

---

# Stata 学术论文专题 视频教程

---

使用说明和 **Stata do-files**

连玉君

(中山大学 岭南学院)

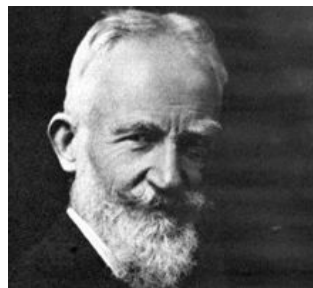
E-mail: [arlionn@163.com](mailto:arlionn@163.com)

Homepage: <http://goo.gl/tRXba>

Weibo: <http://weibo.com/arlionn>

If you have an apple and I have an apple and if we exchange these apples then you and I will still each have one apple. But if you have an idea and I have an idea and we exchange these ideas, then each of us will have two ideas.

—— *George Bernard Shaw*



**重要提示：** 这些资料仅限于本次培训内部使用，请勿散布于网络。

© 2012 连玉君 版权所有

---

## 目 录

1. 课程资料使用说明.....	III
1.1 stata11 软件的基本设定.....	III
1.2 课件的使用方法.....	IV
Q1. 课程配套的资料有哪些? .....	IV
Q2. 如何打开课堂上使用的 do 文档? .....	VI
Q3. 部分命令无法识别或无法执行怎么办? .....	VI
Q4. 如何查看每次执行 Stata 的 log 文件? .....	VII
Q5. 如何正确显示中文字符? .....	VII
Q6. 如何查看帮助文件的 PDF 文件? .....	VIII
Q7. 如何使用 Stata 官方提供的范例数据? .....	VIII
Q8. 使用其他版本的 Stata 软件是否可以? .....	IX
1.3 讨论和建议.....	IX
 2. 研讨论文的 Stata do-files.....	X
 <b>Chang_2009_JCF.do</b> .....	1
Chang, E. C., S. M. L. Wong, 2009, Governance with multiple objectives: Evidence from top executive turnover in China, <i>Journal of Corporate Finance</i> , 15 (2): 230-244.	
<b>Cleary_1999.do</b> .....	22
Cleary, S., 1999, The Relationship between Firm Investment and Financial Status, <i>Journal of Finance</i> , 54 (2): 673-692.	
<b>Faulkender_2006.do</b> .....	33
Faulkender, M., R. Wang, 2006, Corporate Financial Policy and the Value of Cash, <i>Journal of Finance</i> , 61 (4): 1957-1990.	
<b>Fazzari_1988.do</b> .....	44
Fazzari, S., R. Hubbard, B. Petersen, A. Blinder, J. Poterba, 1988, Financing Constraints and Corporate Investment, <i>Brookings Papers on Economic Activity</i> , 1988 (1): 141-206.	
<b>Flannery_2006.do</b> .....	53
Flannery, M. J., K. P. Rangan, 2006, Partial adjustment toward target capital structures, <i>Journal of Financial Economics</i> , 79 (3): 469-506.	
<b>Hansen_1999.do</b> .....	73
Hansen, B., 1999, Threshold Effects in Non-dynamic Panels: Estimation, Testing, and Inference, <i>Journal of Econometrics</i> , 93 (2): 345-368.	
<b>Kumb_2009_SFA.do</b> .....	85
Kumbhakar, S., F. Christopher, 2009, The effects of bargaining on market outcomes: Evidence from buyer and seller specific estimates, <i>Journal of Productivity Analysis</i> , 31 (1): 1-14.	
<b>Lian_2009_SFA.do</b> .....	90
连玉君, 苏治, 2009, 融资约束、不确定性与上市公司投资效率, 管理评论, 1: 19-26.	

---

<b>Lian_2011_PSM.do</b> .....	<b>105</b>
Lian, Y., Z. Su, Y. Gu, 2011, Evaluating the effects of equity incentives using PSM: Evidence from China, <i>Frontiers of Business Research in China</i> , 5 (2): 266-290.	
<b>Lian_NFJJ2007.do</b> .....	<b>130</b>
连玉君, 钟经樊, 2007, 中国上市公司资本结构动态调整机制研究, 南方经济, (1): 23-38.	
<b>Love_2006.do</b> .....	<b>141</b>
Love, I., L. Zicchino, 2006, Financial development and dynamic investment behavior: Evidence from panel VAR, <i>Quarterly Review of Economics and Finance</i> , 46 (2): 190-210.	
<b>Lu_Lian_JYJY2011.do</b> .....	<b>155</b>
卢洪友, 连玉君, 卢盛峰, 2011, 中国医疗服务市场中的信息不对称程度测算, 经济研究, (4): 94-106.	
<b>Opler_1999.do</b> .....	<b>162</b>
Opler, T., L. Pinkowitz, R. Stulz, R. Williamson, 1999, The Determinants and Implications of Corporate Cash Holdings, <i>Journal of Financial Economics</i> , 52(1): 3-46.	
<b>Ye_Lian_JJYJ2012.do</b> .....	<b>179</b>
叶德珠, 连玉君, 黄有光, 李东辉, 2012, 消费文化、认知偏差与消费行为偏差, 经济研究, (2): 80-92.	
<b>z01_Logit.do</b> .....	<b>188</b>
连玉君, 钟经樊, 2012, 计量分析与 Stata 应用(书稿), 第十五章: Logit 模型.	
<b>z02_Event_Study.do</b> .....	<b>195</b>
连玉君, 钟经樊, 2012, 计量分析与 Stata 应用(书稿), 第二十章: 事件研究法.	

## 1. 课程资料使用说明

### 1.1 stata11 软件的基本设定

(1) stata11 软件的放置。将 stata11.rar 解压后放置于 D 盘根目录，即 D:\stata11（注意：是 D:\stata11，而不是 D:\stata11\stata11，另外，stata11 是小写）。如果你的电脑中没有 D 盘，请阅读第 (5) 条说明。

(2) 打开 stata11：双击图 1 中带有蓝色阴影的图标即可。<sup>1</sup>

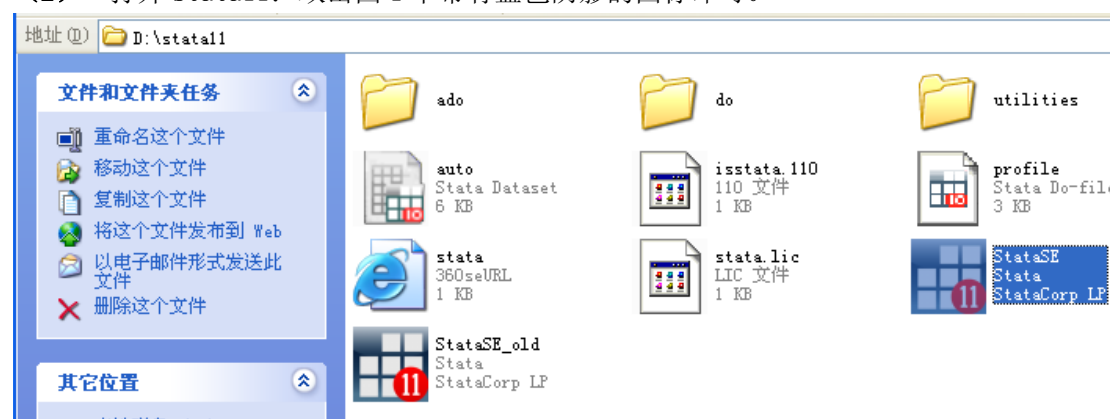


图 1 stata11 放置界面

(3) 关于更新

A、首次打开 stata11，会弹出对话框，询问你是否需要更新。如果不需要更新，请按下图进行设定：

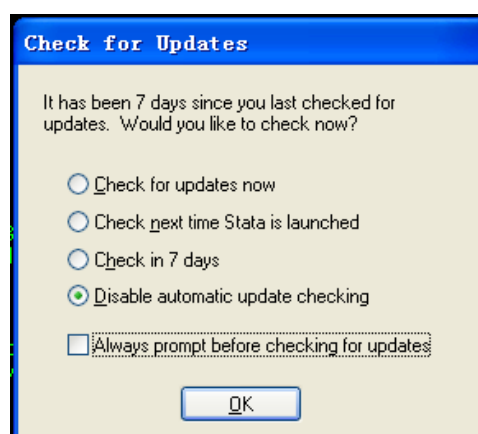


图 2 首次打开 stata11 时的更新设定

B、（为了节省时间，这一步可以跳过，对于一般用户而言，更新与否不会影响你的使用效果）如需把 stata11 更新到最新版本，请在 command 窗口中输入 `update all` 命令。

<sup>1</sup> 你也可以右击这个图标，然后选择“附加到开始菜单”。这样 stata11 的 logo 就会出现在开始菜单中，每次启动 stata11 就只需从“开始”菜单中单击这个 logo 即可。

更新过程大概需要几分钟到几十分钟不等。特别注意的是，更新完毕后，你一定要输入如下命令，才算是真正完成了 stata 的更新：`update swap`。这个命令的作用是用新下载的文件覆盖旧文件。忘记这一步，往往会导致你的 stata 丢失变量或出现一些奇怪的现象。

- (4) 关于 profile.do 文件。每次启动 stata11 时，它会自动执行 D:\stata11 文件夹下的 profile.do 文件，该文件中包含了一系列命令，用于设定 stata 所占用的内存，各种文件路径的位置等信息。如果你是按照上述要求放置 stata11 文件的，那么启动 stata11 后，屏幕上应该显示如下信息（否则你要检查文件名的拼写是否正确）：

```
running D:\stata11\profile.do ...
```

如果想更改 profile.do 文件中的设定，可输入如下命令打开该文件进行编辑，完成后保存，在下次启动 stata11 时，相应的设定即可生效：

```
doedit D:\stata11\profile.do
```

- (5) 如果你的计算机分区中没有 D 盘，你也可以把 stata11 放置于其他盘符下，但需要做一些微小的调整。这里以 F 盘为例，也可以是 C，H，K 等其他盘

①. 将 stata11 解压后放置于 F 盘下，即 F:\stata10。注意：stata 是小写。

②. 修改 profile.do 文件，具体方法为：

Step1: 输入 `doedit F:\stata11\profile.do` 命令，打开 profile.do 文件；

Step2: 将第六行中的 `local D "D"` 命令修改为 `local D "F"`；

Step3: 保存 profile.do 文件，退出 stata11，然后重启即可。

③. 若上述设定无误，则在重新打开 stata11 后，屏幕第一行会显示如下信息：

```
running F:\stata11\profile.do ...
```

同时，输入 `sysdir` 命令，屏幕上会呈现如下信息（此时你能正确使用外部命令）：

```
. sysdir
      Stata:  F:\stata11\
  UPDATES:  F:\stata11\ado\updates\  （存放 stata 自带命令的升级文件）
      BASE:  F:\stata11\ado\base\    （存放 stata 自带命令的原始文件）
      SITE:  F:\stata11\ado\site\
      PLUS:  F:\stata11\ado\plus\    （存放用户下载的外部命令）
  PERSONAL:  F:\stata11\ado\personal\ （存放用户的个人文件）
```

**特别提示：**不按照上述方法设定你的 stata11 文件路径，可能导致部分命令（尤其是外部命令和自编命令）无法识别，常见的错误信息如下（参见 2.2 小节 Q3）：

```
unrecognized command: xxxx
```

## 1.2 课件的使用方法

### Q1. 课程配套的资料有哪些？

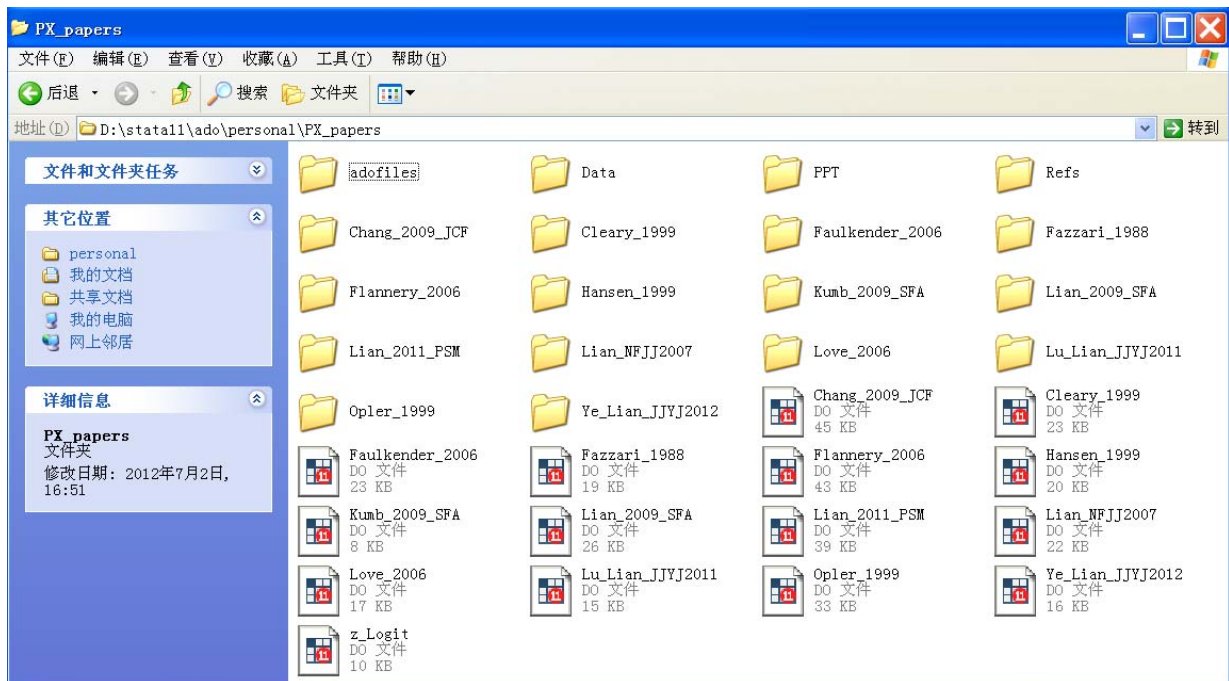
A: 资料清单如表 1 所示，论文的 Stata 实现过程和相关的数 据、程序文件统一存放于 PX\_papers 文件夹下，详情参见图 1。

表 1 Stata 学术论文视频教程资料清单

资料名称	简要说明	使用方法
stata11.rar	stata11 软件包	参见第 III 页, 1.1 小节
PX_papers.rar	论文的 Stata 实现过程相关文件: do-file、数据、程序等	参见 Q2 的解答
	PX_papers\PPT (子文件夹)	课程中使用 PPT 文件
	PX_papers\Refs (子文件夹)	课程中涉及的论文的 PDF 文件
Stata_学术论文_论文集	论文集	主讲论文的合集(建议打印)
Stata_学术论文_使用说明	课件使用方法 所有论文的 Stata do-files	本文件

说明: (1) PDF 讲义为彩色, 若需打印, 请设置“将彩色打印为黑白”或“print color as black”, 以便获得最佳的打印效果; (2) 观看视频教程时, 请使用耳机, 获得最佳音质的同时, 也避免影响他人。

图 1: PX\_papers 文件夹下的内容



说明:

- (1) adofiles (adofiles 文件夹): 存放了课程中使用的自编程序, 例如双边随机边界分析估计程序 SFA2tier.ado, 面板门槛模型估计程序 xtthres.ado 等;
- (2) Data (Data 文件夹): 存放了课程中使用的一些公共数据, 包括 GTA2008.dta (包含了中国上市公司 1991-2008 年的财务变量、公司治理变量等信息)、China\_MI.dta (中国市场化进程指数数据, 存放于 Data\mkt\_index 文件夹下);
- (3) PPT (PPT 文件夹): 课程中每篇文章对应的 PPT 文件;
- (4) Refs (Refs 文件夹): 课程中每篇文章的 PDF 原文, 以及相关文献的 PDF 原文;
- (5) 其他文件配对出现: 例如 Chang\_2009\_JCF.do 文件 ( Chang\_2009\_JCF) 为 Chang and Wong (2009) 这篇文章的 Stata 实现过程的 do-file, 而 Chang\_2009\_JCF 文件夹 ( Chang\_2009\_JCF) 则存放了该文实证分析过程中使用的数据, 产生的统计和回归表格等文件。

## Q2. 如何打开课堂上使用的 do 文档?

A: 请将 PX\_papers.rar 压缩包解压到 D:\stata11\ado\personal 文件夹中, 即 D:\stata11\ado\personal\PX\_papers。若希望练习 Chang and Wong (2009) 这篇文章中的 Stata 实现过程, 可依次执行如下命令:


**Step1:** 在 Stata 命令窗口中输入 `cd D:\stata11\ado\personal\PX_papers` 命令, 定义当前工作路径 (会显示在 Stata 屏幕左下角);


**Step2:** 输入 `doedit Chang_2009_JCF` 命令, 即可打开 Chang\_2009\_JCF.do 文件。

上述过程可以用一条命令完成:

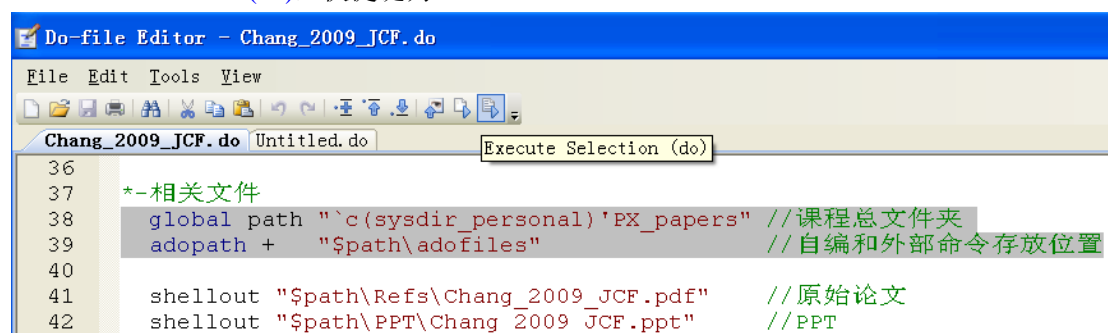
```
doedit D:\stata11\ado\personal\PX_papers\Chang_2009_JCF
```

当然, 我们也可以通过点击菜单的方式完成上述操作, 步骤如下:

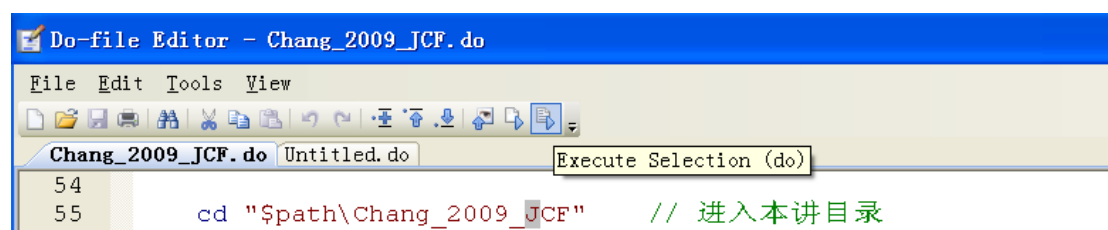
**Step1:** 在 Stata 主菜单 (第二行) 中点击 “New do-file Editor” 图标 ;

**Step2:** 在第一步中弹出的 “Do-file Editor” 窗口中点击 “Open” 图标 , 然后到 D:\stata11\ado\personal\PX\_papers 文件夹下, 双击 “Chang\_2009\_JCF.do” 即可。

B: 在练习之前, 请先执行如下命令, 以便进入本讲所在目录, 本讲中使用的所有数据文件、相关文档, 以及课程中所使用的外部命令的存放路径都通过这些命令来设定。执行命令的方法如下: 选中下图中第 38-39 行的命令, 点击菜单条中第二行中带有蓝色阴影的按钮 (Execute Selection (do), 快捷键为 Ctrl+D)。



进入本讲目录, 需要进一步执行第 55 行的命令:



需要说明的是, 在 do-file 中执行某行命令时, 无需选中该行的所有命令, 只需选中一个以上的字符即可。

## Q3. 部分命令无法识别或无法执行怎么办?

A: 当屏幕上显示红色错误提示信息 “unrecognized command: xxx” 时, 意味着命令 “xxx” 无法识别。课件中涉及到的命令, 主要包括三类: 第一类是 stata11 自带的, 统一存放于 “D:\stata11\ado\base” 文件夹下,<sup>2</sup> 这些命令通常不会出现无法识别的问题; 第二类

<sup>2</sup> 若您的 stata 更新过, 则有部分被更新的命令会统一存放于 “D:\stata11\ado\updates” 文件夹下。

是我自行下载的外部命令，统一存放于“D:\stata11\ado\plus”文件夹中，若您已经严格按照第 2.1 小节的方法设定了 stata 的文件路径，则不会出现命令无法识别的问题；<sup>3</sup>第三类是我自己编写的命令，统一存放于“D:\stata11\ado\personal\PX\_papers\adofiles”文件夹中，要使这些命令正确执行，需要使用 `adopath +` 命令将上述路径加入 stata11 的搜索范围内，详情参见 Q2 的解答。

这里需要简要说明一下 stata 执行命令的原理。当我们输入 `describe` 命令时，stata 需要找到并执行一个名称为 `describe.ado` 的程序文件（ado 文件）。需要注意的是，为了提高执行速度，stata 仅在几个预先约定的文件夹中搜索该文件，一旦找到，就会将该文件中的内容读入内存，进而执行其中的命令，并将结果呈现在屏幕上。搜索的优先顺序为：`base`→`updates`→`plus`→`other path`。这些所谓的“预先约定”的文件夹可以使用 `adopath` 命令列示出来，例如：

```
. adopath
[1] (UPDATES) "D:\stata11\ado\updates/"
[2] (BASE)    "D:\stata11\ado\base/"
[3] (SITE)    "D:\stata11\ado\site/"
[4]          "."
[5] (PERSONAL) "D:\stata11\ado\personal/"
[6] (PLUS)    "D:\stata11\ado\plus/"
```

因此，当您要执行的命令存放于其他文件夹中时，<sup>4</sup>就会出现“unrecognized”的提示信息。此时，有两种解决方法：

其一，采用 `adopath +` 命令把这个文件路径变成“预先约定”路径，例如：

```
adopath + D:\stata11\ado\personal\PX_papers\adofiles
```

其二，把我们自己编写的命令放入“D:\stata11\ado\base\”文件夹中，<sup>5</sup>此时该命令的身份与 stata 自带命令么有任何差异，总是会被 stata 优先搜索到。这是一种一劳永逸的办法，缺陷是不易管理和共享这些自己编写的命令（因为存放的过于分散）。

#### Q4. 如何查看每次执行 Stata 的 log 文件？

A: 我在 `profile.do` 文件中进行了设定，每次启动 Stata11 后，屏幕上呈现的所有结果都会被自动记录到一个日志文件中。这些日志文件被自动保存在 D:\stata11\do 文件夹下，文件名称为打开 Stata 时的时间，例如，名称为 `s22Apr2010111857.txt` 的日志文件是我在 2010 年 4 月 22 日 11 点 18 分 57 秒打开 Stata11 时自动生成的。

#### Q5. 如何正确显示中文字符？

A: 在默认设定下，首次打开 Stata11 时，所有的窗口都是白色的，此时 Stata11 无法正常显示中文字符（多为乱码）。解决方法很简单：

① 要保证 [Results] 窗口中的中文字符能够正常显示，只需做如下设定：

依次点击菜单 Edit → Preference → General Preference → Results Colors → Color scheme，在

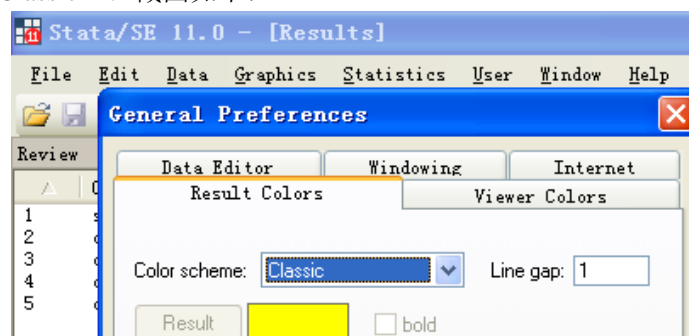
<sup>3</sup> 您可以使用 `findit` 命令查找新的外部命令，例如 `findit panel data` 可以搜索出很多有关 Panel Data 模型的命令，按照指引可以点击下载到 plus 文件夹中。

<sup>4</sup> 如，本课程中使用的所以自编命令都存放于“D:\stata11\ado\personal\PX\_papers\adofiles”。

<sup>5</sup> 建议将自编命令放入 base 文件夹下以该命令首字母命令的子文件夹中。例如，若自编命令为 `myprog.ado`，则应放入 D:\stata11\ado\base\m 文件夹中。

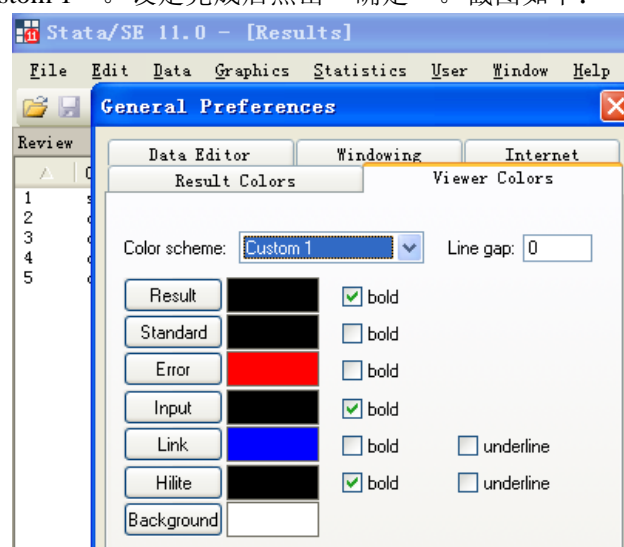


下拉菜单中选择“Classic”。截图如下：



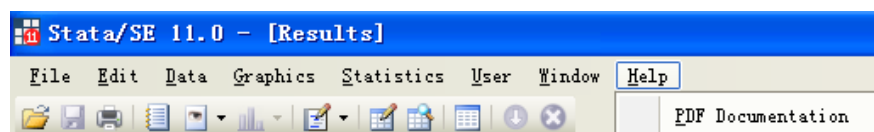
② 要保证帮助文件能够以合适的字体显示，只需作如下设定即可：

依次点击菜单 Edit → Preference → General Preference → **V**iewer Colors → Color scheme，在下拉菜单中选择“Custom 1”。设定完成后点击“确定”。截图如下：



#### Q6. 如何查看帮助文件的 PDF 文件？

A: Stata11 配备了完整的 PDF 格式的使用手册，统一存放于“D:\stata11\utilities”文件夹下。你可以直接进入该文件夹查看相应的 PDF 文件，亦可在 Stata 内部直接打开相应的 PDF 文件。要打开完整的 PDF 使用手册，依次点击菜单 Help → PDF Documentation 即可：



若想查看特定命令的 PDF 使用说明，可以先在命令窗口中输入该命令，如“help regress”，浏览该帮助文件的最下方，点击下图中的 [R] regress 即可打开 regress 的 PDF 帮助文件。


Also see

Manual: [\[R\] regress](#)

#### Q7. 如何使用 Stata 官方提供的范例数据？

A: 多数情况下，教程中提供的资料已经足够大家练习所需了。当然，为了满足个别学员的需求，本教程还进一步提供了 Stata 官方范例数据，统一存放于 D:\stata11\ado\Example

---

文件夹中。例如，你可以输入 `use "D:\stata11\ado\Examples\XTFiles\abdata.dta", clear` 命令导入范例数据 `abdata.dta`。亦可点击 Stata 主界面中的  按钮（快捷键为 Ctrl+o），到上述目录下选择你想要练习的数据文件。

**Q8. 使用其他版本的 Stata 软件是否可以？**

**A:** Stata 视频教程中涉及的多数命令都是基本命令，对用户所使用的 Stata 版本并不敏感。如果您正在使用 stata12，那么上文提到的设置方法都依然适用，唯一需要改动的就是把前文中的“stata11”统一修改成“stata12”。对于正在使用 stata10 或更低版本的用户而言，视频教程中涉及的部分命令可能无法执行，此时，建议您使用课程附带的 stata11.rar 软件。由于该软件包无需安装，你可以把它放在任何一个盘符下，如 D 盘、E 盘或 F 盘，相关设定方法参见 2.1 小节第 (5) 点的说明。

---


## 1.3 讨论和建议

---

我们为大家提供了多种讨论和沟通的方式，我本人非常希望 Stata 网络课堂的学员们能够形成一个协作联盟，大家互通有无。

(1) 人大经济论坛 Stata 专版：<http://www.pinggu.org/bbs/index.asp?boardid=67>。

(2) 人大经济论坛“统计软件培训班 VIP 答疑区”。这里汇集了大量 stata 应用的实例和问题：<http://www.pinggu.org/bbs/forum-114-1.html>。大家在学习过程中如若遇到问题，可以在这里集中讨论。

(3) 连玉君的博客和新浪微博：我的博客（<http://blog.cnfol.com/arlion>）中提供了大量 Stata 资源；通过我的新浪微博（<http://weibo.com/arlionn>, ），大家可以实现及时沟通。

(4) 在网络不畅或特殊情况下，大家也可以发电子邮件给我（[arlionn@163.com](mailto:arlionn@163.com)）。

---

## 2. 研讨论文的 Stata do-files

---

Chang_2009_JCF.do .....	1
Cleary_1999.do .....	22
Faulkender_2006.do .....	33
Fazzari_1988.do .....	44
Flannery_2006.do .....	53
Hansen_1999.do .....	73
Kumb_2009_SFA.do .....	85
Lian_2009_SFA.do .....	90
Lian_2011_PSM.do .....	105
Lian_NFJJ2007.do .....	130
Love_2006.do .....	141
Lu_Lian_JYJY2011.do .....	155
Opler_1999.do .....	162
Ye_Lian_JJYJ2012.do .....	179
z01_Logit.do .....	188
z02_Event_Study.do .....	195

```

1
2  *      -----
3  *      -----
4  *
5  *      Stata 学术论文专题
6  *
7  *      -----
8  *      -----
9
10
11  *      主讲人：连玉君 副教授
12  *
13  *      单 位：中山大学岭南学院金融系
14  *      电 邮：arlionn@163.com
15  *      博 客：http://blog.cnfol.com/arlionn
16  *      主 页：http://goo.gl/tRXba
17  *      微 博：http://weibo.com/arlionn
18
19
20  *-----
21  * -Chang&Wong,2009,JCF -
22  *   Chang, E. C., Wong, S. M. L., 2009.
23  *   Governance with multiple objectives:
24  *   Evidence from top executive turnover in china.
25  *   Journal of Corporate Finance, 15(2): 230-244.
26  *-----
27
28  *-主要方法
29  *   Logit models: help logit
30  *   help tabulate
31  *   help merge
32
33  *-涉及的主题
34  *   公司治理：经理人激励；经理人更换
35  *   预算软约束
36
37  *-相关文件
38  global path "`c(sysdir_personal)'PX_papers" //课程总文件夹
39  adopath + "$path\adofiles" //自编和外部命令存放位置
40
41  shellout "$path\Refs\Chang_2009_JCF.pdf" //原始论文
42  shellout "$path\PPT\Chang_2009_JCF.ppt" //PPT
43
44
45
46
47
48
49
50
51  *=====
52  *   Part I  CEO or Chairman 变更数据的处理
53  *=====
54
55  cd "$path\Chang_2009_JCF" // 进入本讲目录
56
57  clear all
58  set memory 80m // 设定 Stata 占用的内存空间
59
60  use TurnOver_orig.dta, clear //调入变更数据
61
62  rename stkcd id
63  label var id "证券代码"
64  label var annodt "公告日期"
65  label var repttype "公告类型"
66  label var chgdt "变更日期"
67  label var position "变更职位"
68  label var changtyp "变更类型"

```

```

69     label var    name      "姓名"
70     label var    edca      "教育背景"
71     label var    prti      "职称"
72     label var    dimreas   "离职原因"
73     label var    dimage    "离职年龄"
74     renam years  dyears
75     label var    dyears    "任职年限"
76     label var    entele    "继任来源"
77     label var    ifagent   "继任后是否代理"
78     label var    cirplurm   "继任后是否兼任董事长与总经理"
79     label var    indbocha   "继任后独立性"
80     label var    selected   "候选选择"
81
82     gen year = real(substr(chgdt,1,4)) //从变更日期中提取年度数据
83     order id year
84
85     *-处理部分变量的缺漏值
86     replace dimreas=99 if dimreas==.    // 离任原因缺失的归类为"其他"
87
88     *-----
89     *-1.1 定义数字-文字对应表 label values
90     *-----
91
92     *-变更职位
93     label define position_a 1 "董事长" 2 "总经理"
94     label value position position_a
95     *-变更类型
96     label define changtyp_a 1 "离任" 2 "继任"
97     label value changtyp changtyp_a
98     *-变更者的教育程度
99     label define edca_a 1 "中专及以下" 2 "大专" ///
100     3 "本科" 4 "硕士" 5 "博士" 6 "其它"
101     label value edca edca_a
102     *-变更原因
103     label define dimreas_a ///
104     1 "工作调动" 2 "退休" 3 "任期届满" 4 "控股权变动" 5 "辞职" ///
105     6 "解聘" 7 "健康原因" 8 "个人" 9 "完善公司法人治理结构" ///
106     10 "涉案" 99 "其他" 12 "结束代理"
107     label value dimreas dimreas_a
108     *-继任来源
109     label define entele_a 1 "内部" 2 "外部"
110     label value entele entele_a
111     *-是否代理
112     label define ifagent_a 1 "是代理" 2 "不是代理"
113     label value ifagent ifagent_a
114     *-是否兼任总经理或董事长
115     label define cirplurm_a 1 "不兼任" 2 "兼任"
116     label value cirplurm cirplurm_a
117     *-遴选方式
118     label define selected_a ///
119     1 "控股股东产生董事长" 2 "控股股东产生总经理" 3 "职业经理人"
120     label value selected selected_a
121
122     gen dimreas_E = dimreas    // 英文标签
123     #delimit ;
124     label define dimreas_E_a
125     1 "Change of job"
126     2 "Retirement"
127     3 "Contract expiration"
128     4 "Change in controlling shareholders"
129     5 "Resignation"
130     6 "Dismissal"
131     7 "Health"
132     8 "Personal reasons"
133     9 "Corporate governance reform"
134     10 "Legal disputes"
135     12 "Completion of acting duties"
136     99 "No reason given"

```

```

137         ;
138         #delimit cr
139         label value dimreas_E dimreas_E_a
140
141         *-列示所有的 数字-文字对应表
142         labelbook
143
144
145         *-----
146         *-1.2 数据的进一步处理
147         *-----
148
149         compress                                //精简数据的存储类型
150         keep if year>=2000&year<=2007          // 仅保留 2000-2007 年样本
151         duplicates tag name, gen(rep_names) //标注第一次出现的人名
152         keep if changtyp==1 // 仅保留[离任样本],日后可以分析继任样本
153
154
155         *-----
156         *-1.3 一些基本统计分析 (Not reported in Chang&Wong's Paper)
157         *-----
158
159         *-(1)-各个年度上的总经理和董事长变更情况
160         label list position_a
161         tab year position
162
163         *-(2)-一年中只有总经理变更的公司数量
164         preserve
165             keep if position==2 // CEO 变更样本
166             tab year position // 允许一年多次变更
167             duplicates drop id year, force
168             tab year position // 仅考虑每年的第一次变更
169         restore
170
171         *-(3)-一年中只有董事长变更的公司数量
172         preserve
173             keep if position==1 // 董事长变更样本
174             tab year position // 允许一年多次变更
175             duplicates drop id year, force
176             tab year position // 仅考虑每年的第一次变更
177         restore
178
179
180         *-----
181         *-仅保留 CEO 变更的样本
182         *-----
183         keep if position==2 // CEO变更样本
184
185
186         *-----
187         *-1.4 产生 Table 1 所需变量
188         *-----
189
190         *-每家公司每年变更的次数
191         dropvars nchg_yr nchg_id
192         bysort id year: egen nchg_id = count(changtyp)
193         label var nchg_id "Duplicate times of Turnovers in one year"
194
195         *-各年度变更的总次数(可重复)
196         bysort year: egen nchg_yr_dup = count(changtyp)
197         label var nchg_yr_dup "Total number of turnovers"
198
199         *-各年度变更的总次数(不可重复)
200         egen unique_chg = tag(id year)
201         bysort year: egen nchg_yr_uniq = count(changtyp) if unique_chg==1
202         label var nchg_yr_uniq "Total number of turnovers after consolidation"
203
204         *-Note: 上述变量将参与后续运算,并计算出 Table 1 中的结果

```

```

205
206
207 *-----
208 *-1.5 Table 2 [ 输出 Table 2 ]
209 *-----
210
211 *-左侧的表格
212 tab dimreas
213 *-输出 Excel(英文)
214 logout, save(Results/Table02_left) excel replace fix(20): ///
215 tab dimreas_E
216
217 *-右侧的表格
218 tab dimreas if nchg_yr_uniq!=.
219 *-输出 Excel(英文)
220 logout, save(Results/Table02_right) excel replace fix(20): ///
221 tab dimreas_E if nchg_yr_uniq!=.
222
223 *-Note: 亦可选中表格后, 右击, 选择"Copy
table", 贴入Excel整理好, 再贴入word即可。
224
225
226 *-----
227 *-1.6 离任年龄分布
228 *-----
229
230 set scheme slmono
231 histogram dimage, discrete freq xtitle("Dismissal Age")
232 graph export Results/dimage_his.wmf, replace // 保存图片
233
234 *-屏幕呈现: 中文
235 tabstat dimage, by(dimreas_E) s(mean sd min max q) format(%3.0f)
236
237 *-输出结果: 英文
238 logout, save(Results/Table03a_age) excel replace fix(20): ///
239 tabstat dimage, by(dimreas_E) s(mean sd min max q) format(%3.0f)
240
241
242 *-----
243 *-1.7 任职年限分布
244 *-----
245
246 histogram dyears, freq xlabel(1(1)15) xtitle("Duration (years)")
247 graph export Results/dyears_his.wmf, replace // 保存图片
248
249 *-屏幕呈现: 中文
250 tabstat dyears, by(dimreas) s(mean sd min max q) format(%3.1f)
251
252 *-输出结果: 英文
253 logout, save(Results/Table03b_tenure) excel replace fix(20): ///
254 tabstat dyears, by(dimreas_E) s(mean sd min max q) format(%3.1f)
255
256
257 *-----
258 *- Consolidate sample (合并一年变更多次的样本)
259 *-----
260 *-后续分析都采用合并样本, 即同年多次 CEO 变更视为一次
261
262 duplicates drop id year, force
263
264
265 *-----
266 *-1.8 定义被迫离职(forced turnover)虚拟变量
267 *-----
268
269 *-----
270 * follow Chang and Wong (2009, JCF, p.234 最后一段)
271 * Two type: forced v.s. unforced

```



```

272
273     label list dimreas_a // 在写作过程中，对照此表进行说明
274
275     dropvars CT_Chang
276     gen CT_Chang = .
277
278     *-Normal Turnover
279     replace CT_Chang=1 if inlist(dimreas,2,3,4,7,9,10,12)
280
281     *-Forced Turnover
282     replace CT_Chang=2 if CT_Chang!=1
283     count if dimreas==2&dimage<55 //退休年龄未满55岁，视为被迫离职
284                                     //参见 Chang&Wong2009, p.235
285     replace CT_Chang=2 if (dimreas==2&dimage<55)
286
287     label define CT_Chang_a 1 "Normal" 2 "Forced"
288     label value CT_Chang CT_Chang_a
289
290     tab CT_Chang // 列表统计
291     tab CT_Chang year // 年度分布
292
293
294     *-----
295     *-一个简单的统计
296
297     tsset id year
298     xtides, p(30) // Why I am so smart? Haha ...
299
300     * 结果：在 2000-2007 区间，
301     *      有 37.48% 的公司只发生过一次 CEO 变更 // 如何看出
302
303
304     *-----
305     *-1.9 保存 CEO 变更资料（随后与财务资料合并）
306
307     save TurnOver_dealed.dta, replace
308
309
310
311
312
313
314
315     *-----
316     *      Part II 财务数据和公司治理数据处理
317     *-----
318     *-由于前面的数据已经保存，可以直接从这里开始
319
320     *-----
321     *-2.1 合并财务数据和样本筛选
322     *-----
323
324     global path "`c(sysdir_personal)'PX_papers" //课程总文件夹
325     cd "$path\Chang_2009_JCF" //进入本讲目录
326     use "$path\data\GTA2008.dta", clear //调入财务数据
327
328     *-----
329     *-IPO year 对于缺漏值，采用样本中的第一年的年份代替
330     gen IPO_yr = real(substr(ipodate,1,4)) //IPO日期
331     bysort id: egen Tmin = min(year)
332     browse id year IPO_yr Tmin //查看资料形态
333     replace IPO_yr = Tmin if IPO_yr==.
334
335     *-----
336     *-合并两份数据(merge with Chair/CEO turnover data)
337     merge 1:1 id year using TurnOver_dealed.dta
338
339     *-剔除金融类企业

