

第五届 Stata 中国用户大会

一个 Stata 用户的若干思考

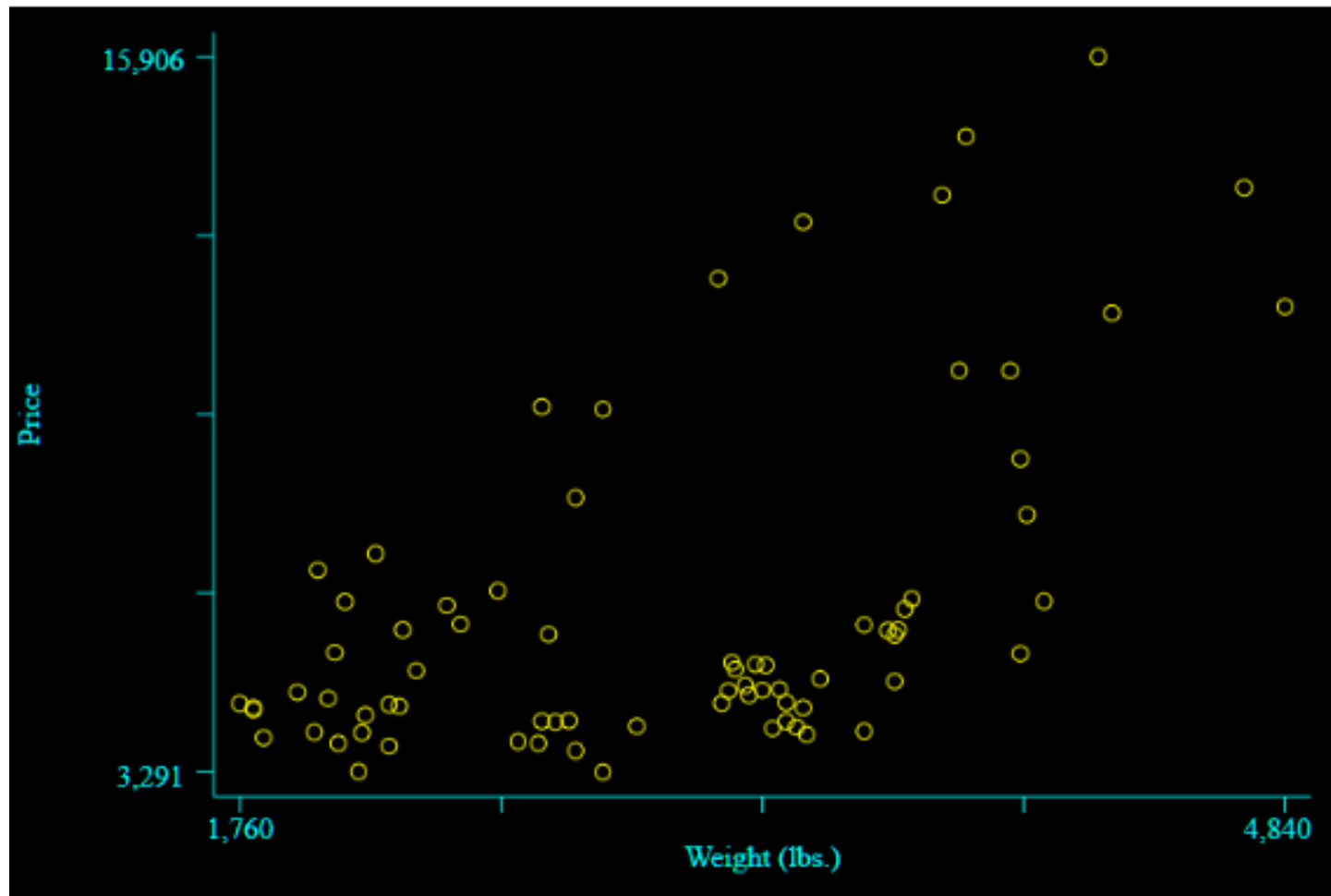
连玉君 (中山大学)

arlionn@163.com

👉 幻灯片下载:

