jun2022" //Chen
```

```
global dofiles= "$root/dofiles"  
global logfiles= "$root/logfiles"  
global raw_data= "$root/Data/raw_data"  
global temp_data= "$root/Data/temp_data"  
global working_data= "$root/Data/working_data"  
global tables= "$root/tables"  
global figures= "$root/figures"
```

```
log using $logfiles/merge_year_region4d_cic2d_220831.log, replace
```

```
/******
```

团队成员只需要修改root这里的路径就可以复制你的结果



4.使用相对路径控制input/output flow

- ◆ 固定每一个do文件的开头，以宏（global）来控制输入输出

```
/******  
Industrial Development in China  
Merging Datasets:  
year-prefecture-cic2d level  
This version: 2022/8/31 Fanghao Chen  
*****/  
  
clear  
set more off  
capture log close  
  
global root="C:\Users\Chinh\Documents\Research\ChinaSpatial\DataCode_Package_22Jun2022" //Chen  
global dofiles= "$root/dofiles"  
global logfiles= "$root/logfiles"  
global raw_data= "$root/Data/raw_data"  
global temp_data= "$root/Data/temp_data"  
global working_data= "$root/Data/working_data"  
global tables= "$root/tables"  
global figures= "$root/figures"  
  
log using $logfiles/merge_year_region4d_cic2d_220831.log, replace  
  
/******
```

通过宏来建立和文件夹之间的联系



4.使用相对路径控制input/output flow

- ◆ 固定每一个do文件的开头，以宏（global）来控制输入输出

```
/******  
Industrial Development in China  
  
Merging Datasets:  
year-prefecture-cic2d level  
  
This version: 2022/8/31 Fanghao Chen  
  
*****/  
  
clear  
set more off  
capture log close  
  
global root="C:\Users\Chinh\Documents\Research\ChinaSpatial\DataCode_Package_22Jun2022" //Chen  
  
global dofiles= "$root/dofiles"  
global logfiles= "$root/logfiles"  
global raw_data= "$root/Data/raw_data"  
global temp_data= "$root/Data/temp_data"  
global working_data= "$root/Data/working_data"  
global tables= "$root/tables"  
global figures= "$root/figures"  
  
log using $logfiles/merge_year_region4d_cic2d_220831.log, replace  
  
/******
```

日志文件和do命名保持一致，用于做version control，保持可追溯性



4.使用相对路径控制input/output flow

- ◆ 读取文件：
 - use \$raw_data/custom.dta, clear
- ◆ 生成中间数据：
 - save \$temp_data/firm_export, replace
- ◆ 生成最终数据：
 - save \$working_data/firm.dta, replace
- ◆ 输出表格：
 - outreg2 using \$tables/baseline_ols_220914.xls, replace
- ◆ 输出图片：
 - graph export \$figures/220622/exportdumumy_vs_firmage.png, replace

4.使用相对路径控制input/output flow

- ◆ 读取文件：
 - use \$raw_data/custom.dta, clear
- ◆ 生成中间数据：
 - save \$temp_data/firm_export, replace
- ◆ 生成最终数据：
 - save \$working_data/firm.dta, replace
- ◆ 输出表格：
 - outreg2 using \$tables/baseline_ols_220914.xls, replace
- ◆ 输出图片：
 - graph export \$figures/220622/exportdumumy_vs_firmage.png, replace

数据不做version control，节省空间。利用replace 来自动覆盖更新结果（用do文件来控制数据版本）

4.使用相对路径控制input/output flow

- ◆ 读取文件：
 - use \$raw_data/custom.dta, clear
- ◆ 生成中间数据：
 - save \$temp_data/firm_export, replace
- ◆ 生成最终数据：
 - save \$working_data/firm.dta, replace
- ◆ 输出表格：
 - outreg2 using \$tables/baseline_ols_220914.xls, replace
- ◆ 输出图片：
 - graph export \$figures/220622/exporthdumy_vs_firmage.png, replace

表格和图片仍然做
version control
(便于结果的迭代
和追溯)。其产出
和主do文件
main_yymmdd.do
对应。

5.建立良好的代码风格

- ◆ 代码的分区分级
- ◆ 不要吝啬空行和缩进

```
/** year-prefecture-industry level */  
  
/** worker size */  
    use $raw_data/TotalWorker_82-15_cic2d.dta, clear  
  
/** similarity & suitability */  
    merge 1:1 year pref_code cic_2d using $working_data/L-based_Sim_allsector.dta  
    drop if _merge==2  
    drop _merge  
  
    merge 1:1 year pref_code cic_2d using $working_data/L-based_Suitability_allsector.dta  
    drop if _merge==2  
    drop _merge  
  