```



```

340     drop if sicmen_str == "I"
341
342     *-保留 2000-2007 年以后的数据 (Turnover 数据的取值区间)
343     keep if (year>=2000 & year<=2007)
344
345
346     *-----
347     *-2.2 Table 1          [ 输出 Table 1 ]
348     *-----
349
350     *-历年上市公司数量
351     bysort year: egen nlist = count(id)
352     label var nlist "Number of listed firms"
353
354     *-离职率
355     gen Annual_tr_dup = nchg_yr_dup/nlist
356     label var Annual_tr_dup "Annual turn rate"
357
358     *-离职率(合并同一年的多次CEO变更为一次)
359     gen Annual_tr_uniq = nchg_yr_uniq/nlist
360     label var Annual_tr_dup "Annual turn rate after consolidation"
361
362
363     *-----begin-----
364     preserve // 执行时要同时选中 -preserve- 和 -restore- 之间的部分
365     keep year nlist nchg_yr* Annual*
366     order year nlist nchg_yr_dup Annual_tr_dup nchg_yr_uniq
367     replace Annual_tr_dup = int(Annual_tr_dup*10000)/100 //百分数,保留两位小数
368     replace Annual_tr_uniq= int(Annual_tr_uniq*10000)/100 //百分数,保留两位小数
369     drop if nchg_yr_dup==.
370     duplicates drop year, force
371     list,sep(0) abbreviate(16) //屏幕显示
372     xpose, clear //转置
373     list,sep(0) //转置后的结果, 变量名称丢失了
374     gen vname = "Year"
375     replace vname = "Number of listed firms" in 2
376     replace vname = "Total number of CEO turnovers" in 3
377     replace vname = "Annual turn rate" in 4
378     replace vname = "Total number of CEO turnovers after consolidation" in 5
379     replace vname = "Annual turn rate after consolidation" in 6
380     order vname
381     egen v2000_2007 = rowtotal(v1-v8) //各列求和
382     egen rmean = rmean(v1-v8)
383     replace v2000_2007 = rmean in 4 //第四行和第六行是平均值, 而非加总值
384     replace v2000_2007 = rmean in 6
385     replace v2000_2007 = . in 1
386     drop rmean
387     xmlsave Results/Table01, doctype(excel) replace // 输出 Table 01
388     restore
389     *-----over-----
390
391     shellout "Results/Table01.xml" // 查看结果
392
393
394     *-----
395     *-2.3 指标构建 有些指标是多余的, 便于大家扩展分析
396     *-----
397
398     rename a001000000 TA // 总资产
399     rename a002000000 TD // 总负债
400
401     *-----
402     *-CEO 变更虚拟变量 (不区分变更类型)
403     gen CT_yes = (changtyp!=.)
404     label var CT_yes "1=Change; 0=No Change"
405
406     *-----

```

```

407 * -2.3.1 performance proxy ROA, ROS, IROA, IROS
408
409 * -重新定义行业分类 (Chang&Wong,2009 的分类不合理)
410 * c 保留两位; 其他一位
411 dropvars industry
412 gen industry = sicmen_str
413 replace industry = substr(sicda_str,1,2) if sicmen_str=="C"
414 replace industry = "C3" if industry=="C2"
415 tab industry //观察值的行业分布
416 panels id: tabulate industry //公司的行业分布
417 panels id: tabulate industry year //公司的行业-年度分布
418
419 * -定义业绩指标
420 gen roa = b002000000/TA //净利润/总资产
421 gen roe = b002000000/a003000000 //净利润/所有者权益合计
422 gen ros = b002000000/b001101000 //净利润/主营业务收入
423 tsset id year
424 gen ln_sale = ln(b001101000)
425 gen sale_gr = D.ln_sale //主营业务收入增长率 Firth (2006, p.1302)
426 gen loss = b001300000<0 //营业利润, Firth (2006, p.1302)
427 label var roa "净利润/总资产"
428 label var roe "净利润/所有者权益合计"
429 label var ros "净利润/主营业务收入"
430 label var sale_gr "sale growth"
431 label var loss "1=营业利润为负,0=otherwise"
432
433 * -下面会进一步生成行业调整后的指标
434
435
436 *-----
437 * -2.3.2 Control variables
438
439 * -核心控制变量
440
441 * -Age
442 * -Keypoint: 离职前的年龄要用离职当年的年龄填充
443 dropvars age* t ct* CT_yes_miss ct
444 tsset id year
445 gen CT_yes_miss = 1 if CT_yes==1
446 bysort id: gen ct123 = sum(CT_yes_miss) //累加
447 gen ct3 = ct123-CT_yes
448 browse id year CT_yes CT_yes_miss ct123 ct3
449 bysort id ct3: egen age = min(dimage) //离任年龄
450 replace age = d04 if age==.
451 * 缺漏的部分采用"董事、监事和高级管理人员平均年龄"代替
452 qui tsset id year
453 * 最后一次离职不是发生在 2007 年, 则会产生 age 缺漏
454 replace age = L.age if age==. // 屏幕上显示了修改的样本个数
455 replace age=69 if age>69
456
457 * -Tenure
458 dropvars tenure*
459 bysort id ct3: egen tenure = min(dyears)
460 bysort id: egen tenure_av = mean(tenure)
461 replace tenure = tenure_av if tenure==. // 缺漏值采用公司均值代替
462 bysort industry year: egen tenure_av_sic = mean(tenure)
463 replace tenure = tenure_av_sic if tenure==.
464 br id year dyears chgdt CT_yes CT_yes_miss ct123 ct3 tenure //查验处理过程
465
466 * -Leverage, size, list
467 gen lev = TD/TA // 负债合计/总资产
468 gen size = ln(TA) // ln(总资产)
469 gen list = 2007 - IPO_yr // Chang&Wong(2009,JCF,p.235)
470
471
472 * -股权结构变量
473
474 * 高管报酬 (王克敏, 2007, 管理世界)

```

```

475     gen gpay= ln(d02)           // 前三名高管年薪总额
476
477     * 股权激励
478     clonevar magstk = d17       // 高管持股比例
479
480     * 董事长持股比例
481     clonevar chair_stk = d18
482
483     * 总经理(CEO)持股比例
484     clonevar ceo_stk = d20
485
486     * 高管控制权
487     gen duality = (d07==1) // 总经理与董事长为同一人则取1, 否则取0
488
489     * 股权结构特征
490     rename d27 topone           // 第一大股东持股比例
491     gen topone2 = topone^2
492     rename d25 hhi5            // 前五大股东持股比例平方和
493     rename d24 zi2            // 第一大股东比例/第二大股东持股比例
494     gen top4_1 = (d28+d29+d30+d31)/topone
495                               // 第二大股东-第五大股东持股之和/第一大股东持股比例
496
497     *-董事会、独立董事相关变量
498     gen independent = d09/d08 // 独立董事比例=独立董事总人数/董事会规模
499     qui tsset id year
500     replace independent = F.independent if id==600083&year==2000 //该观察值缺失
501     clonevar boardsize = d08 // 董事会的规模(boardsize)
502     clonevar boardshare = d11 // 董事会持股比例(boardshare)
503
504
505     *-控制类型虚拟变量(分5组)
506     rename finalcontrol Fcontrol
507     labelbook final_lab           // 列示最终控制人类型
508     tab Fcontrol,gen(dum_control)
509     gen private = (Fcontrol==3) // 民营控股
510     gen state = (Fcontrol!=3&Fcontrol!=5) // 国有控股
511     drop if dum_control5==1
512     drop dum_control5
513     egen miss2 = rmiss(dum_control*)
514     drop if miss2!=0
515
516     *-定义 地方政府控股 和 中央政府控股 虚拟变量
517     gen local = inlist(Fcontrol,1,2)
518     gen central = inlist(Fcontrol,4,6)
519     label var local "地方政府控股(地方国资委和地方国有企业)"
520     label var central "中央政府控股(国资委和中央国有企业)"
521
522
523     *-----
524     *-2.3.3 删除缺漏值
525
526     dropvars miss
527     local v1 "roa ros roe list size lev boardsize topone" //age tenure
528     local v2 "independent topone ceo_stk chair_stk duality"
529     order id year CT* dimreas changtyp `v1' `v2'
530     sum id year CT* dimreas changtyp `v1' `v2'
531     egen miss = rmiss(`v1' `v2')
532     drop if miss>0
533
534
535     *-----
536     *-2.3.4 Winsor 主要变量
537
538     local x1 "roa roe ros lev sale_gr"
539     foreach v of varlist `x1'{
540         rename `v' `v'_a
541         winsor `v'_a, gen(`v') p(0.015)
542         drop `v'_a

```

```

543     }
544     drop if lev>1&CT_yes==0
545     replace lev=0.98 if lev>0.98
546
547
548 *-----
549 *-2.3.5 部分变量进行行业中位数调整和 Moving Average
550
551     local X "roa roe ros"
552     foreach v of varlist `X'{
553         bysort industry year: egen `v'_ind_med = median(`v') //各年度行业中位数
554         gen i`v' = `v' - `v'_ind_med // 经过 年度-行业中位数 调整
555         label var i`v' "经行业-年度中位数调整后的`v'"
556         bysort id tenure: egen m`v' = mean(`v')
557         label var m`v' "经理人任期内的平均 `v'"
558         bysort id tenure: egen mi`v' = mean(i`v')
559         label var mi`v' "经理人任期内的平均 i`v'"
560     }
561     dropvars *_ind_med
562
563 *-半年滞后期的调整 (Chang&Wong, 2009, p.236, line 1-4)
564 * 规则: 上半年离职者,根据前一年的业绩来评价;
565 *       下半年离职者,根据当年的业绩来评价.
566 *-职位变更月份
567     dropvars mt1 mt2 mt3
568     split chgdt, gen(mt) parse(-)
569     browse id year chgdt mt1 mt2 mt3
570     gen chgm = real(mt2) // 经理人并更月份
571     label var chgm "Chair 职位变更月份"
572 *-调整
573     local y "roa iroa mroa miroa" // 业绩指标
574     foreach v of varlist `y'{
575         qui tsset
576         replace `v' = L.`v' if chgm<=6 // see Chang&Wong(2009, p.236)
577     }
578
579
580 *-----
581 *-2.3.6 增加变量标签
582
583     label var roa "净利润/总资产"
584     label var roe "净利润/所有者权益合计"
585     label var ros "净利润/主营业务收入"
586     label var sale_gr "sale growth"
587     label var loss "1=营业利润为负,0=otherwise"
588     label var age "离职年龄"
589     label var tenure "任职年限"
590     label var size "公司规模_ln总资产"
591     label var lev "总负债率=负债总额/资产总额"
592     label var sale_gr "主营业务收入成长率"
593     label var list "2007-IPO年份"
594
595     label var gpay "ln前三名高管年薪总额"
596     label var magstk "高管持股比例"
597     label var duality "总经理与董事长为同一人则取1, 否则取0"
598
599     label var topone "第一大股东持股比例"
600     label var topone2 "第一大股东持股比例的平方"
601     label var hhi5 "前五大股东持股比例平方和"
602     label var zi2 "第一大股东比例/第二大股东持股比例"
603     label var private "民营控股虚拟变量"
604     label var state "国有控股虚拟变量"
605     label var independ "独立董事总人数/董事会规模"
606     label var top4_1 "第2-5大股东持股之和/第一大股东持股比例"
607     label var local "地方政府控股(地方国资委和地方国有企业)"
608     label var central "中央政府控股(国资委和中央国有企业)"
609

```

```

610  *--定义年度虚拟变量
611      dropvars yr_*
612      xi i.year, prefix(dum)
613      renvars dumyear*, presub(dumyear yr)
614
615
616  *-----
617  *--2.3.7 进一步筛选样本
618
619      drop if TA==0
620      drop if b001101000<=0 // 主营业务收入小于零
621      drop if a003000000 <0 // 所有者权益小于零, see Chang p.237 第一段
622      *drop if private==1 // 删除民营企业, see Chang p.237 第一段
623
624
625  *--重新排序变量 (这个处理的必要性不大, 只是为了方便而已)
626      local v1 "ro? *roa* *ros* *roe* list size lev sale_gr age tenure boardsize"
627      local v2 "independent top* ceo_stk chair_stk duality"
628      local v3 "private state yr*"
629      local v4 "industry dimreas changtyp age tenure"
630      order id year CT* `v1' `v2' `v3' `v4'
631
632  *--样本分布特征
633      xtides
634
635
636
637
638
639
640  *=====
641  *          Part III 实证分析
642  *=====
643
644
645  *-----
646  *--Table 4: Summary Statistics of Variables
647  *-----
648
649      local Panel_A "list age tenure duality lev size state"
650      local Panel_B "roa iroa mroa miroa"
651      global vv `Panel_A' `Panel_B'
652      *--删除缺漏值
653          dropvars miss2
654          egen miss2 = rmiss($vv)
655          drop if miss2>0
656
657      *--变量的定义和计算方法
658          des $vv
659
660      *--输出统计表格
661          logout, save(Results/var_statistics) excel replace: ///
662              tabstat $vv, stat(N mean sd min p50 max) format(%6.3f) c(s)
663
664
665  *-----
666  *--相关矩阵
667  *-----
668
669      adopath + "$path\adofiles" //自编和外部命令存放位置
670
671      logout, save(Results/var_correlation_x) excel replace: ///
672          pwcorr_a $vv, star1(0.01) star5(0.05) star10(0.1)
673
674      pwcorr_c $vv, star(0.05) format(%7.2f) //比较简洁的呈现方法
675
676      help pwcorr_c //Given by Lian Yujun
677

```



```

678      *-VIF (see PP.237, section 4.2 第一段)
679      local Panel_A "list age tenure duality lev size state"
680      qui reg roa `Panel_A' // 把 roa 替换为 iroa miroa 结果变化不大
681      estat vif
682
683
684      *-----
685      *-Table 5: Turnovers rates and performance
686      *-----
687      *-这里仅统计被迫变更的比率
688
689      label list CT_Chang_a // 变更类型
690      tab CT_Chang // 列表统计
691      gen TO_force = 0
692      replace TO_force=1 if (CT_Chang==2) //是否被迫离职
693      label var TO_force "1=forced turnover; 0=other"
694
695      *-----
696      *-过程解析
697
698      *-按 年度-preformance 分五组(原文未说明是否按年度划分)
699      *-这里以 ROA 变量为例进行说明 Panel A 的计算过程
700      *-----begin-----
701      global x "roa" // 待分析的变量, 可以替换为 iroa, mroa, miroa
702
703      *-All firms
704      cap drop q_x
705      bysort year: quantiles $x, gen(q_x) n(5) //每个年度上, 按ROA分成五组
706      tabstat $x, s(median) by(q_x) f(%4.3f) //Tab5.Panel_A.Line1
707      ttest $x if inlist(q_x,1,5), by(q_x) //t-test performance
708      tabstat TO_force, s(mean) by(q_x) f(%4.3f) //Tab5.Panel_A.Line2
709      ttest TO_force if inlist(q_x,1,5), by(q_x) //t-test turnover rate
710
711      *-Profit-making firms
712      cap drop q_x
713      bysort year: quantiles $x if loss==0, gen(q_x) n(5) //新增: if loss==0
714      tabstat $x, s(median) by(q_x) f(%4.3f)
715      ttest $x if inlist(q_x,1,5)&(loss==0), by(q_x)
716      tabstat TO_force, s(mean) by(q_x) f(%4.3f)
717      ttest TO_force if inlist(q_x,1,5), by(q_x)
718
719      *-Loss-making firms
720      cap drop q_x
721      bysort year: quantiles $x if loss==1, gen(q_x) n(5) //新增: if loss==1
722      tabstat $x, s(median) by(q_x) f(%4.3f)
723      ttest $x if inlist(q_x,1,5), by(q_x)
724      tabstat TO_force, s(mean) by(q_x) f(%4.3f)
725      ttest TO_force if inlist(q_x,1,5), by(q_x)
726      *-----over-----
727      *-Note: 虽然明了, 但还需手动逐个粘贴到 Excel 表格中
728
729
730      *-----
731      *-优化处理方法: 调入封装好的程序
732
733      adopath + "$path\Chang_2009_JCF"
734      doedit ChangTab5.ado // 查看或编辑源程序, 尾部列出了代码
735      run ChangTab5.ado // 程序代码读入内存
736
737      ChangTab5 roa A // Panel A
738      mat SA = S
739      ChangTab5 iroa B // Panel B
740      mat SB = S
741      ChangTab5 mroa C // Panel C
742      mat SC = S
743      ChangTab5 miroa D // Panel D
744      mat SD = S
745

```

```

746     mat dot = J(1,11,.) // 区分各个 panels
747     mat SS = (SA \ dot \ SB \ dot \ SC \ dot \ SD)
748
749     mat list SS, format(%4.3f) nodotz noheader // 这就是表5中的所有结果
750
751
752
753
754     *-----
755     *-Table 6: Logit regression of turnover-performance links in China's
756     *                               profit-making firms
757     *-----
758     *-Logit 模型简介: See: z_Logit.do
759     doedit "$path\z_Logit.do"
760
761     *-参考资料:
762     shellout "$path\Refs\Lian_15_Logit.pdf" //我的讲义
763     shellout "$path\Refs\Long_2001_Logit.pdf" //极力推荐
764     *-Long, J., J. Freese.
765     * Regression models for categorical dependent variables using Stata.
766     * Stata Press, 2001.
767
768     dropvars ln_age
769     gen ln_age = ln(age)
770
771     global xx "list ln_age tenure duality lev size state" // lnage, age
772
773     *-(1)
774     logit TO_force $xx roa if loss==0, vce(cluster id)
775     est store m1
776     *-(2)
777     logit TO_force $xx iroa if loss==0, vce(cluster id)
778     est store m2
779     *-(3)
780     logit TO_force $xx mroa if loss==0, vce(cluster id)
781     est store m3
782     *-(4)
783     logit TO_force $xx miroa if loss==0, vce(cluster id)
784     est store m4
785
786     *-呈现结果:
787     local s "using Results\Tab06_profit.csv"
788     local m "m1 m2 m3 m4"
789     esttab `m' `s', scalar(N r2_p) b(%6.3f) nogap ///
790             star(* 0.1 ** 0.05 *** 0.01) replace
791
792
793     *-----
794     *-Table 7: Logit regression of turnover-performance links in China's
795     *                               loss-making firms
796     *-----
797
798     global xx "list ln_age tenure duality lev size state"
799
800     *-(1)
801     logit TO_force $xx roa if loss==1, vce(cluster id)
802     est store m1
803     *-(2)
804     logit TO_force $xx iroa if loss==1, vce(cluster id)
805     est store m2
806     *-(3)
807     logit TO_force $xx mroa if loss==1, vce(cluster id)
808     est store m3
809     *-(4)
810     logit TO_force $xx miroa if loss==1, vce(cluster id)
811     est store m4
812
813     *-呈现结果:

```

```

814     local s "using Results\Tab07_Loss.csv"
815     local m "m1 m2 m3 m4"
816     esttab `m' `s', scalar(N r2_p) b(%6.3f) nogap ///
817           star(* 0.1 ** 0.05 *** 0.01) replace
818
819
820 *-----
821 *-Comments:
822 *
823 *(1) 这里的结果与 Chang&Wong 有何区别?
824 *(2) 若采用业绩的滞后项 L.roa, 结果如何?
825 *(3) 若采用其他的标准误计算方法, 如 vce(robust), vce(bootstrap), 结果如何?
826 *(4) 时间虚拟变量是否需要考虑?
827 *(5) 边际效果?
828 *(6) ROE, ROS 也可以作为稳健性检验.
829 *(7) 如何 winsor 关键变量对结果有很大的影响。 (important)
830
831
832 *-----being-----
833 *-修改后的程序
834
835     global xx "list ln_age tenure duality lev size state"
836     global if "" // 全样本 // 三个 if 选一个
837     global if "if loss==0" // profit-making firms
838     global if "if loss==1" // Loss-making firms
839     local L "L." // 采用滞后项
840     local rrr "roe" // 选择业绩指标: roa, roe, ros
841     *local yr "yr_2002-yr_2007" // 年度效应
842
843     qui tsset
844 *(1)
845     logit TO_force $xx `L'`rrr' `yr' $if, vce(cluster id)
846     est store m1
847 *(2)
848     logit TO_force $xx `L'i'`rrr' `yr' $if, vce(cluster id)
849     est store m2
850 *(3)
851     logit TO_force $xx `L'm'`rrr' `yr' $if, vce(cluster id)
852     est store m3
853 *(4)
854     logit TO_force $xx `L'mi'`rrr' `yr' $if, vce(cluster id)
855     est store m4
856
857 *-呈现结果:
858 *local s "using Results\Tab06_mfx.csv"
859 *local m "m1 m2 m3 m4"
860 *esttab `m' `s', scalar(N r2_p) b(%6.3f) nogap ///
861 *      star(* 0.1 ** 0.05 *** 0.01) replace ///
862 *      addnotes("$if")
863 *-----over-----
864 *
865 *-Comments:
866 *-
867 *(1) Loss==1 时,
868 *     L.roa 或 L.roe 等指标是显著的, 表明上期业绩差是导致离职的主因;
869 *     然而, L.ros 却不显著, 一个可能的原因是, ROS 不是评价 CEO 业绩的指标?
870 *(2) 其他情况下, 你会发现结果并不稳健,
871 *     相对于发在 JF 上的那些文章, 这是个问题;
872 *(3) 进一步的研究可以把 "盈余管理" 考虑进来.
873
874
875 *-----
876 *-Table 8: Logit regression of turnover-performance links in China's
877 *           with interactions
878 *-----
879
880     global xx "list ln_age tenure duality lev size state"
881     dropvars *roa_x_*

```



```

882     gen roa_x_local  = roa*local
883     gen roa_x_state  = roa*state
884     gen iroa_x_local = iroa*local
885     gen iroa_x_state = iroa*state
886
887     *-(1)
888     logit TO_force $xx roa  roa_x_*  if loss==0, vce(cluster id)
889     est store m1
890     *-(2)
891     logit TO_force $xx iroa iroa_x_* if loss==0, vce(cluster id)
892     est store m2
893     *-(3)
894     logit TO_force $xx roa  roa_x_*  if loss==1, vce(cluster id)
895     est store m3
896     *-(4)
897     logit TO_force $xx iroa iroa_x_* if loss==1, vce(cluster id)
898     est store m4
899
900     *-呈现结果:
901     local s  "using Results\Tab08_Loss.csv"
902     local m  "m1 m2 m3 m4"
903     local mt "Profit Profit Loss Loss"
904     esttab `m' `s', scalar(N r2_p) b(%6.3f) nogap replace ///
905           mtitle(`mt') star(* 0.1 ** 0.05 *** 0.01)
906
907     *-----
908     *-Comments:
909     *
910     *-(1) 交叉项都不显著, 怎么办? (其实, Chang&Wong 的结果也很微弱)
911     *-(2) 采用业绩及其交叉项的滞后项, 看看结果是否有所变化;
912     *-(3) 引入其他变量来构造交叉项 (关键在于如何讲故事)
913
914     *   save beforeTable9.dta, repalce
915
916
917
918     *-----
919     *-Table 9
920     * Changes in post-turnover performance in China's listed firms
921     *-----
922
923     *-See: pp.236, Section 3.3; pp.242, Section 4.4
924
925     *-难点: ROA[t=1] - ROA[t=0] = 5%
926     *              = 3% + 2%
927     *              = 6% - 1%
928     *              = 0% + 5%
929
930     *-Note: 执行后续命令之前, 请确保 315-912 行的命令都能正确执行
931     *-      或者, 可以在执行完上述命令之后, 执行 914 行的命令, 然后执行如下命令:
932     *-      (要去掉星号)
933
934     *   use beforeTable9.dta, clear
935
936     *-Chang & Wong 文中的潜在问题
937     *-(1) 对于样本区间内发生了多次 CEO 变更的公司, 如何处理?
938     *-(2) 对于在样本初始期(2000年)发生变更的公司, 无法追踪 t=-1;
939     *      对于在样本结束期(2007年)发生变更的公司, 无法追踪 t=+1,+2,+3
940
941     *-我们的处理方式(借鉴 Event Study 研究的思路): 仅保留如下公司
942     *-(1) 样本区间内仅发生了一次 CEO 变更的公司;
943     *-(2) 在 t=[-1,3] 区间内都有观察值的公司, 即至少要有 5 年的观察值;
944     *-(3) 下面的例子中, 仅考虑 t=[0,3] 的情形, 对于 t=-1 的留给大家做练习
945
946     *-----
947     *-9.1 删除样本区间内发生多次 CEO 变更的公司
948
949     dropvars TO_times

```

```

950     bysort id: egen TO_times = sum(TO_force)
951     browse id year TO_force TO_times
952     tab TO_times year    //各个年度上变更次数统计
953     drop if TO_times>1
954     tab TO_times year    //变更组与非变更组的年度分布
955
956
957     *-----
958     *-9.2 仅保留样本区间内有 4 年以上资料的公司
959
960     *-9.2.1 定义 t
961
962     *-思路: 离职当年 t=0, 此前一年为 t=-1, 此后一年 t=+1, 以此类推
963     dropvars t1 t123 t1a t
964     qui tsset id year
965     browse id year TO_force
966     bysort id: gen t123=_n    //定义新的日期变量, 取值为 1,2,3,...
967     bysort id: gen t1=t123 if TO_force==1 //标定变更日
968     egen t1a = min(t1), by(id) //扩充 t1 的观察值
969     gen t = t123 -t1a //以事件日0为基准定义日期
970     browse id year TO_force t1 t123 t1a t
971
972     *-9.2.2 定义事件窗口 [0, 3]
973
974     dropvars window window_obs
975     bysort id: gen window=1 if (t>=0 & t<=3)
976     bysort id: egen window_obs=count(window)
977     browse id year TO_force t *window*
978
979     *-删除事件窗口观察值不足 4 年 的公司
980     tab window_obs year
981     drop if (window_obs>0 & window_obs<4)
982
983
984     *-----
985     *-9.3 寻找配对公司
986
987     *-规则:  $0.8(T) \leq ROA(C) \leq 1.2(T)$  C:配对组; T:CEO变更组
988     *-寻找范围: 同年度、同行业
989     dropvars yesT
990     gen yesT = (t!=.) // 标示变更组(=1)和潜在配对组(=0)
991
992     *-9.3.1 配对组公司至少要有 4 年[t=0,1,2,3]以上的资料, 否则删除
993
994     dropvars T_obs
995     bysort id: egen T_obs = count(id)
996     drop if (T_obs<4 & yesT==0)
997
998     *-9.3.2 开始配对
999
1000     dropvars IND //数字类型的行业代码
1001     encode industry, gen(IND) label(IND_lab)
1002     tab IND
1003     label list IND_lab
1004
1005     local r "roa" // [填入] 业绩指标
1006     dropvars nn Control ratio_* n_match match_id
1007     gen Control = . //最关键的变量, 随后填入 1(表示配对对象), 0(表示变更公司)
1008     gen ratio_roa = . //业绩匹配规则变量
1009     gen ratio_size = . //公司规模匹配规则变量
1010     gen n_match = . //仅考虑 roa[-20%,+20%] 规则时, 符合条件的公司数目
1011     gen match_id = . //匹配公司的代码
1012     gsort -TO_force
1013     gen nn = _n if TO_force==1
1014     browse id year TO_force nn roa
1015     qui sum nn
1016     local N=r(N) //待匹配的公司数目
1017     forvalues i=1/`N'{

```

```

1018     local roa_t0 = `r'[`i']    //变更公司的业绩
1019     local IND_t0 = IND[`i']    //变更公司所处的行业
1020     local size_t0 = size[`i']  //变更公司的公司规模
1021     local yr_t0 = year[`i']    //变更年份
1022     local id_t0 = id[`i']      //变更公司的代码
1023     replace ratio_roa = `r'/'`roa_t0'    ///
1024         if (year==`yr_t0' & IND==`IND_t0' & yesT==0)
1025     dis in red "count"
1026     count if (ratio_roa>0.8&ratio_roa<1.2) & ///
1027         (year==`yr_t0') & (IND==`IND_t0') //配对公司个数
1028     replace n_match = r(N) if nn==`i'
1029     * set trace on          //解析程序运行过程-begin-
1030     if r(N)>1{
1031         replace ratio_size = abs(size/'`size_t0'-1) ///
1032             if (ratio_roa>0.8&ratio_roa<1.2) & (year==`yr_t0') & ///
1033                 (yesT==0) & (IND==`IND_t0')
1034         sum ratio_size    ///
1035             if (ratio_roa>0.8&ratio_roa<1.2) & (year==`yr_t0') & ///
1036                 (yesT==0) & (IND==`IND_t0') //找到最小值
1037         replace Control=1 if ratio_size==r(min) & (year==`yr_t0') & (Control!=1)
1038             *-(Control!=1) 已经设为配对组的公司不会被覆盖掉
1039         replace match_id=`id_t0'    ///
1040             if ratio_size==r(min) & (year==`yr_t0')
1041     }
1042     else{
1043         replace Control=1    ///
1044             if (ratio_roa>0.8&ratio_roa<1.2) & (year==`yr_t0') & ///
1045                 (IND==`IND_t0') & (Control!=1)
1046         replace match_id=`id_t0'    ///
1047             if (ratio_roa>0.8&ratio_roa<1.2) & (year==`yr_t0') & ///
1048                 (IND==`IND_t0')
1049     }
1050     * set trace off          //解析程序运行过程-over-
1051 }
1052
1053
1054 *-9.3.3 配对情况统计
1055
1056     *-仅考虑 roa[-20%,+20%] 规则时, 符合条件的公司数目
1057     sort n_match id year
1058     list id year TO_force n_match if TO_force==1, sepby(n_match)
1059
1060     *-谁和谁成亲了? (Control-Treat组公司代码对应关系)
1061     gsort -Control id year
1062     list id year Control match_id if Control==1
1063
1064     *-有多少变更公司找到了匹配对象?
1065     replace Control=0 if TO_force==1 //CEO强制变更当年, Control=0
1066     table Control TO_force
1067     table Control TO_force year
1068
1069
1070 *-----
1071 *-9.4 定义 Control 组的时间点
1072
1073     *-思路: 选定为Control当年 t=0, 此后一年 t=+1, 以此类推
1074     dropvars t1C t1Ca tC    //加 C, 表示 Control 组
1075     qui tsset id year
1076     bysort id: gen t1C=t123 if Control==1 //标定配对年度
1077         //t123 此前已经定义好的变量, 取值为 1,2,3,...
1078     br id year Control t1C t123
1079     egen t1Ca = min(t1C), by(id) //扩充 t1 的观察值
1080     gen tC = t123 -t1Ca    //以事件年0为基准定义相对年份 -2,-1,0,+1,+2
1081     browse id year Control t1C t123 t1Ca tC t
1082     replace t = tC if t==. //CEO变更组和配对组都有了相对年份标识
1083
1084     *-扩充 Control=0 的观察值
1085     replace Control=0 if yesT==1 //CEO更换公司(所有年度), Control=0

```

```

1086
1087     *-扩充 Control=1 的观察值
1088     replace Control=1 if tC!=.      //配对公司(所有年度), Control=1
1089
1090
1091     /* 这个酌情选择处理
1092     *-删除窗口内观察值不足四年的公司 [0, +3]
1093     dropvars window_new window_new_obs
1094     bysort id: gen window_new=1 if (t>=0 & t<=3)
1095     bysort id: egen window_new_obs=count(window_new)
1096     browse id year Control TO_force t *window_new*
1097     drop if window_new_obs<4 // 删除窗口内观察值个数不足4年的公司
1098     */
1099
1100     *-----
1101     *-9.5 Table 9: Panel A (原始样本, 不配对)
1102     *-----
1103
1104     *-中位数差异检验简介
1105     help ranksum // Wilcoxon rank-sum test
1106     help median // Nonparametric equality-of-medians test
1107
1108     tabstat roa, by(loss) s(p50) f(%6.3f) //范例
1109     ranksum roa, by(loss)
1110
1111     *-----
1112     *-Column (1): Profit making firms
1113     *-----
1114
1115     * [+1 to 0]
1116     *-中位数差异 计算
1117     tabstat roa if (loss==0 & inlist(t,1,0) & Control==0), by(t) s(p50) save
1118     mat diff = r(Stat2) - r(Stat1)
1119     local diff = diff[1,1]
1120     dis "Median performance diff [t=+1 to t=0] = " `diff'
1121     *-中位数差异 检验
1122     ranksum roa if (loss==0 & inlist(t,1,0) & Control==0), by(t)
1123
1124     * [+2 to 0]
1125     *-中位数差异 计算
1126     tabstat roa if (loss==0 & inlist(t,2,0) & Control==0), by(t) s(p50) save
1127     mat diff = r(Stat2) - r(Stat1)
1128     local diff = diff[1,1]
1129     dis "Median performance diff [t=+2 to t=0] = " `diff'
1130     *-中位数差异 检验
1131     ranksum roa if (loss==0 & inlist(t,2,0) & Control==0), by(t)
1132
1133     * [+3 to 0]
1134     *-中位数差异 计算
1135     tabstat roa if (loss==0 & inlist(t,3,0) & Control==0), by(t) s(p50) save
1136     mat diff = r(Stat2) - r(Stat1)
1137     local diff = diff[1,1]
1138     dis "Median performance diff [t=+3 to t=0] = " `diff'
1139     *-中位数差异 检验
1140     ranksum roa if (loss==0 & inlist(t,3,0) & Control==0), by(t)
1141
1142     *-----
1143     *-Column (2): Loss making firms
1144     *-----
1145
1146     * [+1 to 0]
1147     *-中位数差异 计算
1148     tabstat roa if (loss==1 & inlist(t,1,0) & Control==0), by(t) s(p50) save
1149     mat diff = r(Stat2) - r(Stat1)
1150     local diff = diff[1,1]
1151     dis "Median performance diff [t=+1 to t=0] = " `diff'
1152     *-中位数差异 检验
1153     ranksum roa if (loss==1 & inlist(t,1,0) & Control==0), by(t)

```

```

1154
1155     * [+2 to 0]
1156     *-中位数差异 计算
1157     tabstat roa if (loss==1 & inlist(t,2,0) & Control==0), by(t) s(p50) save
1158     mat diff = r(Stat2) - r(Stat1)
1159     local diff = diff[1,1]
1160     dis "Median performance diff [t=+2 to t=0] = " `diff'
1161     *-中位数差异 检验
1162     ranksum roa if (loss==1 & inlist(t,2,0) & Control==0), by(t)
1163
1164     * [+3 to 0]
1165     *-中位数差异 计算
1166     tabstat roa if (loss==1 & inlist(t,3,0) & Control==0), by(t) s(p50) save
1167     mat diff = r(Stat2) - r(Stat1)
1168     local diff = diff[1,1]
1169     dis "Median performance diff [t=+3 to t=0] = " `diff'
1170     *-中位数差异 检验
1171     ranksum roa if (loss==1 & inlist(t,3,0) & Control==0), by(t)
1172
1173     *-----
1174     *-Column (3): P value of difference change between (1) and (2)
1175     *-----
1176
1177     *-计算 t=+1,+2,+3 时点时的业绩相对于 t=0 时点业绩的差异
1178     qui tsset id t
1179     dropvars roa_chg*
1180     gen roa_chg10 = F1.roa - roa // [t=+1 to t=0]
1181     gen roa_chg20 = F2.roa - roa // [t=+2 to t=0]
1182     gen roa_chg30 = F3.roa - roa // [t=+3 to t=0]
1183     browse id t roa roa_ch*
1184
1185     *-中位数差异 检验
1186     * [+1 to 0]
1187     tabstat roa_chg10 if (t==0 & Control==0), by(loss) s(p50)
1188     ranksum roa_chg10 if (t==0 & Control==0), by(loss)
1189     * [+2 to 0]
1190     tabstat roa_chg20 if (t==0 & Control==0), by(loss) s(p50)
1191     ranksum roa_chg20 if (t==0 & Control==0), by(loss)
1192     * [+3 to 0]
1193     tabstat roa_chg30 if (t==0 & Control==0), by(loss) s(p50)
1194     ranksum roa_chg30 if (t==0 & Control==0), by(loss)
1195
1196     *-Comments:
1197     *(1) 要计算 Panel B 中的结果, 需要把 [9.3]-[9.5] 节中的 ROA 都替换成 IROA
1198     *(2) 可以参考 Table 5 中的处理方法, 封装成程序, 并采用矩阵输出结果
1199     *(3) 配对过程以 t=0 为 basetime, 可以修改后扩展到 t=-1 的情形
1200
1201
1202     *-----
1203     *-9.6 Table 9: Panel C (配对样本调整后的结果)
1204
1205     keep if Control!=.
1206     keep if t>=0&t<=3
1207     browse id year Control t roa
1208     panels id: tab Control year
1209
1210     *-CROA 指标
1211     dropvars p50_roa_C* croa
1212     bysort year: egen p50_roa_C = median(roa) if Control==1
1213     bysort year: egen p50_roa_C_fill = min(p50_roa_C)
1214     gen croa = roa - p50_roa_C_fill
1215     br id year Control roa p50* croa
1216
1217
1218     *-----
1219     *-Column (5): Profit making firms
1220     *-----
1221     *-(只需把上面的命令修改一下即可)

```



```

1222
1223     global roa "croa"    //填入业绩变量
1224
1225     * [+1 to 0]
1226     *-中位数差异 计算
1227     tabstat $roa if (loss==0 & inlist(t,1,0) & Control==0), by(t) s(p50) save
1228     mat diff = r(Stat2) - r(Stat1)
1229     local diff = diff[1,1]
1230     dis "Median performance diff [t=+1 to t=0] = " `diff'
1231     *-中位数差异 检验
1232     ranksum $roa if (loss==0 & inlist(t,1,0) & Control==0), by(t)
1233
1234     * [+2 to 0]
1235     *-中位数差异 计算
1236     tabstat $roa if (loss==0 & inlist(t,2,0) & Control==0), by(t) s(p50) save
1237     mat diff = r(Stat2) - r(Stat1)
1238     local diff = diff[1,1]
1239     dis "Median performance diff [t=+2 to t=0] = " `diff'
1240     *-中位数差异 检验
1241     ranksum $roa if (loss==0 & inlist(t,2,0) & Control==0), by(t)
1242
1243     * [+3 to 0]
1244     *-中位数差异 计算
1245     tabstat $roa if (loss==0 & inlist(t,3,0) & Control==0), by(t) s(p50) save
1246     mat diff = r(Stat2) - r(Stat1)
1247     local diff = diff[1,1]
1248     dis "Median performance diff [t=+3 to t=0] = " `diff'
1249     *-中位数差异 检验
1250     ranksum $roa if (loss==0 & inlist(t,3,0) & Control==0), by(t)
1251
1252     *-----
1253     *-Column (6): Loss making firms
1254     *-----
1255
1256     * [+1 to 0]
1257     *-中位数差异 计算
1258     tabstat $roa if (loss==1 & inlist(t,1,0) & Control==0), by(t) s(p50) save
1259     mat diff = r(Stat2) - r(Stat1)
1260     local diff = diff[1,1]
1261     dis "Median performance diff [t=+1 to t=0] = " `diff'
1262     *-中位数差异 检验
1263     ranksum $roa if (loss==1 & inlist(t,1,0) & Control==0), by(t)
1264
1265     * [+2 to 0]
1266     *-中位数差异 计算
1267     tabstat $roa if (loss==1 & inlist(t,2,0) & Control==0), by(t) s(p50) save
1268     mat diff = r(Stat2) - r(Stat1)
1269     local diff = diff[1,1]
1270     dis "Median performance diff [t=+2 to t=0] = " `diff'
1271     *-中位数差异 检验
1272     ranksum $roa if (loss==1 & inlist(t,2,0) & Control==0), by(t)
1273
1274     * [+3 to 0]
1275     *-中位数差异 计算
1276     tabstat $roa if (loss==1 & inlist(t,3,0) & Control==0), by(t) s(p50) save
1277     mat diff = r(Stat2) - r(Stat1)
1278     local diff = diff[1,1]
1279     dis "Median performance diff [t=+3 to t=0] = " `diff'
1280     *-中位数差异 检验
1281     ranksum $roa if (loss==1 & inlist(t,3,0) & Control==0), by(t)
1282
1283     *-----
1284     *-Column (7): P value of difference change between (1) and (2)
1285     *-----
1286
1287     *-计算 t=+1,+2,+3 时点时的业绩相对于 t=0 时点业绩的差异
1288     qui tsset id t
1289     dropvars roa_chg*