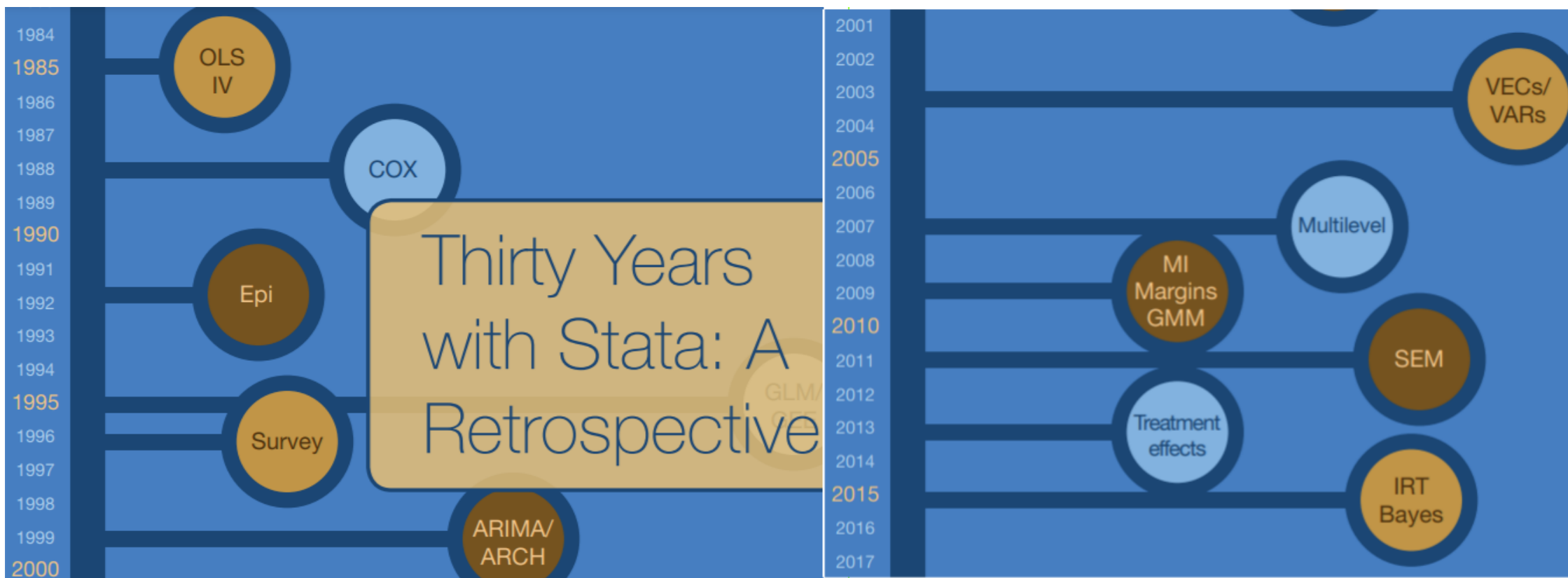
<https://gitee.com/arlionn/md>

```
sysuse "auto.dta", clear  
version 7.0      // 古董  
graph price wei
```



Stata 的前世今生

- [History of Stata](#), `help version`
- Stata Press, 2015, [Thirty Years with Stata: A Retrospective](#)



提纲

- 用户
- 老师
- Stata 公司

1. 对用户的建议

- Stata 的角色
- 基础
- 资源
- 习惯

Stata: 16 → 17 想到的

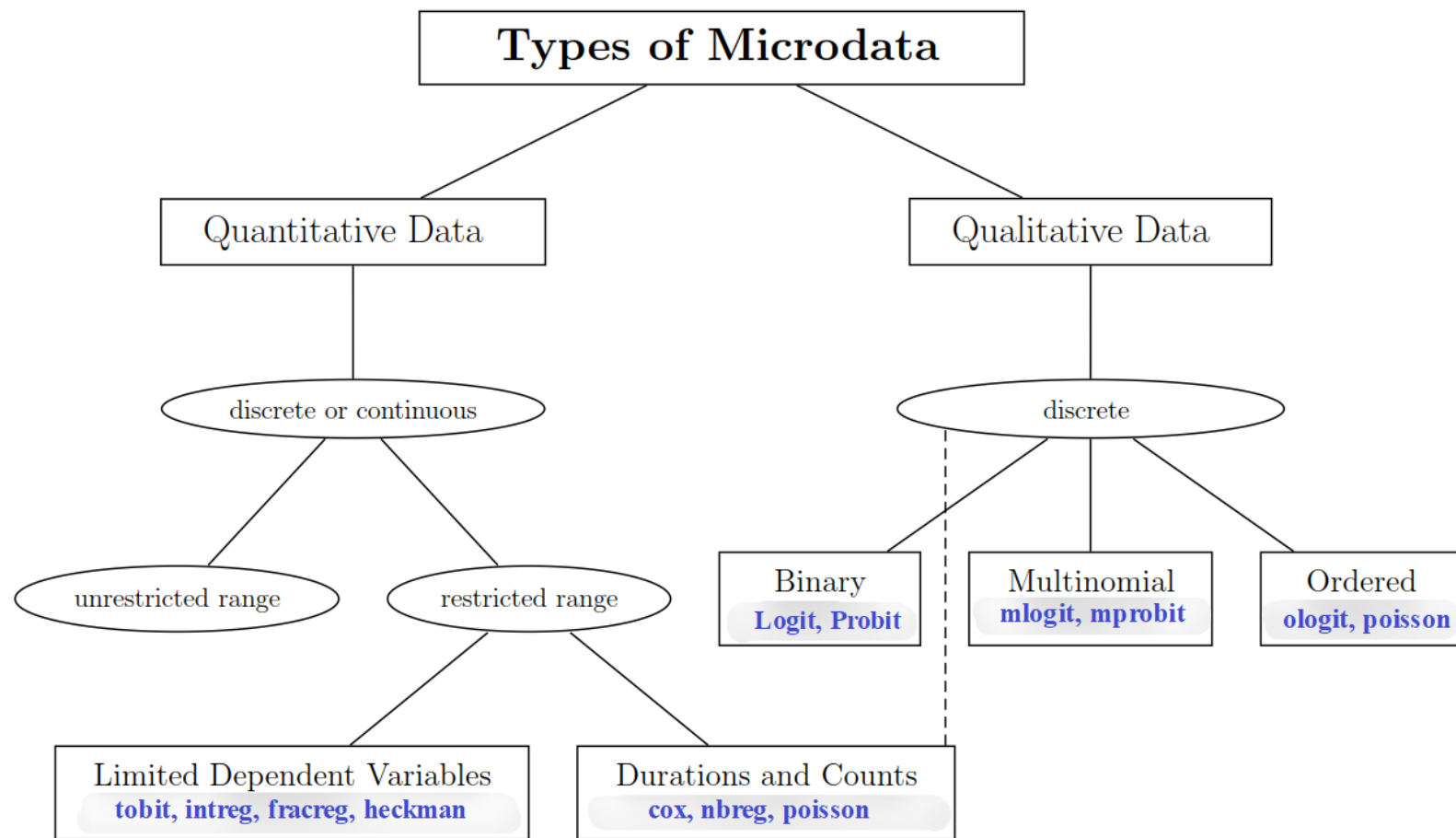
- `help whatsnew16to17` [New in Stata17](#)
- [Jupyter Notebook with Stata](#)
- [pystata: Example 1; API](#)
- Machine Learning - [H2O.ai](#) → [H2O integration](#)
- [Java integration](#) (Java: 电子交易系统; 高频交易; 安全)