/** specialization & diversity */  
    merge 1:1 year pref_code cic_2d using $temp_data/SpecializationDiversity_82-15_Nonagr.dta  
    drop if _merge==2  
    drop _merge  
  
/** Bartik IV */
```

5.建立良好的代码风格

◆ 缩进

```
\begin{frame}{5.规范代码的几条tips}{  
  \begin{itemize}  
    \item 使用tab进行缩进  
  \end{itemize}  
  
  \begin{figure}[htbp]  
    \centering  
    \includegraphics[width=10cm]{figures/line.png} % requires thgraphicx package  
    % \caption[]{} % adds name for the figure  
    % \label{fig2} % adds cross-reference  
  \end{figure}  
\end{frame}
```

◆ 分行

```
twoway  
  (line CumulShare_dist2sea dist2sea if year==1982,lwidth(thick) lpattern(dash))  
  (line CumulShare_dist2sea dist2sea if year==1990,lwidth(thin) lpattern(dash))  
  (line CumulShare_dist2sea dist2sea if year==2000,lwidth(thin) lpattern(solid))  
  (line CumulShare_dist2sea dist2sea if year==2005,lwidth(thick) lpattern(solid))  
  title("Aggregate Employment")  
  xtitle("kilometers to seaboard")  
  ytitle("cumulative fraction of national employment")  
  xlabel(0(500)3500)  
  legend(order(1 "1982" 2 "1990" 3 "2000" 4 "2005"))  
  legend(row(1))  
  legend(region(lcolor(none)))  
  );  
#delimit cr
```

5.建立良好的代码风格

- ◆ 注释：重要的步骤、意义不明的位置有注释
 - 利于自己，也利于团队

```
drop if _ID==14 | _ID==28 | _ID==38 | _ID==39 | _ID==43 |  
_ID==45 | _ID==46
```



```
drop if ///  
_ID==14 |    ///  
_ID==28 |    ///  
_ID==38 |    ///  
_ID==39 |    ///  
_ID==43 |    ///  
_ID==45 |    ///  
_ID==46      ///  
              // Puerto Rico  
              // Alaska  
              // American Samoa  
              // United States Virgin Islands  
              // Hawaii  
              // Guam  
              // North Mariana Islands
```



6.在 do文件之间建立连接

- ◆ 用00_main_yymmdd.do记录下来所有文件的运行顺序，一键生成结果
- ◆ yymmdd 代表结果的版本，和其他图表、文稿等文件相对应，便于追溯

```
/** Preparing Data **/  
  
/** market potential (it) **/  
do $dofiles/preparation/clean_pref_MP_220622.do  
  
/** specialization (ijt) & diversity (it) **/  
do $dofiles/preparation/clean_SpecializationDiversity_220817.do  
  
/** employment-based sectoral suitability indices (ijt) **/  
do $dofiles/preparation/clean_L-based_Sim_220622.do //similarity indices  
do $dofiles/preparation/clean_L-based_GlobalSuitability_220622.do //suitability indices  
  
/** Bartik IV **/  
do $dofiles/preparation/clean_bartik_220831.do //Bartik-type IV  
  
/** Regression **/  
  
/** descriptive regressions **/  
  
do $dofiles/analysis/merge_year_region4d_cic2d_220831.do //merging datasets at year-prefecture-cic2d level  
do $dofiles/analysis/sumstat_220622.do //summary statistics & descriptive regressions  
  
/** regressions **/  
  
do $dofiles/analysis/reg_year_region4d_cic2d_220622.do //regression at year-prefecture-cic2d level
```



6.在 do文件之间建立连接

- ◆ 每个文件都应该能够独立运行！
 - 便于debug阶段微调设定
- ◆ 撰写readme文件
 - 打包给别人的时候，告诉别人从何下手
 - 理解重要文件的功能与意义
 - 文件与文件之间是如何关联起来的



进一步提高：学习资料

◆ 计量方法及代码实现：

- 《高级计量经济学及Stata运用》（陈强）
- Stata Journal
- Help文件/ssc new/ssc hot/ssc install xxxxx, replace
- 公众号：计量经济圈

◆ 规范与技巧：

- [Replication File](#)
- 公众号：连玉君、爬虫俱乐部

进一步提高：日常积累

- ◆ 勤百度基本上可以解决
 - You are not alone in encountering any problems
- ◆ 好记性不如烂笔头：整理自己的应用笔记
- ◆ 建立模板
 - Sublime的Snippet可以一键调用

使用代码的两个阶段：exploratory and explanatory

- ◆ 在最终呈现结果之前，一个研究者必会经历一个探索阶段（**exploratory stage**），需要频繁修改与测试代码。
- ◆ 有一些小技巧可以服务于这个阶段，提高效率

探索阶段的小技巧1：测试不同设定的代码

◆ If 1==0{

<代码>

}

◆ 选定内容批量注释：Ctrl+/

◆ /*

<代码>

*/

探索阶段的小技巧1：测试不同设定的代码

◆ 给代码设置开关

*这三行是用来定义「开关」的