```

```

1290     gen roa_chg10 = F1.$roa - $roa // [t=+1 to t=0]
1291     gen roa_chg20 = F2.$roa - $roa // [t=+2 to t=0]
1292     gen roa_chg30 = F3.$roa - $roa // [t=+3 to t=0]
1293     browse id t $roa roa_ch*
1294
1295     *-中位数差异 检验
1296     * [+1 to 0]
1297     tabstat roa_chg10 if (t==0 & Control==0), by(loss)
1298     ranksum roa_chg10 if (t==0 & Control==0), by(loss)
1299     * [+2 to 0]
1300     tabstat roa_chg20 if (t==0 & Control==0), by(loss)
1301     ranksum roa_chg20 if (t==0 & Control==0), by(loss)
1302     * [+3 to 0]
1303     tabstat roa_chg30 if (t==0 & Control==0), by(loss)
1304     ranksum roa_chg30 if (t==0 & Control==0), by(loss)
1305
1306
1307     *-----
1308     *-Comments:
1309
1310     *-(1) 更为合理的处理方法: DID + PSM;
1311     help diff
1312
1313     *-(2) 也可以采用 treatment effect models:
1314     help treatreg
1315
1316
1317
1318
1319     *-----
1320     *-Appendix: ChangTab.ado 源程序
1321     *-----
1322
1323     *! Program used to present the results in
1324     *! Chang and Wong (2009) Table 5
1325     *! 命令格式: ChangTab5 roa A
1326
1327     capture program drop ChangTab5
1328     program define ChangTab5
1329     version 11.2
1330
1331     args x A // 命令格式: ChangTab5 roa A
1332
1333     mat S = J(6,11,.) // 存储结果的矩阵
1334
1335     *-All firms
1336     cap drop q_x
1337     qui bysort year: quantiles `x', gen(q_x) n(5)
1338     qui sum q_x
1339     mat S[1,1] = r(N) // 全样本的观察值个数
1340     qui tabstat `x', s(median) by(q_x) f(%4.3f) save //Tab5.Panel_A.Line1
1341     mat S[1,2] = (r(Stat1),r(Stat2),r(Stat3),r(Stat4),r(Stat5))
1342     mat S[1,8] = r(Stat5) - r(Stat1) // Performance spread (5)-(1)
1343     qui ttest `x' if inlist(q_x,1,5), by(q_x) //t-test performance
1344     mat S[1,10]= (r(t), r(p)) // p-value 和 t-value
1345
1346     qui tabstat TO_force, s(mean) by(q_x) f(%4.3f) save //Tab5.Panel_A.Line2
1347     mat S[2,2] = (r(Stat1),r(Stat2),r(Stat3),r(Stat4),r(Stat5),r(StatTotal))
1348     mat S[2,9] = r(Stat5) - r(Stat1) // Turnover spread (5)-(1)
1349     qui ttest TO_force if inlist(q_x,1,5), by(q_x) //t-test turnover rate
1350     mat S[2,10]= (r(t), r(p))
1351
1352     *-Profit-making firms
1353     cap drop q_x
1354     global if "if loss==0"
1355     qui bysort year: quantiles `x' $if, gen(q_x) n(5)
1356     qui sum q_x
1357     mat S[3,1] = r(N) // Profit-making 公司的观察值个数

```

```

1358     qui tabstat `x', s(median) by(q_x) f(%4.3f) save //Tab5.Panel_A.Line1
1359     mat S[3,2] = (r(Stat1),r(Stat2),r(Stat3),r(Stat4),r(Stat5))
1360     mat S[3,8] = r(Stat5) - r(Stat1) // Performance spread (5)-(1)
1361     qui ttest `x' if inlist(q_x,1,5), by(q_x) //t-test performance
1362     mat S[3,10] = (r(t), r(p)) // p-value 和 t-value
1363
1364     qui tabstat TO_force, s(mean) by(q_x) f(%4.3f) save //Tab5.Panel_A.Line2
1365     mat S[4,2] = (r(Stat1),r(Stat2),r(Stat3),r(Stat4),r(Stat5),r(StatTotal))
1366     mat S[4,9] = r(Stat5) - r(Stat1) // Turnover spread (5)-(1)
1367     qui ttest TO_force if inlist(q_x,1,5), by(q_x) //t-test turnover rate
1368     mat S[4,10] = (r(t), r(p))
1369
1370     *-Loss-making firms
1371     cap drop q_x
1372     global if "if loss==1"
1373     qui bysort year: quantiles `x' $if, gen(q_x) n(5)
1374     qui sum q_x
1375     mat S[5,1] = r(N) // Loss-making 公司的观察值个数
1376     qui tabstat `x', s(median) by(q_x) f(%4.3f) save //Tab5.Panel_A.Line1
1377     mat S[5,2] = (r(Stat1),r(Stat2),r(Stat3),r(Stat4),r(Stat5))
1378     mat S[5,8] = r(Stat5) - r(Stat1) // Performance spread (5)-(1)
1379     qui ttest `x' if inlist(q_x,1,5), by(q_x) //t-test performance
1380     mat S[5,10] = (r(t), r(p)) // p-value 和 t-value
1381
1382     qui tabstat TO_force, s(mean) by(q_x) f(%4.3f) save //Tab5.Panel_A.Line2
1383     mat S[6,2] = (r(Stat1),r(Stat2),r(Stat3),r(Stat4),r(Stat5),r(StatTotal))
1384     mat S[6,9] = r(Stat5) - r(Stat1) // Turnover spread (5)-(1)
1385     qui ttest TO_force if inlist(q_x,1,5), by(q_x) //t-test turnover rate
1386     mat S[6,10] = (r(t), r(p))
1387
1388     *-矩阵名称和结果呈现
1389     local cc "ATO PF(5)-(1) TO(5)-(1) t p-value"
1390     mat colnames S = obs (1)Low (2) (3) (4) (5)High `cc'
1391     mat rownames S = All_`x' turnover Prof_`x' turnover Loss_`x' turnover
1392
1393     dis _n(2) in yellow _dup(52) "_"
1394     dis in red "Panel `A': Observations are sorted according to `x'"
1395     dis in yellow _dup(52) "_"
1396     mat list S, format(%4.3f) nodotz noheader
1397
1398     end
1399
1400

```



```

1
2 * -----
3 * -----
4 *
5 *           Stata 学术论文专题
6 *
7 * -----
8 * -----
9
10
11 *           主讲人：连玉君 副教授
12 *
13 *           单 位：中山大学岭南学院金融系
14 *           电 邮：arlionn@163.com
15 *           博 客：http://blog.cnfol.com/arlionn
16 *           主 页：http://goo.gl/tRXba
17 *           微 博：http://weibo.com/arlionn
18
19
20 *-----
21 * Cleary, S., 1999,
22 *   The Relationship between Firm Investment and Financial Status,
23 *   Journal of Finance, 54(2):673-692.
24 *-----
25 *-密切关联：Fazzari_1988.do
26 *-----
27 *-主要方法
28 *-----
29 *-(1) discriminant analysis
30 *   help discrim
31 *-(2) Bootstrap           see B9_MC_BS.do(Stata高级视频)
32 *   help bootstrap
33
34
35 *-设定课程存储路径
36   global path "`c(sysdir_personal)'PX_papers" //请先执行该命令
37
38 *-原始论文
39   shellout "$path\Refs\Cleary_1999.pdf"
40 *-PPT
41   shellout "$path\PPT\Cleary_1999.ppt"
42
43 *-----
44 *-相关文献
45 *-----
46 *
47 *-Fazzari et al.(1988): Fazzari_1988.do
48 *-Fazzari, S., R. Hubbard, B. Petersen, 1988,
49 *   Financing Constraints and Corporate Investment,
50 *   Brookings papers on economic activity, 1988(1):141-206.
51   shellout "$path\Refs\Fazzari_1988.pdf" // PDF 文件
52
53 *-Kaplan, S., L. Zingales, 1997,
54 *   Do investment-cash flow sensitivities provide
55 *   useful measures of financing constraints?,
56 *   Quarterly Journal of Economics, 112(1):169-215.
57   shellout "$path\Refs\Kaplan_Zingales_1997.pdf" // PDF 文件
58
59 *-Fazzari, S., R. Hubbard, B. Petersen, 2000,
60 *   Investment-Cash Flow Sensitivities are Useful:
61 *   A Comment on Kaplan and Zingales,
62 *   Quarterly Journal of Economics, 115(2):695-705.
63   shellout "$path\Refs\Fazzari_2000.pdf" // PDF 文件
64
65 *-Kaplan, S., L. Zingales, 2000,
66 *   Investment-Cash Flow Sensitivities are not Valid
67 *   Measures of Financing Constraints,
68 *   Quarterly Journal of Economics, 115(2):707-712.

```

```

69      shellout "$path\Refs\Kaplan_Zingales_2000.pdf" // PDF 文件
70
71      *-Cleary, S., P. Povel, M. Raith, 2007,
72      *   The U-Shaped Investment Curve: Theory and Evidence,
73      *   Journal of Financial and Quantitative Analysis, 42(1):1-40.
74      shellout "$path\Refs\Cleary_2007.pdf"
75
76      *-Vogt, S., 1994,
77      *   The cash flow/investment relationship:
78      *   evidence from US manufacturing firms,
79      *   Financial Management, 23(2):3-20.
80      shellout "$path\Refs\Vogt_1994.pdf"
81
82
83
84      *=====
85      * 样本筛选
86      *=====
87
88      adopath + "$path\adofiles" //自编程序(这条命令一定要执行)
89
90      cd "$path\Cleary_1999" //进入本文目录
91
92      use "$path\Data\GTA2008_my.dta", clear //调入数据
93
94      keep id year tobin size t1 TA TD ll sgr tagr ///
95      div_cash div_yes sicda sicmen tang pyear ///
96      b001101000 c001000000 a001212000 ///
97      a001101000 c002006000 a001100000 ///
98      a002100000 a001101000 a001123000 ///
99      b002000000 a003000000 a001111000 ///
100     b001300000 a001109000 a002101000
101
102     rename a001212000 K
103
104     keep if year>1998 // 1998年以后采用现金流量表
105
106
107     *-----
108     *-定义变量
109     *-----
110
111     *-股利变化指标
112
113     replace div_cash=0 if div_cash==.
114     gen Ddiv = D.div_cash // 股利的变化：用一阶差分表示
115     gen g_div = 1 if Ddiv>0 // 增加股利的公司-年度
116     replace g_div = 2 if Ddiv<0 // 减少股利的公司-年度
117     replace g_div = 3 if Ddiv==0 // 股利维持不变的公司-年度
118
119     *-定义"数字-文字对应表"
120     label define g_div_lab 1 "increased DPS" ///
121     2 "decreased DPS" ///
122     3 "no change in DPS"
123     label value g_div g_div_lab
124
125     tab g_div
126
127     * 说明：只有采用TA标准化，才能得到与 Cleary 相似的结果，
128     * 采用L.TA、L.K或K，都无法得到这一结果
129     * Clear(1999) 文中采用 L.K 进行标准化
130     * 详见：Fazzari et al. (1988)：Fazzari_1988.do
131
132     local fm "TA" // 采用总资产标准化
133     gen invt = c002006000/`fm' // 购建固定资产、无形资产和其他长期资产支付的现金/fm
134     gen cflow= c001000000/`fm' // 经营活动产生的现金流净额/fm
135     gen cash = a001101000/`fm' // 货币资金/fm
136     gen sale = b001101000/`fm' // 营业收入/fm

```

```

137     gen roa = b002000000/TA // 净利润/总资产
138     gen roe = b002000000/a003000000 // 净利润/所有者权益合计
139     gen slack = (a001101000+a001109000+0.5*a001123000+0.7*a001111000-a002101000)/TA
140     // 现金+短期投资+0.5存货+0.7应收账款-短期银行借款
141     gen ni_margin = b001300000/b001101000 // 营业利润/营业收入
142     gen cr = a001100000/a002100000 // current ratio 流动资产/流动负债
143     gen age = 2006-pyear // 公司年龄
144     gen lnage = ln(age)
145
146     *-删除离群值
147     drop if t1>1 // 这个处理未必合理
148
149     *-删除缺漏值
150     drop if missing(invt,cflow,tobin,cash,sale,tang,cr,ni_margin)
151
152     *-处理为平行面板(可选)
153     qui tsset
154     * drop if slack==. // 若增加该变量,则只能用 1999-2006 的资料
155     * xtbalance, range(1999 2008) ///
156     * miss(invt cflow tobin cash sale tang cr ni_margin )
157     *
158     *-讨论: 进行 xtbalance 处理有何缺陷?
159
160     *-Winsor 处理
161     foreach v of varlist invt-cr{
162         rename `v' `v'_a
163         winsor `v'_a, gen(`v') p(0.01)
164         drop `v'_a
165     }
166     *-Cleary(1999) 文中采用了个别处理的方式,详见 p.678, 以及 Footnote 5
167
168     tsset id year
169     order id year invt cflow tobin sale cash ni_margin cr slack ///
170         div_cash div_yes g_div Ddiv tang t1 roa roe
171
172     *-year dummy
173     qui tab year, gen(dum_yr)
174     drop dum_yr1
175
176
177     *-定义融资约束分组指标 (我们可以采用这种传统的分组方式作对比)
178
179     *-根据公司规模分组
180     * 根据样本公司在 1998-2008 期间的平均 TA 分组,
181     * 这样可以保证同一家公司始终处于同一组中
182     * Fazzari et al. (1988): Fazzari_1988.do 采用了这种处理方式
183     * 我们将样本公司分成 3 组,
184     * 其中第一组定义为 FC 组, 第五组定义为 NFC 组
185     bysort id: egen av_size = mean(TA)
186     sort av_size
187     gen gsize = group(3)
188     *-数字文字对应表
189     label define gsize_lab 1 "Small" 3 "Large"
190     label value gsize gsize_lab
191     *-定义融资约束类别变量
192     gen FC_size = gsize==1
193     gen NFC_size = gsize==3
194     gen Other_size = 1-FC_size-NFC_size
195
196     *-Comment: 按照 Cleary 的逻辑, 应该允许公司的 FC 状态发生变化
197     *-按照公司规模分年度分成三组, 分别检验
198     dropvars sizeq33 sizeq66 qsize
199     bysort year: egen sizeq33 = pctlile(size) if size!=., p(33)
200     bysort year: egen sizeq66 = pctlile(size) if size!=., p(66)
201     gen qsize = 2
202     replace qsize=1 if size<=sizeq33
203     replace qsize=3 if size>=sizeq66
204     *-数字-文字对应表

```

```

205     label define qsize_lab 1 "Small(<q33)" 2 "Middle" 3 "Large(>q66)"
206     label value qsize qsize_lab
207
208     tabstat cash cflow t1 tang size, by(qsize) f(%4.3f)
209
210     *-根据股利支付率分组(see Fazzari_1988.do)
211     * Num_div 是在1998-2008期间, 每家公司支付股利的次数
212     * 若Num_div<=2,则定义为FC 组;
213     * 若Num_div>=8,则定义为NFC组;
214     dropvars Num_div FC_div NFC_div Other_div
215     bysort id: egen Num_div = sum(div_yes)
216     recode Num_div (min/2=1) (2/8=2) (8/max=3), gen(gg_div)
217     gen FC_div = Num_div<=2
218     gen NFC_div = Num_div>8
219     gen Other_div = 1 - FC_div - NFC_div
220
221     *-另存一份精简的数据, 以备后用
222     save Cleary_1999_data, replace
223
224
225     *-----
226     *- Z_FC index      discrim analysis (判别分析)
227     *-----
228
229     *-discrim analysis: Introduction
230
231     *use http://www.stata-press.com/data/r11/lawnmower2, clear
232     use lawnmower2, clear //(Johnson and Wichern (2007) Table 11.1)
233
234     *-特征变量的密度函数
235     local x "income"
236     local x "lotsize"
237     twoway (kdensity `x' if owner==0) ///
238            (kdensity `x' if owner==1), ///
239            xtitle("`x'") ///
240            legend(label(1 NonOwner) label(2 Owner))
241
242     *-散点图
243     twoway (scatter lotsize income if owner==0) ///
244            (scatter lotsize income if owner==1), ///
245            legend(label(1 NonOwner) label(2 Owner))
246
247
248     *-判别分析的原理
249     *-----
250     #delimit ;
251     twoway (scatter lotsize income if owner==0)
252            (scatter lotsize income if owner==1)
253            (function y = 31.7-0.125*x, range(60 140)),
254            legend(label(1 NonOwner) label(2 Owner)
255                   label(3 discrim_line) row(1)) ;
256     #delimit cr
257     *-----
258     * lotsize = 31.8 -0.135*income
259     * 应用: 若某人 (income,lotsize)=(100,15), 则为 owner
260
261     *-其他设定:
262     *   y = 31.7-0.125*x
263     *   y = 31.8-0.135*x
264
265     *-Z-score(Stata操作)
266     discrim lda income lotsize, group(owner)
267     dropvars Zscore Zgroup
268     predict Zscore, dscore // Z-score
269     predict Zgroup // predicted group
270     gen str1 miss="x" if owner!=Zgroup // 误判
271
272     *-权重和判别函数

```

```

273     estat loadings, unstandardized
274     gen myscore = -11.96094 +0.0484468*income + 0.3795228*lotsize //判别函数
275
276     *-Z-score 及其应用
277     twoway (kdensity Zscore if owner==0) ///
278            (kdensity Zscore if owner==1), ///
279            xtitle("Zscore") ///
280            legend(label(1 NonOwner) label(2 Owner))
281
282     sum Zscore if owner==0
283     sum Zscore if owner==1
284     #delimit ;
285     twoway (kdensity Zscore if owner==0)
286            (kdensity Zscore if owner==1),
287            xline( 0.5646, lp(dash) lc(red*1.6))
288            xline(-0.6175, lp(dash) lc(blue*2.5))
289            xtitle("Zscore")
290            legend(label(1 NonOwner) label(2 Owner));
291     #delimit cr
292
293
294     *-----
295     *-Cleary(1999) 文中的 Z_FC index
296     *-----
297
298     use Cleary_1999_data, clear
299     qui tsset
300     dropvars L*
301     foreach v of varlist cr ni_margin sgr t1 size slack{
302         gen L`v' = L.`v'
303     }
304
305     discrim lda Lcr Lni_margin Lsgr Ltl Lsize ///
306                dum_yr3-dum_yr10, group(g_div)
307
308     dropvars Zscore Zgroup
309     predict Zscore, dscore // Z-score
310
311     predict Zgroup // predicted group
312     *-定义"数字-文字对应表"
313     label define Zgroup_lab 1 "Predict: increase DPS" ///
314                            2 "predict: decrease DPS" ///
315                            3 "Predict: nochange DPS"
316     label value Zgroup Zgroup_lab
317
318     *-权重
319     estat loadings, unstandardized
320
321
322
323     *-----
324     *- Table I Sample Summary Statistics (1998-2008)
325     *-----
326
327     *-样本分布
328     xtides
329     panels id
330     panels id: tab sicmen // 行业分布, (p.678)
331
332     capture drop SH
333     gen SH = id>=600000
334     panels id: tab SH // 深市/沪市分布
335
336
337     *-Panel A (pp.679)
338
339     *-基于股利增减分组结果

```

```

340     tabstat size cr tl ni_margin tobin sgr cflow invt, ///
341           by(g_div) f(%6.2f)
342
343     *-采用其它分组指标(我们自己设定的分组标准)
344
345     *-gsize: 根据公司在样本区间内的平均 size 分组, FC 状态不随时间变化
346     tabstat size cr tl ni_margin tobin sgr cflow invt, ///
347           by(gsize) f(%6.2f)
348
349     *-qsize: 每年分组, 允许公司的 FC 状态发生变化
350     tabstat size cr tl ni_margin tobin sgr cflow invt, ///
351           by(qsize) f(%6.2f)
352
353     *-Panel B
354     *-屏幕呈现
355     tabulate g_div year, column
356
357     *-输出到 Excel 表格中
358     label var year "Year" // logout 命令不支持中文
359     logout, save(Table_01_B) fix(4) excel replace: ///
360     tab g_div year, col
361
362
363     *-----
364     *- Table II Correlation among variables
365     *-----
366
367     pwcorr cflow cr tl invt tobin ni_m sgr Zscore, star(0.05)
368
369     *-连玉君自行定义的命令(存放于 $path/adofiles 文件夹下)
370     pwcorr_c cflow cr tl invt tobin ni_m sgr Zscore, ///
371           format(%6.2f) star(0.05)
372
373     pwcorr_a cflow cr tl invt tobin ni_m sgr Zscore
374
375
376
377     *-----
378     *- Table III Selected Financial Ratio Means p.683
379     *-----
380
381     *-Column 1-2
382     tabstat size cr tl ni_margin tobin sgr cflow invt Zscore, ///
383           by(Zgroup) labelwidth(25) f(%6.2f) // 效果并不好
384
385     *-Column 3-5
386     *-每个年度上, 根据 Zscore 值将样本公司分成三组
387     dropvars Zq33 Zq66 g_FC
388     bysort year: egen Zq33 = pctlile(Zscore) if Zscore!=., p(33)
389     bysort year: egen Zq66 = pctlile(Zscore) if Zscore!=., p(66)
390     gen g_FC = 2
391     replace g_FC=1 if (Zscore<=Zq33)
392     replace g_FC=3 if (Zscore>=Zq66&Zscore!=.)
393     *-数字-文字对应表
394     label define g_FC_lab 1 "FC" 2 "PFC" 3 "NFC"
395     label value g_FC g_FC_lab
396     *-分组统计分析
397     tabstat size cr tl ni_margin tobin sgr cflow invt Zscore, ///
398           by(g_FC) f(%6.2f)
399
400
401
402     *-----
403     *- Table IV Regression Results for the Total Sample p.686
404     *-----
405
406     *-----
407     *-A: Cleary99 的设定方法

```



```

408
409     global xx  "tobin cflow"    // 解释变量
410     global opt  ""              // 常规标准误
411     global opt  "cluster(id)"    // 稳健性标准误
412     qui tsset
413     xi: xtreg invt $xx i.year, fe
414     est store Full
415     xi: xtreg invt $xx i.year if g_FC==1, fe $opt    // FC 组
416     est store FC
417     xi: xtreg invt $xx i.year if g_FC==2, fe $opt
418     est store PFC
419     xi: xtreg invt $xx i.year if g_FC==3, fe $opt    // NFC 组
420     est store NFC
421
422     *-结果汇总
423     local ss "using Table_IV.csv"
424     local mm "Full FC PFC NFC"
425     esttab `mm' `ss', mtitle(`mm') replace compress nogaps ///
426             b(%6.3f) s(N r2_w r2_b r2_o) drop(*_I*)    ///
427             star(* 0.1 ** 0.05 *** 0.01)
428
429
430     *-----
431     *-B: 其他模型设定方式
432
433     global xx "tobin L.tobin cflow"    // 参见 Fazzari_1988.do
434     global xx "tobin sgr cflow"        // 同时控制近期和远期投资机会
435     global xx "L(0/1).tobin L.cflow"    // 采用解释变量的滞后项
436     qui tsset
437     xi: xtreg invt $xx i.year, fe
438     est store Full
439     xi: xtreg invt $xx i.year if g_FC==1, fe $opt    // FC 组
440     est store FC
441     xi: xtreg invt $xx i.year if g_FC==2, fe $opt
442     est store PFC
443     xi: xtreg invt $xx i.year if g_FC==3, fe $opt    // NFC 组
444     est store NFC
445
446     *-结果汇总
447     local mm "Full FC PFC NFC"
448     esttab `mm', mtitle(`mm') replace compress nogaps ///
449             b(%6.3f) s(N r2_w r2_b r2_o) drop(*_I*)    ///
450             star(* 0.1 ** 0.05 *** 0.01)
451
452
453     *-----
454     *-C: 采用公司规模分组  gsize v.s. qsize
455
456     global xx "L.tobin L.cflow"    // 采用解释变量的滞后项
457     global xx "tobin cflow"        // 采用解释变量的当期值
458     local group "gsize"            // 分组指标1: 公司 FC 状态不随时间变化
459     local group "qsize"            // 分组指标2: 公司 FC 状态随时间变化
460     qui tsset
461     xi: xtreg invt $xx i.year, fe
462     est store Full
463     xi: xtreg invt $xx i.year if `group'==1, fe $opt    // FC 组
464     est store FC
465     xi: xtreg invt $xx i.year if `group'==2, fe $opt
466     est store PFC
467     xi: xtreg invt $xx i.year if `group'==3, fe $opt    // NFC 组
468     est store NFC
469
470     *-结果汇总
471     local mm "Full FC PFC NFC"
472     esttab `mm' , mtitle(`mm') replace compress nogaps ///
473             b(%6.3f) s(N r2_w r2_b r2_o) drop(*_I*)    ///
474             star(* 0.1 ** 0.05 *** 0.01)
475

```

```

476    *-Comment: 得到的结论与 Cleary 相似,
477    *           即, NFC 组的投资-现金流敏感性反而更强.
478
479
480    *-----
481    *-采用 Bootstrap 获取组间系数差异    pp.648
482    *-----
483
484    *-----
485    *-A: 一些技术细节的说明
486    *-----
487
488    *-----
489    *-A1: 随机抽样: 从 4 个观察值中随机抽取三个
490
491    preserve
492    clear
493    set obs 4
494    gen x = _n
495    list x    // 原始的样本顺序
496    gen u = runiform() // u~U(0,1)
497    list x u  // 原始的样本顺序
498    sort u    // 打乱原始的排序方式
499    list x u  // 打乱后的样本顺序
500    restore
501
502    *-----
503    *-A2: 无放回 v.s. 有放回 抽样方法
504    help sample
505    help bsample
506
507    *-无放回
508    preserve
509    clear
510    set obs 6
511    gen x = _n
512    list, sep(0)
513    sample 3, count    // 无放回抽样
514    list
515    restore
516
517    *-有放回
518    preserve
519    clear
520    set obs 6
521    gen x = _n
522    list, sep(0)
523    bsample 3    // 有放回抽样(sample with replacement)
524    list
525    restore
526
527    *-----
528    *-A3: p 值的含义
529
530    twoway (function y=normalden(x), range(-4 4) lc(blue)) ///
531           (function y=normalden(x), range(-1.96 1.96) ///
532            recast(area) color(black*0.4)),          ///
533           xlabel(-4 -1.96 0 1.96 4) ///
534           text(0.2 0 "95%") ///
535           ytitle("") xtitle("") legend(off)
536
537    preserve
538    clear
539    set obs 10000
540    gen h = 170 + 12*rnatural()
541    sum h, detail
542    *-H10: 姚明的身高与中国男性平均身高无差异
543    count if h>226

```



```

544         dis "p-value = " %6.4f r(N)/10000
545     *--H20: 连玉君的身高与中国男性平均身高无差异
546         count if h<175
547         dis "p-value = " %6.4f r(N)/10000
548     restore
549
550
551 *-----
552 *--A4: Bootstrap 组间差异检验: 一个简单的例子
553
554     *--1: 数据生成过程
555     preserve
556         clear
557         set obs 100
558         set seed 1234567
559         gen e = rnormal() //干扰项, e~N(0,1)
560         gen x = 3*normal() //解释变量, x~N(0,9)
561         gen y = .
562         replace y = 10 + 0.4*x + e in 1/50 //被解释变量的数据生成过程(DGP)
563         replace y = 10 + 0.2*x + e in 51/100
564         twoway (scatter y x in 1/50) (lfit y x in 1/50) ///
565             (scatter y x in 51/100) (lfit y x in 51/100) ///
566             , legend(off)
567         save bs_eg.dta, replace
568     restore
569
570     *--2: 组间系数的真实差异
571     preserve
572         use bs_eg.dta, clear
573         reg y x in 1/50
574         local b1 = _b[x]
575         est store m1
576         reg y x in 51/100
577         local b2 = _b[x]
578         est store m2
579         esttab m1 m2, nogap r2
580         global d_true = `b1' - `b2'
581         dis "ture difference: " $d_true
582     restore
583
584     *--3: 检验组间差异
585     preserve
586         clear matrix
587         set matsize 1000
588         set seed 1358989 //设定种子值
589         mat D = J(1000,1,.) //存储系数差异的矩阵
590         forvalues i = 1/1000{ //抽样1000次
591             use bs_eg.dta, clear
592             gen u = runiform()
593             qui reg y x if u<0.5 // 随机抽取 50% 的观察值, 作为第一组
594             local b1 = _b[x]
595             qui reg y x if u>=0.5 // 剩下的 50% 的观察值, 作为第二组
596             local b2 = _b[x]
597             mat D[`i',1] = `b1' - `b2' //两组之间的系数差异
598         }
599         mat list D
600         svmat D, names(d_bs)
601         sort d_bs
602         list d_bs in 1/1000 //bootstrap 得到的系数差异
603         dis "ture difference: " $d_true
604     *--图示
605         twoway (kdensity d_bs) (function x=$d_true,xline($d_true))
606     *--Empirical p-value (经验 p 值)
607         count if d_bs>$d_true
608         local pvalue = r(N)/1000
609         dis "Empirical p-value = " %4.3f `pvalue'
610     restore
611

```

```

612      *-----
613      *-几点说明
614
615      *-检定力(Power)
616      *-请将第 511 行中的  $y = 10 + 0.2 \cdot x$  依次改写为如下 DGP, 结果有何差异?
617      *  $y = 10 + 0.30 \cdot x$ 
618      *  $y = 10 + 0.35 \cdot x$ 
619
620      *-检验结果具有一定的随机性, 因为抽样是随机的
621      *-请改变种子值, 即 set seed 13579 语句
622
623
624      *-----
625      *-B: 文中的组间差异检验实现过程
626      *-----
627
628      *-另存一份比较精简的数据
629      preserve
630          keep id year invt tobin cflow
631          save "Cleary_BS_temp.dta", replace
632      restore
633
634      *-记录真实系数差异
635
636      global xx "tobin cflow"
637      local gFC "g_FC" // 参考 Cleary, 基于 Z-score 分组
638      *local gFC "qsize" // 采用公司规模分组, 年度 FC 状态可变
639      *local gFC "gsize" // 采用公司规模分组, 年度 FC 状态不可变
640      qui tsset
641      xi: xtreg invt $xx i.year if `gFC'==1, fe // FC 公司
642          est store FC
643          mat b1 = e(b)
644      xi: xtreg invt $xx i.year if `gFC'==3, fe // NFC 公司
645          est store NFC
646          mat b3 = e(b)
647      mat D0 = b1 - b3
648      mat list D0, title(系数估计值的真实差异)
649      global cflow_diff0 = D0[1,2] // cflow 的真实系数差异
650
651      *-----
652      preserve
653          local reps = 100 // 抽样次数
654          local c = colsof(D0)
655          mat D = J(`reps', `c', .) // 存储结果的矩阵
656          forvalues j = 1/`reps'{
657              qui use "Cleary_BS_temp.dta", clear
658              dis _c "."
659              gen random = uniform()
660              sort random
661              bysort year: gen gg = group(3) // 将样本随机等分为三组
662              qui tsset
663              qui xi:xtreg invt $xx i.year if gg==1,fe //第一组, 视为 FC公司
664              matrix b1 = e(b)
665              est store FC
666              qui xi:xtreg invt $xx i.year if gg==3,fe //第三组, 视为 NFC公司
667              matrix b3 = e(b)
668              est store NFC
669              matrix diff = b1 - b3
670              mat D[`j',1] = diff
671          }
672      mat list D
673      svmat D, names(diff)
674      * diff1-diff2 对应 tobin, cflow 的系数差异
675      mat P = J(2,2,.) /*记录系数真实差异和实证P值的矩阵*/
676      forvalues j = 1/2{
677          local diff0_`j' = D0[1,`j']
678          qui count if (diff`j'>=`diff0_`j'')&(diff`j'!= .)
679          local p = `r(N)'/`reps'

```

```

680         if `p'>0.5{
681             local p = 1-`p' // 双尾检验
682         }
683         mat P[`j',1] = (`diff0_`j'' , `p')
684     }
685
686     mat colnames P = 系数真实差异 经验P值
687     mat rownames P = tobin cflow
688     mat list P
689     restore
690     *-----
691
692     *-结果对比重现
693     local mm "FC NFC"
694     esttab `mm' , mtitle(`mm') replace compress nogaps ///
695         b(%6.3f) s(N r2_w r2_b r2_o) drop(*_I*) ///
696         star(* 0.1 ** 0.05 *** 0.01)
697
698     *-Bootstrap 结果
699     mat colnames D = tobin cflow yrdumy
700     mat list D // coll-col2 对应 tobin,cflow 的系数差异
701
702     *-图解 empirical p-value: kdensity
703     dropvars *diff*
704     svmat D, name(diff)
705     rename diff2 cflow_diff_bs
706     kdensity cflow_diff_bs, xline($cflow_diff0)
707
708     *-图解 empirical p-value: scatter
709     sort cflow_diff_bs
710     dropvars ii
711     gen ii = 5
712     scatter cflow_diff_bs ii in 1/100, yline($cflow_diff0) ///
713         ylabel(-0.05(0.01)0.05, angle(0)) msymbol(dh)
714
715
716     *-----
717     *-Appendix: PPT 中的图片
718     *
719     use lawnmower2, clear
720     replace lotsize = lotsize +12
721     replace income = income/10+5
722     label var income "年收入(万元)"
723     label var lotsize "年龄"
724     twoway (scatter lotsize income if owner==0,msymbol(T)) ///
725         (scatter lotsize income if owner==1), ///
726         legend(label(1 无车) label(2 有车))
727     *-----
728
729
730

```

```

1
2  *      -----
3  *      -----
4  *
5  *      Stata 学术论文专题
6  *
7  *      -----
8  *      -----
9
10
11  *      主讲人：连玉君 副教授
12  *
13  *      单 位：中山大学岭南学院金融系
14  *      电 邮：arlionn@163.com
15  *      博 客：http://blog.cnfol.com/arlionn
16  *      主 页：http://goo.gl/tRXba
17  *      微 博：http://weibo.com/arlionn
18
19
20  *-----
21  *-Faulkender, 2006, JF-
22  *   Faulkender, M., R. Wang, 2006,
23  *   Corporate Financial Policy and the Value of Cash,
24  *   Journal of Finance, 61(4): 1957-1990.
25  *-----
26
27  *-----
28  *-与本专题中的关联论文：
29  *   Opler et al. (1999,JFE)      || Opler_1999.do
30  *   Fazzari et al. (1988,BPEA)  || Fazzari_1988.do
31  *   Cleary (1999,JF)           || Cleary_1999.do
32
33  *-----
34  *-主要方法
35  *   OLS: regress
36  *   FE (Fixed Effect Model): xtreg, fe
37  *   交叉项的设定和解释
38  *   构造 5x5 投资组合，求取组合收益率
39  *   如何进行稳健性分析
40
41  *-----
42  *-设定课程存储路径
43  *   global path "`c(sysdir_personal)'PX_papers" //请先执行该命令
44
45  *-----
46  *-原始论文
47  *   shellout "$path\Refs\Faulkender_2006.pdf"
48
49  *-----
50  *-PPT
51  *   shellout "$path\PPT\Faulkender_2006.ppt"
52
53
54  *-----
55  *-相关文献
56  *-----
57  /*
58  Fama, E., K. French, 1998,
59  Taxes, Financing Decisions, and Firm Value,
60  Journal of Finance, 53(3): 819-843.
61  shellout "$path\Refs\Fama_French_1998.pdf" // PDF原文
62  Pinkowitz, L., R. Stulz, R. Williamson, 2006,
63  Does the Contribution of Corporate Cash Holdings and Dividends to Firm Value
64  Depend on Governance A Cross-country Analysis,
65  Journal of Finance, 61(6): 2725-2751.
66  shellout "$path\Refs\Pinkowitz_2006_JF.pdf" // PDF原文
67  Dittmar, A., J. Mahrt-Smith, 2007,
68  Corporate Governance and the Value of Cash Holdings,

```

```

69     Journal of Financial Economics, 83(3): 599-634.
70     shellout "$path\Refs\Dittmar_2007_JFE.pdf"      // PDF原文
71     Denis, D., V. Sibilkov, 2009,
72     Financial Constraints, Investment, and the Value of Cash Holdings,
73     Review of Financial Studies, 23(1): 247-269.
74     shellout "$path\Refs\Denis_2009_RFS.pdf"        // PDF原文
75     Frésard, L., C. Salva, 2010,
76     The Value of Excess Cash and Corporate Governance:
77     Evidence from U.S. Cross-listings,
78     Journal of Financial Economics, 98(2): 359-384.
79     shellout "$path\Refs\Fresard_2007_JFE.pdf"      // PDF原文
80     Lee, E., R. Powell, 2011,
81     Excess Cash Holdings and Shareholder Value,
82     Accounting & Finance, 51(2): 549-574.
83     沈艺峰, 况学文, 聂亚娟, 2008,
84     终极控股股东超额控制与现金持有量价值的实证研究,
85     南开管理评论, (1): 15-23.
86     杨兴全, 张照南, 2008,
87     制度背景, 股权性质与公司持有现金价值,
88     经济研究, (12): 111-123.
89     罗琦, 秦国楼, 2009,
90     投资者保护与公司现金持有,
91     金融研究, (10): 162-178.
92     */
93     * 如何追踪这些信息? Google学术搜索: http://scholar.google.com/
94
95
96
97
98     *=====
99     * 样本筛选
100    *=====
101
102    cd "$path\Faulkender_2006"          // 进入本讲目录
103
104    clear all
105    set memory 80m                      // 设定 Stata 占用的内存空间
106
107    use "$path\data\GTA2008.dta", clear //调入数据
108
109
110    *=====
111    * 定义变量 pp.1968-1969
112    *=====
113
114    *-I:
115        tsset id year
116        gen I = b001211000             //财务费用
117        label var I "利息支出"        //只能用此替代了
118        gen DI = D.I/L.mv_total
119
120    *-D
121        replace div_cash=0 if div_cash==. //通常都是如此处理,又如 R&D
122        gen D = div_cash*nshra         //每股现金股利*A股流通股本
123        label var D "现金股利"
124        gen DD = D.D/L.mv_total
125
126    *-DC, LC
127        gen DC = D.a001101000/L.mv_total // 货币资金
128        gen LC = L.a001101000/L.mv_total
129
130    *-DNA
131        gen NA = (a001000000-a001101000) // net_Asset = Asset-Cash
132        gen DNA= D.NA/L.mv_total
133
134        gen L   = a002000000/(a002000000+mv_total) //负债合计
135
136        gen E   = b001000000+ b001211000 //利润总额+财务费用

```

```

137     gen DE = D.E/L.mv_total
138
139     rename a16 tobin
140
141     rename a001000000 TA // 总资产
142     gen size = ln(TA) // ln(总资产)
143
144     gen capExp = c002006000/L.mv_total
145     // 购建固定资产、无形资产和其他长期资产支付的现金/市值
146
147     gen nwc = (a001100000-a002100000-a001101000)/L.mv_total
148     // [(营运资本=流动资产-流动负债)-货币资金]/市值
149
150     gen cflow = c001000000/L.mv_total // 经营活动产生的现金流净额
151
152     gen shorttl = a002100000/L.mv_total // 流动负债
153
154     gen coverage= (b001000000+a001101000)/b001211000
155     // (利润总额+货币资金)/财务费用 = 利润加现金除以利息
156
157     gen ln_sale = ln(b001101000) // 营业收入
158
159     /* 如果硬盘紧张，可以执行这段语句
160     keep id year sicda sicmen_str TA rr div_cash I ///
161         D E DE DI DD DC NA DNA L LC E cflow size ///
162         tobin div_cash div_yes capExp shorttl nwc ///
163         mv_atshr mv_total coverage ln_sale
164     */
165
166
167     *=====
168     * 样本筛选
169     *=====
170
171     *-剔除金融类企业和公用事业企业 (pp.1969)
172     drop if sicmen_str == "I"
173     *drop if sicmen_str == "F"
174
175     drop if NA<0
176     drop if mv_total<0
177     drop if D<0
178     drop if year<1998
179
180     *-计算个股年均回报率
181     gen return = D.mv_atshr/L.mv_atshr
182
183     *-Winsor 处理
184     global x1 "DI DD LC DC DNA L DE return"
185     global x2 "nwc cflow shorttl size tobin capExp"
186     foreach v of varlist $x1 $x2{
187         rename `v' `v'_a
188         winsor `v'_a, gen(`v') p(0.01)
189         drop `v'_a
190     }
191
192     *-交乘项
193     gen LCxDC = LC*DC
194     gen LxDC = L*DC
195
196     *-----
197     *-计算 5x5 组合收益率
198     *-----
199     dropvars MV_size g_MV g_tobin
200     gen MV_size = ln(mv_total)
201
202     *-每个年度内，按 Market Value 分成五组
203     qui tsset
204     bysort year: quantiles MV_size, gen(g_MV) n(5)