Python Java

Stata Jupyter Notebook R

Matlab C++ Fortran

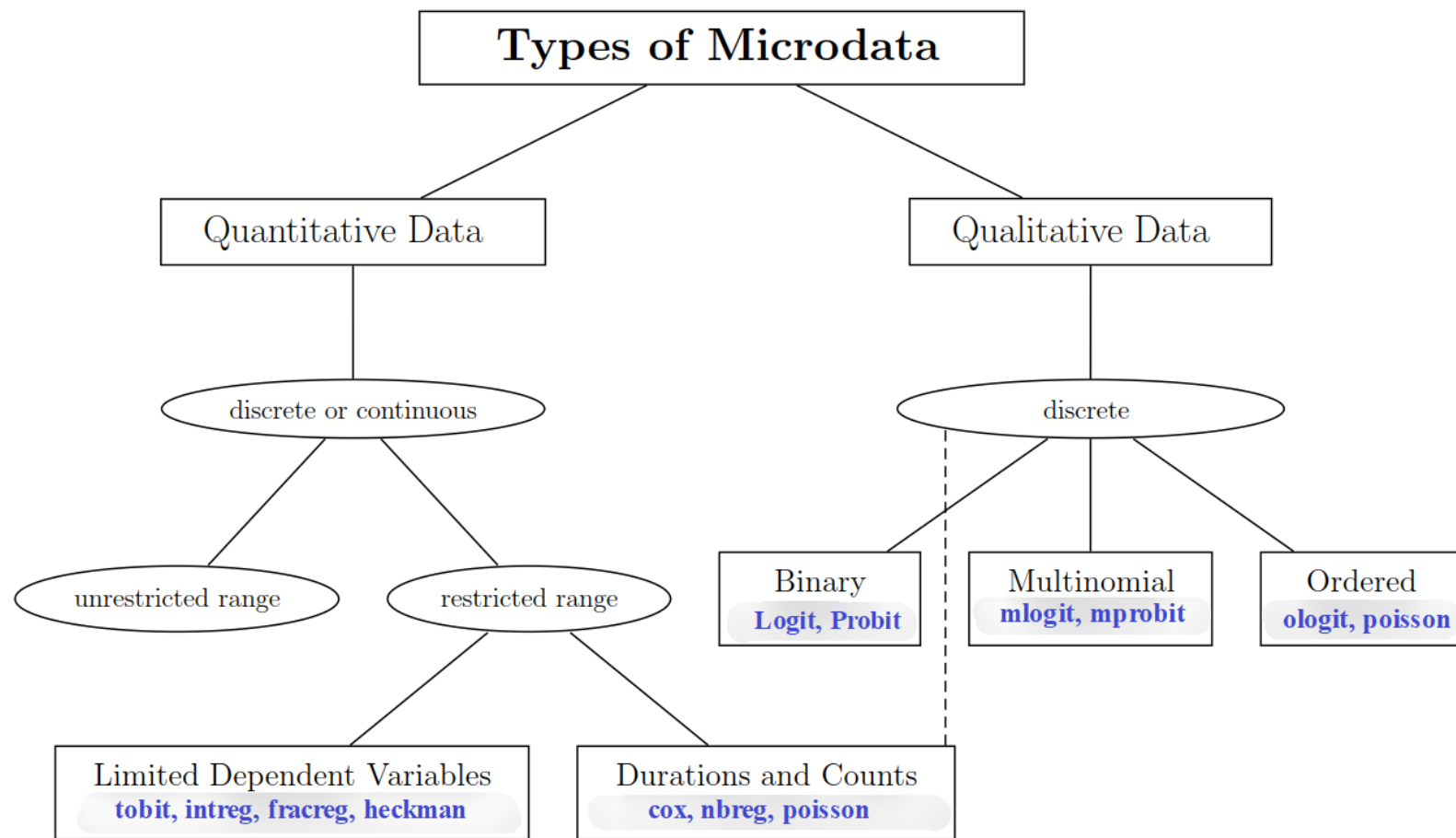
学习/使用 Stata: 我们的猎物



学习 Stata

- 基本知识架构: OLS, GMM, MLE → 因果推断 → 机器学习?
- Conditional Expectation function (CEF):
 - $E(y|\mathbf{x}) = f(\mathbf{x}; \beta) = \alpha + x_1\beta_1 + x_2\beta_2 + \varepsilon$
 - Marginal Expectation Effects (MEE)
 - OLS → Dummy → FE, RE, DID, RDD
 - GMM → IV, 2SLS, Dynamic Panel Data
- Conditional Probability Function (CPF):
 - $P(y|\mathbf{x}; \beta) = g(\mathbf{x}; \beta) = \Phi(\mathbf{x}'\beta)$ or $g(\mathbf{x}; \beta) = \frac{\exp(\mathbf{x}'\beta)}{1+\exp(\mathbf{x}'\beta)}$
 - Marginal Probability Effects (MPE)
 - MLE → Probit/Logit, Count Data, Tobit, Heckman, Treatment