```
local A    1 // 第 1 个开关
```

```
local B    0 // 第 2 个开关
```

```
if `A' == 1 {
```

```
<第一种设定>
```

```
}
```

```
if `B' == 1 {
```

```
<第二种设定>
```

```
}
```

探索阶段的小技巧2：用global设定变量组

```
global Control_Full z1 z2 z3 i.prov_id
```

```
global Control_WithoutProvFE z1 z2 z3
```

```
reg y x $Control_Full
```

```
reg y x $Control_WithoutProvFE
```


探索阶段的小技巧3：批量输出回归结果

```
global Control_Full z1 z2 z3 i.prov_id

global Control_WithoutProvFE z1 z2 z3

local s=0

foreach y of varlist y1 y2 y3{
    foreach x of varlist x1 x2 x3{

        Local s=`s'+1
        reg `y' x $Control_Full
        est store o`s'

        Local s=`s'+1
        reg `y' x $Control_WithoutProvFE
        est store o`s'

    }

}

outreg2 [s*] using "$tables/reg_220914.xls", replace

\\ \\ stats(coef se) level(95) dec(3)
```



参考资料

- ◆ Matthew Gentzkow, Jesse Sharpiro., Code and Data for the Social Sciences: A Practitioner's Guide, 2014.
- ◆ Asjad Naqvi, The Stata Workflow Guide, 2021.



谢 谢

- ◆ 陈方豪 助理教授
- ◆ 经济学院 特区港澳经济研究所
- ◆ 2022年毕业于北京大学国家发展研究院，获经济学博士学位，多伦多大学访问学者；主要研究领域为产业经济学、城市经济学与国际经济学；现有学术论文发表或待刊于《管理世界》、《经济学（季刊）》、《中国经济学》等权威期刊。他的研究主线是开放经济下的产业变迁，尤其关注中国的产业发展与中国不断深化参与的国际贸易与跨国投资之间的关系。当前的研究重心为改革开放后作为先驱投资者的海外侨商在中国大陆投资的历史演变、决定因素与长期影响。
- ◆ 邮箱：chenfanghao@jnu.edu.cn 或 fhchen2017@nsd.pku.edu.cn
- ◆ 个人网页：<https://fanghaochen.github.io/homepage/>
- ◆ 知乎号：Chinhogo



为什么使用STATA?

- ◆ 消极的理由：已经成为应用微观计量的通用“语言”
 - 经济学者分享的数据/代码基本以STATA的dta文件和do文件为主
- ◆ 积极的理由：真的很好用！
 - 相比于鼠标流软件（如SPSS）：以代码控制操作，透明可追溯、灵活易修改
 - 相比于一般的编程软件（如Python/matlab）：model-based，代码简洁清晰，入门门槛低；围绕计量经济学核心的回归分析功能得到优化，功能全而强大



重要的学习资源：Replication File

◆ <https://dataverse.harvard.edu/dataverse/qje>

The screenshot displays the Harvard Dataverse interface. At the top, the Harvard Dataverse logo is on the left, and navigation links (Add Data, Search, About, User Guide, Support, Sign Up, Log In) are on the right. Below the header, a table lists files. The first four files are visible:

1 to 10 of 328 Files		Download
<input type="checkbox"/>	<p>01_Shopping_List.do replication/code-build/ Stata Syntax - 832 B Published 2022年8月29日 2 Downloads</p>	
<input type="checkbox"/>	<p>02_Geoprocessing.py replication/code-build/ Python Source Code - 15.9 KB Published 2022年8月29日 2 Downloads</p>	
<input type="checkbox"/>	<p>03_Maps.do replication/code-build/ Stata Syntax - 12.4 KB Published 2022年8月29日 2 Downloads</p>	
<input type="checkbox"/>	<p>04_City_Books.do replication/code-build/ Stata Syntax - 5.2 KB Published 2022年8月29日 2 Downloads</p>	

重要的学习资源：Replication File

◆ <https://dataverse.harvard.edu/dataverse/restat>



对文件按一定顺序排列

- 01_setup_v4.do
- 02_merge_v2.do
- 03_tables_v11.do
- 04_regressions_v15.do
- 05_figures_v2.do
- 06_master_v1.do