```



```

205     *-每个年度内, 按 Market-to-Book ratio (Tobin) 分成五组
206     qui tsset
207     bysort year: quantiles tobin, gen(g_tobin) n(5)
208
209     /* 旧版本的做法: 使用恰当的命令可以大幅提高效率呀!
210     foreach v of varlist MV_size tobin{
211         forvalues i=1999(1)2008{
212             qui{
213                 centile `v' if year==`i', centile(20(20)100)
214                 qui replace `v'_id=1 if `v'<=r(c_1)& year==`i'
215                 qui replace `v'_id=2 if `v'>r(c_1) & `v'<=r(c_2) & year==`i'
216                 qui replace `v'_id=3 if `v'>r(c_2) & `v'<=r(c_3) & year==`i'
217                 qui replace `v'_id=4 if `v'>r(c_3) & `v'<=r(c_4) & year==`i'
218                 qui replace `v'_id=5 if `v'>r(c_4)& `v'<=r(c_5) & year==`i'
219             }
220         }
221     }
222     */
223
224     *-每个年度内, 计算 5x5 组合的市值加权平均收益 = 10*5*5 组基准收益
225     cap drop bench
226     gen bench = .
227     forvalues i=1999(1)2008{
228         forvalues j=1(1)5{
229             forvalues k=1(1)5{
230                 qui sum return [weight=mv_atshr] ///
231                 if (year==`i' & g_MV==`j' & g_tobin==`k')
232                 qui replace bench = r(mean) ///
233                 if (year==`i' & g_MV==`j' & g_tobin==`k')
234             }
235         }
236     }
237     sort year g_MV g_tobin
238     br year g_MV g_tobin return bench
239     *-计算超额收益率
240     gen overRet = return-bench // 超额收益率
241
242     *-winsor at 1% tails
243     rename overRet overRet0
244     winsor overRet0, gen(overRet) p(0.01)
245     drop overRet0
246
247
248     *=====
249     * Table 1 : Summary Statistics
250     *=====
251
252     *-删除缺漏值, 以保证 Tab.1 与后续表格中的观察值数目相同
253     local v "overRet DC LC DE DNA DI DD L"
254     sum `v' //可以粗略看出缺漏值情况
255     qui reg `v'
256     keep if e(sample)
257     sum `v' //处理后的观察值情况
258
259     *-基本统计量
260     local v "overRet DC LC DE DNA DI DD L"
261     logout, save(Table01) excel replace: ///
262     tabstat `v', s(mean q sd min max) f(%10.4f) c(s)
263
264
265     *=====
266     * Table 2 : Regression Results for the Whole Sample
267     *=====
268     *-Note: Col.3 is not reported due to data limit
269
270     *-----
271     *-S.E. 的设定 see: help reg; help vce_option
272

```

```

273 *-White heteroscedastic-consistent standard errors, corrected for
274 * correlation across observations of a given firm (White (1980)).
275
276     global robust "vce(cluster id)" //本文采用这个
277
278 *-White(1980) heteroscedastic-consistent standard errors, no cluster
279
280 * global robust "vce(robust)" //普通稳健性标准误
281 * global robust "vce(ols)" //不做任何调整
282
283 *-----
284 *--年度虚拟变量 pp.1974, Footnote 20
285     tsset id year
286     dropvars yr_dum*
287     qui tab year, gen(yr_dum)
288     drop yr_dum1
289
290     local yr "*yr_dum*" // Robust Check, year effects, pp.1974, Footnote.20
291                        // 验证 Table.2 的结果时, 不要包含此行
292
293 *-(1)
294     reg overRet DC DE DNA DI LC L `yr', $robust
295     est store m_21_OLS
296     xtreg overRet DC DE DNA DI LC L `yr', $robust fe
297     est store m_21_fe
298
299 *-(2)
300     reg overRet DC DE DNA DI LC L `yr' LCxDC LxDC, $robust
301     est store m_22_OLS
302     xtreg overRet DC DE DNA DI LC L `yr' LCxDC LxDC, $robust fe
303     est store m_22_fe
304
305 *--结果呈现
306     local s "using Table02.csv" // 控制是否输出 Excel 表格
307     local m "m_21_OLS m_22_OLS m_21_fe m_22_fe"
308     local mt "OLS OLS FE FE"
309     esttab `m' `s', mtitle(`mt') b(%6.3f) replace ///
310             compress nogaps scalar(N r2_a r2_w) ///
311             star(* 0.1 ** 0.05 *** 0.01) drop(`yr')
312
313 *-----
314 *--Comments:
315 *-----
316 *
317 *-(1) 采用 OLS 估计更为合理, 因为所有变量基本上都是采用 D.x 的形式
318 *
319 * (2) DI 不显著, 有一个可能的原因是我们的指标存在严重的衡量偏误
320 *
321 *-(3) 此前没有处理 LCxDC, LxDC 的离群值时, 二者的推断结论恰好相反,
322 *     DC 前面的系数与此处报告的也有差异
323
324 *-(4) Year effect: 执行时包含第 290 行即可 pp.1974, Footnote 20
325
326 *-(5) 文中 Col.(2) Cash value 的计算: pp.1972, 5%cash v.s. 15%cash
327     reg overRet DC DE DNA DI LC L LCxDC LxDC
328 *--The marginal value of cash for a firm with 5% Cash-to-Equity
329     local vc_5per = _b[DC] + 0.05*_b[LCxDC]
330     dis in g "The marginal value of Cash (at 5%) = " in y %6.3f `vc_5per'
331 *--The marginal value of cash for a firm with 15% Cash-to-Equity
332     local vc_15per = _b[DC] + 0.15*_b[LCxDC]
333     dis in g "The marginal value of Cash (at 15%) = " in y %6.3f `vc_15per'
334
335 *-(6) 文中 Col.(2) Cash value 的计算: pp.1974, $0.94
336     sum LC
337     local mean_LC = r(mean)
338     sum L
339     local mean_L = r(mean)
340     qui reg overRet DC DE DNA DI LC L LCxDC LxDC

```

```

341     local vc = _b[DC] + `mean_LC'*_b[LCxDC] + `mean_L'*_b[LxDC]
342
343     dis in g "The marginal value of Cash = " in y %6.3f `vc'
344
345
346     *=====
347     * Table 3      : Regressions with Alternative Definitions of the
348     *                  Expected Change in Cash Holdings
349     *=====
350
351     *-(1): Normal Cash: 5*5 组合的平均 Cash
352
353     cap drop av_cash_bech
354     bysort year g_MV g_tobin: egen av_cash_bech = mean(DC)
355     dropvars *DC1*
356     gen DC1      = DC - av_cash_bech
357     gen LCxDC1   = LC*DC1
358     gen LxDC1    = L*DC1
359
360     reg overRet DC1 DE DNA DI LC L LCxDC1 LxDC1, $robust
361     est store m_31
362
363
364     *-(2) Norm Cash: Almeida et al.(2004,JF) 精简模型为基础
365     shellout "$path\Refs\Almeida_2004.pdf" // PDF, 查看模型
366
367     global xx "L.cflow L.tobin L.size"
368     dropvars *DC2*
369     gen DC2 = .
370     forvalues t = 2001/2008{
371         qui tsset
372         *qui reg DC $xx if year==`t' // 无行业虚拟变量
373         qui xi: reg DC $xx i.sicda if year==`t' // 有行业虚拟变量
374         cap drop res
375         qui predict res, res
376         qui replace DC2 = res if year==`t'
377     }
378
379     gen LCxDC2 = LC*DC2
380     gen LxDC2  = L*DC2
381
382     reg overRet DC2 DE DNA DI LC L LCxDC2 LxDC2, $robust
383     est store m_32
384
385
386     *-(3) Norm Cash: Almeida et al.(2004,JF) 完整模型为基础
387
388     global xx "L.cflow L.tobin L.size L.capExp D.nwc D.shortttl"
389     dropvars *DC3*
390     gen DC3 = .
391     forvalues t = 2001/2008{
392         *qui reg DC $xx if year==`t' // 无行业虚拟变量
393         qui xi: reg DC $xx i.sicda if year==`t' // 有行业虚拟变量
394         cap drop res
395         qui predict res, res
396         replace DC3 = res if year==`t'
397     }
398
399     gen LCxDC3 = LC*DC3
400     gen LxDC3  = L*DC3
401
402     reg overRet DC3 DE DNA DI LC L LCxDC3 LxDC3, $robust
403     est store m_33
404
405     *-结果呈现
406     local s "using Table03.csv"
407     local m "m_31 m_32 m_33"
408     esttab `m' `s', replace compress nogaps b(%6.3f) ///

```

```

409          mtitle(Portf.Ave ACW(1) ACW(2))          ///
410          star(* 0.1 ** 0.05 *** 0.01)            ///
411          scalar(N r2) order(DC1 DC2 DC3)
412
413
414
415  *=====
416  * Table 4 : Summary Statistics for Constrained and Unconstrained Groups
417  *=====
418
419  *-----
420  *-定义融资约束分组指标
421  *-----
422
423  *-根据公司规模(ln_sale)分组
424
425  * 参见 Faulkender and Wang(2006, p.1971)
426  * 根据样本公司在1998-2008期间各年度的 ln_sale 排序分组,
427  * 同一家公司在不同年度所属的组别可能不同;
428  * 我们将样本公司分成 3 个组,
429  * 其中第一组定义为 FC 组, 第三组定义为 NFC 组
430  *
431  *-Also see: Cleary_1999.do (Line 177-206)
432
433  dropvars g_size FC_size NFC_size Other_size
434  qui tsset
435  bysort year: quantiles ln_sale, gen(g_size) n(3)
436  //可将 ln_sale 替换为 size 作为稳健性检验
437  gen FC_size = g_size==1
438  gen NFC_size = g_size==3 // 如此定义便于后续循环操作
439  gen Other_size = 1-FC_size-NFC_size
440
441  *-分组计算基本统计量 (size=ln_sale)
442  *
443  *-参见 Opler_1999.do (Table 3, Line 437-486)
444
445  local vars "DC LC DE DNA DI DD L ln_sale"
446  local nv = wordcount("`vars'")*2
447  local g_var "g_size"
448  mat aa = J(`nv', 2, 0)
449  local i = 1
450  foreach var of varlist `vars'{
451      local c = 1
452      forvalues j=1(2)3{
453          qui sum `var' if `g_var'==`j', d
454          mat aa[`i', `c++'] = (r(mean) \ r(p50))
455      }
456      local i = `i'+2
457  }
458  mat rownames aa = DC p50 LC p50 DE p50 DNA p50 DI p50 DD p50 L p50 sale p50 ///
459  mat colnames aa = Size(C) Size(U)
460  capture mat drop aa_size
461  mat rename aa aa_size
462  mat list aa_size, format(%6.4f)
463
464
465  *-根据股利支付率分组 (div_Num)
466  *
467  *-参见: Fazzari_1988.do (Line 179-194)
468
469  * Num_div 是在1998-2008期间, 每家公司支付股利的次数
470  * 若Num_div<=2,则定义为FC 组;
471  * 若Num_div>=8,则定义为NFC组;
472  dropvars Num_div FC_div NFC_div Other_div g_div
473  bysort id: egen Num_div = sum(div_yes)
474  *recode Num_div (min/2=1) (2/8=2) (8/max=3), gen(g_div)
475  gen FC_div = Num_div<=2

```

```

477     gen NFC_div = Num_div>=8
478     gen Other_div = 1 - FC_div - NFC_div
479     gen g_div = 3
480     replace g_div=1 if FC_div ==1 // FC
481     replace g_div=2 if NFC_div ==1 // NFC
482
483     *-分组计算基本统计量 (div_Num)
484     local vars "DC LC DE DNA DI DD L ln_sale"
485     local nv = wordcount("`vars'")*2
486     local g_var "g_div"
487     mat aa = J(`nv', 2, 0)
488     local i = 1
489     foreach var of varlist `vars'{
490         forvalues j=1(1)2{
491             qui sum `var' if `g_var'==`j', d
492             mat aa[`i',`j'] = (r(mean) \ r(p50))
493         }
494         local i = `i'+2
495     }
496     mat rownames aa = DC p50 LC p50 DE p50 DNA p50 DI p50 //
497                     DD p50 L p50 sale p50
498     mat colnames aa = Div_num(C) Div_num(U)
499     mat aa_Div_num = aa
500     mat list aa_Div_num, format(%6.4f)
501
502
503     *-根据股利变化指标 (参考 Cleary,1999,JF)
504     *
505     *-参见: Cleary_1999.do (Line 111-125)
506
507     dropvars Ddiv NFC_divchg FC_divchg g_divchg
508     tsset id year
509     replace div_cash=0 if div_cash==.
510     gen Ddiv = D.div_cash
511     gen NFC_divchg = 1 if Ddiv>0 // 增加股利的公司-年度
512     gen FC_divchg = 1 if Ddiv<0 // 减少股利的公司-年度
513     gen g_divchg = 3
514     replace g_divchg = 1 if FC_divchg==1
515     replace g_divchg = 2 if NFC_divchg==1
516
517     *-分组计算基本统计量 (divchg)
518     local vars "DC LC DE DNA DI DD L ln_sale"
519     local nv = wordcount("`vars'")*2
520     local g_var "g_divchg"
521     mat aa = J(`nv', 2, 0)
522     local i = 1
523     foreach var of varlist `vars'{
524         forvalues j=1(1)2{
525             qui sum `var' if `g_var'==`j', d
526             mat aa[`i',`j'] = (r(mean) \ r(p50))
527         }
528         local i = `i'+2
529     }
530     mat rownames aa = DC p50 LC p50 DE p50 DNA p50 DI p50 //
531                     DD p50 L p50 sale p50
532     mat colnames aa = Div_chg(C) Div_chg(U)
533     mat aa_Div_chg = aa
534     mat list aa_Div_chg, format(%6.4f)
535
536
537     *-综合列表 (Final Results)
538     mat Table4 = (aa_size, aa_Div_num, aa_Div_chg)
539     *-屏幕呈现
540     mat list Table4, format(%6.4f)
541     *-输出到 Excel
542     logout, save(Table04) excel replace: ///
543     mat list Table4, format(%6.4f)
544

```

```

545  *--Comments-
546  *
547  *-(1) 融资约束程度的划分是文献中颇具争议的问题;
548  *-(2) 就我本人看来, ln(Assets), ln(Sales) 是不错的分组指标
549  *-(3) 虽然英文文献中经常使用 dividend payment 分组, 但对于中国上市公司基本上不适用
550  *-(4) 在上表中, L 在两组的统计结果有时是相反的.
551
552
553
554  *=====
555  * Table 5 : Regressions for Constrained and Unconstrained Groups
556  *=====
557
558  *--Goals: 比较一下 FC v.s NFC 公司的现金价值有何差异?
559
560      global xx  "overRet DC DE DNA DI LC L LCxDC LxDC"
561
562  *--Size(ln_sale)
563      reg $xx if FC_size==1, $robust
564      est store FC_size
565      reg $xx if NFC_size==1, $robust
566      est store NFC_size
567
568  *--Div_num
569      reg $xx if FC_div==1, $robust
570      est store FC_divNum
571      reg $xx if NFC_div==1, $robust
572      est store NFC_divNum
573
574  *--Div_chg
575      reg $xx if FC_divchg==1, $robust
576      est store FC_divchg
577      reg $xx if NFC_divchg==1, $robust
578      est store NFC_divchg
579
580  *--呈现结果
581      local s  "using Table05.csv"  //屏幕呈现时, 无需选中本行
582      local m  "FC_size NFC_size FC_divNum NFC_divNum FC_divchg NFC_divchg"
583      local mt "Size(C) Size(U) Payout(C) Payout(U) Divchg(C) Divchg(U)"
584      esttab `m' `s', replace b(%6.3f) compress nogaps  ///
585      mtitle(`mt') scalar(N r2_a) star(* 0.1 ** 0.05 *** 0.01)
586
587  *-----
588  *--Comments
589  *-----
590  *-(1) 局限1: 无法直接对比 c 和 u 组现金持有边际市场价值的大小;
591  *-(2) 局限2: 组间差异检验没有执行;
592
593
594  *=====
595  * Table 5 : 完整结果, 进一步呈现 Cash 的平均边际价值
596  *=====
597
598      global xx  "overRet DC DE DNA DI LC L LCxDC LxDC"
599
600  *-(1) 把边际市场价值的计算结果直接加入回归结果中
601      reg $xx if FC_size==1, $robust
602      sum LC if e(sample)
603      local mean_LC = r(mean)
604      sum L  if e(sample)
605      local mean_L  = r(mean)
606      local vc = _b[DC] + `mean_LC'*_b[LCxDC] + `mean_L'*_b[LxDC]
607      estadd scalar mvc = `vc'  // marginal value of cash
608      est store FC_size
609      esttab FC_size, nogap s(r2_a mvc) // 注意最后一行的变化
610
611  *--更为简洁的处理方法: 把上面的代码包装成一个小程序
612  *--使用方法: 选中[begin]-[over]之间的语句, 点击[run], 快捷键: Ctrl+R

```



```

613 *-----begin-----
614     cap program drop mvc
615     program define mvc
616     version 11.2
617         sum LC if e(sample)
618         local mean_LC = r(mean)
619         sum L   if e(sample)
620         local mean_L   = r(mean)
621         local vc = _b[DC] + `mean_LC'*_b[LCxDC] + `mean_L'*_b[LxDC]
622         dis _n in g "The marginal value of Cash = " in y %6.3f `vc'
623         estadd scalar mvc = `vc' // marginal value of cash
624                                 // 添加到 Stata 的返回值列表中
625     end
626 *-----over-----
627
628 *-(2) 检验组间系数差异
629     reg $xx if FC_size==1 // FC 公司
630     est store FC_size
631     reg $xx if NFC_size==1 // NFC 公司
632     est store NFC_size
633     suest FC_size NFC_size, $roburst
634     test [FC_size_mean]DC = [NFC_size_mean]DC
635     return list // 内存中存储的返回值
636
637 *-封装成一个小程序
638 *-使用方法: 选中[begin]-[over]之间的语句, 点击[run], 快捷键: Ctrl+R
639 *-----begin-----
640     cap program drop test_b
641     program define test_b, rclass
642     version 11.2
643         args groupvar testvar // 如, groupvar=FC_size; testvar=DC
644         qui reg $xx if `groupvar'==1 // FC 公司
645         est store FC
646         qui reg $xx if N`groupvar'==1 // NFC 公司
647         est store NFC
648         qui suest FC NFC, $roburst
649         test [FC_mean]`testvar' = [NFC_mean]`testvar'
650         return scalar chi2v = r(chi2) // 把返回值存在内存中
651         return scalar p_value = r(p)
652     end
653 *-----over-----
654 *
655 *-应用范例:
656     test_b FC_size DC // 检验基于 size 分组时, DC 的组间系数差异
657     return list // 查看返回值
658     test_b FC_div LC // 检验基于 div_num 分组时, LC 的组间系数差异
659     return list
660
661 *-Note: 亦可采用 Bootstrap 执行组间差异检验,
662 *      参见 Cleary_1999.do (Line 480-624-713)
663
664 *-----
665 *-Size(ln_sale)
666 *-----
667 *-组间差异检验
668     test_b FC_size DC //原文 Table V, 第一个 [p-value(C-U!=0)]
669     global p_DC = r(p_value)
670     test_b FC_size LC //原文 Table V, 第二个 [p-value(C-U!=0)]
671     global p_LC = r(p_value)
672     est drop FC NFC
673 *-估计结果
674     reg $xx if FC_size==1, $roburst // FC 公司
675     mvc // 执行上面的程序, 计算现金的边际市场价值
676     est store FC_size
677     reg $xx if NFC_size==1, $roburst // NFC 公司
678     mvc
679     estadd scalar p_DC = $p_DC
680     estadd scalar p_LC = $p_LC

```

```

681         est store NFC_size
682
683     *-----
684     *-Div_num
685     *-----
686     *-组间差异检验
687         test_b FC_div DC
688         global p_DC = r(p_value)
689         test_b FC_div LC
690         global p_LC = r(p_value)
691         est drop FC NFC
692     *-估计结果
693         reg $xx if FC_div==1, $robust    // FC 公司
694         mvc
695         est store FC_divNum
696         reg $xx if NFC_div==1, $robust    // NFC 公司
697         mvc
698         estadd scalar p_DC = $p_DC
699         estadd scalar p_LC = $p_LC
700         est store NFC_divNum
701
702     *-----
703     *-Div_chg
704     *-----
705     *-组间差异检验
706         test_b FC_divchg DC
707         global p_DC = r(p_value)
708         test_b FC_divchg LC
709         global p_LC = r(p_value)
710         est drop FC NFC
711     *-估计结果
712         reg $xx if FC_divchg==1, $robust    // FC 公司
713         mvc
714         est store FC_divchg
715         reg $xx if NFC_divchg==1, $robust    // NFC 公司
716         mvc
717         estadd scalar p_DC = $p_DC
718         estadd scalar p_LC = $p_LC
719         est store NFC_divchg
720
721     *-----
722     *-呈现结果
723         local s "using Table05_full.csv"
724         local m "FC_size NFC_size FC_divNum NFC_divNum FC_divchg NFC_divchg"
725         local mt "Size(C) Size(U) Payout(C) Payout(U) Divchg(C) Divchg(U)"
726         esttab `m' `s', replace b(%6.3f) compress nogaps ///
727         mtitle(`mt') star(* 0.1 ** 0.05 *** 0.01)    ///
728         scalar(N r2_a mvc p_DC p_LC)    ///
729         addnote("Note: mvc = _b[DC] + mean_LC*_b[LCxDC] + mean_L*_b[LxDC]")
730
731     *-----
732     *-Comments
733
734     *(1) 融资约束指标的选取对结果有很大的影响, 尽量以前期文献为依据, 考虑样本特征;
735     *      练习: 采用 ln(TA) 进行分组; 采用总负债率进行分组(see Hansen_1999.do)
736
737     *(2) 可以进一步做稳健性检验: 用 DC1-DC3 替代此处的 DC (参见原文
pp.1982, 第二段)
738
739     *(3) 可以进一步细化分组方式, 参见原文后续部分
740
741
742
743
744     *-----以下部分(略)-----
745
746

```

```

1
2  *      -----
3  *      -----
4  *
5  *      Stata 学术论文专题
6  *
7  *      -----
8  *      -----
9
10
11  *      主讲人：连玉君 副教授
12  *
13  *      单 位：中山大学岭南学院金融系
14  *      电 邮：arlionn@163.com
15  *      博 客：http://blog.cnfol.com/arlionn
16  *      主 页：http://goo.gl/tRXba
17  *      微 博：http://weibo.com/arlionn
18
19
20  *-----
21  * -Fazzari,1988,BPEA -
22  *   Fazzari, S., R. Hubbard, B. Petersen, 1988,
23  *   Financing constraints and corporate investment,
24  *   Brookings Papers on Economic Activity, 1988(1):141-206.
25  *-----
26  *-主要方法
27  *- 提出了一种简便的绕开衡量偏误的问题：分组检验(这个非常重要!)
28  *- OLS: regress
29  *- FE (Fixed Effect Model): xtreg, fe
30  *- 衡量偏误的处理方法：差分
31  *- 如何进行稳健性分析(模型设定、时间跨度等)
32
33  *-相关文献：
34  * Cleary (1999):      Cleary_1999.do
35  * Hansen(1999):      Hansen_1999.do
36  * Opler et al.(1999): Opler_1999.do
37  * Faulkender and Wang(2006): Faulkender_2006.do
38
39
40  *-设定课程存储路径
41  global path "`c(sysdir_personal)'PX_papers" //请先执行该命令
42
43  *-PPT
44  shellout "$path\PPT\Fazzari_1988.ppt"
45
46  *-原始论文
47  shellout "$path\Refs\Fazzari_1988.pdf"
48
49
50
51  *=====
52  * 样本筛选
53  *=====
54  *
55  *-原文数据处理过程：pp.191 (Appendix B)
56  *
57  *-Data source: annual Value Line data base
58  *-Sample firms: manufacturing firms (two-digit SIC codes #20-#39)
59  *-Balanced Panel Data: 1969-1984
60  *-Why begin 1969?: it is the staring point of inventory data
61  *-Why end 1984?: few firms with observations in 1985
62  *
63  *-Observations dropped:
64  *
65  *-M&A: mergers valued >10% of their capital stock
66  *-Missing values
67  *-The definitions of variables are discussed in detail
68

```

```

69
70     global path "`c(sysdir_personal)'PX_papers"    //课程总文件夹
71
72     clear all
73     set memory 80m    // 设定 Stata 占用的内存空间
74
75     cd "$path\Fazzari_1988"                        // 进入本讲目录
76     use "$path\data\GTA2008_my.dta", clear        //调入数据
77
78     keep id year tobin size t1 TA TD ll sgr tagr ///
79         div_cash div_yes sicda sicmen            ///
80         b001101000 c001000000 a001212000        ///
81         a001101000 c002006000 a001100000        ///
82         a002100000 a001101000 a001123000        ///
83         b002000000 a003000000
84     rename a001212000 K
85
86     keep if year>1997    // 1998年以后采用现金流量表
87
88     *-----
89     *-定义变量
90
91     * 说明：只有采用 TA 标准化，才能得到与 Fazzari 相似的结果，
92     *       采用 L.TA、L.K 或 K，都无法得到这一结果
93
94     local fm "TA"        // 采用期初总资产标准化
95     *local fm "K"        // 采用期初固定资产净值标准化
96     gen cflow= c001000000/`fm'    // 经营活动产生的现金流净额/fm
97     gen cash = a001101000/`fm'    // 货币资金/fm
98     gen invt = c002006000/`fm'    //
99     构建固定资产、无形资产和其他长期资产支付的现金/fm
100    gen sale = b001101000/`fm'    // 营业收入/fm
101    gen wcmi = ((a001100000-a002100000)-a001123000)/`fm'
102
103    //(营运资本-存货净额)/总资产，营运资本=流动资产-流动负债
104    gen roa = b002000000/TA        // 净利润/总资产
105    gen roe = b002000000/a003000000 // 净利润/所有者权益合计
106    tsset id year
107
108    order id year invt cflow tobin sale cash wcmi
109
110    *-----
111    *-负债率的离群值
112    sum t1, detail
113    count if t1>1
114    drop if t1>1    //trim 处理
115    *replace t1=1 if t1>1 //winsor 处理
116
117    *-----
118    *-M&A (兼并收购)
119    des tagr
120    sum tagr, detail
121    histogram tagr
122    count if tagr>1.5
123    drop if (tagr>1.5&tagr!=.)
124    *-Note: 这个完全是主观判断了，这个处理对结果有何影响？
125
126    *-----
127    *-处理为平行面板
128    xtbalance, range(1999 2008) miss(invt cflow tobin cash sale wcmi)
129
130    *-如果不采用平行面板，可以采用如下方式删除缺漏值
131    * local v "invt cflow tobin cash sale wcmi"
132    * qui reg `v'
133    * keep if e(sample)
134
135    *-----
136    *-Winsor 处理

```

```

135     foreach v of varlist invt-wcml{
136         rename `v' `v'_a
137         winsor `v'_a, gen(`v') p(0.01)
138         drop `v'_a
139     }
140
141
142     *-----
143     *-定义融资约束分组指标
144     *-----
145
146     *-----
147     *-原文中的划分方法(pp.158-159)
148     *-----
149     *
150     *-Class1 (FC):      D/I<10% for at least 10 years --> 49 家
151     *-Class2(PFC): 10%<D/I<20% for at least 10 years --> 39 家
152     *-Class3(NFC): All other firms --> 334家
153     *-D/I: Dividend-Income ratio
154     *
155     *-Comments:
156     *
157     *-(1) cut pions: 10% and 20%, why?
158     *-(2) 49 < 334
159     *-(3) the Footnote.#c in Table 2 is incorrect
160
161
162     *-----
163     *-根据公司规模分组
164     *-----
165     * 根据样本公司在1998-2008期间的平均TA分组,
166     * 这样可以保证同一家公司始终处于同一组中
167     * 我们将样本公司分成5个组,
168     * 其中第一组定义为 FC 组, 第五组定义为NFC组
169     bysort id: egen av_size = mean(TA)
170     sort av_size
171     gen g_size = group(5)
172     gen FC_size = g_size==1
173     gen NFC_size = g_size==5
174     gen Other_size = 1-FC_size-NFC_size
175
176     label define g_size_lab 1 "Small" 5 "Large"
177     label value g_size g_size_lab
178
179     *-----
180     *-根据股利支付率分组
181     *-----
182     * Num_div 是在1998-2008期间, 每家公司支付股利的次数
183     * 若Num_div<=2,则定义为FC 组;
184     * 若Num_div>=8,则定义为NFC组;
185     dropvars Num_div FC_div NFC_div Other_div
186     bysort id: egen Num_div = sum(div_yes)
187     recode Num_div (min/2=1) (2/8=2) (8/max=3), gen(g_div)
188     gen FC_div = Num_div<=2
189     gen NFC_div = Num_div>8
190     gen Other_div = 1 - FC_div - NFC_div
191
192     label var g_div "Payout times"
193     label define g_div_lab 1 "[0,2]" 2 "(2,8]" 3 "(8,oo)"
194     label value g_div g_div_lab
195
196
197     *-----
198     *- Table 2 基本分组统计量
199     *-----
200
201     bysort id: egen sd_cflow = sd(cflow)
202     bysort id: egen sd_invt = sd(invt)

```

```

203
204  *--根据股利支付情况分组结果
205      logout, save(Tab02_A) excel replace: ///
206          tabstat Num_div sgr invt cflow sd_invt sd_cflow cash size tl ll, ///
207          by(g_div) stat(mean) f(%4.3f) save
208
209  *--根据公司规模分组结果
210      logout, save(Tab02_B) excel replace: ///
211          tabstat Num_div sgr invt cflow sd_invt sd_cflow cash size tl ll, ///
212          by(g_size) stat(mean) f(%4.3f) save
213
214  *--分组公司的行业分布
215      tab sicmen g_size if year==2008
216      tab sicmen g_div if year==2008
217
218
219
220  *----- pp.167
221  *-- Table 4 : Effects of Q and Cash Flow on Investment
222  *-----
223
224      qui tsset
225      cap rename tobin tobin_a
226      cap gen tobin = L.tobin_a // 若采用L.tobin, 则增加这两条命令
227                               // 否则在该命令前增加 * 号
228                               // pp.165,采用 L.Tobin
229
230  *--Note: 由于使用了 local, 执行时必须一次性选中`begin'-`over'之间的命令
231  *
232  *-----begin-----
233      local g "_div" // 根据股利支付分组
234      local g "_size" // 根据公司规模情况分组
235
236  *--1998-2002
237      local ifyear "year<=2002"
238      xi:xtreg invt tobin cflow i.year if (FC`g'&`ifyear'), fe
239      est store FC`g'9802
240      xi:xtreg invt tobin cflow i.year if (Other`g'&`ifyear'), fe
241      est store Other`g'9802
242      xi:xtreg invt tobin cflow i.year if (NFC`g'&`ifyear'), fe
243      est store NFC`g'9802
244
245  *--1998-2005
246      local ifyear "year<=2005"
247      xi:xtreg invt tobin cflow i.year if (FC`g'&`ifyear'), fe
248      est store FC`g'9805
249      xi:xtreg invt tobin cflow i.year if (Other`g'&`ifyear'), fe
250      est store Other`g'9805
251      xi:xtreg invt tobin cflow i.year if (NFC`g'&`ifyear'), fe
252      est store NFC`g'9805
253
254  *--1998-2008
255      xi:xtreg invt tobin cflow i.year if FC`g', fe
256      est store FC`g'9808
257      xi:xtreg invt tobin cflow i.year if Other`g', fe
258      est store Other`g'9808
259      xi:xtreg invt tobin cflow i.year if NFC`g', fe
260      est store NFC`g'9808
261
262
263  *--结果汇总
264  *--1998-2002
265      local mm "FC`g'9802 Other`g'9802 NFC`g'9802"
266      local keep "tobin cflow"
267      esttab `mm' using Tab04`g'.csv, mtitle(`mm') b(%6.3f) ///
268          nogaps keep(`keep') replace /// //replace
269          star(* 0.1 ** 0.05 *** 0.01) s(N r2_w r2_b r2_o) ///
270          addnote("by `g'")

```



```

271
272 *-1998-2005
273     local mm "FC`g'9805 Other`g'9805 NFC`g'9805"
274     local keep "tobin cflow"
275     esttab `mm' using Tab04`g'.csv, mtitle(`mm') b(%6.3f)      ///
276         nogaps keep(`keep') append                          /// //append
277         star(* 0.1 ** 0.05 *** 0.01) s(N r2_w r2_b r2_o)      ///
278         addnote("by `g'")
279
280 *-1998-2008
281     local mm "FC`g'9808 Other`g'9808 NFC`g'9808"
282     local keep "tobin cflow"
283     esttab `mm' using Tab04`g'.csv, mtitle(`mm') b(%6.3f)      ///
284         nogaps keep(`keep') append                          /// //append
285         star(* 0.1 ** 0.05 *** 0.01) s(N r2_w r2_b r2_o)      ///
286         addnote("by `g'")
287 *-----over-----
288 *
289 *-Comments:
290 *
291 *(1) 采用不同的分组方法得到的结果是不同的，这在后续的文献中也经常出现；
292 *(2) 有些文献专门探讨 FC 的分组方法；
293 *(3) 用"肉眼"能明确判断组间差异是否显著吗？
294 *(4) 采用"平行面板(Balanced Panel Data)"合理吗？
295
296
297
298 *-随后的分析都采用 Div 分组
299
300 *----- pp.169
301 *- Table 5 : Effects of Q and Cash Flow on Investment:
302 *----- Consideration of Measurement Error
303
304     local aa "FC Other NFC"
305     foreach v of local aa{
306         qui{
307             tsset
308             xi: xtreg invt tobin i.year if `v'_div, fe
309                 est store OLS_`v'
310             xi: xtreg invt tobin cflow i.year if `v'_div, fe      // add cflow
311                 est store Cflow_`v'
312             xi: xtivreg invt cflow (tobin=L.tobin) i.year if `v'_div, fe // IV
313                 est store iv_`v'
314             xi: xtreg D.invt D.tobin D.cflow i.year if `v'_div, fe // FD
315                 est store FD1_`v'
316             xi: xtreg D2.invt D2.tobin D2.cflow i.year if `v'_div, fe // SOD
317                 est store FD2_`v'
318         }
319     }
320
321 *-对于 IV 估计的解释
322
323 *-----
324 *-对于 FD 估计的解释
325 *
326 * 衡量偏误(measurement error in the regressor)
327 * 对于模型  $y = x_0*b + e$  (1)
328 * 假设  $x = x_0 + me$  ( $x$  是  $x_0$  的样本观测值)
329 * 即  $x_0 = x - me$  (2)
330 * 模型的正确形式应为:
331 *  $y = a + x_0*b + e$ 
332 *  $= a + (x - me)*b + e$ 
333 *  $= a + x*b + (e - me*b)$ 
334 * 用于实证分析的模型
335 *  $y = a + x*b + e_2$  (3)
336 * 其中,  $e_2 = (e - me*b)$ 
337 * 显然,
338 *  $Corr(x, e_2) \neq 0 \implies$  (3) 式中  $b$  的 OLS 估计值是有偏的;

```

```

339
340 *-----
341 *-呈现结果
342 *-----
343 *-class 1: FC group
344     local ss "using Tab05.csv"
345     local mm "OLS_FC Cflow_FC iv_FC FD1_FC FD2_FC"
346     local order "tobin cflow D.tobin D.cflow D2.tobin D2.cflow"
347     esttab `mm' `ss', mtitle(`mm') s(N r2_w) ///
348             compress nogaps b(%6.3f) replace /// //replace
349             star(* 0.1 ** 0.05 *** 0.01) ///
350             keep(`order') order(`order')
351
352 *-class 2: Other group
353     local ss "using Tab05.csv"
354     local mm "OLS_Other Cflow_Other iv_Other FD1_Other FD2_Other"
355     local order "tobin cflow D.tobin D.cflow D2.tobin D2.cflow"
356     esttab `mm' `ss', mtitle(`mm') s(N r2_w) ///
357             compress nogaps b(%6.3f) append /// //append
358             star(* 0.1 ** 0.05 *** 0.01) ///
359             keep(`order') order(`order')
360
361 *-class 3: NFC group
362     local ss "using Tab05.csv"
363     local mm "OLS_NFC Cflow_NFC iv_NFC FD1_NFC FD2_NFC"
364     local order "tobin cflow D.tobin D.cflow D2.tobin D2.cflow"
365     esttab `mm' `ss', mtitle(`mm') s(N r2_w) ///
366             compress nogaps b(%6.3f) append /// //append
367             star(* 0.1 ** 0.05 *** 0.01) ///
368             keep(`order') order(`order')
369
370 *-----
371 *-整体呈现
372 *-----
373 *-pooling class1-class3
374     local ss "using Tab05_All.csv"
375     local mm "Cflow_FC Cflow_Other Cflow_NFC iv_FC iv_Other iv_NFC"
376     local order "tobin cflow"
377     esttab `mm' `ss', mtitle(`mm') compress nogaps b(%6.3f) ///
378             s(N r2_w) keep(`order') order(`order') replace /// //replace
379             star(* 0.1 ** 0.05 *** 0.01)
380
381 *-pooling class1-class3 (continue)
382     local ss "using Tab05_All.csv"
383     local mm "FD1_FC FD1_Other FD1_NFC FD2_FC FD2_Other FD2_NFC"
384     local order "D.tobin D.cflow D2.tobin D2.cflow"
385     esttab `mm' `ss', mtitle(`mm') compress nogaps b(%6.3f) ///
386             s(N r2_w) keep(`order') order(`order') append /// //append
387             star(* 0.1 ** 0.05 *** 0.01)
388
389
390
391 *----- pp.171
392 *- Table 6 : Effects of Q and Cash Flow on Investment:
393 *----- Alternative Specifications
394 *
395 *-add L(0/2).cflow terms: time-to-build effect
396 *
397 *-add L.Q term: Q is a too forward-looking variable
398 * see: Alti(2003)
399 * Alti, A., 2003.
400 * How sensitive is investment to cash flow when financing is frictionless?.
401 * Journal of Finance, 58 (2): 707-722.
402 * shellout "$path\Refs\Alti_2003_JF.pdf" //打开 PDF 文件
403
404 local aa "FC Other NFC"
405 foreach v of local aa{
406     *-A: model with additional cash flow lags

```

```

407     qui xi: xtreg invt tobin L(0/2).cf flow i.year if `v'_div, fe
408     est store Tab6_1_`v'
409     *-B: model including lagged Q
410     qui xi: xtreg invt tobin L.tobin cf flow i.year if `v'_div, fe
411     est store Tab6_2_`v'
412 }
413 *-Note: regress on full sample range
414
415 *-Panel A: model with additional cash flow lags
416 local ss "using Tab06.csv"
417 local mm "Tab6_1_FC Tab6_1_Other Tab6_1_NFC"
418 local order "tobin cf flow L.cf flow L2.cf flow"
419 esttab `mm' `ss', mtitle(`mm') b(%6.3f) nogaps ///
420     star(* 0.1 ** 0.05 *** 0.01) s(N r2_w) ///
421     keep(`order') order(`order') replace // replace
422
423 *-Panel B: model including lagged Q
424 local ss "using Tab06.csv"
425 local mm "Tab6_2_FC Tab6_2_Other Tab6_2_NFC"
426 local order "tobin L.tobin cf flow"
427 esttab `mm' `ss', mtitle(`mm') b(%6.3f) nogaps ///
428     star(* 0.1 ** 0.05 *** 0.01) s(N r2_w) ///
429     keep(`order') order(`order') append // append
430
431
432
433
434 *----- pp.175
435 *- Table 7 : Effects of Sales and Cash Flow on Investment
436 *-----
437 *-
438 *-设定依据: Sales Accelerator Investment Demand Models
439 *-Also see: Footnote 45, pp.174
440
441 local aa "FC Other NFC"
442 foreach v of local aa{
443     *-A: model with sales-capital ratio
444     qui xi: xtreg invt cf flow L(0/3).sale i.year if `v'_div, fe
445     est store Tab7_1_`v'
446     *-B: model with sales-capital ratio and Q
447     qui xi: xtreg invt tobin cf flow L(0/3).sale i.year if `v'_div, fe
448     est store Tab7_2_`v'
449 }
450 *-Note: regress on full sample range
451 *-
452 *-Panel B: why Q is added? Cash flow 的双重功能
453
454 *-Panel A: model with additional cash flow lags
455 local ss "using Tab07.csv"
456 local mm "Tab7_1_FC Tab7_1_Other Tab7_1_NFC"
457 local order "cf flow sale L.sale L2.sale L3.sale"
458 esttab `mm' `ss', mtitle(`mm') b(%6.3f) nogaps ///
459     star(* 0.1 ** 0.05 *** 0.01) s(N r2_w) ///
460     keep(`order') order(`order') replace
461
462 *-Panel B: model including lagged Q
463 local ss "using Tab07.csv"
464 local mm "Tab7_2_FC Tab7_2_Other Tab7_2_NFC"
465 local order "tobin cf flow sale L.sale L2.sale L3.sale"
466 esttab `mm' `ss', mtitle(`mm') b(%6.3f) nogaps ///
467     star(* 0.1 ** 0.05 *** 0.01) s(N r2_w) ///
468     keep(`order') order(`order') append
469
470 *-Comments:
471 *
472 *- (1) for Thesis: 需要把一个问题研究的很细致, 这里提供了很好的思路;
473 * 不过, 你要说清楚, 为什么有必要加入 L*.sale 作为新的控制变量;
474 *