再看我们的猎物：CEF v.s CPF



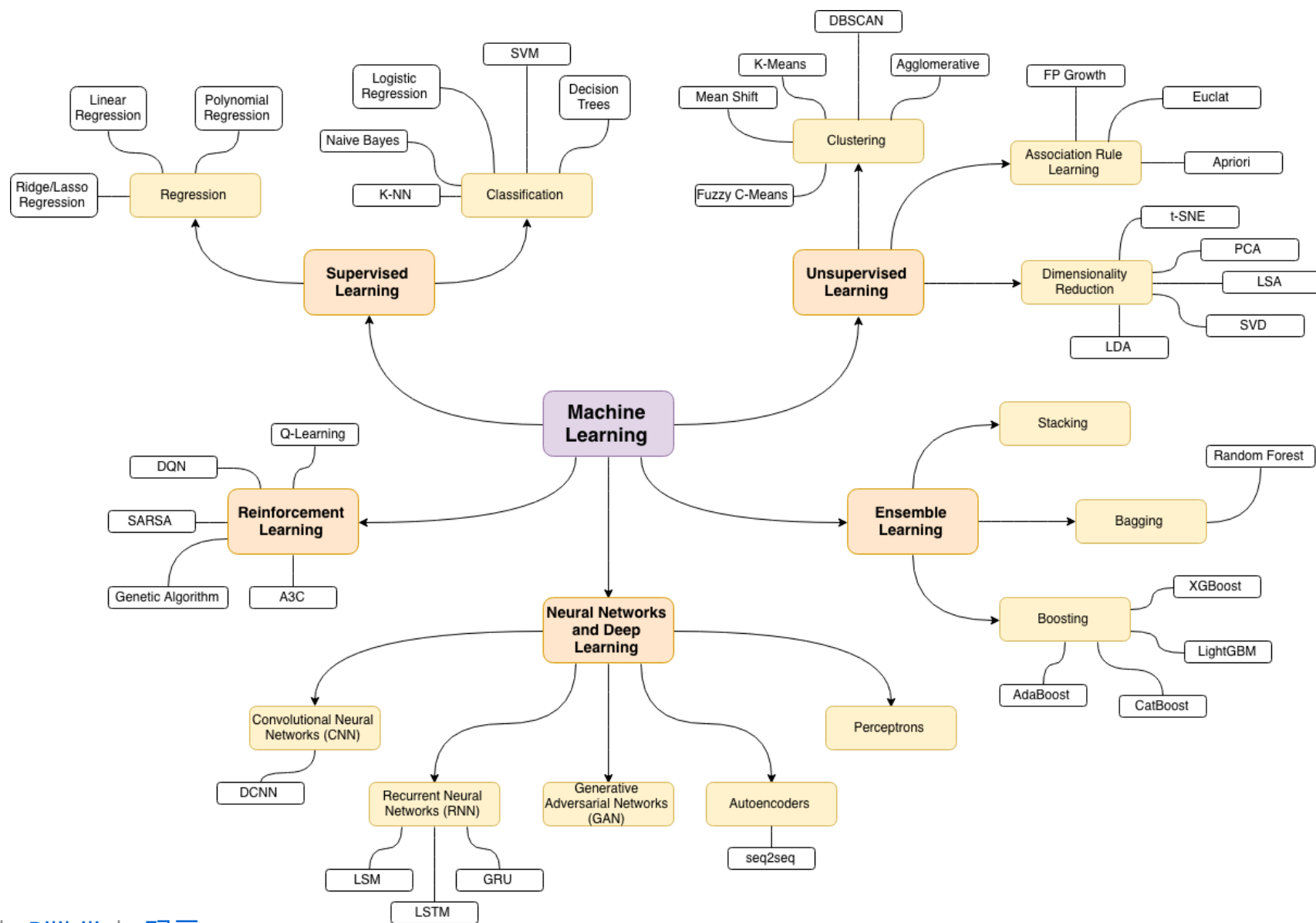
经验总结

- 打基础：Prabability, Matrix, Mata, Bayes
 - **慢下来才能快起来**
 - 基本概念、核心原理 (不变) → 应万变
 - 理解了 CPF → Probit, Tobit, Heckman, Count data
 - 煮肉：一次性煮熟 Econometrica 背后运球
- 细嚼慢咽 + 反刍
 - 一个主题 3-5 本书 → 整理笔记
 - 读文献、干中学 (What is Bootstrap ?)
 - 讲出来、写出来

工具：协作+分享 | 资源

- [github](#) → [my Github](#) [github-Stata](#)
 - 团队协作: [worldbank/stata](#), [mostly-harmless](#), [stata-fundamentals](#)
 - 代码分享: [reghdfe](#), [ivreghdfe](#), [moremata](#), [gtools](#), [xtdcce2](#)
 - 教程和代码集合: [ML](#) (18.2k Star), [PythonDataScienceHandbook](#) (30.5k Star)
- [gitee](#) ([码云](#))
 - [五分钟 Markdown](#) [synth](#) [论文复现-28篇](#)
 - [lianxh](#), [songbl](#)

Source: ML (18.2k Star)



工具：Github → Gitee → 星选集

- 我的 星选集：<https://gitee.com/arlionn/collections>
- 操作：Fork Github 仓库 → 进入 Gitee
→ 导入 Github 仓库 → Star, 拖入「星选集」

★ 教科书Data_dofile

❤️ 3

9 个仓库



✎ 编辑 🗑 删除

★ 论文重现代码

❤️ 26

一些经典的论文重现代码和数据

28 个仓库

✎ 编辑 🗑 删除

🔒 <https://gitee.com/arlionn/collections> ☆

★ DID

❤️ 4

4 个仓库

Stata连享会

@arlionn

连享会主页: lianxh.cn

✎ 编辑 🗑 删除

★ 合成控制法

❤️ 7

标准合成控制法 (Abadie et al., 2010, 2015); 合成DID; 一般化SCM; 贝叶斯SCM; Lasso-SCM; 等

10 个仓库

✎ 编辑 🗑 删除

🔗 [Stata连享会 / alternative-synthetic-control-sparsity](#)

An alternative to synthetic control for models with many covariates under