```

```

475  *-(2) 如果研究中发现这个结果很糟糕，该怎么办？
476  *      其实，这里呈现出来的结果就很微妙：
477  *      0.046*** > 0.058*   ???
478
479
480
481  *-----
482  *-      Table 8   Not follow
483  *-----
484  *
485  *-Have no data on prices
486
487
488
489  *-----
490  *-      Table 9   pp.179
491  *-----
492
493  *-行业分布
494      tab sicmen if year==2008
495      tab sicmen g_div if year==2008 //各个行业的 FC,NFC 公司分布
496
497  *-分行业回归 (Self-done)
498  *
499  *-How to do?
500  *-What about the balanced panel data?
501
502
503  *----- pp.180
504  *-      Table 10   : Effects of Balance Sheet Variables on Investment
505  *-----
506
507  local aa "FC Other NFC"
508  foreach v of local aa{
509      *-A: model including cash and equivalents variable
510      qui xi: xtreg invt tobin cflow L.cash i.year if `v'_div, fe
511      est store Tab10_1_`v'
512      *-B: model including working capital
513      qui xi: xtreg invt tobin cflow L.wcml i.year if `v'_div, fe
514      est store Tab10_2_`v'
515      *-C: model including current and lagged cash and sales
516      xi: xtreg invt tobin L.cash L(0/2).cflow L(0/3).sale ///
517          i.year if `v'_div, fe
518      est store Tab10_3_`v'
519  }
520  *-Note: regress on full sample range
521
522  *-Panel A: model including cash and equivalents variable
523      local ss "using Tab10.csv"
524      local mm "Tab10_1_FC Tab10_1_Other Tab10_1_NFC"
525      local order "tobin cflow L.cash"
526      esttab `mm' `ss', mtitle(`mm') b(%6.3f) nogaps ///
527          star(* 0.1 ** 0.05 *** 0.01) s(N r2_w) ///
528          keep(`order') order(`order') replace
529
530  *-Panel B: model including lagged Q
531      local ss "using Tab10.csv"
532      local mm "Tab10_2_FC Tab10_2_Other Tab10_2_NFC"
533      local order "tobin cflow L.wcml"
534      esttab `mm' `ss', mtitle(`mm') b(%6.3f) nogaps ///
535          star(* 0.1 ** 0.05 *** 0.01) s(N r2_w) ///
536          keep(`order') order(`order') append
537
538  *-Panel C: model including current and lagged cash and sales
539      local ss "using Tab10.csv"
540      local mm "Tab10_3_FC Tab10_3_Other Tab10_3_NFC"
541      local order "tobin L.cash cflow L.cflow L2.cflow sale L.sale L2.sale L3.sale"
542      esttab `mm' `ss', mtitle(`mm') b(%6.3f) nogaps ///

```

```
543          star(* 0.1 ** 0.05 *** 0.01) s(N r2_w) ///
544          keep(`order') order(`order') append
545
546      *-Comments:
547      *
548      *-(1) 加入 Sale 后, 结果发生了很大的变化;
549      *      在Panel C 中, NFC组的 Invt-Cflow 敏感性反而较高。
550      *      有两种可能的解释: 一是前面的分析可能存在遗漏变量的偏误;
551      *      二是加速器模型并不适于解释中国上市公司的投资行为
552      *      存在严重的共线性吗?
553      pwcorr_a tobin cflow cash wcmi sale
554
555      *-(2) 这张表引发了 Cash holding 领域的研究,
556      *      Opler et al. (1999)
557      *      Opler, T., Pinkowitz, L., Stulz, R., Williamson, R., 1999.
558      *      The determinants and implications of corporate cash holdings.
559      *      Journal of Financial Economics, 52(1): 3-46.
560      shellout "$path\Refs\Opler_1999.pdf" //打开 PDF 文件
561
562      *      Almeida et al. (2004)
563      *      Almeida, H., Campello, M., Weisbach, M., 2004.
564      *      The cash flow sensitivity of cash.
565      *      Journal of Finance, 59(4): 1777-1804.
566      shellout "$path\Refs\Almeida_2004.pdf" //打开 PDF 文件
567
568      *      Faulkender and Wang (2006)
569      *      Faulkender, M., Wang, R., 2006.
570      *      Corporate financial policy and the value of cash.
571      *      Journal of Finance, 61 (4): 1957-1990.
572      shellout "$path\Refs\Faulkender_2006.pdf" //打开 PDF 文件
573
```

```

1
2  *      -----
3  *      -----
4  *
5  *      Stata 学术论文专题
6  *
7  *      -----
8  *      -----
9
10
11  *      主讲人：连玉君 副教授
12  *
13  *      单 位：中山大学岭南学院金融系
14  *      电 邮：arlionn@163.com
15  *      博 客：http://blog.cnfol.com/arlionn
16  *      主 页：http://goo.gl/tRXba
17  *      微 博：http://weibo.com/arlionn
18
19
20  *-----
21  * -Flannery & Rangan, 2006-
22  *      Flannery, M. J., K. P. Rangan, 2006,
23  *      Partial adjustment toward target capital structures,
24  *      Journal of Financial Economics, 79(3): 469-506.
25  *-----
26  * -Ref: 连玉君和钟经樊(2007,南方经济)
27
28  * -主要方法
29  *   OLS: regress
30  *   FE (Fixed Effect Model): xtreg, fe
31  *   Dynamic Panel Data: xtabond
32  *   Fama-Macbench: xtftmb
33  *   IV-Panel: xtivreg
34  *   Measurement Error
35  *
36  *-----
37  * - 这篇文章的实证分析做的非常精彩，值得反复研读！
38  *-----
39
40  * -设定课程存储路径
41  *   global path "`c(sysdir_personal)'PX_papers" //请先执行该命令
42
43  * -原始论文
44  *   shellout "$path\Refs\Flannery_2006.pdf"
45
46  * -PPT
47  *   shellout "$path\PPT\Flannery_2006.ppt"
48
49
50
51
52  *=====
53  * - Part I: 样本筛选和基本统计分析
54  *=====
55
56  *   cd "$path\Flannery_2006" // 进入课程目录
57  *   adopath + "$path\adofiles" // 自编程序
58
59  *   clear all
60  *   set memory 80m // 设定 Stata 占用的内存空间
61  *   use "$path\data\GTA2008_my.dta", clear //调入数据
62
63
64  *-----
65  * -定义变量
66
67  *   tsset id year
68  *   renvars tl tobin size / BDR MB LnTA // 更改变量名称

```



```

69
70     gen MV = MB*TA //公司市值
71     gen MDR = TD/MV //负债总额/公司市值
72     gen EBIT_TA = b001000000/TA //利润总额/总资产
73     gen DEP_TA = d000103000/TA //固定资产折旧/总资产
74     gen FA_TA = a001212000/TA //固定资产净额/总资产
75     qui tsset
76     gen L3MDR =(L1.MDR+L2.MDR+L3.MDR)/3 //前三年平均负债率
77
78
79     *-----
80     *-Ind_Median 指标：负债率的行业中位数
81     *-----
82
83     *-Flannery 文中(pp.475)采用的是 Fama&French(2002) 的行业分类方法
84
85     * bysort sicda year: egen Ind_Median = median(BDR) //粗略定义
86     panels id: tab sicda //有些行业中的公司数目太少
87
88     *-行业划分(制造业细分到次类，其他行业采用门类)
89     *-定义行业分类方法：黄梅,夏新平(2009) 南开管理评论
90     clonevar sic2 = sicda_str
91     order id year sic2
92     replace sic2 = substr(sic2,1,1) if substr(sic2,1,1)!="C"
93     replace sic2 = substr(sic2,1,2) if substr(sic2,1,1)=="C"
94     replace sic2 = "C9" if sic2=="C2" // 将 C2 并入 C9
95
96     *-每年度每个行业至少保留 10 家公司 (酌情选择)
97     bysort sic2 year: egen num_sic_year = count(id)
98     keep if num_sic_year >=10 // Roychowdhury(2006, p.349, 15家)
99
100    *-基本统计分析
101    panels id : tab sic2 //各行业的公司数目
102    panels id : tab sic2 year //各行业-年度的公司数目
103    tab sic2 year //各行业-年度的观察值数目
104
105    *-定义指标
106    bysort sic2 year: egen Ind_Median = median(MDR)
107
108
109    *-----
110    *-MB_EFWA 指标：market timing proxy
111    *-----
112    *
113    * 参见：
114    * Baker, M., Wurgler, J., 2002.
115    * Market timing and capital structure. Journal of Finance 57, 1-32
116    *
117    shellout "$path\Refs\Baker_2002_JF.pdf"
118
119    *-中文：
120    * [1] 才静涵,刘红忠. 市场择时理论与中国市场的资本结构.
121    * 经济科学,2006,(4):59-69.
122    * [2] 胡志强,卓琳玲. IPO 市场时机选择与资本结构关系研究.
123    * 金融研究,2008,(10):136-149.
124
125    rename a003000000 book_equity // 所有者权益合计
126    rename a003102000 retain_earn // 资本公积(留存收益)
127    drop if (book_equity<=0|retain_earn<=0) //讨论合理性?
128
129    qui tsset
130    gen e = D.book_equity + D.retain_earn
131    gen d = D.TD // 负债总额
132    gen sum_ed = e+d // 各年度之和
133    bysort id: gen SUM_ed = sum(e+d) // 分母: Baker(2002, Eq.3, pp.12)
134    gen weight = sum_ed/SUM_ed // 权重: Eq.3, 求和符号内部部分
135    replace weight=0 if weight<0 // 权重不能小于零, Baker(2002,pp.12)
136    gen weight_MB = weight*MB

```

```

137
138     bysort id: gen MB_EFWA = sum(weight_MB)
139
140     gen MB_EFWA0 = MB_EFWA // 备份一份，留待后用
141
142     *-处理离群值(非常重要)
143     sum MB_EFWA, detail // 存在严重的离群值
144     panels id if (MB_EFWA>10) // 统计存在此问题的公司数目
145     replace MB_EFWA=. if MB_EFWA>10 // Baker(2002,p.12, Footnote 8)
146     sum MB_EFWA, detail // 此时还有离群值吗?
147     graph box MB_EFWA // 箱型图
148
149
150     *-----
151     *-FINDEF 指标
152     *-----
153     *
154     * 参见: pp.229, Table 2
155     * Frank, M.Z., Goyal, V.K., 2003
156     * Testing the pecking order theory of capital structure.
157     * Journal of Financial Economics, 67(2): 217-248.
158     *
159     *shellout "`c(sysdir_personal)'PX_papers\Refs\Frank_2003_JFE.pdf"
160     *
161     * FINDEF = dividend payments + investments + change in working capital
162     *           - internal cashflow
163
164     qui tsset
165     gen divpay = div_cash*nshra // 每股现金分红*A股流通股数
166     replace divpay=0 if divpay==. // div_cash 中有很多缺漏值，视为 0
167     gen invest = c002006000 //
购建固定资产、无形资产和其他长期资产支付的现金
168     gen wc = a001100000-a002100000 // 营运资本=流动资产-流动负债
169     gen intercflo = c001000000 // 经营活动产生的现金流量净额
170     gen FINDEF = (divpay + invest + D.wc - intercflo)/TA
171
172
173     *-----
174     *-负债率的离群值
175
176     *drop if BDR>1 | MDR>1
177     replace BDR=0.999 if BDR>=1
178     replace MDR=0.999 if MDR>=1
179     *-之所以做 winsor 而非 trim 处理，是因为：
180     des bankd
181     sum bankd, detail
182
183
184     *-----
185     *-重新排序变量
186     order id year MDR BDR EBIT_TA MB DEP_TA LnTA ///
187           FA_TA Ind_Median L3MDR FINDEF MB_EFWA
188
189
190     *-----
191     *-Winsor 处理
192     foreach v of varlist MDR-FA_TA L3MDR FINDEF {
193         local a: var lab `v' // 记录变量的标签
194         rename `v' `v'_a
195         winsor `v'_a, gen(`v') p(0.01)
196         drop `v'_a
197         label var `v' "`a'" // 定义变量标签
198     }
199     *-Note: Ind_Median 没有 winsor
200     * MB_EFWA 此前已经 winsor 过了
201
202     *-----
203     *-再次排序变量

```

```

204     order id year MDR BDR EBIT_TA MB DEP_TA LnTA   ///
205           FA_TA Ind_Median L3MDR FINDEF MB_EFWA
206
207
208     *-----
209     *-仅保留连续两年以上有样本的公司
210     xtpattern, gen(pp)           // 记录每家公司的样本形态
211     gen p2 = strpos(pp, "11")
212     drop if p2==0
213
214
215
216     *-----
217     *-   Table 1 基本统计量
218     *-----
219
220     *-----
221     *-样本分布   pp.476
222
223     xtides
224     xtides, p(20)   //呈现前 20 种模式
225
226     panels id           // 整体分布, pp.476, Section 3 Data, Footnote 3
227     panels id: tab sicmen // 行业门类分布(文中没有报告)
228     panels id: tab sicda  // 行业大类分布(文中没有报告)
229     panels id: tab sic2   // 重新归类后的行业划分方式
230
231     capture drop SH
232     gen SH = id>=600000
233     panels id: tab SH     // 深市/沪市分布
234
235     *-----
236     *-表 1 结果   pp.474
237
238     logout, save("Table01") excel replace:   ///
239     tabstat BDR MDR EBIT_TA MB DEP_TA LnTA FA_TA Ind_Median ///
240           MB_EFWA L3MDR FINDEF, ///
241           s(N mean p50 sd min max) c(s) f(%10.4f)
242
243     *-Note: MB_EFWA 中可能还存在严重的离群值问题
244
245
246     *-----
247     *   Fig 1
248     *-----
249     *-Step1: 每个年度, 根据 Q1,Q2,Q3 将 (MDR*-MDR) 分成四组;
250     *-Step2: 每个年度, 计算 BDR[t+1]-BDR[t] 的平均值和中位数;
251     *-Step3: 以 (MDR*-MDR) 的四个分位数为横轴变量, 绘制柱状图。
252
253     qui tsset
254     gen F_MDR = F.MDR
255     gen F_BDR = F.BDR
256
257     *-Note: Fig.1, pp.473, 纵轴为 Change in (Book Debt Ratio)
258     preserve
259     qui xi: xtreg F_MDR EBIT MB DEP LnTA FA_TA Ind_Median i.year, fe
260     predict DR_star, xbu           // 目标负债率的估计值
261     qui tsset
262     gen D_BDR = D.F_BDR           // 纵轴变量
263     gen dis_MDR = DR_star - MDR   // 横轴变量
264     drop if dis_MDR==.           // 删除缺漏值, 否则会影响分组
265     bysort year: quantiles dis_MDR, gen(q_dis_MDR) nquant(4) //分成四组
266     bysort q_dis_MDR: egen mean_dis_MDR = mean(dis_MDR) // 横轴平均值
267     bysort q_dis_MDR: egen mean_D_BDR = mean(D_BDR) // 纵轴平均值
268     bysort q_dis_MDR: egen med_D_BDR = median(D_BDR) // 纵轴中位数
269     keep q_dis_MDR mean_dis_MDR mean_D_BDR med_D_BDR
270     duplicates drop
271     foreach v of varlist mean_dis_MDR mean_D_BDR med_D_BDR{

```

```

272     replace `v' = `v'*100           // 转换成百分数
273     replace `v' = int(`v'*10)/10    // 保留小数点后一位
274 }
275 #delimit ;
276     graph bar mean_D_BDR med_D_BDR, over(mean_dis_MDR, gap(*3))
277         scheme(s2mono)                //图形模板, 控制颜色和整体风格
278         ylabel(-2(2)8, angle(0))      //纵轴刻度
279         title("Fig. 1. Susequent year's change in book debt ratio.")
280         b2title("Mean Distance From Target (MDR*-MDR) in year t -1") //横轴标题
281         l1title("Change in Book Debt Ratio") //纵轴标题
282         bar(1, fcolor(black*0.8) lcolor(black) lw(medthick)) //bar的颜色
283         bar(2, fcolor(black*0.2) lcolor(black) lw(medthick))
284         blabel(bar, position(outside) format(%3.1f) color(black) size(*1.01))
285         legend(label(1 "Mean") label(2 "Median") row(2) ring(0) position(10))
286         yline(-2(2)8, lc(black*0.6)) //附加的横线
287         caption("Notes: Numbers are in percentage") //注释语句
288     ;
289 #delimit cr
290     graph export Fig01_book.wmf, replace //保存图片, 可直接贴入 word
291 restore
292
293 *--Note: Fig.1 我们自己做的另一个版本,纵轴为 Change in (Market Debt Ratio)
294 preserve
295     qui xi: xtreg F_MDR EBIT MB DEP LnTA FA_TA Ind_Median i.year, fe
296     predict DR_star, xbu // 目标负债率的估计值
297     qui tsset
298     gen D_MDR = D.F_MDR
299     gen dis_MDR = DR_star - MDR
300     drop if dis_MDR==.
301     bysort year: quantiles dis_MDR, gen(q_dis_MDR) nquant(4) //分成四组
302     bysort q_dis_MDR: egen mean_dis_MDR = mean(dis_MDR) // 横轴平均值
303     bysort q_dis_MDR: egen mean_D_MDR = mean(D_MDR) // 纵轴平均值
304     bysort q_dis_MDR: egen med_D_MDR = median(D_MDR) // 纵轴中位数
305     keep q_dis_MDR mean_dis_MDR mean_D_MDR med_D_MDR
306     duplicates drop
307     foreach v of varlist mean_dis_MDR mean_D_MDR med_D_MDR{
308         replace `v' = `v'*100           // 转换成百分数
309         replace `v' = int(`v'*10)/10    // 保留小数点后一位
310     }
311 #delimit ;
312     graph bar mean_D_MDR med_D_MDR, over(mean_dis_MDR, gap(*3))
313         scheme(s2mono)                //图形模板, 控制颜色和整体风格
314         ylabel(-8(2)10, angle(0))      //纵轴刻度
315         title("Fig. 1a. Susequent year's change in market debt ratio.")
316         b2title("Mean Distance From Target (MDR*-MDR) in year t -1") //横轴标题
317         l1title("Change in Market Debt Ratio") //纵轴标题
318         bar(1, fcolor(black*0.8) lcolor(black) lw(medthick)) //bar的颜色
319         bar(2, fcolor(black*0.2) lcolor(black) lw(medthick))
320         blabel(bar, position(outside) format(%3.1f) color(black) size(*1.01))
321         legend(label(1 "Mean") label(2 "Median") row(2) ring(0) position(4))
322         yline(-8(2)10, lc(black*0.6))
323         caption("Notes: Numbers are in percentage") ;
324 #delimit cr
325     graph export Fig01_market.wmf, replace
326 restore
327
328
329 *-----
330 * Fig 2
331 *-----
332 *--Note: Fig.2, pp.473
333 preserve
334     local y "MDR" //选择纵轴变量
335     local y "BDR"
336     local x "MDR" //选择横轴变量
337     local x "BDR"
338     if "`y'"=="MDR"{
339         local ytext "Market" // 后面的图形标题可以自动变化

```

```

340     }
341     else{
342         local ytext "Book"
343     }
344     qui tsset
345     gen D_`y' = D.F_`y' //纵轴变量
346     bysort year: quantiles `x', gen(q_`x') nquant(4) //分成四组
347     bysort q_`x': egen mean_`x' = mean(`x') //横轴平均值
348     bysort q_`x': egen mean_D_`y' = mean(D_`y') //纵轴平均值
349     bysort q_`x': egen med_D_`y' = median(D_`y') //纵轴中位数
350     keep q_`x' mean_`x' mean_D_`y' med_D_`y'
351     duplicates drop
352     foreach v of varlist mean_`x' mean_D_`y' med_D_`y' {
353         replace `v' = `v'*100 // 转换成百分数
354         replace `v' = int(`v'*10)/10 // 保留小数点后一位
355     }
356     #delimit ;
357     graph bar mean_D_`y' med_D_`y', over(mean_`x', gap(*3))
358         scheme(s2mono) //图形模板, 控制颜色和整体风格
359         ylabel(-4(2)6, angle(0)) //纵轴刻度
360         title("Fig. 2. Mean reversion in `ytext' leverage")
361         b2title("Absolute `x'") //横轴标题
362         l1title("Change in `ytext' Debt Ratio") //纵轴标题
363         bar(1, fcolor(black*0.8) lcolor(black) lw(medthick)) //bar的颜色
364         bar(2, fcolor(black*0.2) lcolor(black) lw(medthick))
365         blabel(bar, position(outside) format(%3.1f) color(black) size(*1.01))
366         legend(label(1 "Mean") label(2 "Median") row(2) ring(0) position(2))
367         yline(-4(2)6, lc(black*0.6)) //附加的横线
368         caption("Notes: Numbers are in percentage") //注释语句
369     ;
370     #delimit cr
371     graph export Fig02_`y'_`x'.wmf, replace //保存图片, 可直接贴入 word
372     restore
373
374     *-Comments:
375     *-1: Fig.1 也可以采用这种灵活的方式来绘制, 因此, 要学会使用 local;
376     *-2: 上述两个图形都表明, 公司的负债率存在 Mean reversion 行为,
377     * 这也对 trade off 理论提供了初步的经验支持;
378     *-3: 其实, 上述图形中也隐含着非对称调整特征;
379
380
381     *-----
382     * Fig 3
383     *-----
384     *-Step1: 每个年度, 根据 Q1,Q2,Q3 将 (MDR[t]*-MDR[t]) 分成四组: g1-g4;
385     *-Step2: 每个年度, 计算 g1-g4 组中 BDR[t+1]-BDR[t] 的平均值和中位数;
386     *-Step3: 以 (MDR*-MDR) 的四个分位数为横轴变量, 分别绘制 g1-g4 组的柱状图。
387     *-Comment: 其实就是分组绘制 Fig.1
388
389     qui tsset
390
391     *-Note: Fig.3 pp.481, 纵轴为 Change in (Book Debt Ratio)
392     preserve
393     qui xi: xtreg F_MDR EBIT MB DEP LnTA FA_TA Ind_Median i.year, fe
394     predict DR_star, xbu // 目标负债率的估计值
395     qui tsset
396     gen D_BDR = D.F_BDR // 纵轴变量
397     gen dis_MDR = DR_star - MDR // 横轴变量
398     drop if dis_MDR==. // 删除缺漏值, 否则会影响分组
399     *-依据(MDR*-MDR)分成四组
400     bysort year: quantiles dis_MDR, gen(q_dis_MDR) nquant(4)
401     *-依据 MDR 分成四组
402     bysort year: quantiles MDR, gen(q_MDR) nquant(4)
403     *-计算各组 D.BDR 的平均值和中位数
404     bysort q_dis_MDR q_MDR: egen mean_dis_MDR = mean(dis_MDR) //横轴平均值
405     bysort q_dis_MDR q_MDR: egen mean_D_BDR = mean(D_BDR) //纵轴平均值
406     bysort q_dis_MDR q_MDR: egen med_D_BDR = median(D_BDR) //纵轴中位数
407     keep q_dis_MDR q_MDR mean_dis_MDR mean_D_BDR med_D_BDR

```



```

408     duplicates drop
409     foreach v of varlist mean_dis_MDR mean_D_BDR med_D_BDR{
410         replace `v' = `v'*100           // 转换成百分数
411         replace `v' = int(`v'*10)/10    // 保留小数点后一位
412         format `v' %3.1f
413         gen str6 `v'_s = string(`v')+ "%" // "14.2" --> "14.2%"
414         replace `v'_s = "0"+`v'_s if (`v'<1&`v'>0) // ".4%" --> " 0.4%"
415         replace `v'_s = substr(`v'_s, "-.", "-0.", .) // "-.3%" --> "-0.3%"
416     }
417
418     *-开始绘图
419     #delimit ;
420     dropvars posit1 posit2;
421     gen posit1 = 3;
422     replace posit1 = 12 if (q_MDR==1&q_dis_MDR==4);
423     replace posit1 = 4 if (q_MDR==1&q_dis_MDR==3);
424     gen posit2 = 12;
425     twoway
426         (connect mean_D_BDR mean_dis_MDR if q_MDR==1,
427             lw(*1.2) mlabel(mean_D_BDR_s) mlabvposition(posit1) msymbol(S))
428         (connect mean_D_BDR mean_dis_MDR if q_MDR==2, lp(dash) msymbol(Dh))
429         (connect mean_D_BDR mean_dis_MDR if q_MDR==3, lp("-."))
430         (connect mean_D_BDR mean_dis_MDR if q_MDR==4, lw(*1.2)
431             lw(*1.2) mlabel(mean_D_BDR_s) mlabvposition(posit2) mlabgap(*2) )
432     ,
433     scheme(s2mono) //图形模板, 控制颜色和整体风格
434     ylabel(-3(1)7, angle(0)) //纵轴刻度
435     xlabel(-15(3)15, angle(0)) //纵轴刻度
436     subtitle("Fig.3 Susequent year's change in book debt ratio.")
437     xtitle("Mean Distance From Target (MDR*-MDR) in year t -1 (%)") //横轴标题
438     xscale(titlegap(2)) //坐标轴与坐标轴标题间距
439     ytitle("Change in Book Debt Ratio (%)") //纵轴标题
440     legend( label(1 "Lowest MDR") label(2 "Low MDR") //图例
441             label(3 "High MDR") label(4 "Highest MDR")
442             ring(0) col(1) position(10) size(*0.8) ) //图例的位置,大小
443     ylabel(-3(1)-1 1(1)7, lc(black*0.4)) //附加的横线
444     ylabel(0, lc(black) lw(*1.2) lp(longdash))
445     ;
446     #delimit cr
447     restore
448
449     *-Comments:
450     * (1) 有关二维图的选项, 参见 help twoway
451     * (2) 有关上图图例的选项, 参见 help scatter##marker_options
452     * (3) 绘图技巧, 先画最基本图形, 再逐步修饰 --> "素描"
453
454
455
456
457     *=====
458     *- Part II: Regression results
459     *=====
460
461     *-----
462     *- Table 2
463     *-----
464
465     global y "BDR" //账面负债率
466     global y "MDR" //市场负债率
467     global xx "EBIT_TA MB DEP_TA LnTA FA_TA Ind_Median" //公共解释变量
468
469     *-Col(1) FM
470     qui tsset
471     xtfrm F_`y' `y' `xx'
472     estadd scalar speed = 1 - _b[`y']
473     estadd scalar Half_Life = ln(2)/(1-_b[`y']) // pp.477, 倒数三行
474     est store FM
475

```



```

476      *-Col(2)  FE panel
477      xtreg F_ $y $y $xx, fe
478      estadd scalar speed = 1 - _b[$y]
479      estadd scalar Half_Life = ln(2)/(1-_b[$y])
480      est store FE
481
482      *-Col(3)  方法：先去除个体效应，再执行 FM 估计
483      preserve
484      keep id year F_ $y $y $xx
485      xtdata, fe clear
486      replace year = int(year)
487      xtfrm F_ $y $y $xx
488      estadd scalar speed = 1 - _b[$y]
489      estadd scalar Half_Life = ln(2)/(1-_b[$y])
490      est store FM_demean
491      restore
492
493      *-Col(4)
494      xi:xtreg F_ $y $y $xx i.year, fe
495      estadd scalar speed = 1 - _b[$y]
496      estadd scalar Half_Life = ln(2)/(1-_b[$y])
497      est store FE_year
498
499      *-Col(5)
500      xi: xtivreg F_ $y $xx i.year ($y=L.BDR $xx), fe
501      estadd scalar speed = 1 - _b[$y]
502      estadd scalar Half_Life = ln(2)/(1-_b[$y])
503      est store IV
504
505      *-Col(6)
506      dropvars p25 p75
507      bysort year: egen p25=pctile($y), p(25)
508      bysort year: egen p75=pctile($y), p(75)
509      tabstat p25 p75, s(mean) f(%6.3f) by(year) //pp.479,倒数第二段
510      qui tsset
511      *-IV-middle 50th percentile
512      xi: xtivreg F_ $y $xx i.year ($y=L.BDR $xx) ///
513      if ($y>p25&$y<p75), fe //上行括号中设定工具变量
514      estadd scalar speed = 1 - _b[$y]
515      estadd scalar Half_Life = ln(2)/(1-_b[$y])
516      est store IV_p50
517      *-FE-middle 50th percentile
518      xi: xtreg F_ $y $y $xx i.year ///
519      if ($y>p25&$y<p75), fe
520      estadd scalar speed = 1 - _b[$y]
521      estadd scalar Half_Life = ln(2)/(1-_b[$y])
522      est store FE_p50
523
524      *-Col(7)
525      * the Rated data is not aviable, thus this column cannot be done
526
527      *--结果汇总
528      local save "using Tab02.csv" //执行时选中此行可以把结果输出到 Excel 中
529      local mm "FM FE FM_demean FE_year IV IV_p50 FE_p50"
530      local keep "$y $xx"
531      esttab `mm' `save', mtitle(`mm') b(%6.3f) nogaps replace ///
532      star(* 0.1 ** 0.05 *** 0.01) keep(`keep') compress ///
533      scalar(N r2 r2_a r2_w speed Half_Life) ///
534      addnotes("Speed = 1-_b[$y]" "Half-Life = ln2/Speed (Years)")
535
536
537
538      *-----
539      * Table 3 : other specification of target leverage
540      *-----
541
542      global y "BDR" //账面负债率
543      global y "MDR" //市场负债率

```

```

544     global xx "EBIT_TA MB DEP_TA LnTA FA_TA " //公共解释变量
545           //这里去掉了Ind_Median, 原因参见 pp.482, Footnote 10
546
547     *-Col(1): Pooled OLS Static model
548     xi: reg F_$y $xx i.year
549     est store OLS_st
550
551     *-Col(2): Pooled OLS Dynamic model
552     xi: reg F_$y $y $xx i.year
553     estadd scalar speed = 1 - _b[$y]
554     est store OLS_dy
555
556     *-Col(3): FE Dynamic model
557     xi: xtreg F_$y $y $xx i.year, fe
558     estadd scalar speed = 1 - _b[$y]
559     est store FE_yr
560
561     *-Col(4): TDR_OLS as target
562     qui xi: reg F_$y $xx i.year
563     cap drop TDR_OLS
564     predict TDR_OLS if e(sample) // important
565     xi: reg F_$y $y TDR_OLS i.year
566     estadd scalar speed = _b[TDR_OLS]
567     estadd scalar LongE = _b[TDR_OLS]/(1-_b[$y]) //pp.482 最后两行
568     est store TDR_OLS
569     dis "Long-run elasticity = " _b[TDR_OLS]/(1-_b[$y])
570     *-理论上, TDR_OLS 与 MDR 的系数之和应为 1
571
572     *-Col(5): L3DR as target
573     qui tsset
574     cap drop L3DR
575     gen L3DR = (L.$y+L2.$y+L3.$y)/3
576     xi: reg F_$y $y L3DR i.year
577     estadd scalar speed = _b[L3DR]
578     estadd scalar LongE = _b[L3DR]/(1-_b[$y]) //pp.482 最后两行
579     est store L3DR
580     dis "Long-run elasticity = " _b[L3DR]/(1-_b[$y])
581     *-理论上, L3DR 与 MDR 的系数之和应为 1
582
583     *--结果汇总
584     local save "using Tab03.csv" //选中执行则输出 Excel 表格
585     local mm "OLS_st OLS_dy FE_yr TDR_OLS L3DR"
586     local keep "$y $xx TDR_OLS L3DR"
587     esttab `mm' `save', mtitle(`mm') b(%6.3f) nogaps replace ///
588           star(* 0.1 ** 0.05 *** 0.01) keep(`keep') compress ///
589           order(`keep') scalar(N r2 r2_a r2_w speed LongE) ///
590           addnotes("Speed = 1-_b[$y]" "LongE: Long-run elasticity, pp.482")
591
592
593
594
595     *-----
596     * Table 4 : Effect of Adding Noise to the Target Debt
597     Ratio
598     *-----
599
600     *-基本想法:
601     *
602     *-在模型(3)中, 分析 MDR* 的衡量偏误对调整速度估计结果的影响,
603     * 通过 MC 产生随机数, 加入 MDR* 之中
604     *
605     *-Model (3):
606     *
607     *  $MDR_{t+1} = (1-Speed)MDR_t + Speed(MDR^*_{t+1}) + u_{t+1}$ 
608     *
609     *-Model (3a): adding noise
610     *
611     *  $MDR_{t+1} = (1-Speed)MDR_t + Speed(MDR^*_{t+1} + Noise_t) + u_{t+1}$ 

```

```

611      *
612      *
613
614      *-回归方法设定
615      local xt "xt" //若选中这两行执行, 则采用 xtreg,fe
616      local fe ",fe" //否则, 为 OLS, 即 reg
617
618      *-变量设定
619      global y "BDR" //账面负债率
620      global y "MDR" //市场负债率
621      global xx "EBIT_TA MB DEP_TA LnTA FA_TA Ind_Median"
622
623      *-估计 MDR*
624      qui xi: xtreg F_$y $xx i.year, fe
625      cap drop DR_star
626      predict DR_star, xbu // MDR*, 其中已经考虑了个体效应
627
628      *-Col(1) 0%
629      dropvars noise DR_starNois
630      gen noise = invnorm(uniform())*(0) //Noise~N(0,0)
631      gen DR_starNois = DR_star + noise
632      xi: `xt'reg F_$y $y DR_starNois i.year `fe'
633      estadd scalar LongR = _b[DR_starNois]/(1-_b[$y])
634      est store E0
635
636      *-Col(2) 5%
637      replace noise = invnorm(uniform())*(0.05) //Noise~N(0,0.05^2)
638      replace DR_starNois = DR_star + noise
639      xi: `xt'reg F_$y $y DR_starNois i.year `fe'
640      estadd scalar LongR = _b[DR_starNois]/(1-_b[$y])
641      est store E5
642
643      *-Col(3) 10%
644      replace noise = invnorm(uniform())*(0.1) //Noise~N(0,0.1^2)
645      replace DR_starNois = DR_star + noise
646      xi: `xt'reg F_$y $y DR_starNois i.year `fe'
647      estadd scalar LongR = _b[DR_starNois]/(1-_b[$y])
648      est store E10
649
650      *-Col(4) 20%
651      replace noise = invnorm(uniform())*(0.2) //Noise~N(0,0.2^2)
652      replace DR_starNois = DR_star + noise
653      xi: `xt'reg F_$y $y DR_starNois i.year `fe'
654      estadd scalar LongR = _b[DR_starNois]/(1-_b[$y])
655      est store E20
656
657      *-Col(5) 25%
658      replace noise = invnorm(uniform())*(0.25) //Noise~N(0,0.25^2)
659      replace DR_starNois = DR_star + noise
660      xi: `xt'reg F_$y $y DR_starNois i.year `fe'
661      estadd scalar LongR = _b[DR_starNois]/(1-_b[$y])
662      est store E25
663
664      *-Col(6) 50%
665      replace noise = invnorm(uniform())*(0.5) //Noise~N(0,0.5^2)
666      replace DR_starNois = DR_star + noise
667      xi: `xt'reg F_$y $y DR_starNois i.year `fe'
668      estadd scalar LongR = _b[DR_starNois]/(1-_b[$y])
669      est store E50
670
671      *-----
672      *-结果汇总
673      local save "using Tab04_$y.csv"
674      local mm "E0 E5 E10 E20 E25 E50" //模型存储名称
675      local mt "0% 5% 10% 20% 25% 50%" //模型标题
676      local keep "$y DR_starNois"
677      local note "= _b[DR_starNois]/(1-_b[$y])"
678      esttab `mm' `save', mtitle(`mt') b(%6.3f) nogaps ///

```

```

679      star(* 0.1 ** 0.05 *** 0.01) replace      ///
680      scalar(N r2 r2_a r2_w LongR) compress      ///
681      keep(`keep') order(`keep')      ///
682      addnotes("LongR: Long-run effect of target `note'")
683
684      *-----
685      *-Smart! See P.484 倒数第二段
686
687      *-一个 Target 的 Good Proxy 应该具有与 MDR 相似的分布特征
688      *-基本统计量比较
689      tabstat DR_star $y, s(mean sd median skew kurto q) ///
690                      c(s) f(%6.3f)
691      *-分布函数比较
692      twoway (kdensity DR_star, lw(*1.3) lc(red))      ///
693            (kdensity $y, lw(*1.3) lc(blue)), ///
694            legend(label(1 DR_star) label(2 $y))
695
696      *-----
697      *-Comments
698      *
699      *-(1) Stata 提供了很多产生随机数的函数
700      *      help random_number_functions
701      *-(2) 文中 Table 4 第二行的标注方式容易让人产生误解, 如下方式更好:
702      *      S.D. of ss: 0.01 0.05 0.10 0.2 0.25 0.5
703      *-(3) 若 y=BDR, 则得到的结果与文中非常相似
704      *-(4) 若不采用 OLS, 而是采用 FE 估计, 则结果差别很大, 为何?
705      *      winthin-transformation can remove the Mesurement Error
706      *      虽然 t 期和 t 期都存在 ME, 但由于程度相当,
707      *      所以 WG 时会被删掉
708
709
710
711
712      *-----
713      * Table 5 Pecking Order and Market Timing
714      *-----
715
716      *-目的: 检验 PO 和 MT 相对于 Trade-off 的解释能力
717      *
718      *-方法: 把反应 PO 和 MT 的指标加入 Trade-off 模型中, 即 Eq.(4)
719      *      参见正文中的 Eq.(7) 和 Eq.(8)
720      *
721      *-模型:
722      *
723      * Col(1),(2), and (4)
724      *
725      *  $BDR_{t+1} = (\text{Lamda} * \text{Beta})X_t + (1 - \text{Lamda})BDR_t + \text{Gama} * Z_t + e_{t+1}$ 
726      *
727      * Col(3) 正文中 P.488 此处表述有误, should be "In column (3)"
728      *
729      *  $D.BDR_{t+1} = (\text{Lamda} * \text{Beta})X_t - \text{Lamda} * BDR_t + \text{Gama} * Z_t + e_{t+1}$ 
730      *
731      * Col(6):
732      *
733      *  $D.MDR_{t+1} = (\text{Lamda} * \text{Beta})X_t - \text{Lamda} * MDR_t + \text{Gama} * Z_t + e_{t+1}$ 
734      *
735      *-MB_EFWA 可能存在严重的离群值
736
737      *-----
738      *- Table 5: Panel A
739      *-----
740
741      global xx "EBIT_TA MB DEP_TA LnTA FA_TA Ind_Median" //公共解释变量
742
743      *-Col(1) BDR, FE
744      xi: xtreg F_BDR BDR $xx i.year, fe
745      est store m1_BDR
746

```

```

747  *--Col(2)  BDR, FE  MB_EFWA
748      xi: xtreg F_BDR BDR $xx MB_EFWA i.year, fe
749      est store m2_BDR
750      mfx
751  *--Col(3)  D.BDR, FE  FINDEF
752      qui tsset
753      xi: xtreg D.F_BDR BDR $xx FINDEF i.year, fe
754      est store m3_D_BDR
755
756  *--Col(4)  BDR, FE MB_EFWA FINDEF
757      xi: xtreg F_BDR BDR $xx MB_EFWA FINDEF i.year, fe
758      est store m4_BDR
759
760  *--Col(1M)  MDR, FE
761      xi: xtreg F_MDR MDR $xx i.year, fe
762      est store m1a_MDR
763
764  *--Col(5)  MDR, FE  MB_EFWA
765      qui tsset
766      xi: xtreg F_MDR MDR $xx MB_EFWA i.year, fe
767      est store m5_MDR
768
769  *--Col(6)  D.MDR, FE  FINDEF
770      qui tsset
771      xi: xtreg D.F_MDR MDR $xx FINDEF i.year, fe
772      est store m6_D_MDR
773
774  *--Col(7)  MDR, FE MB_EFWA FINDEF
775      xi: xtreg F_MDR MDR $xx MB_EFWA FINDEF i.year, fe
776      est store m7_MDR
777
778  *--结果汇总
779      local save "using Tab05_A.csv"  //选中此行执行,结果输出到 Excel
780      local mm "m1_BDR m2_BDR m3_D_BDR m4_BDR m1a_MDR m5_MDR m6_D_MDR m7_MDR"
781      local keep "BDR MDR $xx MB_EFWA FINDEF"
782      esttab `mm' `save', mtitle(`mm') b(%6.3f) nogaps replace ///
783          star(* 0.1 ** 0.05 *** 0.01) keep(`keep') compress ///
784          scalar(N r2_o r2_b r2_w) order(`keep')
785
786  *--Comments
787  *
788  * (1) 采用不同的方式处理 MB_EFWA 变量的离群值,得到的结果差异很大;
789  *
790  * (2) 采用这种方式比较 Trade-off, Pecking-order, Market-timing 合理吗?
791
792  *--对 MB_EFWA 进行 1%,99% winsor 处理的结果
793      tabstat MB_*,s(mean sd q min max) c(s) f(%6.3f)
794      doedit Table05_check.do //打开文档, see My Appendix B.1 (Boottom)
795      do      Table05_check.do //执行文档
796
797
798  *-----
799  *-- Table 5: Panel B
800  *-----
801
802  *--Goal: 采用标准化系数比较不同解释变量的相对重要性
803
804  *--Method: 对每个变量执行标准化处理,进而执行 xtreg,fe 估计
805  *          或,采用 areg, beta
806
807      qui tsset
808      dropvars DBDR DMDR
809      gen DBDR = D.BDR
810      gen DMDR = D.MDR
811
812  *--估计 BDR* and MDR*
813      dropvars BDR_star MDR_star
814      qui xi: xtreg F_BDR $xx i.year, fe