最近更新: 5个月前

🔗 [Stata连享会 / pensynth](#)

Penalized Synthetic Control w/ Alberto Abadie

最近更新: 5个月前

🔗 [Stata连享会 / synth_runner](#)

A tool to run a pool of synthetic controls, conduct inference, and produce

最近更新: 5个月前

🔗 [Stata连享会 / ElasticSynth](#)

Synthetic Controls for Causal Analysis using Elastic Net Regression

最近更新: 5个月前

🔗 [Stata连享会 / SytheticControlMethod](#)

Simulation study of Synthetic control method

最近更新: 2个月前

🔗 [Stata连享会 / microsynth](#)

Synthetic controls for micro-level data

最近更新: 5个月前

🔗 [Stata连享会 / augsynth](#)

Augmented Synthetic Control Method

最近更新: 5个月前

🔗 [Stata连享会 / tidysynth](#)

A tidy implementation of the synthetic control method in R

最近更新: 5个月前

笔记工具 - Markdown

- 五分钟 Markdown: <https://gitee.com/arlionn/md>, 新浪视频

```
9  ## 2. 原文
10
11 > 写在前面: Markdown 是简单标记语言
12
13 ## 简介
14 花五分钟介绍个好东西——Markdown! .....
15
16 ---
17
18 ### 什么是 Markdown?
19 - 一种简单的标记语言
20 | -- 用 `#` 表示一级标题, `##` 标记二级标题;
21 | -- 用 `-` 标记一个条目, 缩进两格 - 表示二级条目;
22 - 代码高亮和代码块
23 | `regress` (我高亮了) 命令很好用:
24
25 ```stata
26 . sysuse "auto.dta", clear
27 . regress price weight length
28 . regfit
29 price = 10386.54 + 4.70*weight - 97.96*length
30 | (4308.16) (1.12) (39.17)
31 | N = 74, R2 = 0.35, adj-R2 = 0.33
32 ```
```

2. 原文

写在前面: Markdown 是简单标记语言

简介

花五分钟介绍个**好东西**——*Markdown*!

什么是 Markdown?

- 一种简单的标记语言
 - 用 # 表示一级标题, ## 标记二级标题;
 - 用 - 标记一个条目, 缩进两格 - 表示二级条目;
- 代码高亮和代码块 `regress` (我高亮了) 命令很好用:

```
. sysuse "auto.dta", clear
. regress price weight length
. regfit
price = 10386.54 + 4.70*weight - 97.96*length
      (4308.16) (1.12)      (39.17)
      N = 74, R2 = 0.35, adj-R2 = 0.33
```

笔记工具 - Markdown

- Stata 17 也支持 Markdown 语法了
- Jupyter Notebook 主要用 Markdown 来码字
- [VScode](#) + [Typora](#) + [web.marp.app](#)

笔记工具 - LaTeX

- 剑桥大神 [Dexter Chua](#) 的 [Cambridge Notes](#) 从大一到大四总计 4000 多页
 - [github](#) → [下载](#)

dec41.user.srcf.net/notes/

MICHAELMAS TERM

Differential Equations (2014, M. G. Worster)

[HTML](#) [PDF](#) [PDF \(trim\)](#) [PDF \(defs\)](#) [PDF \(thm\)](#) [PDF \(thm+proof\)](#) [TEX](#) [Example Sheet](#) [Official Notes](#)

Groups (2014, J. Goedecke)

[HTML](#) [PDF](#) [PDF \(trim\)](#) [PDF \(defs\)](#) [PDF \(thm\)](#) [PDF \(thm+proof\)](#) [TEX](#) [Example Sheet](#) [Official Notes](#)

Numbers and Sets (2014, A. G. Thomason)

[HTML](#) [PDF](#) [PDF \(trim\)](#) [PDF \(defs\)](#) [PDF \(thm\)](#) [PDF \(thm+proof\)](#) [TEX](#) [Example Sheet](#) [Official Notes](#)

Vectors and Matrices (2014, N. Peake)

[HTML](#) [PDF](#) [PDF \(trim\)](#) [PDF \(defs\)](#) [PDF \(thm\)](#) [PDF \(thm+proof\)](#) [TEX](#) [Example Sheet](#) [Official Notes](#)

笔记工具 - LaTeX

TeX Live + TeX Studio + Mathpix

安装指南

The screenshot shows the TeX Studio interface with the following content:

Source Code (Left Panel):

```

40
41 \section{随机趋势模型 (Random Trend Model)}
42
43 \boxtext[参考]{0.7}{
44   Wooldridge (2010, pp. 374-381),
45
46   11.7 Models with Individual-Specific Slopes
47
48   11.7.1 Random Trend Model
49 }
50 % (Wooldridge 2010, pp. 374-381)
51 % 11.7 Models with Individual-Specific Slopes
52 % 11.7.1 Random Trend Model
53
54 Consider the following extension of the standard
55 unobserved effects model:
56 \begin{equation}\label{eq:panel-Woold2010-11-64}
57   y_{i t}=c_{i}+g_{i} t+\mathbf{x}_{i t}
58   \boldsymbol{\beta}+u_{i t}, \quad t=1,2, \dots, T
59 \end{equation}
60 This is sometimes called a \textbf{random trend model},
61 as each individual, firm, city, and so on is allowed to
62 have its own time trend. The individual-specific trend
  
```

Rendered Output (Right Panel):

7.2 随机趋势模型 (Random Trend Model)

参考

Wooldridge (2010, pp. 374-381),
11.7 Models with Individual-Specific Slopes
11.7.1 Random Trend Model

Consider the following extension of the standard unobserved effects model:

$$y_{it} = c_i + g_i t + \mathbf{x}_{it} \boldsymbol{\beta} + u_{it}, \quad t = 1, 2, \dots, T \quad (7-1)$$