```

```

815     predict BDR_star, xbu // BDR*, 其中已经考虑了个体效应
816     qui xi: xtreg F_MDR $xx i.year, fe
817     predict MDR_star, xbu // MDR*
818
819     *-定义存储结果的矩阵
820     mat tab5b = J(3,8,.)
821
822     *-标准化处理
823     dropvars s_* sx_*
824     global yy "BDR MDR DBDR DMDR MB_EFWA FINDEF BDR_star MDR_star"
825     * ssc install center, replace //下载或更新该命令
826     bysort id: center $yy, prefix(s_) standardize
827     *-Note: 上述处理已经去掉了个体效应
828
829     *-Col(2) BDR, FE MB_EFWA (OLS 估计即可)
830     xi: reg F.s_BDR s_BDR s_BDR_star s_MB_EFWA i.year
831     mat tab5b[1,1] = _b[s_BDR_star]
832     qui sum BDR if e(sample)
833     mat tab5b[1,2] = _b[s_BDR_star]/r(sd)
834     mat tab5b[2,1] = _b[s_MB_EFWA]
835     mat tab5b[2,2] = _b[s_MB_EFWA]/r(sd)
836
837     *-Col(3) D.BDR, FE FINDEF
838     qui tsset
839     xi: reg F.s_DBDR s_BDR s_BDR_star s_FINDEF i.year
840     mat tab5b[1,3] = _b[s_BDR_star]
841     qui sum F.s_DBDR if e(sample)
842     mat tab5b[1,4] = _b[s_BDR_star]/r(sd)
843     mat tab5b[3,3] = _b[s_FINDEF]
844     mat tab5b[3,4] = _b[s_FINDEF]/r(sd)
845
846     *-Col(5) MDR, FE MB_EFWA
847     qui tsset
848     xi: reg F.s_MDR s_MDR s_MDR_star s_MB_EFWA i.year
849     mat tab5b[1,5] = _b[s_MDR_star]
850     qui sum MDR if e(sample)
851     mat tab5b[1,6] = _b[s_MDR_star]/r(sd)
852     mat tab5b[2,5] = _b[s_MB_EFWA]
853     mat tab5b[2,6] = _b[s_MB_EFWA]/r(sd)
854
855     *-Col(6) D.MDR, FE FINDEF
856     qui tsset
857     xi: reg F.s_DMDR s_MDR s_MDR_star s_FINDEF i.year
858     mat tab5b[1,7] = _b[s_MDR_star]
859     qui sum F.s_DBDR if e(sample)
860     mat tab5b[1,8] = _b[s_MDR_star]/r(sd)
861     mat tab5b[3,7] = _b[s_FINDEF]
862     mat tab5b[3,8] = _b[s_FINDEF]/r(sd)
863
864     mat rownames tab5b = Est_Target MB_EFWA FINDEF
865     local t1 "Abs Pct_BDR_sd Abs Pct_DBDR_sd"
866     local t2 "Abs Pct_MDR_sd Abs Pct_DMDR_sd"
867     mat colnames tab5b = `t1' `t2'
868
869     *-呈现结果
870     mat list tab5b, f(%6.4f)
871
872     *-输出到 Excel (执行后 matrix 会消失)
873     logout, save("Tab05b") excel fix replace: ///
874     mat list tab5b, f(%6.4f)
875
876
877     *-----
878     * Table 6&7 (omitted)
879     *-----
880
881
882

```



```

883  *=====
884  *- Part III: Robustness Check
885  *=====
886
887  *-----
888  * Table 8 : Estimating over differing forecast
horizons
889  *-----
890
891  *-目的：随着时间的拉长，调整速度是否符合理论上预期的几何级数关系？
892  *
893  *      s2 = [1-(1-s1)^2]  (s1:一年期调整速度；s2:两年期调整速度)
894  *
895  *-方法：拉长分析的时间跨度 k = 2,3,4,5 年
896  *
897  *-模型：
898  *
899  *   MDR[t+k]=(Lamda*Beta)X[t] + (1-Lamda)*MDR[t] + e[t+1]
900
901  global y  BDR
902  global y  MDR
903  global xx "EBIT_TA MB DEP_TA LnTA FA_TA Ind_Median"
904
905  *-k=1
906  xi: xtreg F1.$y $y $xx i.year, fe
907  est store k1
908  *-k=2
909  xi: xtreg F2.$y $y $xx i.year, fe
910  est store k2
911  *-k=3
912  xi: xtreg F3.$y $y $xx i.year, fe
913  est store k3
914  *-k=4
915  xi: xtreg F4.$y $y $xx i.year, fe
916  est store k4
917  *-k=5
918  xi: xtreg F5.$y $y $xx i.year, fe
919  est store k5
920
921  *-结果汇总
922  local save "using Tab08_$y" // 输出结果
923  local mm "k1 k2 k3 k4 k5"
924  local keep "$y $xx"
925  esttab `mm' `save', mtitle(`mm') b(%6.3f) replace ///
926  compress nogaps order(`keep') keep(`keep') ///
927  star(* 0.1 ** 0.05 *** 0.01) ///
928  scalar(N r2_o r2_b r2_w)
929
930  *-Comments:
931  *
932  *- (1) k>2 时，估计出的调整速度已经大于 1 了，可能是样本区间过短所致；
933  *- (2) MDR 和 BDR 的分析结果存在较大差异，尤其是 R2 和 adj-R2
934
935
936  *-----
937  *-完整版本
938  *-----
939  *global y  BDR
940  global y  MDR
941  global xx "EBIT_TA MB DEP_TA LnTA FA_TA Ind_Median"
942
943  *-k=1
944  xi: xtreg F1.$y $y $xx i.year, fe
945  local s0 = 1-_b[$y]
946  estadd scalar s = 1-_b[$y]
947  estadd scalar Half_Life = ln(2)/(1-_b[$y])
948  est store k1
949  *-k=2

```

```

950     local k=2
951     xi: xtreg F`k'.$y $y $xx i.year, fe
952     estadd scalar s = 1-_b[$y]
953     estadd scalar s_p = 1-(1-`s0')^`k' //理论上的调整速度, pp.496
954     estadd scalar Half_Life = ln(2)/(1-_b[$y])
955     est store k`k'
956 *--k=3
957     local k=3
958     xi: xtreg F`k'.$y $y $xx i.year, fe
959     estadd scalar s = 1-_b[$y]
960     estadd scalar s_p = 1-(1-`s0')^`k'
961     estadd scalar Half_Life = ln(2)/(1-_b[$y])
962     est store k`k'
963 *--k=4
964     local k=4
965     xi: xtreg F`k'.$y $y $xx i.year, fe
966     estadd scalar s = 1-_b[$y]
967     estadd scalar s_p = 1-(1-`s0')^`k'
968     estadd scalar Half_Life = ln(2)/(1-_b[$y])
969     est store k`k'
970 *--k=5
971     local k=5
972     xi: xtreg F`k'.$y $y $xx i.year, fe
973     estadd scalar s = 1-_b[$y]
974     estadd scalar s_p = 1-(1-`s0')^`k'
975     estadd scalar Half_Life = ln(2)/(1-_b[$y])
976     est store k`k'
977
978 *--结果汇总
979     local save "using Tab08a_$y" // 输出结果
980     local mm "k1 k2 k3 k4 k5"
981     local keep "$y $xx"
982     esttab `mm' `save', mtitle(`mm') b(%6.3f) replace ///
983             compress nogaps order(`keep') keep(`keep') ///
984             star(* 0.1 ** 0.05 *** 0.01) ///
985             scalar(N r2_o r2_b r2_w s s_p Half_Life) ///
986             addnotes("s: observed speed" "s_p: theoretical speed")
987
988
989
990 *-----
991 * Table 9: Stability across firm sizes
992 *-----
993
994 *--目的: 前文结果在不同规模的公司中是否稳定?
995 *--方法:
996 *   s1: 每年将所有公司按规模(equity MV)分成 10 组;
997 *   s2: 分别估计 Smallest-->Largest 组的调整速度;
998
999 *--Equity Market Value
1000     dropvars EMV
1001     gen EMV = MV-TD //总市值-债务账面价值
1002
1003 *--定义相关变量(通过这里的组合, 可以检验多种结果)
1004     global y BDR //选择被解释变量
1005     global y MDR
1006     *global gvar "LnTA" //按总资产分组
1007     global gvar "EMV" //按 Equity market value 分组
1008     *global gvar "$y" //新增分组方式, 按照负债率高低
1009     global xx "EBIT_TA MB DEP_TA LnTA FA_TA Ind_Median"
1010
1011 *--按 q10-->q90 分成十组
1012     dropvars q gg
1013     bysort year: quantiles $gvar, gen(q) n(10) //分成十组
1014
1015 *--统计各组的财务特征
1016     table q
1017     tabstat $gvar $y $xx, s(mean) f(%6.3f) by(q)

```

```

1018
1019    *-重新定义组别   pp.498 Col_1-Col_4
1020    gen gg = .
1021    replace gg=1 if inlist(q,9,10) //deciles 1&2 Largest 2 groups
1022    replace gg=2 if inlist(q,7,8) //deciles 3&4
1023    replace gg=3 if inlist(q,4,5,6) //deciles 5,6&7
1024    replace gg=4 if inlist(q,1,2,3) //deciles 8,9,10
1025
1026    *-回归分析
1027    qui tsset
1028    forvalues g=1/4{
1029        xi:xtreg F_{$y} $y $xx i.year if gg==`g', fe
1030        estadd scalar speed = 1-_b[{$y}]
1031        estadd scalar Half_Life = ln(2)/(1-_b[{$y}])
1032        est store g`g'
1033    }
1034
1035    *-结果汇总
1036    local y "{$y}"
1037    local save "using Tab9_`y'_$gvar.csv" // 输出结果
1038    local mm "g1 g2 g3 g4"
1039    local mt "Deciles1_2 Deciles3_4 Deciles5_6_7 Decile8_9_10"
1040    local keep "{$y} $xx"
1041    local n1 "Notes: (1) Depend variable: $y"
1042    local n2 "          (2) Grouped by: $gvar"
1043    local n3 "          (3) Decile #1 is largest;"
1044    local n4 "          --> Decile #10 is smallest"
1045    esttab `mm' `save', mtitle(`mt') b(%6.3f) ///
1046            nogaps order(`keep') keep(`keep') ///
1047            star(* 0.1 ** 0.05 *** 0.01) replace ///
1048            scalar(N r2_o r2_b r2_w speed Half_Life) ///
1049            addnotes("`n1'" "`n2'" "`n3'" "`n4'")
1050
1051    *-----
1052    *-Comments & Discussions:
1053    *
1054    * (1) 我们的检验结果似乎与Flannery不同, 如何解释?
1055    * (2) pp.497, Ft.25 MB in Table 9 is positive,
1056    *      "We have no good explanation for this result"
1057    *      large firms: rely more on public debt;
1058    *      small firms: rely more on private (bank) debt
1059    *      public debt has few covenants, thus slow adjustment
1060    *      如果是你, 该如何作答?
1061
1062
1063
1064
1065    *-----
1066    * Table 10: Stability over time
1067    *-----
1068
1069    global y BDR
1070    global y MDR
1071    global xx "EBIT_TA MB DEP_TA LnTA FA_TA Ind_Median"
1072
1073    *-我们的样本区间
1074    qui reg $y $xx
1075    xtides if e(sample)
1076
1077    *-1998-2002
1078    xi: xtreg F_{$y} $y $xx i.year if (year>=1998&year<=2002), fe
1079    estadd scalar speed = 1 - _b[{$y}]
1080    estadd scalar Half_Life = ln(2)/(1-_b[{$y}])
1081    est store m_9802
1082
1083    *-1998-2006
1084    xi: xtreg F_{$y} $y $xx i.year if (year>=1998&year<=2006), fe
1085    estadd scalar speed = 1 - _b[{$y}]

```

```

1086     estadd scalar Half_Life = ln(2)/(1-_b[$y])
1087     est store m_9806
1088
1089     *-2003-2006
1090     xi: xtreg F_$y $y $xx i.year if (year>=2003&year<=2006), fe
1091     estadd scalar speed = 1 - _b[$y]
1092     estadd scalar Half_Life = ln(2)/(1-_b[$y])
1093     est store m_0306
1094
1095     *-2003-2008
1096     xi: xtreg F_$y $y $xx i.year if (year>=2003&year<=2008), fe
1097     estadd scalar speed = 1 - _b[$y]
1098     estadd scalar Half_Life = ln(2)/(1-_b[$y])
1099     est store m_0308
1100
1101     *-结果汇总
1102     local save "using Tab10_$y.csv" //执行时选中此行可以把结果输出到 Excel 中
1103     local mm "m_9802 m_9806 m_0306 m_0308"
1104     local mt "1998-2002 1998-2006 2003-2006 2003-2008"
1105     local keep "$y $xx"
1106     esttab `mm' `save', mtitle(`mt') b(%6.3f) nogaps replace ///
1107     star(* 0.1 ** 0.05 *** 0.01) keep(`keep') compress ///
1108     scalar(N r2_o r2_b r2_w speed Half_Life) ///
1109     addnotes("Speed = 1-_b[$y]" ///
1110             "Half-Life = ln2/Speed (Years)")
1111
1112
1113
1114     *-----
1115     * Appendix A.1: Alternative Methods for Dynamic Panel Data
1116     *-----
1117
1118     *-目的: 采用其它方法估计模型
1119     * IV: Anderson and Hsiao (1982)
1120     * FD-GMM: Arellano and Bond (1991)
1121     * SYS-GMM: Arellano and Bover(1995)
1122     * Blundell and Bond (1998) New add
1123
1124     global y MDR
1125     global xx "EBIT_TA MB DEP_TA LnTA FA_TA Ind_Median"
1126
1127     *-Pooled OLS
1128     xi: reg F_$y $y $xx i.year
1129     estadd scalar speed = 1 - _b[$y]
1130     estadd scalar Half_Life = ln(2)/(1-_b[$y])
1131     est store OLS
1132
1133     *-FE
1134     xi: xtreg F_$y $y $xx i.year, fe
1135     estadd scalar speed = 1 - _b[$y]
1136     estadd scalar Half_Life = ln(2)/(1-_b[$y])
1137     est store FE
1138
1139     *-IV instruments: MDR[t-1] for D.MDR[t]
1140     *-IV 估计 (also see: help xtivreg2)
1141     *-序列相关检验
1142     preserve
1143     dropvars cx_* dx_*
1144     center F_$y MDR, prefix(cx_) //去除个体效应
1145     center $xx, prefix(dx_)
1146     gen Lcx_MDR = L.cx_MDR
1147     qui tab year, gen(yr)
1148     ivreg2 cx_F_$y (cx_$y=Lcx_MDR) dx_* yr*
1149     abar //序列相关检验
1150     local ar1 = r(ar1)
1151     local ar1p = r(ar1p)
1152     restore
1153     *-IV 估计

```

```

1154     xi: xtivreg F_{$y} ($y=L.MDR) $xx i.year, fe
1155     estadd scalar speed = 1 - _b[$y]
1156     estadd scalar Half_Life = ln(2)/(1-_b[$y])
1157     estadd scalar ar1 = `ar1'
1158     estadd scalar arlp = `arlp'
1159     est store IV_MDR
1160
1161     *-IV instruments: BDR[t-1] for D.MDR[t]
1162
1163     *-序列相关检验
1164     preserve
1165     dropvars cx_* dx_*
1166     center F_{$y} $y BDR, prefix(cx_) //去除个体效应
1167     center $xx, prefix(dx_)
1168     gen Lcx_BDR = L.cx_BDR
1169     qui tab year, gen(yr)
1170     ivreg2 cx_F_{$y} (cx_{$y}=Lcx_BDR) dx_* yr*
1171     abar //序列相关检验
1172     local ar1 = r(ar1)
1173     local arlp = r(arlp)
1174     restore
1175     *-IV 估计
1176     xi: xtivreg F_{$y} ($y=L.BDR) $xx i.year, fe
1177     estadd scalar speed = 1 - _b[$y]
1178     estadd scalar Half_Life = ln(2)/(1-_b[$y])
1179     estadd scalar ar1 = `ar1'
1180     estadd scalar arlp = `arlp'
1181     est store IV_BDR
1182
1183     *-FD-GMM: Arellano and Bond (1991)
1184
1185     xi: xtabond F.{$y} $xx i.year, maxldep(5)
1186     estadd scalar speed = 1 - _b[$y]
1187     estadd scalar Half_Life = ln(2)/(1-_b[$y])
1188     *-模型设定合理性检验
1189     xi: xtabond F.{$y} $xx i.year, twostep maxlags(5)
1190     estat abond // AR test 序列相关检验
1191     estadd scalar ar2 = e(arm2)
1192     estadd scalar ar2p = 2*normal(e(arm2))
1193     estat sargan // Sargan test 过度识别检验
1194     estadd scalar sargan_p = 1-chi2(e(zrank)-e(rank), e(sargan))
1195     est store FD_GMM
1196
1197     *-SYSGMM: Arellano and Bover (1995), BB1998
1198
1199     xi: xtdpdsys F.{$y} $xx i.year, maxlags(5) maxldep(5)
1200     estadd scalar speed = 1-_b[$y]
1201     estadd scalar Half_Life = ln(2)/(1-_b[$y])
1202     *-模型设定合理性检验
1203     xi: xtdpdsys F.{$y} $xx i.year, twostep maxlags(5) maxldep(5)
1204     estat abond // AR test 序列相关检验
1205     estadd scalar ar2 = e(arm2)
1206     estadd scalar ar2p = 2*normal(e(arm2))
1207     estat sargan // Sargan test 过度识别检验
1208     estadd scalar sargan_p = 1-chi2(e(zrank)-e(rank), e(sargan))
1209     est store SYS_GMM
1210
1211
1212     *-结果汇总
1213     local save "using TabA1.csv"
1214     local mm "OLS FE IV_MDR IV_BDR FD_GMM SYS_GMM"
1215     local keep "$y $xx"
1216     esttab `mm' `save', mtitle(`mm') b(%6.3f) compress nogaps ///
1217     star(* 0.1 ** 0.05 *** 0.01) keep(`keep') ///
1218     scalar(N r2 r2_a r2_w ar1 arlp ar2 ar2p sargan sargan_p)
1219
1220     *-----
1221     * Final Comments

```

```

1222  *-----
1223  *
1224  *-选题的视角很重要(个体效应);
1225  *
1226  *-稳健性检验的途径(没有固定的模式);
1227  *
1228  *-计量模型和估计方法的合理应用(如 Measurement Error, Mean reverse);
1229  *
1230  *-一个疑点: why not GMM ?
1231
1232
1233
1234
1235  *-----
1236  *   My Appendix B.1
1237  *-----
1238
1239  *-----
1240  *   Table 5       Another Check
1241  *-----
1242
1243  *-目的: 对 MB_EFWA0 进行 1%,99% winsor 处理的结果
1244  *
1245  * 正文中的处理方式: 删除 MB_EFWA0>10 的观察值
1246  *   replace MB_EFWA0=. if MB_EFWA0>10 // Baker(2002,p.12, Footnote 8)
1247  *
1248  *
1249  *   sum MB_EFWA0, detail
1250  *   replace MB_EFWA0=r(p1) if MB_EFWA0<r(p1)
1251  *   replace MB_EFWA0=r(p99) if MB_EFWA0>r(p99)
1252  *
1253  *   global xx "EBIT_TA MB DEP_TA LnTA FA_TA Ind_Median" //公共解释变量
1254  *
1255  *-Col(1)  BDR, FE
1256  *   xi: xtreg F_BDR BDR $xx i.year, fe
1257  *   est store m1a_BDR
1258  *
1259  *-Col(2)  BDR, FE MB_EFWA0
1260  *   xi: xtreg F_BDR BDR $xx MB_EFWA0 i.year, fe
1261  *   est store m2a_BDR
1262  *
1263  *-Col(3)  D.BDR, FE FINDEF
1264  *   qui tsset
1265  *   xi: xtreg D.F_BDR BDR $xx FINDEF i.year, fe
1266  *   est store m3a_D_BDR
1267  *
1268  *-Col(4)  BDR, FE MB_EFWA0 FINDEF
1269  *   xi: xtreg F_BDR BDR $xx MB_EFWA0 FINDEF i.year, fe
1270  *   est store m4a_BDR
1271  *
1272  *-Col(5)  MDR, FE
1273  *   xi: xtreg F_MDR MDR $xx i.year, fe
1274  *   est store m5a_MDR
1275  *
1276  *-Col(6)  MDR, FE MB_EFWA0
1277  *   qui tsset
1278  *   xi: xtreg F_MDR MDR $xx MB_EFWA0 i.year, fe
1279  *   est store m6a_MDR
1280  *
1281  *-Col(7)  D.MDR, FE FINDEF
1282  *   qui tsset
1283  *   xi: xtreg D.F_MDR MDR $xx FINDEF i.year, fe
1284  *   est store m7a_D_MDR
1285  *
1286  *-Col(8)  MDR, FE MB_EFWA0 FINDEF
1287  *   xi: xtreg F_MDR MDR $xx MB_EFWA0 FINDEF i.year, fe
1288  *   est store m8a_MDR
1289

```



```

1290    *--结果汇总
1291    local save "using Tab05a.csv" //选中此行执行,结果输出到 Excel
1292    local m1 "m1a_BDR m2a_BDR m3a_D_BDR m4a_BDR"
1293    local m2 "m5a_MDR m6a_MDR m7a_D_MDR m8a_MDR"
1294    local keep "BDR MDR $xx MB_EFWA0 FINDEF"
1295    esttab `m1' `m2' `save', mtitle(`m' `m2') b(%6.3f) nogaps ///
1296    star(* 0.1 ** 0.05 *** 0.01) keep(`keep') compress ///
1297    scalar(N r2 r2_a r2_w) order(`keep') replace ///
1298    addnotes("winsor MB_EFWA0 at 1% and 99%")
1299
1300
1301    *-----
1302    * Table 5 Panel B Column(2).... 的另一种计算方法
1303    *-----
1304    *--Col(2) BDR, FE MB_EFWA
1305    xi: xtreg F_BDR BDR $xx MB_EFWA i.year, fe
1306    qui sum BDR
1307
1308    *--计算 BDR_star
1309    dropvars BDR_star*
1310    local lamda = 1- _b[BDR] //lamda
1311    gen BDR_star_gross = _b[_cons] //((lamda*beta)X_it = lamda*BDR*_it
1312    foreach v of varlist $xx{
1313        replace BDR_star_gross = BDR_star_gross + _b[`v']*`v' if e(sample)
1314    }
1315    replace BDR_star = BDR_star/`lamda' //(beta)X_it = BDR*_it
1316
1317    sum BDR BDR_star if e(sample)
1318
1319    *--计算 BDR 变动 (s.d.)
1320    sum BDR_star if e(sample)
1321    local sigma_BDR_star = r(sd)
1322    local mean_BDR_star = r(mean)
1323    local chg_BDR = `lamda'*r(sd)
1324    qui sum BDR if e(sample)
1325    local chg_BDR_pct = `chg_BDR'/r(sd)*100
1326
1327    dis in g "Impact on BDR (Absolute): " in y %6.4f `chg_BDR'
1328    dis in g "Impact on BDR (% of BDR's Std. Dev.): " in y %4.2f `chg_BDR_pct' "%"
1329
1330

```

```

1
2  *      -----
3  *      -----
4  *
5  *      Stata 学术论文专题
6  *
7  *      -----
8  *      -----
9
10
11  *      主讲人：连玉君 副教授
12  *
13  *      单 位：中山大学岭南学院金融系
14  *      电 邮：arlionn@163.com
15  *      博 客：http://blog.cnfol.com/arlionn
16  *      主 页：http://goo.gl/tRXba
17  *      微 博：http://weibo.com/arlionn
18
19
20  *-面板门槛模型
21  *-----
22  *   Hansen, B., 1999.
23  *   Threshold effects in non-dynamic panels:
24  *   Estimation, testing, and inference.
25  *   Journal of Econometrics, 93(2):345-368.
26  *-----
27
28
29  *-主要方法
30  *   Panel Threshold Model: xtthres
31  *   Bootstrap
32  *   FE (Fixed Effect Model): xtreg, fe
33
34  global path "`c(sysdir_personal)'PX_papers" //定义课程目录
35
36  cd "$path\Hansen_1999" // 进入本讲目录
37  adopath + "$path\adofiles" // 自编程序
38
39
40  *-原始论文
41  shellout "$path\Refs\Hansen_1999.pdf"
42
43  *-中文介绍
44  shellout "$path\Refs\连玉君_08_Panel_Data.pdf"
45
46
47
48
49  *-----
50  *-> 面板门槛模型
51  *-----
52
53  *-----
54  *-1.1 简介
55
56  *-----
57  *-1.1.1 结构突变问题
58
59  *-----
60  *-结构变化问题非常普遍
61
62  *-人力资本与经济增长      _| 型, 扁 S 型
63
64  *-第一大股东持股比例与公司价值      倒 U 型
65
66  *-资本结构与公司价值      倒 U 型
67
68  *-----

```

```

69      *-典型的处理方法
70
71      * 加入解释变量的二次项, 如 reg y x x^2
72
73      * 加入虚拟变量和交乘项, 如 reg y x d d*x (d=1 if top1>0.5)
74
75      * 分组回归
76
77      *-----
78      * 存在的问题: x 与 x^2 往往是高度共线性的;
79      *              如何确定分组界点? 错误的界点会导致严重的偏误
80
81
82      *-----
83      *-1.1.2 面板门槛模型的设定
84
85      shellout "$path\Refs\Hansen_1999.pdf"
86
87      *      { u[i] + b1*X[it] + e[it]   if q[it]<=gamma
88      *  y[it] = {
89      *      { u[i] + b2*X[it] + e[it]   if q[it]> gamma      (1)
90      *
91      *- y[it] -- 被解释变量
92      *- X[it] -- 解释变量
93      *- q[it] -- 门槛变量
94
95      *-----
96      *-Key points:
97      *
98      * (1) gamma 是未知的, 将通过数据的特征自动估计出来;
99      * (2) 可以很方便地扩展多多个门槛的情形
100
101      *-另一种表述方式:
102      *
103      *  y[it] = u[i] + b1*X[it]*I(q[it]<=gamma)      (2)
104      *          + b2*X[it]*I(q[it]>gamma) + e[it]
105
106      *-估计方法:
107      * 在样本区间内遍试所可能的 q[it] 值, 拟合的最好的就是所需的届点
108      *
109      *-检验方法:
110      * Bootstrap, 因为在原假设下, 参数 gamma 是未知的;
111      * 详见: Stata 高级视频, B9_MC_BS.do
112      *
113      *-命令: xtthres 和 xttr_graph
114
115      *-----
116      *-1.1.3 应用
117
118      *-连玉君, 程建.
119      * 不同成长机会下资本结构与经营绩效之关系研究, 当代经济科学, 2006年02期.
120
121      *-李平, 许家云.
122      * 国际智力回流的技术扩散效应研究—基于中国地区差异及门槛回归的实证分析.
123      * 经济学(季刊). 2011年3期
124
125      *-李梅, 柳士昌
126      * 对外直接投资逆向技术溢出的地区差异和门槛效应
127      * 管理世界. 2012年第1期
128
129      *-刘胜强, 刘星.
130      * 董事会规模对企业R&D投资行为的门槛效应分析.
131      * 预测. 2010年6期
132
133
134
135      *-----
136      *-1.2 估计方法: 一份模拟的数据

```

```
137
138 *-----
139 *-模拟数据的生成：单一门槛
140 *-模拟方法详见：Stata 高级视频，B9_MC_BS.do
141
142 clear
143 set obs 100
144 set seed 123456
145 gen e = invnorm(uniform()) // e~N(0,1)
146 gen x = 3*invnorm(uniform()) // x~N(0,3^2)
147 gen t = _n // t=1,2,3...
148 gen y = .
149 tsset t
150 replace y = 1 + 2*x + e
151 replace y = 1 + -2*x + e if t>50
152 save xtthres_sim1.dta, replace
153
154
155 *-----
156 *-基本统计分析
157
158 scatter y t
159 scatter y x
160
161 reg y x
162 est store full
163 reg y x if t<=50
164 est store left
165 reg y x if t>50
166 est store right
167
168 local m "full left right"
169 esttab `m', mtitle(`m')
170
171
172 *-----
173 *-虚拟变量回归：Chow Test 的基本思想
174
175 *-如果知道断点 在 t=50
176 gen d = (t>50)
177 gen xd = x*d
178 reg y x xd
179 est store true50
180
181 *-如果不知道断点
182
183 gen d10 = (t>10) // 假设 t=10
184 gen xd10 = x*d10
185 reg y x xd10
186 est store t10
187
188 gen d20 = (t>20) // 假设 t=20
189 gen xd20 = x*d20
190 reg y x xd20
191 est store t20
192
193 gen d30 = (t>30) // 假设 t=30
194 gen xd30 = x*d30
195 reg y x xd30
196 est store t30
197
198 gen d60 = (t>60) // 假设 t=60
199 gen xd60 = x*d60
200 reg y x xd60
201 est store t60
202
203 local m "full t10 t20 t30 true50 t60"
204 esttab `m', mtitle(`m') s(r2 rss ll) nogap compress
```

```

205
206     *-评论:
207     * (1) 在 t=50 处, RSS 是最小的, LL 和 R2 都是最大的
208     * (2) 在 t!=50处, 模型的参数估计偏差很大,
209     *     这意味着传统的估计方法可能存在严重偏误(事先指定断点位置)
210     *     当然, 离 t=50 越近, RSS 也越小, 这对建模有一定的启示
211     * (3) 更为合理的方法?
212
213
214     *-----
215     *-网格搜索结构突变点
216
217     gen rss = .
218     forvalues i=5/95{
219         gen di = (t>`i')
220         gen xdi = x*di
221         qui reg y x xdi
222         qui replace rss = e(rss) in `i'
223         drop di xdi
224     }
225     sort rss
226     list rss t in 1/5
227
228     *-图示
229     line rss t, xline(50,lp(dash) lw(thick)) ///
230                     lw(thick) xlabel(0(10)100) sort
231
232
233
234
235     *-----
236     *-两个突变点
237
238     *-生成模拟数据
239     clear
240     set seed 1357 // 种子值, 只是为了保证结果可重现
241     set obs 2000
242     gen t = _n // t=1,2,3...
243     gen e = invnorm(uniform()) if t<=100 // e~N(0,1)
244     gen x = 3*invnorm(uniform()) if t<=100 // x~N(0,3^2)
245     gen y = .
246     tsset t
247     replace y = 1 + 2*x + e if t<=100
248     replace y = 1 + -2*x + e if t>30&t<=60
249     replace y = 1 + 4*x + e if t>60&t<=100
250     save xtthres_sim2.dta, replace // 保存一份模拟数据
251
252     scatter y x
253     scatter y t if t<=100 // 看不出规律
254
255     twoway (lfit y x in 1/30) (scatter y x in 1/30) ///
256            (lfit y x in 31/60) (scatter y x in 31/60) ///
257            (lfit y x in 61/100) (scatter y x in 61/100), ///
258            legend(off)
259
260     *-搜索一下
261     timer clear 1
262     timer on 1 // 开始记录用时
263     gen rss = .
264     forvalues i=15/95{ // 只是为了节省搜索时间
265         forvalues j=55/75{
266             qui{
267                 gen d1 = (t>`i')
268                 gen xd1 = x*d1
269                 gen d2 = (t>`j'&t!=`i')
270                 gen xd2 = x*d2
271                 qui reg y x xd1 xd2
272                 qui replace rss = e(rss) in `i'

```

```

273         drop d1 xd1 d2 xd2
274     }
275     dis "x" _c
276 }
277 }
278 timer off 1 // 结束记录用时
279 timer list 1
280 *-Q1: 一共需要循环多少次?
281 *-Q2: 有没有更好的搜索方法?  $K^2 \rightarrow 2*K$ 
282
283     sort rss
284     list rss t in 1/10, clean
285
286 *-图示
287     line rss t if t<100, xline(30 60,lp(dash) lw(thick)) ///
288         xlabel(0(30)90) lw(thick) sort
289
290
291 *-----
292 *-优化搜索方法
293
294     *-思路: 先搜索第一个门槛 t1;
295     *         固定第一个门槛 t1, 搜索第二个门槛 t2;
296     * [可选]固定第二个门槛, 重新搜索第一个门槛 t1_new;
297     *         可以重复上述过程 2-3 次, 即可确定最终的门槛值。
298
299 use xtthres_sim2.dta, clear
300 keep in 1/100
301
302 *-Step1: 搜索第一个门槛值
303     dropvars rss1
304     gen rss1 = . // 记录第一轮 RSS
305     forvalues i=5/95{
306         dropvars di xdi
307         qui gen di = (t>`i')
308         qui gen xdi = x*di
309         qui reg y x xdi
310         qui replace rss1 = e(rss) in `i'
311         dis "." _c
312     }
313     sort rss1
314     global t1 = t[1] // rss 最小值对应的 t 值
315     line rss1 t, lw(thick) xlabel(0(30)90) ///
316         xline(60,lp(dash) lc(red)) sort
317
318 *-Step2: 搜索第二个门槛值(固定住第一个门槛)
319     sort t // 这一步很重要, 保证了 rss 与 t 对应
320     dropvars rss2
321     gen rss2 = . // 记录第一轮 RSS
322     forvalues i=5/95{
323         qui{
324             dropvars d1 d2 xd1 xd2
325             local t_left = min(`i', $t1)
326             local t_right = max(`i', $t1)
327             *gen d1 = (t>`i')
328             *gen d2 = (t>$t1&t!=`i')
329             gen d1 = (t<`t_left')
330
331             gen d2 = (t>`t_left')&(t<`t_right')
332             gen xd1 = x*d1
333             gen xd2 = x*d2
334             reg y x xd1 xd2
335             replace rss2 = e(rss) in `i'
336             dis "0" _c
337         }
338     }
339     sort rss2
340     global t2 = t[1] // 第二个门槛值

```



```

341     line rss2 t, lw(thick) xlabel(0(30)90) sort
342
343     *-Step3: 重新搜索第一个门槛值(固定住第二个门槛)
344     sort t
345     dropvars rss3
346     gen rss3 = . // 记录第一轮 RSS
347     forvalues i=5/95{
348         qui{
349             dropvars d1 d2 xd1 xd2
350             local t_left = min(`i', $t2) // 注意此处的变化
351             local t_right = max(`i', $t2)
352             gen d1 = (t<`t_left')
353             gen d2 = (t>`t_right')
354             gen xd1 = x*d1
355             gen xd2 = x*d2
356             reg y x xd1 xd2
357             replace rss3 = e(rss) in `i'
358             dis "+" _c
359         }
360     }
361     sort rss3
362     global t12 = t[1] // 更新后的第一个门槛值
363     line rss3 t, lw(thick) xlabel(0(30)90) ///
364             xline(60,lp(dash) lc(red)) sort
365
366
367     *-三次搜索过程重现
368     local opt "lw(thick) sort"
369     twoway (line rss1 t, sort lw(*1.3)) ///
370            (line rss2 t, sort lw(*1.5)) ///
371            (line rss3 t, sort lw(*2.0)), ///
372            xline(30 60, lp(dash) lc(blue)) ///
373            legend(row(1)) xlabel(0(30)90)
374
375     *-Q1: 从 Step1 到 Step3, 一共搜索了多少次?
376
377     *-评论:
378     * (1) 对于大样本而言, 上述方法仍然非常耗时;
379     * Hansen(1999, p.349)建议如下:
380     * 只搜索门槛变量 q[it] 中的非重复值;
381     * 对这些非重复值排序, 在[1%th, 99%th]区间内搜索
382     * 如果需要搜索的格点仍然很多, 可以在上述区间内仅搜索特定的百分位数
383     * 如, {1.00%, 1.25%, 1.50%, 1.75%, 2%, 2,99.0}
384     * (2) 为了保证每个区间内都有足够的样本数, q[it] 的最大值和最小值
385     * 不能过于靠近两端;
386
387
388
389
390
391     *-----
392     *-1.3 假设检验
393
394     *-检验 1: 是否存在门槛效应? (以单一门槛为例)
395
396     *
397     * 
$$y[it] = \begin{cases} u[i] + b1*X[it] + e[it] & \text{if } q[it] \leq \gamma \\ u[i] + b2*X[it] + e[it] & \text{if } q[it] > \gamma \end{cases} \quad (1)$$

398     *
399     *
400     *- y[it] -- 被解释变量
401     *- X[it] -- 解释变量
402     *- q[it] -- 门槛变量
403
404     *-H0: b1=b2 ; H1: b1!=b2
405     *
406     * So - S1(gamma)
407     * F = ----- True value: F0
408     * sigma^2

```

```

409      *
410      *- So: 不存在门槛效应时的 OLS 残差平方和, reg y x --> e --> SUM e^2
411      *- S1: 存在门槛效应时的 OLS 残差平方和, reg y x xd --> e --> SUM e^2
412      *- sigma^2 = S1/[n(T-1)]
413
414      *-Bootstrap 步骤
415      * 1. Bootstrap the sample with replace; (以公司为单位)
416      * 2. Use the BS sample, 分别估计 H0 和 H1 下的模型, 计算 F 统计量,
417      * 3. 重复第一步和第二步 K=1000 次, 得到 Fj = F1, F2, F3, ..., F1000
418      * 4. 经验 p 值 = (Fj 大于 F0 的次数)/抽样次数
419
420      use xtthres_sim1.dta, clear
421
422      *-----
423      *-Step 1: 计算F真实值
424      *-----
425      preserve
426      *-1- 估计线性模型(无门槛), 记录 S0
427      qui reg y x
428      local S0 = e(rss)
429      *-2- 估计单一门槛模型, 记录 S1(gamma=50)
430      tempvar d xd
431      gen `d' = t>50
432      gen `xd' = x*`d'
433      qui reg y x `xd'
434      local S1 = e(rss)
435      local sigma2 = `S1'/(100-1) // N-1=100-1
436      *-3- 计算 F 值
437      global F0 = (`S0'-`S1')/`sigma2'
438      dis "F_true = " in y $F0
439      restore
440
441      *-----
442      *-Step 2: Bootstrap
443      *-----
444      global reps = 99 // bootstrap 次数
445      gen F = . // 记录 Bootstrap 得到的 F 值
446      dis _n in g "STATA 自抽样中, 请等待 ... ." _n // 纯属娱乐
447      forvalues j=1/$reps{
448          qui{
449              preserve
450              bsample // Bootstrap 抽样
451              *-计算F值
452              *-1- 估计线性模型(无门槛), 记录 S0
453              qui reg y x
454              local S0 = e(rss)
455              *-2- 估计单一门槛模型, 记录 S1(gamma=50)
456              *-2.1-搜索门槛值
457              gen rss = .
458              forvalues i=5/95{
459                  gen di = (t>`i')
460                  gen xdi = x*di
461                  qui reg y x xdi
462                  qui replace rss = e(rss) in `i'
463                  drop di xdi
464              }
465              sort rss
466              local gamma = t[1] // 找到了门槛值
467              *-2.2-估计单一门槛模型, 计算 BS 样本的 S1
468              gen d = t>`gamma'
469              gen xd= x*d
470              reg y x xd
471              local S1 = e(rss)
472              *-2.3-计算 F 值
473              local F = (`S0'/`S1'-1)*(100-1)
474              restore
475              replace F = `F' in `j' // 记录第 j 轮循环的 F 值
476          } // over qui