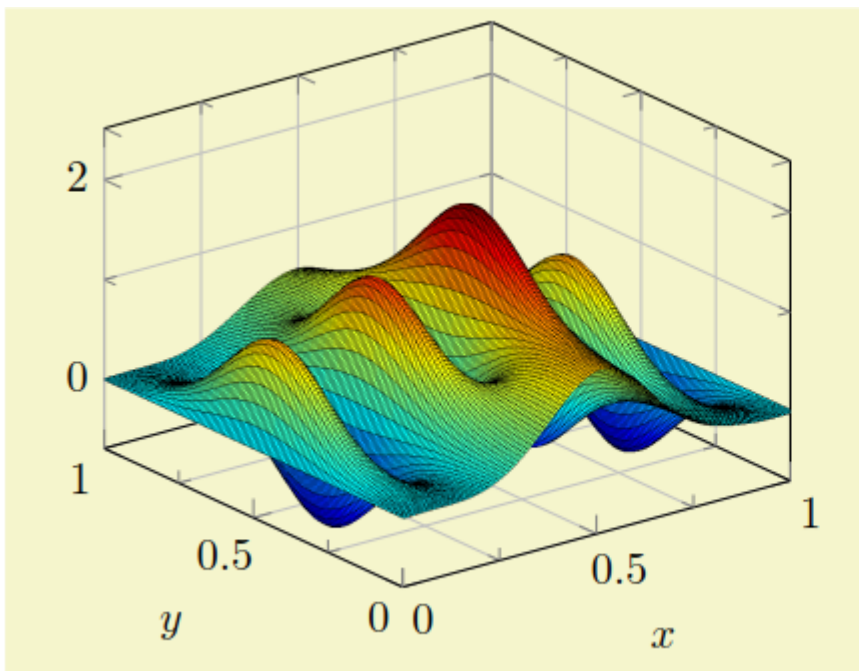
This is sometimes called a **random trend model**, as each individual, firm, city, and so on is allowed to have its own time trend. The individual-specific trend is an additional source of heterogeneity. If y_{it} is the natural log of a variable, as is often the case in economic studies, then g_i is (roughly) the average growth rate over a period (holding the explanatory variables fixed). Then equation (7-1) is referred to a **random growth model**; see, for example, Heckman and Hotz (1989).

In many applications of equation (11.64) we want to allow (c_i, g_i) to be arbitrarily correlated

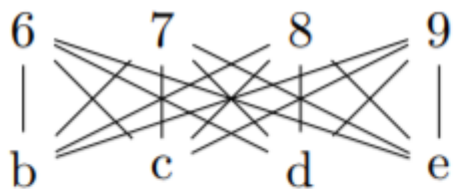
Help: `texdoc pdfplots` (pp.68)

4.3.8 Keys To Configure Plot Graphics

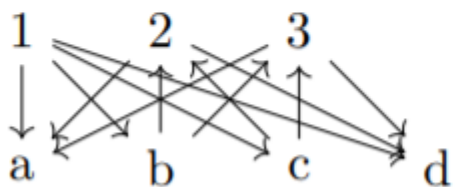
`\usepackage{tikz}` [tikz Examples](#), [tikz 中文手册-1000多页](#)



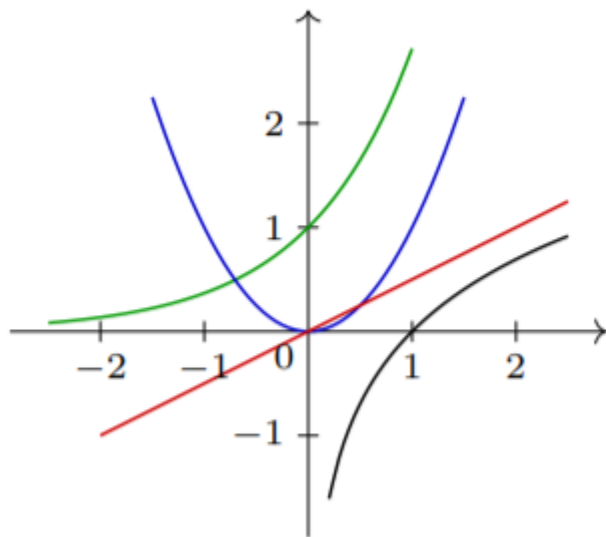
```
% Preamble: \pgfplotsset{width=7cm,compat=1.17}
\begin{tikzpicture}
\begin{axis}[
  grid=both,minor tick num=1,
  xlabel=$x$,ylabel=$y$,
]
  \addplot3 graphics [
    points={% important
      (0,1,0) => (0,207-112)
      (1,0,0) => (446,207-133)
      (0.5546,0.5042,1.825) => (236,207)
      (0,0,0) => (194,207-202)
    },
  ] {plotdata/plotgraphics3dsurf.png};
\end{axis}
\end{tikzpicture}
```



```
\tikz \graph [branch right, grow down]
{ subgraph K_nm [V={6,...,9}, W={b,...,e}] };
```



```
\tikz \graph [simple, branch right, grow down]
{
  subgraph K_nm [V={1,2,3}, W={a,b,c,d}, ->];
  subgraph K_nm [V={2,3}, W={b,c}, <-];
};
```

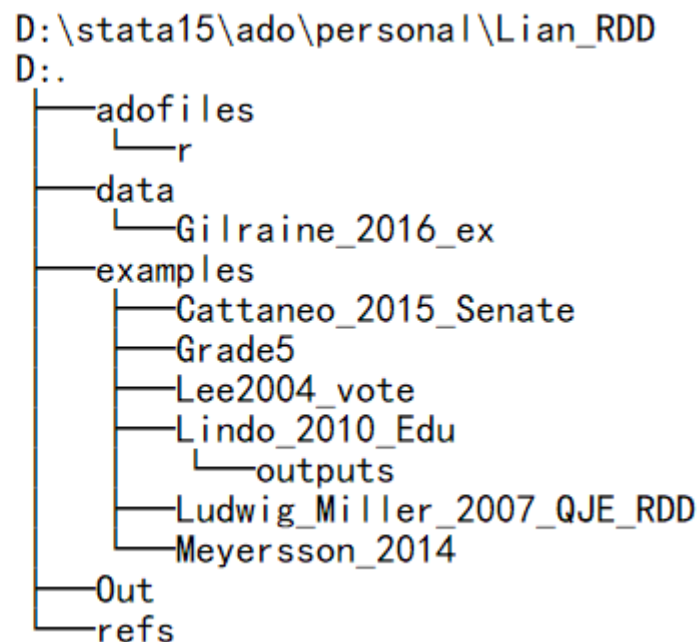
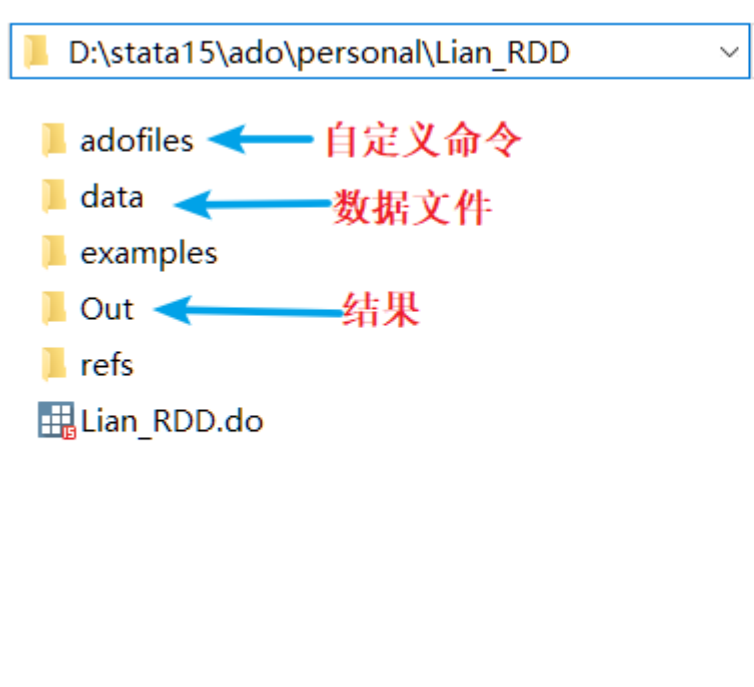


```
\tikz \datavisualization [
  school book axes, all axes={unit length=7.5mm},
  visualize as smooth line/.list={log, lin, squared, exp},
  style sheet=strong colors]
data group {function classes};
```

习惯：可重现

可重复性研究：如何保证你的研究结果可重现？

- 文件夹：一篇文章一个文件夹；设定根路径； `readme.txt`
- 代码： `version 16`、格式、 `help net` → `net set`
- 可重复 → 效率 see [论文复现-28篇](#), `readme.txt` !



```
1 #d ;
2 twoway (histogram x , color( red*0.8%40))
3 (histogram x_tr , color( blue*0.5%40))
4 (histogram x_cen, color( green*0.8%40))
5
6 legend(label(1 "x")
7 label(2 "x_tr")
8 label(3 "x_cen")
9 ring(0) position(3) row(3)) ;
10 #d cr
```

```
1 #d ;
2 twoway (histogram x , color( red*0.8%40))
3 (histogram x_tr , color( blue*0.5%40))
4 (histogram x_cen, color( green*0.8%40))
5
6 legend(label(1 "x")
7 label(2 "x_tr")
8 label(3 "x_cen")
9 ring(0) position(3) row(3)) ;
10 #d cr
```

2. 与老师 (同行) 们的讨论

- 导师：导 v.s 教
 - 测试学生的潜力： `lianxh` , `ihelp` , `cnbeta` , `xtsmooth` , `ttsf`
 - 养成写作的习惯：推文、讲义 (Econometrica)
- 一个导师的自我修炼：
 - Stata + Python (R) +
 - 用 Markdown 或 LaTeX 写讲义
 - 看文献 + 看代码

3. 对 Stata 公司的一点建议

- 简单对比
 - [Stata - SSC](#) (3605 items [Top 25](#))
 - Github ? ? [github-Stata](#) (1,710 repositories)
 - [R packages](#) (18038 packages) | [R github](#) (Repositories 21.7k)
 - [Python - pypi.org](#) (322,106 projects)
- 谁在用 Stata?

- 建议

- 一个易于分享、规范的社区 → packages (不同于 FAQs, Stata List)
- adofiles 发布：模板、规则
 - [Stata Journal](#) → Stata Tips, Stata Blogs → **Stata Packages**
- Markdown 支持：其他编辑器中的语法高亮
- LaTeX 支持：
 - 已有： `sjlatex`, `stata.sty`, `esttab`, `outreg2`, `texdoc`
 - 可以有： `stata.sty` 收录到 TeXLive 发行版中；语法高亮
- R 接口： `rcall`
- Cheaper ^_^

Thanks