```

```

477         dis in y "+" _c
478     } // over reps
479
480     *-----
481     *-Step 3: 计算“经验 p 值”
482     *-----
483     dis "$F0"
484     count if F>$F0&F!=.
485     local p_value = r(N)/$reps
486     dis "Empirical p-value = " `p_value'
487
488
489
490     *-检验 2: 门槛值是否等于真实值?
491
492     *-H0: gamma_hat = gamma0 (Hansen,1999, p.351)
493
494     *-LR test
495     *           S1(gamma) - S1(gamma_hat)
496     *   LR(gamma) = -----
497     *                   sigma^2
498
499     *-Note: gamma_hat 是估计出来的门槛值;
500     *       gamma 是其他可能的门槛值;
501
502     *-关键点: 构造非拒绝域
503
504     *-临界值: Hansen(1999, Eq.15)
505     *
506     *   c(a) = -2log[1-sqrt(1-a)]    a 为显著水平, 如 5%, 10%
507
508     *-e.g.
509     local a = 0.05 // 5% 显著水平
510     local c5 = -2*ln(1-sqrt(1-`a'))
511     dis "5% 临界值 = " `c5'
512
513
514     use xtthres_sim1.dta, clear
515     gen S1 = . // S1(gamma)
516     forvalues i=5/95{ // Note: i=gamma
517         gen di = (t>`i')
518         gen xdi = x*di
519         qui reg y x xdi
520         qui replace S1 = e(rss) in `i'
521         drop di xdi
522     }
523     sort t
524     global S1_hat = S1[50] // S1(gamma_hat), gamma_hat=50
525     dis "S1_hat = " $S1_hat
526     global sigma2 = $S1_hat/(100-1)
527     gen LR = (S1-$S1_hat)/$sigma2 // LR value
528
529     global c5 = -2*ln(1-sqrt(1-0.05)) // 5% 临界值
530
531     *-图示
532     line LR t, xline(50,lp(dash) lw(*1.3) lc(red)) ///
533             yline($c5,lp(dat) lw(*1.5) lc(blue)) ///
534             lw(thick) xlabel(0(10)100) sort
535
536     *-5% 置信区间
537     sort LR
538     list LR if LR<=$c5
539
540
541
542
543
544

```

```

545  *-----
546  *-1.4 范例 A: 采用 xtthres 命令进行估计
547
548  help xtthres           // Panel Threshold model 估计程序
549  which xtthres          // 查看命令的存储位置和提示信息
550  adoedit xtthres.ado    // 编辑源程序文件
551
552  help xttr_graph        // 绘图程序
553  which xttr_graph
554
555
556  *-----begin-----
557  cap log close
558  log using xtthres_test01, text replace
559
560  *-1- 生成数据
561
562  clear
563  local N = 40
564  local T = 5
565  local NT=`N'*`T'
566  set obs `NT'
567  set seed 13579  // 种子值
568
569  *-产生截面和年度标示
570  egen id=seq(),from(1) to(`N') block(`T')
571  sort id
572  by id: gen year = _n + 2000
573  tsset id year
574
575  *-产生个体效应
576  gen ui_1 = invnorm(uniform()) if year==2001
577  bysort id: gen ui = ui_1[1] //个体效应不随时间改变
578  drop ui_1
579
580  *-产生干扰项和x
581  gen e = invnorm(uniform()) // e~N(0,1)
582  gen x = 3+10*invnorm(uniform()) // x~N(3,10^2)
583
584  *-产生门槛变量
585  gen q = uniform() // [0,1] 均匀分布
586
587  *-产生 y
588  gen y = .
589  replace y = ui + 2*x + e if q<0.3
590  replace y = ui + 8*x + e if q>=0.3&q<0.6
591  replace y = ui + 12*x+ e if q>=0.6
592
593  *-保存一份数据
594  save xtthres_sim_xt, replace
595
596  *-Note: 两个门槛, gamma1=0.3, gamma2=0.6
597
598  *-2- 使用 xtthres 命令进行估计
599  use xtthres_sim_xt.dta, clear
600  set seed 1357911
601  xtthres y , thres(q) dthres(x) bs1(30) bs2(30) bs3(20)
602
603  *-3- 绘图
604  xttr_graph // 第一轮搜索结果
605  graph export Figs\xtthres_fig01.wmf, replace
606  xttr_graph, m(22) white // 第二轮, 搜索第二个门槛, 黑白图片
607  graph export Figs\xtthres_fig02.wmf, replace
608  xttr_graph, m(21) // 第二轮, 重新搜索第一个门槛
609  graph export Figs\xtthres_fig03.wmf, replace
610  xttr_graph, m(3) // 第三轮, 第三个门槛
611  graph export Figs\xtthres_fig04.wmf, replace
612

```

```

613  *--4- 呈现估计结果
614      local q1 = e(rhat21)          // 取出门槛值
615      local q2 = e(rhat22)
616      gen d1 = (q<=`q1`)            // 生成虚拟变量
617      gen d2 = (q>`q2`)
618      gen xd1 = x*d1                // 交乘项
619      gen xd2 = x*d2
620      xtreg y x xd1 xd2, fe          // 常规标准误
621      est store fe
622      xtreg y x xd1 xd2, fe robust // 稳健型估计
623      est store fe_robust
624      local m "fe fe_robust"
625      esttab `m', mtitle(`m') nogap s(r2 r2_w N F) ///
626              star(* 0.1 ** 0.05 *** 0.01)
627
628      log close
629  *-----over-----
630
631      shellout xtthres_test01.log    // 日志文件
632
633
634
635
636
637  *-----
638  *-1.5 范例 B: Hansen 论文中的结果
639
640      shellout "$path\Refs\Hansen_1999.pdf"
641
642  *----- begin -----
643      cap log close
644      log using xtthres_Hansen99, text replace
645      clear all
646      set memory 200m                // 设定分配给 Stata 的内存空间
647      set matsize 8000
648      use hansen1999, clear // 调入 Hansen99 数据
649      set seed 1357911              // 第 648-651 行只是为了减少样本数，节省时间
650      gen u = uniform()
651      bysort id: egen mu = mean(u)
652      sort mu
653      order mu
654      keep if mu<0.6 // 随机抽取一些样本
655
656      *-Table 1: Summary statistics
657      tabstat i q1 c1 d1, s(min p25 p50 p75 max) ///
658                  format(%6.3f) c(s)
659
660      *-Estimating
661      xtthres i q1 q2 q3 d1 qd1, th(d1) d(c1) ///
662                  min(120) bs1(3) bs2(3) bs3(2)
663
664      *-Graphing
665      xttr_graph
666      graph export Figs\Hansen_Fig01.wmf, replace
667      xttr_graph, m(22)
668      graph export Figs\Hansen_Fig02.wmf, replace
669      xttr_graph, m(21)
670      graph export Figs\Hansen_Fig03.wmf, replace
671      log close
672  *----- over -----
673
674      shellout xtthres_Hansen99.log // 查看日志文件
675
676  *-几点说明:
677  * (1) 当样本数较大时, 需要慎重设定 xtthres 命令中的 minobs() 选项,
678  *     以保证每个区间内都有较多的样本可参与估计,
679  *     否则搜索出的门槛值将位于两端;
680  *     同时, 可以通过改变该选项中的数值, 来检验结果的稳健性;

```

```

681 * (2) 在初步分析过程中, 可设定较小的 bs1(3) bs2(3) bs3(3) 值,
682 * 以便节省时间, 等模型设定稳定后, 在设定较大的 bs 次数;
683 * (3) 在发表论文过程中, 一定要先设定 set seed #, 即种子值,
684 * 这样以来, 当杂志社需要提供实证结果时, 可以保证一致性;
685 * (4) 门槛变量也可以出现在解释变量中
686
687
688
689
690 *-----
691 *-1.6 范例 c: 资本结构与公司业绩
692
693 shellout "$path\Refs\连玉君_2006_DDJKX.pdf"
694
695 *-----
696 *-模型设定
697
698 * Tobin[it] = u[i] + b0*X[it]
699 *             + b1*d[it]*I(g[it]<=gamma)
700 *             + b2*d[it]*I(g[it]>gamma) + e[it]
701
702 *-----
703 *-估计
704
705 *----- begin -----
706 cap log close
707 log using xtthres_Lian2006, text replace
708 use xtthres_ddjkkx.dta, clear // 调入 Lian2006 数据
709 tsset id year
710
711 *-0- Summary Statistics
712 local x "tobin t1 size tang tshr prof grow"
713 tabstat `x', s(mean sd min p25 p50 p75 max) ///
714         format(%6.3f) c(s)
715
716 *-1- Estimating
717 set seed 1357911
718 xtthres tobin size tang tshr prof, thres(grow) dthres(t1) ///
719         min(50) bs1(3) bs2(3) bs3(3)
720 *-Note: 这里为了节省时间, 设定了 bs1(3) bs2(3) bs3(3)
721 *       实际分析过程中, 要设定 bs1(500) bs2(500) bs3(300)
722
723 *-2- Graphing
724 xttr_graph
725 graph export Figs\Lian06_Fig01.wmf, replace
726 xttr_graph, m(22)
727 graph export Figs\Lian06_Fig02.wmf, replace
728 xttr_graph, m(21)
729 graph export Figs\Lian06_Fig03.wmf, replace
730
731 *-3- 估计结果
732 dis e(rhat21)
733 dis e(rhat22)
734 global q1 = min(e(rhat21),e(rhat22)) // 取出门槛值
735 global q2 = max(e(rhat21),e(rhat22))
736 dis "$q1" // 较小的门槛值
737 dis "$q2" // 较大的门槛值
738 dropvars d1 d2 t1_x_*
739 gen d1 = (grow<=$q1) // 生成虚拟变量
740 gen d3 = (grow> $q2)
741 gen t1_x_grow1 = t1*d1 // 交乘项
742 gen t1_x_grow3 = t1*d3
743 local x "size tang tshr prof" // 解释变量
744 xtreg tobin `x' t1_x_grow* t1, fe // 常规标准误
745 est store fe
746 xtreg tobin `x' t1_x_grow* t1, fe robust // 稳健型标准误
747 est store fe_robust
748 local m "fe fe_robust"

```



```

749      esttab `m', mtitle(`m') nogap s(r2 r2_w N F) ///
750          star(* 0.1 ** 0.05 *** 0.01)          ///
751          order(`x' tl_x_grow1 tl tl_x_grow3 )
752      log close
753      *----- over -----
754
755      shellout xtthres_Lian2006.log // 查看日志文件
756
757      *-Note:
758      * (1) 修改 minobs() 选项的取值, 如 min(30), min(150)
759      *      看看结果有何变化? 有何启示?
760      * (2) 设定 bs1(500) bs2(500) bs3(300), 看看结果有何差异?
761
762

```

```

1
2  *      -----
3  *      -----
4  *
5  *      Stata 学术论文专题
6  *
7  *      -----
8  *      -----
9
10
11  *      主讲人：连玉君 副教授
12  *
13  *      单 位：中山大学岭南学院金融系
14  *      电 邮：arlionn@163.com
15  *      博 客：http://blog.cnfol.com/arlionn
16  *      主 页：http://goo.gl/tRXba
17  *      微 博：http://weibo.com/arlionn
18
19
20  *-----
21  * -KP09-
22  *      Kumbhakar, S., F. Christopher, 2009,
23  *      The effects of bargaining on market outcomes:
24  *      Evidence from buyer and seller specific estimates,
25  *      Journal of Productivity Analysis, 31(1):1-14.
26  *-----
27
28  *-主要方法
29  *
30  *   OLS
31  *   Two-tier Stochastic Frontier Model
32  *   help SFA2tier
33
34  global path "`c(sysdir_personal)'PX_papers" //定义课程目录
35  cd      "$path\Kumb_2009_SFA" // 进入课程目录
36  adopath + "$path\adofiles" // 自编程序
37
38  *-原始论文
39  shellout "$path\Refs\Kumbhakar_2009_SFA.pdf"
40
41
42
43  *-----
44  *-> 双边 SFA 模型
45  *-----
46
47  *-----
48  *-简介
49  *-----
50
51  *-基本思想
52  #delimit ;
53  twoway
54  function y=x^0.4,
55  rang(0 5) ylabel(,angle(0)) xlabel(0(1)5)
56  lwidth(thick) lcolor(blue)
57  text(1.75 4.0 "随机边界", size(*1.5) box)
58  text(1.5 1.3 "**", size(*2))
59  text(1.5 1.7 "Over", size(*2) color(red))
60  text(0.8 1.3 "**", size(*2))
61  text(0.8 1.8 "Under", size(*2) color(green))
62  ;
63  #delimit cr
64
65  *-模型设定
66  *-----
67  * y[i] = x[i]*b + e[i] - u[i] + w[i]
68  *-----

```

```

69      *-e[i] 常规意义上的干扰项, e~N(0, sigma_e^2)
70      *-u[i] 低于边界值的部分, u~exp(sigma_u)
71      *-w[i] 高于边界值的部分, w~exp(sigma_w)
72
73
74      *-----
75      *-估计方法: MLE
76      *-----
77
78      help SFA2tier
79
80      *-详细推导过程参见
81      * Kumbhakar and Christopher, 2009, JPA
82
83      shellout "$path\Refs\Kumbhakar_2009_SFA.pdf"
84      shellout "$path\Refs\连玉君_钟经樊_2008.pdf" //博士论文核心内容
85
86
87
88
89      *-----Begin-----
90
91      *-设定 Log 文件
92      cap log close
93      log using KP09_logfile.log, text replace
94
95      *-调入 KP2009 提供的原始数据
96      clear all // 清空内存中的所有元素
97      insheet using "KP09_2tier_SFA_data.txt", clear // 导入文本文件
98
99      *-Data and variable label
100     label data "Dataset in Kumbhakar&Christopher(2009, JPA)"
101     label var lwage "log wage"
102     label var educ "education (years)"
103     label var exper "work experience (years)"
104     label var tenure "tenure (years)"
105     label var iq "IQ(unmeasured ability)"
106     label var urban "dummy: 1=live in urban; 0=otherwise"
107     label var south "dummy: 1=live in south; 0=otherwise"
108     label var marri "dummy: 1=married; 0=otherwise"
109     label var black "dummy: 1=black; 0=othewise"
110     label var sibs "number of siblings (兄弟姐妹)"
111     label var brthord "the birth order of the worker"
112     label var meduc "mother's education (years)"
113     label var feduc "father's education (years)"
114
115     *-数字-文字对应表
116     label define black_lab 1 "Black workers" 0 "White workers"
117     label value black black_lab
118     label define marri_lab 1 "Married workers" 0 "Single worders"
119     label value marri marri_lab
120
121     *-变量的界定和含义
122     des
123
124     *-基本统计量
125     tabstat _all, s(N mean sd med min max) c(s) f(%4.2f)
126
127
128     *-克隆原始变量并生成对应的平方项
129     foreach v of varlist educ exper ten age{
130         gen `v'0 = `v' // 各变量的原始数值
131         gen `v'0_2 = `v'^2 // 各变量的原始数值的平方项
132     }
133
134     *-变换: 除以样本均值
135     foreach v of varlist educ exper age tenure{
136         qui sum `v'

```

```

137     replace `v' = `v'/r(mean)    // 除以样本均值
138     qui sum `v'0_2
139     gen `v'2 = (`v'0^2)/r(mean)  // 除以样本均值
140 }
141 drop age0_2 age2
142
143
144
145 *-----
146 *-Table 1: Estimates of log wage regression function (OLS)
147 *-----
148 *-OLS
149     reg lwage iq educ educ2 exper exper2 tenure tenure2 age married ///
150             south urban black sibs brthord meduc feduc
151     est store OLS
152     predict lwage_hat
153     predict e, residual
154
155
156 *-----
157 *-Table 2: Estimates of log wage regression (Two-tier frontier)
158 *-----
159 *-Twotier SFA
160     version 9 // 这个命令必须加上
161     SFA2tier lwage iq educ educ2 exper exper2 ///
162             tenure tenure2 age married south urban ///
163             black sibs brthord meduc feduc, ///
164             sigmau() sigmaw() check search
165     version 9 // 这个命令必须加上
166     ml max // MLE 估计目标函数
167     est store SFA2tier // 保存结果
168
169
170 *-结果对比
171     local out "using Results_KP09\Table_1_2.csv" // 执行时选中此行可输出到
Excel 中
172     esttab OLS SFA2tier `out', mtitle(OLS SFA2tier) replace compress ///
173             b(%6.3f) star(* 0.1 ** 0.05 *** 0.01) p nogap s(N r2)
174
175 *-----
176 *-方差分析: Bottom of Table 2, p.9 右侧第三段的分析
177 *-----
178     SFA2tier_sigs // 假设检验结果呈现于 SFA2tier 估计结果底部
179
180 *-计算方法:
181     dis exp(-1.660) // sigma_v 估计值
182     dis exp(-1.511) // sigma_u 估计值
183     dis exp(-1.666) // sigma_w 估计值
184
185     viewsource SFA2tier_sigs.ado // 查看程序源文件
186
187
188 *-----
189 *-Table 3: Surplus extracted by firms and workers
190 *-----
191
192 *-估计单边效应—效率分析
193     dropvars u_* w_* uw_* Firms* Worker* N*
194     SFA2tier_eff
195     rename u_hat Firms_E_u // Firm Surplus
196     rename w_hat Workers_E_w // Worker Surplus
197     rename uw_diff Net_E_u_w // Net Surplus
198     rename u_hat_exp Firms
199     rename w_hat_exp Workers
200     rename uw_diff_exp NI
201
202 *-转换成百分数, 便于统计
203     foreach v of varlist Firms_E_u-NI{

```

```

204     replace `v' = `v'*100
205 }
206
207 *-----
208 *-Table 3: Surplus extracted by firms and workers
209     logout, save(Results_KP09/Tabel_3) excel replace: /// //输出 Excel 表格
210     tabstat Firms_E_u-NI, s(mean p25 p50 p75) f(%4.1f) c(s)
211
212     log using KP09_logfile.log, text append
213
214
215 *-分组统计
216
217 *-----
218 *-Table 4: Surplus extracted by firms and workers across race
219 *-----
220
221 *-White v.s. Black workers (Panel 1-2)
222     logout, save(Results_KP09/Tabel_4a) excel replace: ///
223     tabstat Firms_E_u Workers_E_w Net_E_u_w, ///
224     by(black) stat(mean p25 p50 p75) f(%4.1f) c(s) nototal
225
226     log using KP09_logfile.log, text append
227
228 *-White v.s. Black workers (Panel 3-4)
229     logout, save(Results_KP09/Tabel_4b) excel replace: ///
230     tabstat Workers Firms NI, ///
231     by(black) stat(mean p25 p50 p75) f(%4.1f) c(s) nototal
232
233     log using KP09_logfile.log, text append
234
235 *-----
236 *-Table 5: Surplus extracted by firms and workers across marital status
237 *-----
238
239 *-Married v.s. Single workers (Panel 1-2)
240     logout, save(Results_KP09/Tabel_5a) excel replace: ///
241     tabstat Firms_E_u Workers_E_w Net_E_u_w, ///
242     by(married) stat(mean p25 p50 p75) f(%4.1f) c(s) nototal
243
244     log using KP09_logfile.log, text append
245
246 *-Married v.s. Single workers (Panel 3-4)
247     logout, save(Results_KP09/Tabel_5b) excel replace: ///
248     tabstat Workers Firms NI, ///
249     by(married) stat(mean p25 p50 p75) f(%4.1f) c(s) nototal
250
251     log using KP09_logfile.log, text append
252
253
254 *-----
255 *-Distributions of Worker and Firm surplus
256 *-----
257
258     histogram Workers // 默认为彩色图片
259     graph export "Results_KP09\Fig01_Worker.wmf", replace
260
261     histogram Firms, scheme(slmono) // 黑白图片
262     graph export "Results_KP09\Fig02_Firm.wmf", replace
263
264     histogram NI
265     graph export "Results_KP09\Fig01_Net.wmf", replace
266
267 *-----Over-----
268
269     log close // 关闭 Log 文件
270
271     shellout KP09_logfile.log // 查看 log 文件

```

272        cdout  
273



```

1
2  *      -----
3  *      -----
4  *
5  *      Stata 学术论文专题
6  *
7  *      -----
8  *      -----
9
10
11  *      主讲人：连玉君 副教授
12  *
13  *      单 位：中山大学岭南学院金融系
14  *      电 邮：arlionn@163.com
15  *      博 客：http://blog.cnfol.com/arlionn
16  *      主 页：http://goo.gl/tRXba
17  *      微 博：http://weibo.com/arlionn
18
19
20  *  Stochastic Frontier Analysis (SFA) 随机边界分析
21  *-----
22  *  连玉君,苏治, 2009.
23  *    融资约束、不确定性与上市公司投资效率.
24  *    管理评论,(1): 19-26.
25  *-----
26
27  *-设定课程存储路径
28  global path "`c(sysdir_personal)'PX_papers" //请先执行该命令
29  cd "$path\Lian_2009_SFA"
30  adopath + "$path\Lian_2009_SFA\sfa_adofiles" //本讲使用的自编程序存放于此
31
32  *-----
33  *-主要方法
34  *-----
35
36  *-Stochastic Frontier Analysis (SFA)
37  help frontier
38  help xtfreier
39
40  *-异质性 SFA
41  shellout Wang03_manual.pdf
42  *-完整程序下载地址：(王泓仁老师主页)
43  view browse "http://homepage.ntu.edu.tw/~wangh/"
44
45  *-MLE || help ml; help ml##ml_maximize_options
46  *  Gould, W., J. Pitblado, W. Sribney.
47  *    Maximum likelihood estimation with Stata.
48  *    Stata Press, 2006.
49
50
51  *-原始论文
52  shellout "$path\Refs\连玉君_2009_Het_SFA.pdf"
53
54  *-专题中的同类文献:
55  *  Fazzari et al. (1988): Fazzari_1988.do
56  *  Cleary (1999): Cleary_1999.do
57  *  Faulkender and Wang (2006): Faulkender_2006.do
58
59
60  *-----
61  *-相关文献
62  *-----
63  *
64  *  Wang, H., 2003.
65  *    A stochastic frontier analysis of financing constraints on investment.
66  *    Journal of Business and Economic Statistics, 21(3): 406-419.
67  shellout "$path\Refs\WangHR_2003_Het_SFA.pdf" //PDF原文
68

```

```

69 * Wang, H., C. W. Ho, 2010,
70 *   Estimating fixed-effect panel stochastic frontier models by model
transformation,
71 *   Journal of Econometrics, 157(2): 286-296.
72 *   shellout "$path\Refs\WangHR_2003_Het_SFA.pdf" //PDF原文
73
74 * Habib, M., Ljungqvist, A., 2005.
75 *   Firm value and managerial incentives: A stochastic frontier approach.
76 *   Journal of Business, 78(6): 2053-2094.
77 *   shellout "$path\Refs\Habib_2005_JB_Het_SFA.pdf" //PDF原文
78
79 * 苏治, 连玉君, 2011.
80 *   中国上市公司代理成本的估算—基于异质性随机前沿模型的经验分析.
81 *   管理世界, (6): 174-175+188.
82 *   shellout "$path\Refs\连玉君_2011_Het_SFA.pdf" //PDF原文
83
84 *-----
85
86
87
88
89
90 *=====
91 *- Part I: Stochastic Frontier Models (随机边界模型)
92 *=====
93
94
95 * -----
96 * ---- 本节目录 ----
97 * -----
98
99 * 1.1 传统的SFA模型 (Traditional Stochastic Frontier Model)
100 * 1.2 异质性SFA模型 (Heteroscedastic Stochastic Frontier Model)
101
102
103 *-----
104 *-> 1.1 传统的 SFA 模型
105 *-----
106
107     help frontier
108
109 *-SFA 介绍
110     shellout "$path\Refs\Frontier_Stata11_Manual.pdf"
111
112 *-----
113 *-1.1.1 简介
114
115 *-----图示 SFA-----
116     clear
117     set obs 50
118     set seed 13579113
119     gen x = 4*abs(invnorm(uniform()))
120     gen e = abs(invnorm(uniform()))
121     gen y = (x)^0.3 - e
122     drop if y<0
123     #delimit ;
124     twoway
125         (scatter y x)
126         (function y=x^0.3, range(0 10) lcolor(green) lwid(thick))
127     ,
128     text(1.58 3 "随机边界", size(*1.5) box)
129     legend(label(1 "实际产出") label(2 "最大产出"))
130     xlabel(none) ylabel(none)
131     xtitle("x", place(right))
132     ytitle("y", place(top) orientation(horizontal))
133     ;
134     #delimit cr
135 *-----

```

```

136
137
138 *-----
139 *-1.1.2 截面模型
140
141 *-----
142 *-1.1.2.1 模型设定
143
144 * s1-完美假设下的产出函数(假设生产效率为100%)
145 *
146 * -----
147 *       $Y^* = f(z,b)$  (1)
148 * -----
149 *
150 * e.g. C-D 生产函数
151
152 * s2-实际生产中，往往存在效率损失
153
154 * -----
155 *       $Y = f(z,b)*w$  (2)
156 * -----
157
158 *      w 表示【效率水平】，显然， $0 < w \leq 1$ ，如  $w=90\%$ ， $80\%$ 
159 *       $w=1$ ，表示现有生产技术达到了理论上的最优产出水平；
160 *       $w < 1$ ，表示存在效率损失
161
162 * s3-同时，假设生产可能受到随机因素的影响(如天气、运气等)
163 *      这也是“随机”(Stochastic)的含义
164
165 * -----
166 *       $Y = f(z,b)*w*\exp(v)$  (3)
167 * -----
168
169 *      v 为随机干扰的因素，
170 *      采用指数形式  $\exp(v)$  是为了保证  $Y > 0$ 。
171
172 * s4-对(3)式两边取对数，可得：
173
174 * -----
175 *       $\ln Y = \ln\{f(z,b)\} + \ln(w) + v$  (4)
176 * -----
177
178 * s5-转化为计量模型：
179
180 * -----
181 *       $\ln Y[i] = a + x[i]*b + v[i] - u[i]$  (5)
182 * -----
183 *      where,  $u = -\ln(w)$ 
184 *       $w \leq 1 \implies \ln(w) \leq 0 \implies u \geq 0$ 
185 *      因此，需要将 w 设定为单边分布(one-sided distribution)
186
187 * 实证分析中，通常设定如下：
188 *
189 * -----SFA model-----
190 *       $y[i] = a + x[i]*b + v[i] - u[i]$ 
191 *
192 *       $v[i] \sim N(0, \sigma_v^2)$  一般意义上的随机干扰项
193 *
194 *       $u[i] \sim \text{one-sided distribution}$  反映无效率的干扰项，单边分布
195 * -----
196
197
198 *-----
199 *-1.1.2.2 效率估计
200
201 *- 【效率】的含义
202
203 * -----

```

```

204      *      Y = f(z,b)*w                                (2)
205      *      -----
206
207      *      w 表示【效率水平】，显然，0<w<=1
208
209      *      -----
210      *      lnY[i] = a + x[i]*b + v[i] - u[i]            (5)
211      *      -----
212      *      where, u = -ln(w)
213
214      *- 因此，估计出 u 后，可得
215
216      *      -----
217      *      效率水平:  w = exp(-u)                        (6)
218      *      无效率水平: 1 - w
219      *      -----
220
221
222      *-----
223      *-1.1.2.3 常用的【单边分布】函数
224
225      *-----
226      *-半正态分布  u ~ N+(0, sigma_u^2)
227
228      clear
229      set obs 1000
230      gen u = invnorm(uniform()) // 模拟数据产生方法
231      histogram u, normal        // 标准正态分布
232      histogram u if u>0         // 半正态分布
233
234      *-----
235      *-截断型半正态分布  u ~ N+(Eu, sigma_u^2)
236
237      clear
238      set obs 1000
239      gentrun x, left(0.5)
240      sum x
241      histogram x
242
243      gentrun y, right(-0.5)
244      sum y
245      histogram y
246
247      *-----
248      *-指数分布  u ~ Exp(lamda)
249
250      *-模拟数据产生方法: -lamda*ln(uniform())
251      clear
252      rndexp 10000 3           // Exp(3) 分布
253      sum xe                   // E[xe] = Simga[xe] = lamda
254      twoway (histogram xe) (kdensity xe)
255
256      rndexp 10000 120         // Exp(120) 分布
257      twoway (histogram xe) (kdensity xe)
258
259      *-结论:
260      * (1) 若 x 服从自由度为 lamda 的指数分布，即 Exp(lamda)，则
261      *      E[x] = lamda, Var[x] = lamda^2
262      * (2) 指数分布在 0 处截断，向右托尾，是典型的非对称截断型分布，
263      *      如，在随机边界(SFA)模型中,常用于描述非效率成分的分布特征。
264
265      *-----
266      *-伽马分布  invgammamp(a, p)  a 是形状参数, p 表示累积概率
267
268      help density functions
269
270      set scheme s2color
271      clear

```

```

272      set obs 10000
273      gen gama_1 = invgammap(1, uniform())
274      gen gama_3 = invgammap(3, uniform())
275      twoway (kdensity gama_1) (kdensity gama_3)
276
277      gen gama_100 = invgammap(100, uniform())
278      gen gama_500 = invgammap(500, uniform())
279      twoway (kdensity gama_100) (kdensity gama_500)
280
281      tabstat gama*, s(mean sd p25 p50 p75 min max skewness kurtosis) ///
282                  f(%4.2f) c(s)
283
284      *-结论:
285      * (1) a 较小时, Gamma 分布呈现出明显的单边分布特征;
286      *      a 较大时, Gamma 分布逐渐趋近于正态分布;
287      * (2) Gamma 分布的期望值等于 a
288
289      *-----
290      *-总体评论:
291      *-----
292      * (1) 上述单边分布函数中, 指数分布最简单, Gamma 分布最灵活;
293      * (2) 文献中使用最为广泛的是半正态分布或截断型半正态分布;
294      * (3) 具体设定为何种分布形式, 决定于数据的特征和推导的难度。
295
296
297
298      *-----
299      *-1.1.2 SFA 估计方法
300      *-----
301
302      *-MLE (参见 Stata_Frontier_Manual.pdf, p.502-504)
303
304      help frontier
305      help ml // Stata 高级视频教程, B5_MLE.do
306      shellout "$path\Refs\Frontier_Stata11_Manual.pdf"
307
308      * y[i] = a + x[i]*b + v[i] - u[i]
309      * v[i] -- N(0, sigma_v^2) 一般意义上的随机干扰项
310      * u[i] -- N+(Eu, sigma_u^2) 反映无效率的干扰项, 具有半正态分布
311      *
312      * 给定 v[i] 和 u[i] 的分布函数, 可以写出联合密度函数;
313      * 进而采用 MLE 估计
314
315
316      *-----
317      *-半正态分布 u~N+(0,sigma_u^2)
318      *-----
319
320      use frontier1, clear
321
322      *-估计
323      frontier lnoutput lncapital lnlabor
324      est store SFA_hNormal
325
326      *-----
327      *-结果释义
328
329      local sigma_v = e(sigma_v) // v~N(0,sigma_v^2)
330      dis "/lnsig2v = " ln(`sigma_v'^2)
331
332      *-总波动
333      local sigma2 = (e(sigma_v))^2 + (e(sigma_u))^2
334      dis "sigma2 = " `sigma2'
335
336      *- u 的波动占总波动的比重
337      local ratio2 = (e(sigma_u))^2/`sigma2'
338      dis "sigma_u^2/(sigma_v^2+sigma_u^2) = " `ratio2'
339

```

```

340      *-lambda
341      local lambda = e(sigma_u)/e(sigma_v)
342      dis "lambda = " `lambda'
343
344      *-非效率成分是否显著?
345      * LR test (The bottom line)
346
347      *-OLS 估计
348      reg lnoutput lncapital lnlabor
349      est store OLS
350
351      *-结果对比
352      esttab SFA_hNormal OLS, mtitle(SFA_hNormal OLS) nogap
353
354
355      *-----
356      *-估算效率 TE = E{exp(-u)|e}
357
358      frontier lnoutput lncapital lnlabor
359      predict te if e(sample), te
360
361      *-基本统计量
362      sum te, detail
363
364      *-直方图
365      histogram te
366
367      *-95% CI
368      frontier_teci te
369      gen id = _n
370      twoway line te te_l te_u id in 1/30, ///
371      yline(1,lc(red) lp(dash))
372
373      *-----
374      *-效率的影响因素
375
376      reg te size
377      gen lnsize = ln(size)
378      gen lnsize2 = lnsize^2
379      reg te lnsize*
380
381
382      *-----
383      *-截断型半正态分布      u~N+(Eu,sigma_u^2)
384      *-----
385
386      frontier lnoutput lncapital lnlabor, distribution(tnormal)
387      est store SFA_tNormal
388
389      *-结果释义
390      *-Note that: u ~ N+(Eu, sigma_u^2)
391      * Eu = /mu = -0.7811509, insignificant
392      * sigma_u^2 = sigma_u2 = 8.33808
393
394      *-假设检验: The bottom line
395
396      *-效率估算
397      predict te_tNorm if e(sample), te
398
399
400      *-----
401      *-指数分布      u~Exp(sigma_u)
402      *-----
403
404      *-估计
405      frontier lnoutput lncapital lnlabor, d(exponential)
406      est store SFA_exp
407

```



```

408      *-效率估算
409      predict te_exp if e(sample), te
410
411      *-----
412      *-上述模型估计结果对比
413
414      local m "OLS SFA_hNormal SFA_tNormal SFA_exp"
415      esttab `m', mtitle(`m') nogap s(ll N)
416
417      lrtest SFA_tNormal SFA_exp
418
419      *-效率测算结果对比
420      tabstat te te_tNorm te_exp, ///
421      s(mean sd p50 min max) c(s) f(%4.2f)
422
423      *-Note: 在不同的分布假设下, 效率的估算结果存在较大差异
424
425
426
427
428      *-----
429      *-> 3.2 异质性SFA模型
430      *-----
431
432      *-----
433      *-1.2.1 简介
434
435      *-----
436      *-基本思想
437
438      #delimit ;
439      twoway
440      (function y=x^0.4,
441      rang(0 5) ylabel(,angle(0)) xlabel(0(1)5)
442      lwidth(thick) lcolor(blue)
443      text(1.75 4.0 "随机边界", size(*1.2) box)
444      text(1.1 2.8 "效率的影响因素", size(*1.2) color(pink))
445      text(0.8 2.0 "**", size(*2))
446      text(0.7 2.0 "**", size(*2))
447      text(0.73 2.9 "不确定性的影响因素", size(*1.2) color(green)))
448      (function y=x^0.4,range(2 2) recast(dropline) lc(red) lw(*2))
449      (function y=0.8,range(2 2) recast(dropline) lc(white) lw(*3))
450      , legend(off) ;
451      #delimit cr
452
453      *-可以分析影响效率水平的因素
454
455
456      *-----
457      *-模型设定
458
459      *   y[i] = x[i]*b + e[i] - u[i]
460
461      *-e[i] 常规意义上的干扰项,
462      *   e~N(0, sigma_e2[i])
463
464      *-----
465      *-u[i] 非效率部分
466      *   u~N(0, sigma_u2[i])           半正态分布
467      *   u~N(mu[i], sigma_u2[i])      截断型半正态分布
468      *   u~exp(sigma_u2[i])           指数分布
469
470      *-----
471      *-模型设定的基本想法
472      *
473      *   sigma_e2[i] = b0 + z1[i]*b1 + z2[i]*b2 + ... = z[i]*b
474      *
475      *   mu[i] = z[i]*b

```

```

476      *
477      *   sigma_u2[i] = z[i]*b
478      *
479      * 即，把影响效率或不确定性的因素写成线性组合
480
481      *-----
482      *-回顾一些普通线性回归模型的设定思路
483      *
484      *-   y[i] = a    // 估计样本均值
485      *
486      *-   y[i] = b0 + b1*x1 + b2*x2 + ...  // 条件期望值
487
488      *-----
489      *-参考文献
490      * Stata_Frontier_Manual.pdf p.503
491      * 连玉君和苏治(2009，管理评论)
492
493
494      *-----
495      *-1.2.2 估计方法
496
497      * u~N(0, sigma_u2[i]) 半正态分布
498
499      use frontier1, clear
500      frontier lnoutput lncapital lnlabor, uhet(size)
501      *-size 与不确定性负相关
502
503      * u~N(mu[i], sigma_u2[i]) 截断型半正态分布
504      frontier lnoutput lncapital lnlabor, d(tnormal) cm(size)
505      *-size 并未显著影响效率水平
506      *-该命令无法同时分析 cm() 和 uhet()
507
508      * u~exp(sigma_u2[i]) 指数分布
509      frontier lnoutput lncapital lnlabor, d(exp) uhet(size z)
510      frontier lnoutput lncapital lnlabor, d(exp) uhet(size) vhet(size)
511
512
513
514
515
516
517      *=====
518      *- Part II: 连玉君和苏治(2009) 一文的实证分析过程
519      *=====
520
521
522      *=====
523      * 初步样本筛选
524      *=====
525
526      use SFA_data.dta, clear
527      tsset id year
528      xtides
529
530      *-剔除金融类企业
531      panels id if isfinan==1
532      drop if isfinan==1
533
534      *-剔除 ST-PT 公司
535      panels id if (v370st!="")
536      drop if (v370st~="")
537
538      *-保留 2000-01-01 以前上市且仅发行A股的公司1998年以后的数据
539      drop if (id>200000 & id<300000) | id>900000
540      drop if pyear>2000
541      drop if year<2000
542
543

```

```

544 *=====
545 * 指标构建
546 *=====
547
548 *-主要财务指标
549     gen t1      = a200000/a100000
550     gen invt    = d220101/L.a130127
551     gen CFlow   = d100000/L.a130127
552     gen size    = ln(a100000)
553     gen stfn    = a310101 + a310301 // 股权融资 = 股本 + 资本公积金
554     gen dbfn    = a200000          // 债务融资 = 总负债
555     gen EQUI    = D.stfn/L.a100000
556     gen DEBT    = D.dbfn/L.a100000
557     gen sgr     = D.b110101/L.b110101
558     gen tagr    = D.a100000/L.a100000
559     gen cash    = (d630101+d630301)/a100000 // 现金及现金等价物期末余额/总资产
560
561 *-Tobin Q
562     gen aps     = a300000/totalshr // 每股净资产 = 股东权益合计/总股本
563     gen mv      = tshr*p_av_dec+(totalshr-tshr)*aps + a200000 // 股东权益市值+负债合计
564     gen tobin   = mv/a100000
565
566 *-对数转换
567     gen ln_invt = ln(invt)
568     gen ln_tobin= ln(tobin)
569
570 *-变量标签
571     label var invt      投资支出率
572     label var CFlow     现金流量
573     label var size      公司规模
574     label var sgr       主营业务收入成长率
575     label var tagr      总资产成长率
576     label var EQUI      股权融资增加额对数值
577     label var DEBT      债务融资增加额对数值
578     label var cash      现金持有比率
579
580 *=====
581 * 再次样本筛选
582 *=====
583
584 *-去除缺漏值
585     egen miss = rmiss(ln_invt ln_tobin EQUI DEBT size CFlow)
586     drop if miss != 0
587
588 *-删除负债率>100%,总资产成长率或销售额成长率大于200%的公司
589     drop if (t1>1 | sgr>2 | tagr>2)
590
591 *-转换为平行面板
592 * ssc install xtbalance, replace // 若未安装, 请执行该命令
593     xtbalance, rang(2001 2006)
594
595 *-对部分变量进行缩尾处理
596     winsor ln_tobin , gen(ln_tobin_w) p(0.01) lowonly
597     winsor ln_invt , gen(ln_invt_w) p(0.01)
598     winsor EQUI , gen(EQUI_w) p(0.01)
599     winsor DEBT , gen(DEBT_w) p(0.01)
600     winsor CFlow , gen(CFlow_w) p(0.01)
601     winsor cash , gen(cash_w) p(0.01)
602
603     dropvars ln_tobin ln_invt DEBT EQUI CFlow cash
604     local vars ln_tobin ln_invt DEBT EQUI CFlow cash
605     foreach v of local vars{
606         rename `v'_w `v'
607     }
608
609
610
611 *=====

```

```

612 * 表 1 基本统计量
613 *=====
614
615     logout, save(Table01) excel replace:      ///
616     tabstat invt tobin CFlow size EQUI DEBT, ///
617             stat(mean sd min max) col(s) format(%6.3f)
618
619
620 *=====
621 * 表2 异质性随机前沿模型估计及检验结果
622 *=====
623
624 *-----
625 *-模型设定方式
626 *-----
627
628     *-----
629     *   I =  x_it*beta   + v_it - u_it
630     *           -----
631     *           最优投资支出
632     *-----
633     *
634     *   v_it ~ N(0, vsigma^2)    || 常规意义上的随机干扰项
635     *
636     *   u_it ~ N+(mu, usigma^2)  || 无效率部分, 截断型半正态分布
637
638     dropvars dumt*
639     tab year, gen(dumt)
640     drop dumt1
641
642     global y ln_invt                // (5)式被解释变量
643     global xx ln_tobin dumt*        // (5)式中的 x_it
644     global zx CFlow size EQUI DEBT // (6)式左侧 mu = zx[it]*delta
645     global zv CFlow size EQUI DEBT // (6)式右侧 usigma^2 = zv[it]*gamma
646
647 *-----
648 *-模型估计
649 *-----
650
651 *-模型 1: 无约束
652     version 9
653     sfmodel $y, production distribution(truncated) frontier($xx) ///
654             mu($zx) usigmas($zv) vsigmas()
655     ml max, difficult // gtol(0.001)
656     est store m1
657
658     *-命令使用方法介绍:
659     shellout Wang03_manual.pdf
660
661 *-模型 2: gamma=0 Het u_it
662     version 9
663     sfmodel $y, production dist(truncated) frontier($xx) ///
664             mu($zx) usigmas() vsigmas()
665     ml search
666     ml max, diff
667     est store m2
668
669 *-模型 3: delta = 0 Het sigma_u
670     version 9
671     sfmodel $y, production dist(truncated) frontier($xx) ///
672             mu() usigmas($zv) vsigmas()
673     ml search
674     ml max, diff
675     est store m3
676
677 *-Stata 官方命令
678     frontier $y $xx, distribution(hnormal) uhet($zx)
679     est store m3_stata

```

```

680     *--结果对比
681     local m "m3 m3_stata"
682     esttab `m', mtitle(`m') nogap
683
684     *--模型 4: mu_it = 0    Het sigma_u
685     version 9
686     constraint define 1 [mu]_cons = 0
687     ml model d2 sf_trund (frontier: $y=$xx) ///
688         (mu:) (usigmas:$zv) (vsigmas:) , constraint(1)
689     ml search
690     ml max, diff
691     est store m4
692
693     *--模型 5: u_it = 0, without one-sided effect
694     version 9
695     constraint define 1 [mu]_cons = 0
696     constraint define 2 [usigmas]_cons = 0
697     ml model d2 sf_trund (frontier: $y=$xx) ///
698         (mu:) (usigmas:) (vsigmas:), constraint(1 2)
699     ml search
700     ml max, diff
701     est store m5
702
703     *--等价于 OLS
704     reg $y $xx
705     est store m5_OLS
706     *--结果对比
707     local m "m5 m5_OLS"
708     esttab `m', mtitle(`m') nogap
709
710     *-----
711     *--表 2 A: 估计结果呈现
712     *-----
713     local ss "using Table02.csv"
714     local mm "m1 m2 m3 m4 m5"
715     esttab `mm' `ss', mtitle(`mm') compress    ///
716         b(%6.3f) t(%6.2f) drop(dumt*)         ///
717         star(* 0.1 ** 0.05 *** 0.01) nogap     ///
718         scalar(ll N) sfmt(%6.1f) replace
719
720
721     *-----
722     *--表 2 B: LR 检验
723     *-----
724
725     *--LR test 释义
726
727     lrtest m1 m2
728
729     *-----
730     * LR1: LR test based on the null model (m1)
731     *-----
732     local model m1 m2 m3 m4
733     local k : word count `model'
734     mat LR1 = J(2, `k', 0)
735     local j = 1
736     foreach md of local model {
737         qui lrtest m5 `md'
738         mat LR1[1, `j++'] = (r(chi2) \ r(p))
739     }
740     mat colnames LR1 = `model'
741     mat rownames LR1 = chi2 p-value
742     mat list LR1, format(%6.3f)
743
744     *-----
745     * LR2: LR test based on the full model (m1)
746     *-----
747     local model m2 m3 m4 m5

```

```

748     local k : word count `model'
749     mat LR2 = J(2,`k',0)
750     local j = 1
751     foreach m of local model{
752         qui lrtest m1 `m'
753         mat LR2[1,`j++'] = (r(chi2) \ r(p))
754     }
755     mat colnames LR2 = `model'
756     mat rownames LR2 = chi2 p-value
757     mat list LR2, format(%6.3f)
758
759     *-整合上述结果 (表2 最下方的四行结果)
760     mat LR = J(4,5,.)
761     mat LR[1,1] = LR1
762     mat LR[3,2] = LR2
763     mat colnames LR = m1 m2 m3 m4 m5
764     mat rownames LR = LR1 p-value LR2 p-value
765     mat list LR, format(%6.3f)
766
767
768     *=====
769     * 表 3 分地区估计结果
770     *=====
771
772     drop if province == "西藏"
773     cap drop group
774     gen group = 2
775     replace group = 1 if ///
776         prov=="陕西" | prov=="甘肃" | prov=="宁夏" | prov=="青海" | ///
777         prov=="新疆" | prov=="重庆" | prov=="广西" | prov=="内蒙" | ///
778         prov=="四川" | prov=="贵州" | prov=="云南"
779     replace group = 3 if ///
780         prov=="北京" | prov=="天津" | prov=="河北" | prov=="辽宁" | ///
781         prov=="上海" | prov=="江苏" | prov=="浙江" | prov=="福建" | ///
782         prov=="山东" | prov=="广东" | prov=="海南"
783
784     label define group_lab 1 "西部" 2 "中部" 3 "东部"
785     label value group group_lab
786
787     *-----
788     *-1- 西部
789     *-----
790     version 9
791     sfmodel $y if group==1, production dist(truncated) frontier($xx) ///
792         mu($zx) usigmas($zv) vsigmas()
793     ml max, diff
794     est store m_west
795
796     *-----
797     *-2- 中部
798     *-----
799     version 9
800     sfmodel $y if group==2, production dist(truncated) frontier($xx) ///
801         mu($zx) usigmas($zv) vsigmas()
802     ml max, diff iterate(50)
803     est store m_mid
804
805     *-----
806     *-3- 东部
807     *-----
808     version 9
809     sfmodel $y if group==3, production dist(truncated) frontier($xx) ///
810         mu($zx) usigmas($zv) vsigmas()
811     ml max, diff
812     est store m_east
813
814     *-----
815     *-表 3: 呈现结果

```



```

816 *-----
817 local ss "using Table03.csv"
818 local mm "m_west m_mid m_east"
819 esttab `mm' `ss', mtitle(`mm') compress ///
820      b(%6.3f) t(%6.2f) drop(dumt*) ///
821      star(* 0.1 ** 0.05 *** 0.01) nogap ///
822      scalar(11 N) sfmt(%6.1f) replace
823
824
825 *=====
826 * 投资效率估算
827 *=====
828
829 *-Step 1: Estimating the Het-SFA model (表1: 模型1)
830
831 version 9
832 sfmodel $y, production dist(truncated) frontier($xx) ///
833      mu($zx) usigmas($zv) vsigmas()
834 ml max, diff
835
836 *-Step 2: 估算投资效率 (pp.21, Eq.(9))
837 dropvars te
838 sfa_te te
839
840 *-按省份-年度统计投资效率
841 format te %6.3f
842 tab province year if (year==2001|year==2006)&group==1, sum(te) mean
843 tab province year if (year==2001|year==2006)&group==2, sum(te) mean
844 tab province year if (year==2001|year==2006)&group==3, sum(te) mean
845
846 *-----
847 *-图1: 投资效率指数的频数分布
848 *-----
849 histogram te, percent scheme(slmono) ///
850      ytitle("百分比") xtitle("投资效率(IEI)") ///
851      ylabel(,angle(0) labsize(*1.2)) ///
852      xlabel(0.3(0.1)1.0, labsize(*1.2)) ///
853      xscale(titlegap(2) outergap(0)) ///
854      graphregion(margin(1-2 r-0 t-3 b-3))
855
856 *-保存图片
857 graph export Figure01.wmf, replace
858
859
860 *-----
861 *-图2: 投资效率的时序特征
862 *-----
863
864 *-----
865 *-左图: 根据公司规模分组
866 *-----
867 dropvars g_size
868 bysort year: quantiles size, gen(g_size) n(3)
869 mat e = J(6,6,0)
870 forvalues i = 1(1)6{
871     local yr = 2000 + `i'
872     mat e[`i',1] = `yr'
873     forvalues j = 1(1)3{
874         qui sum te if (year==`yr'&g_size==`j') //各个年度上的平均效率
875         mat e[`i',`= `j'+1'] = r(mean)
876     }
877     qui sum te if year == `yr'
878     mat e[`i', 5] = r(mean)
879 }
880
881 dropvars aa* te_* yr
882 svmat e ,names(aa)
883 rename aal yr

```

```

884     rename aa2 te_small
885     rename aa3 te_med
886     rename aa4 te_large
887     rename aa5 te_all
888
889     #delimit ;
890     twoway (connect te_small yr, lp(shortdash))
891            (connect te_med yr, lp(dash))
892            (connect te_large yr, lp(longdash))
893            (connect te_all yr, lp(solid))
894
895     ,
896     title("IEI",position(11)) ytitle("") xtitle("")
897     legend(label(1 "小规模") label(2 "中等规模")
898            label(3 "大规模") label(4 "总体"))
899     rows(1) size(*1.2) bmargin(medlarge))
900     ylabel(0.64(0.02)0.76,angle(0) labsize(*1.2) grid format(%3.2f))
901     xlabel(,labsize(*1.2)) graphregion(margin(1-2 r-0 t=3 b-2))
902     yscale(range(0.64 0.76)) scheme(slmono);
903     #delimit cr
904
905     graph export "Figure02a.wmf", replace
906
907     *-----
908     *-右图：根据地区分组
909     *-----
910     mat e = J(6,6,0)
911     forvalues i = 1(1)6{
912         local yr = 2000 + `i'
913         mat e[`i',1] = `yr'
914         forvalues j = 1(1)3{
915             qui sum te if (year==`yr'&group==`j') //各个年度上的平均效率
916             mat e[`i',`=j'+1'] = r(mean)
917         }
918         qui sum te if year == `yr'
919         mat e[`i', 5] = r(mean)
920     }
921
922     dropvars aa* te_* yr
923     svmat e ,names(aa)
924     rename aa1 yr
925     rename aa2 te_west
926     rename aa3 te_middle
927     rename aa4 te_east
928     rename aa5 te_all
929
930     #delimit ;
931     twoway (connect te_west yr, lp(shortdash))
932            (connect te_middle yr, lp(dash))
933            (connect te_east yr, lp(longdash))
934            (connect te_all yr, lp(solid))
935
936     ,
937     title("IEI",position(11)) ytitle("") xtitle("")
938     legend(label(1 "西部") label(2 "中部")
939            label(3 "东部") label(4 "总体"))
940     rows(1) size(*1.2) bmargin(medlarge))
941     ylabel(0.67(0.01)0.72,angle(0) labsize(*1.2) grid format(%3.2f))
942     xlabel(,labsize(*1.2)) graphregion(margin(1-2 r-0 t=3 b-2))
943     yscale(range(0.67 0.72)) scheme(slmono);
944     #delimit cr
945
946     graph export "Figure02b.wmf", replace
947
948     *-----
949     *-Appendix：原始数据的来源
950     *-----
951     /*
952     use "D:\stata9\ado\data\newdata2006\newdata2006_m", clear

```

```
952
953     keep id year isfinan v370st pyear ///
954     a200000 a100000 a130127 d220101 d100000 a310101 a310301 ///
955     b110101 d630101 d630301 a300000 totalshr ///
956     tshr p_av_dec totalshr province sicmen sicda
957
958     save SFA_data.dta, replace
959 */
960
961
```

```

1
2  *      -----
3  *      -----
4  *
5  *      Stata 学术论文专题
6  *
7  *      -----
8  *      -----
9
10
11  *      主讲人：连玉君 副教授
12  *
13  *      单 位：中山大学岭南学院金融系
14  *      电 邮：arlionn@163.com
15  *      博 客：http://blog.cnfol.com/arlionn
16  *      主 页：http://goo.gl/tRXba
17  *      微 博：http://weibo.com/arlionn
18
19  *-----
20  * -PSM-
21  *   Lian, Y., Z. Su, Y. D. Gu, 2011,
22  *   Evaluating the Effects of Equity Incentives Using PSM:
23  *   Evidence from China,
24  *   Frontiers of Business Research in China, 5(2): 266-290.
25  *-----
26
27  global path "`c(sysdir_personal)'PX_papers" //课程总文件夹
28  cd "$path\Lian_2011_PSM" //进入本讲目录
29
30  *-原始论文
31  shellout "$path\Refs\Lian_2011_PSM.pdf"
32  shellout "$path\PPT\连玉君_2011_PSM_股权激励有效吗.ppt"
33
34  *-主要方法
35  *   PSM: Propensity Score Matching, help psmatch2
36  *   Logit: help logit
37  *   ROC:   help roc
38
39  *-目的：采用 PSM 检验公司实施股权激励后，业绩是否显著提升？
40
41
42
43  *=====
44  *-> 1 倾向得分匹配分析综述    Propensity Score Matching (PSM)
45  *=====
46
47  *-----
48  *-1.1 因果推断综述和进展
49  *-----
50
51  *-Nichols, A., 2007, (非常详细地综述了因果推断相关的文献：PSM, IV-GMM, RDD)
52  *   Causal inference with observational data,
53  *   Stata Journal, 7 (4): 507-541.
54  *   shellout "$path\Refs\Nichols_2007_Causal_inference.pdf"
55
56  *-Ichino, A., 2007 (较为细致地介绍了因果推断问题, PPT)
57  *   The problem of causality in microeconomics.
58  *   University of Bologna and Cepr, PPT
59  *   shellout "$path\PPT\Ichino_2007_ppt_PSM_RDD.pdf"
60
61  *-Imbens, G., J. Wooldridge, 2009, (值得细读!!)
62  *   Recent developments in the econometrics of program evaluation,
63  *   Journal of Economic Literature, 47(1): 5-86.
64  *   shellout "$path\Refs\Imbens_Wooldridge_2009.pdf" //Google: 649次
65
66  *-Pearl, J., 2009, (涉及结构方程等新内容)
67  *   Causal inference in statistics: An overview,
68  *   Statistics Surveys, 3: 96-146.

```

```

69      shellout "$path\Refs\Pearl_2009_Causal_inference.pdf" //Google: 80次
70
71
72      *-----
73      *-1.2  PSM 相关文献
74      *-----
75
76      *-Grilli, L., and C. Rampichini, 2011 (详细介绍了PSM和-psmatch2-的应用)
77      *   Propensity scores for the estimation of average
78      *   treatment effects in observational studies
79      *   shellout "$path\PPT\Grilli_2011_ppt_PSM.pdf" //极力推荐,解释的很到位!
80
81      *-Becker, S., A. Ichino, 2002, (第一篇系统性的介绍Stata应用的文章)
82      *   Estimation of average treatment effects based on propensity scores,
83      *   Stata Journal, 2(4): 358-377.
84      *   相关命令: findit Becker
85      *   shellout "$path\Refs\Becker_2002_PSM.pdf" //Google引用: 1030 次
86
87      *-Dehejia, R. H., S. Wahba, 2002, (较早应用 PSM 分析的文章)
88      *   Propensity score-matching methods for nonexperimental causal studies,
89      *   Review of Economics and Statistics, 84(1): 151-161.
90      *   shellout "$path\Refs\Dehejia_2002_PSM.pdf" //Google引用: 1700 次
91
92      *-Abadie, A., D. Drukker, J. Herr, G. Imbens, 2004,
93      *   Implementing matching estimators for average treatment effects in Stata,
94      *   Stata Journal, 4(3): 290-311.
95      *   help nnmatch //文中使用的命令
96      *   doedit "nnmatch.do" //文中范例的 do 文档
97      *   shellout "$path\Refs\Abadie_2004_PSM.pdf" //Google引用: 508 次
98
99      *-Caliendo, M., S. Kopeinig, 2008, (很好的综述和应用细节介绍)
100     *   Some practical guidance for the implementation of propensity score matching,
101     *   Journal of Economic Surveys, 22(1): 31-72.
102     *   shellout "$path\Refs\Caliendo_2008_PSM.pdf" //Google引用: 843 次
103
104     *-Stuart, E. A., 2010, (对 PSM 的发展历程和应用细节进行了细致的综述)
105     *   Matching Methods for Causal Inference: A Review and a Look Forward,
106     *   Statistical Science, 25(1): 1-21.
107     *   shellout "$path\Refs\Stuart_2010_PSM.pdf" //Google引用: 67 次
108     *-Software for implementing matching methods and propensity scores
109     view browse "http://www.biostat.jhsph.edu/%7Eestuart/propensityscoresoftware.html"
110
111
112     *-----
113     *-1.3  Stata 中的 PSM 命令
114     *-----
115
116     *-----
117     *-1.3.1 PSM 相关命令
118
119     help psmatch2
120     help nnmatch // Abadie et al., 2004
121     help psmatch
122     help pscore // Becker and Ichino, 2002
123
124     *-持续获取最新的 PSM 信息和程序
125     findit propensity score
126     findit matching
127
128
129     *-----
130     *-1.3.2 应用 psmatch2.ado 的论文
131
132     *-psmatch2.ado 命令的引用方式
133     *
134     *-Leuven, E., and Sianesi, B. (2003). || Google: 900次
135     *   PSMATCH2: Stata Module to Perform Full Mahalanobis
136     *   and Propensity Score Matching, Common Support Graphing,

```

```

137 * and Covariate Imbalance Testing. Version 4.0.5,
138 * http://ideas.repec.org/c/boc/bocode/s432001.html.
139 *- 最新命令下载方法:
140 ssc install psmatch2, replace
141 *- 查看版本号的命令,引用时请填写相应版本
142 which psmatch2
143
144 *-Arnold, J. M., K. Hussinger, 2005,
145 * Export behavior and firm productivity in German manufacturing:
146 * A firm-level analysis,
147 * Review of World Economics, 141(2): 219-243.
148 shellout "$path\Refs\Arnold_2005_Export_PSM.pdf"
149
150 *-Heyman, F., F. Sjöholm, P. G. Tingvall, 2007,
151 * Is there really a foreign ownership wage premium?
152 * Evidence from matched employer-employee data,
153 * Journal of International Economics, 73(2): 355-376.
154 *-See pp.361 for panel data (逐年匹配 + D-in-D)
155 *-Figure 1 非常值得借鉴
156 shellout "$path\Refs\Heyman_2007_PSM.pdf"
157
158 *-Volpe Martincus, C., J. Carballo, 2008,
159 * Is export promotion effective in developing countries?
160 * Firm-level evidence on the intensive and the extensive margins of exports,
161 * Journal of International Economics, 76(1): 89-106.
162 shellout "$path\Refs\Martincus_2008_PSM.pdf"
163
164 *-Wagstaff, A., M. Lindelow, G. Jun, X. Ling, Q. Juncheng, 2009,
165 * Extending health insurance to the rural population:
166 * An impact evaluation of China's new cooperative medical scheme,
167 * Journal of Health Economics, 28(1): 1-19.
168 shellout "$path\Refs\Wagstaff_2009_China_PSM.pdf"
169
170 *-Bockerman, P., P. Ilmakunnas, 2009,
171 * Unemployment and self-assessed health: Evidence from panel data,
172 * Health Economics, 18 (2): 161-179.
173 *-完全应用了 psmatch2 命令,同时进行了 D-in-D 分析
174 shellout "$path\Refs\Bockerman_2009_PSM_DiD.pdf"
175
176 *-Lin, S., H. Ye, 2007,
177 * Does inflation targeting really make a difference?
178 * Evaluating the treatment effect of inflation targeting
179 * in seven industrial countries,
180 * Journal of Monetary Economics, 54(8): 2521-2533.
181 shellout "$path\Refs\Lin_Ye_2007_PSM.pdf"
182 *-Lin, S., H. Ye, 2009,
183 * Does inflation targeting make a difference in developing countries?,
184 * Journal of Development Economics, 89(1): 118-123.
185 shellout "$path\Refs\Lin_Ye_2009_PSM.pdf"
186 *-Lin 的两篇文章把最新版的 psmatch2 命令提到的匹配方法都用了一遍
187
188 *-Ivashina, V., V. B. Nair, A. Saunders, N. Massoud, R. Stover, 2009,
189 * Bank debt and corporate governance,
190 * Review of Financial Studies, 22(1): 41-77.
191 shellout "$path\Refs\Ivashina_2009_PSM.pdf"
192
193
194
195 *-----
196 *-1.4 一些最新的进展
197 *-----
198
199 *-----
200 * CEM Coarsened Exact Matching
201 *-----
202 * Iacus, S. M., G. King, G. Porro, 2012,
203 * Causal Inference without Balance Checking: Coarsened Exact Matching,
204 * Political Analysis, 20(1): 1-24.

```



---

```

205     shellout "$path\Refs\Iacus_2012_cem.pdf" //Google: 100次
206 *=Stata 实现
207 * Blackwell, M., S. M. Iacus, G. King, G. Porro, 2009,
208 * cem: Coarsened exact matching in Stata,
209 * Stata Journal, 9(4): 524-546.
210     shellout "$path\Refs\Blackwell_2009_cem.pdf" //Google: 34次
211 *=Application
212 * Lewis, B., J. Walls, G. Dowell, 2012,
213 * Difference in Degrees:
214 * CEO Characteristics and Firm Responses to Pressures for Disclosure,
215 * SSRN working paper.
216     shellout "$path\Refs\Lewis_2012_cem.pdf"
217
218 *-----
219 * QTE Quantile Treatment Effects Models (分位数处理效应模型)
220 *-----
221 *-Frolich, M., B. Melly, 2010, (四种方法: quantile treatment effects)
222 * Estimation of quantile treatment effects with Stata,
223 * Stata Journal, 10(3): 423-457.
224     shellout "$path\Refs\Frolich_2010_ivqte.pdf" //Google: 29次
225
226 *-----
227 * PSM+DID PSM + Difference in Difference
228 *-----
229     help diff
230
231
232
233
234
235 *=====
236 *-> 2 PSM 简介
237 *=====
238
239
240 *-----
241 *-2.1 为何使用 PSM ?
242 *-----
243
244     *-股权激励有效吗? //证监会的规定
245
246     *-读北大有助于提高收入吗? //多数考上北大的本身能力就强
247
248     *-出口有助于提高技术水平吗?
249
250     *-CEO 更换有助于提高公司经营绩效吗?
251
252
253 *-----
254 *-2.2 PSM 的基本思想
255 *-----
256
257 *-----
258 *-2.2.1 匹配方法
259
260     *-传统的配对方法(精确匹配: Exact Matching)
261
262         *-同行业 (一维配对)
263         *-同行业、规模相当 (二维配对)
264         *-同行业、规模相当、股权结构相当、..... (多维匹配) 咋办?
265
266     *-PSM: 把多个维度的信息浓缩成一个(降维: 多维-->一维)
267
268
269 *-----
270 *-2.2.2 一个简单的例子: 上北大有助于提高收入吗?
271
272     *-(1) 自我选择偏误: 考上北大的孩子本身就很出色(聪明,有毅力,能力强...)

```

---

```

273
274  *-(2)解决办法: 配对
275  *          双胞胎? 很难
276  *          高\矮\胖\瘦\IQ\父母Edu\一线城市\.... (PSM可以做到这一点)
277  *  高考分数 = M1 + M2 + M3 + M4 + M5
278  *  可以用高考分数作为“配对”的依据
279
280  *-(3)配对过程中的两个核心问题
281
282  *  Q1: 哪一种样本更好一些?
283  *
284  *  -----
285  *  上北大的同学(Treated)  v.s.  没上北大的同学(Control)
286  *  -----
287  *  Samp1:          ~650-670~          ~530-620~
288  *  Samp2:          ~650-670~          ~610-680~
289  *  -----
290  *
291  *-A: Samp2较好: 比较容易满足 [Common Support Assumption] 共同支撑假设
292
293
294  *  Q2: Stu C1,C2,C3 三人中, 谁是 Stu PK 的最佳配对对象?
295  *  -----
296  *          语文   数学   英语   物理   化学   总分   毕业当年年薪
297  *  -----
298  *  Stu PK: 120   138   130   142   140   = 670   12W
299  *  -----
300  *  Stu C1: 120   128   135   137   150   = 670   10W
301  *  Stu C2: 95    150   128   147   150   = 670   14W
302  *  Stu C3: 119   132   135   139   143   = 668   9W
303  *  -----
304  *
305  *-A: Stu C3 是最佳配对对象
306  *  比较容易满足 [Balancing Assumption] 平行假设
307
308  *-(4)ATT: Average Treatment effect on the Treated
309  *
310  *  ATT = E[Y(1)-Y(0) | T=1]
311  *
312  *  Y(1): Stu PK 上北大后的年薪      || 可以直接观察到
313  *  Y(0): Stu PK 假如不上北大的年薪  || 无法直接观察到,采用配对者的收入来替代
314  *
315  *-No man ever steps in the same river twice
316  *  --- Heraclitus (赫拉克利特,约公元前530年—前470年)
317  *
318  *-上例中: ATT = 12W - 9W = 3W
319  *
320  * 可能的偏误:
321  *  直接用北大毕业生的平均年薪减去其他学校学生的平均年薪
322  *  bias: 高估了北大的作用;
323
324
325  *-----
326  *-2.3   PSM 的分析过程
327  *-----
328
329  *-Source: Grilli, L., and C. Rampichini, 2011
330  *  shellout "$path\PPT\Grilli_2011_ppt_PSM.pdf"
331
332  *-Step1: Logit regression (specification)
333  *  Treat = Logit(a0 + a1*X1 + a2*X2 + ..... )
334  *  Stata command: logit Treat X1-Xk
335
336  *-Step2: Calculate Propensity Score (倾向得分值, PS)
337  *  Proability(Treat)
338  *  Note: Treat(0/1) --> (0<=Proability<=1)
339  *  Stata command: predict Treat if e(sample)
340

```

```

341  *--Step3: Restrict samples to ensure common support.
342  *   Stata option: psmatch2 ....., common / trim
343  *   [common] imposes a common support by dropping
344  *           treatment observations whose pscore is
345  *           higher than the maximum or less than
346  *           the minimum pscore of the controls.
347  *   [trim(integer)] imposes common support by dropping
348  *           # percent of the treatment observations
349  *           at which the pscore density of the control
350  *           observations is the lowest.
351
352  *--Step4: 根据 PS 值进行配对(双胞胎、多胞胎)
353
354  *--Step5: Check the balancing: test that the means of each covariate
355  *           do not differ between treated and control units.
356  *           if balancing is not achieved, repeat steps 1-4
357  *   Stata command: pstest
358
359  *--Step6: 比较 Treat组 与 Control组的差异, 进行统计推断
360
361  *-----
362  *--Comments
363  *-----
364  *--A. The algorithm is recursive:
365  *   if in step 5 balancing is not satisfactory
366  *   then repeat steps 1-4
367  *       (1) modify the matching algorithm (step 4)
368  *       (2) and/OR modify the propensity score model (step 1).
369  *--B. Note that the outcome plays no role in the algorithm for the
370  *   estimation of the propensity score
371
372
373  *-----
374  *--2.4   PSM 的主要假设条件
375  *-----
376
377  *--A1: 共同支撑假设(Common Support Assumption, Overlap)
378
379  twoway (function Control = normalden(x,.2,.1), range(0.0 0.5)) ///
380        (function Treated = normalden(x,.7,.1), range(0.4 1.0)), ///
381        xtitle("propensity score") xlabel(-0.1(0.1)1.0)
382  do files/common_1.do // common support 区域
383
384  twoway (function Control = normalden(x,.3,.1), range(0.0 0.6)) ///
385        (function Treated = normalden(x,.6,.1), range(0.3 0.9)), ///
386        xtitle("propensity score") xlabel(0(0.1)0.9)
387  do files/common_2.do // common support 区域
388
389  twoway (function Control = normalden(x,.40,.1), range(0.1 0.7)) ///
390        (function Treated = normalden(x,.45,.1), range(0.2 0.8)), ///
391        xtitle("propensity score") xlabel(0(0.1)0.9)
392  do files/common_3.do // common support 区域
393
394
395  *--A2: 平行假设(Balancing Assumption)
396
397  *--在接受 Treat 之前, 两组之间没有差异
398  *--Treat 组的变化完全来自 Treatment (如股权激励, 上大学, 接受补贴等)
399
400  *--Example
401  *   Data Source: webuse nlswork, clear
402  use nlswork, clear
403  describe
404  xi i.race
405  global x "coll age tenure not_s c_city south nev_mar _I*"
406  psmatch2 union $x, outcome(ln_wage)
407  pstest $x, both graph
408

```

```

409  *--Comments: PSM 与 OLS 的区别
410  *
411  *--(1) 如果二者都满足上述两个条件, 则结果不存在明显差异;
412  *      关键是很多情况下, OLS 分析并不考虑上述两个假设
413  *--(2) 模型设定问题
414  *
415  *--参见: pp.17-25; pp.72-73
416  shellout "$path\PPT\Grilli_2011_ppt_PSM.pdf"
417
418
419  *-----
420  *-2.5  获取 PS 值
421  *-----
422
423  *-基本思路: Logit/Probit 回归得到的概率值
424
425  *-Logit 模型简介: See: z_Logit.do
426  doedit "$path\z01_Logit.do"
427
428  *-参考资料:
429  shellout "$path\Refs\Lian_15_Logit.pdf" //我的讲义
430  shellout "$path\Refs\Long_2001_Logit.pdf" //极力推荐
431  *-Long, J., J. Freese.
432  *   Regression models for categorical dependent variables using Stata.
433  *   Stata Press, 2001.
434
435  *-----
436  *-Example 1: PS 值的计算方法
437
438  * w=1(treated); w=0(Control)
439  clear
440  input ///
441      i      w      x      y
442      1      0      2      7
443      2      0      4      8
444      3      0      5      6
445      4      0      3      5
446      5      1      2      8.9
447      6      1      3      6.5
448      7      1      1      8.2
449  end
450
451  logit w x
452  predict p_logit //概率值
453  *-手动计算
454  gen w_hat = 3.674253 -1.478705*x // exp(xb)
455  gen p_w = exp(w_hat)/(1+exp(w_hat)) //anti-Logit变换: Pr(xb)= -----
456  list w x p* w_hat, sep(0) // 1 + exp(xb)
457
458  *-----
459  *-Example 2: 加入工会的倾向
460
461  *-union=1(Treated); union=0(Control)
462  use nlswork, clear
463  describe
464  xi i.race
465  global x "coll age tenure not_s c_city south nev_mar _I*"
466  sum $x
467
468  logit union $x // Logit 回归
469  predict ps_logit if e(sample)
470
471  probit union $x // Probit 回归
472  predict ps_probit if e(sample)
473
474  tabstat ps*, s(mean sd min max q) f(%6.3f) c(s) //基本统计量对比
475
476  twoway (kdensity ps_logit) (kdensity ps_probit) //密度函数对比

```

```

477
478 *-----
479 *-Comments:
480 *
481 *(1) ps 值能实现降维的目的, 把多个变量的信息浓缩为一个概率值 — ps;
482 * 可以根据 ps 值确定样本是否满足[共同支撑]假设;
483 * ps[T]min <= ps <= ps[C]max
484 *(2) 在采用 ps 值进行配对之前, 要首先检验[平行假设]是否满足,
485 * 否则要重新设定模型, 如, 加入 age^2, tenure*age 等高阶项;
486
487
488 *-----
489 *-2.6 平均处理效应的衡量: ATE, ATC, ATT
490 *-----
491
492 *-ATT: the Average Treatment effect for the Treated
493 *
494 *  $ATT = E[Y(1)-Y(0)|Treat=1]$ 
495
496 *-ATC: the Average Treatment effect for the Controls
497
498 *-ATE: the Average Treatment Effect
499 *
500 * 
$$ATE = ATT \frac{N1}{(N1+N0)} + ATU \frac{N0}{(N1+N0)}$$

501
502 *-对于 ATC 的解释:
503 * 假设 Treat1=1 (上大学), Treated group
504 * Treat1=0 (不上大学) Control group
505 * 其实可以反过来定义另一个变量:
506 * Treat2=1(不上大学)
507 * Treat2=0(上大学)
508 * 此时得到的 ATT 就是第一种定义(Treat1)下的 ATC
509 * 当大学普及时, “不上大学”产生的影响就变成一个很有意思的研究话题了
510
511 *-In case of ATE, all observations are matched to their
512 * nearest m neighbors of the opposite treatment group.
513 *-In estimating the ATT or ATC,
514 * only the treated or controls, respectively, are matched.
515
516
517
518 *-----
519 *-Example 3: ATT, ATE, ATC 的计算
520
521 * w=1(treated); w=0(Control)
522 clear
523 input ///
524     i    w    x    y
525     1    0    2    7
526     2    0    4    8
527     3    0    5    6
528     4    0    3    5
529     5    1    2    8.9
530     6    1    3    6.5
531     7    1    1    8.2
532 end
533
534 psmatch2 w x, outcome(y) //默认条件下只计算 ATT
535 sort _id
536 format _pscore %4.3f
537 list w _treated x y _pscore _weight _y _id _n1 _nn _pdif, sep(4)
538
539 *-ATT (Average Treatment effect for the Treated)
540 sum y if _treated==1 & _support==1, mean
541 local m1t = r(mean)
542 local N1 = r(N) // Treat 组的样本数
543 sum _y if _treated==1 & _support==1, mean
544 local m0t = r(mean)

```

```

545     local att = `m1t' - `m0t'
546     dis "ATT = " `att'
547     *-手动计算
548     sum y if _treated==1  //[fw=_weight]
549     sum _y if _treated==1  //[fw=_weight]
550
551     *-ATE 和 ATC
552     psmatch2 w x, outcome(y) ate  //附加 ate 选项会同时计算 ATT, ATE 和 ATC
553     sort _id
554     format _pscore %4.3f
555     list w _treated x y _pscore _weight _y _id _nl _nn _pdif, sep(4)
556
557     *-ATC (ATU) (Average Treatment effect for the Controls\Untreated)
558     sum _y if _treated==0 & _support==1, mean
559     local mlu = r(mean)
560     sum y if _treated==0 & _support==1, mean
561     local m0u = r(mean)
562     local N0 = r(N)  // Control 组的样本数
563     local atu = `mlu' - `m0u'
564     dis "ATU = " `atu'
565
566     *-ATE (Average Treatment Effect)
567     local ate = `att'*`N1'/(`N0'+`N1') + `atu'*`N0'/(`N0'+`N1')
568     dis "ATE = " `ate'
569     *-Note: ATE 其实就是 ATT 和 ATC 的加权平均
570
571     *-test [反向定义 treat 得到的 ATT 就是原始 treat 顶一下的 ATC]
572     gen w2 = (w==0)  //反向定义 treat 变量
573     psmatch2 w2 x, outcome(y)
574     sort _id
575     list w* x y _pscore _treated _weight _y _id _nl _nn _pdif, sep(0)
576     *-Source: psmatch2.ado
577
578
579     *-----
580     *-2.7 匹配方法
581     *-----
582
583     shellout "PSM_pair.ppt"  // 图示
584
585     *-----
586     *-最近邻匹配 (k-Nearest neighbors matching)
587
588     * 一夫一妻 1:1 ; 一夫多妻 1:n
589
590     *-----
591     *-半径匹配 (Radius matching)
592
593     *-----
594     *-核匹配 (Kernel matching) 游戏中的美女
595
596     *-----
597     *-其他匹配方法
598
599     *-广义精确匹配(Coarsened Exact Matching) || help cem
600     *-局部线性回归匹配 (Local linear regression matching)
601     *-样条匹配 (Spline matching)
602     *-马氏匹配 (Mahalanobis matching)
603
604
605     *-----
606     *-2.8 敏感性分析 (Sensitivity analysis)
607     *-----
608
609     *-两个各科分数都相同的学生同时进入北大经济学院，但毕业的收入差距悬殊。
610     * why?
611
612     *-对应于 PSM 分析而言，配对变量(x)都是相同的，

```



```

613 * 但有些不可观测的因素在影响他们是否能读北大,或者毕业后的择业
614
615 *-DiPrete, T., and M. Gangl. 2004.
616 *   Assessing bias in the estimation of causal effects:
617 *   Rosenbaum bounds on matching estimators and
618 *   instrumental variables estimation with imperfect instruments.
619 *   Sociological Methodology 34: 271-310.
620   help rbounds
621
622 *-Becker, S., G. Munich, M. Caliendo, G. Berlin, 2007,
623 *   Sensitivity analysis for average treatment effects,
624 *   Stata Journal, 7(1): 71-83.
625   help mhbounds
626   shellout "$path\refs\Becker_2007_PSM.pdf"
627
628 *-Nannicini, T., 2007,
629 *   Simulation-based sensitivity analysis for matching estimators,
630 *   Stata Journal, 7(3): 334-350.
631   help sensatt // Stata Journal, 7(3):334-350
632   shellout "$path\refs\Nannicini_2007_PSM.pdf"
633
634 *-Imbens, G., J. Wooldridge, 2009, (值得细读!!)
635 *   Recent developments in the econometrics of program evaluation,
636 *   Journal of Economic Literature, 47(1): 5-86.
637   shellout "$path\Refs\Imbens_Wooldridge_2009.pdf" //pp.24-25
638
639
640
641
642 *=====
643 *-> 3   PSM 应用举例: 加入工会有助于获得更高的工资吗?
644 *=====
645
646 *-----
647 *-3.1 基本统计分析
648 *-----
649
650   use nlswork, clear
651   describe // 变量定义
652
653 *-----
654 *-Union 与 non-Union 特征对比
655
656   *-均值
657   tabstat ln_wage age msp collgrad ttl_exp tenure hours, ///
658           by(union) f(%4.2f)
659   *-中位数
660   tabstat ln_wage age msp collgrad ttl_exp tenure hours, ///
661           by(union) f(%4.2f) s(p50)
662
663   *-密度函数对比
664   twoway (kdensity ln_wage if union) ///
665          (kdensity ln_wage if ~union), ///
666          xtitle("ln_wage") ///
667          legend(label(1 "Union") label(2 "Non-Union"))
668
669 *-----
670 *-OLS regression
671
672   xi i.race i.ind i.occ // 创建虚拟变量
673   global x "coll age ten not_s c_city south nev_m _I*"
674   regress ln_w union
675   est store r0
676   regress ln_w union $x
677   est store r1
678   esttab r0 r1, nogap
679
680   *-Note: 回归分析可以从一定程度上控制两组人群的差异,但并不纯粹

```

```

681      *-Keypoint: Common Support & Balancing, 详见:
682      shellout "$path\PPT\Grilli_2011_ppt_PSM.pdf" //pp.18-23 中的例子
683
684
685  *-----
686  *-3.2 PSM 分析
687  *-----
688
689      *- ssc install psmatch2, replace // 请及时更新
690
691  *-----
692  *-3.2.1 匹配变量的筛选
693
694      *-原则:
695      * (1) 理论和前期经验研究;
696      * (2) 统计检验(help pstest, imbalance, prop_sel)
697      *      e.g. Heckman et al.,1998,
698      *           Heckman and Smith, 1999,
699      *           Black and Smith, 2004
700      * (3) The model for the PS does not need a behavioral interpretation.
701      *      我们的目的只是为了找到匹配对象
702
703      *-初步设定
704      logit union $x // 可以考虑在 x 中包含其他的变量, 但要以理论为基础
705      *-Note: 很多变量都不显著
706
707      *-逐步回归法 (建议由多到少)
708      stepwise, pr(0.1): logit union $x
709
710      *-取出显著的变量
711      mat b = e(b)
712      global x: colnames b // 取出符合标准的变量的名称
713      global x: dis subinword("$x","_cons","",.) // 去除 _cons, 常数项名称
714      dis "`$x'"
715
716      *-删除缺漏值
717      egen miss = rmiss($x)
718      drop if miss>0
719
720      *-另存一份数据 (视分析目的而定)
721      keep id year union ln_wage collgrad $x
722      compress
723      save nlswork_simple.dta, replace
724
725      *-其它筛选模型的方法:
726      * Goodness-of-fit pseudo-R2; || help logit
727      * ROC, AUC || help roc, help roctab
728
729
730  *-----
731  *-3.2.2 最近邻匹配 k-Nearest neighbors matching
732
733      use nlswork_simple.dta, clear
734      local v1 "_Iocc_code_8 _Iocc_code_7 tenure not_smsa c_city south collgrad"
735      local v2 "nev_mar _Irace_2 _Iocc_code_11 _Iocc_code_13 _Iocc_code_4"
736      local v3 "_Iind_code_4 _Iind_code_5 _Iind_code_6 _Iocc_code_6"
737      local v4 "_Iocc_code_5 _Iind_code_12 _Iocc_code_2 _Iind_code_11"
738      global x "`v1' `v2' `v3' `v4'"
739
740      cap safedrop u
741      generate u = uniform() // 截面数据, 必须随机排序
742      sort u
743      psmatch2 union $x, out(ln_wage) neighbor(2) ate // 1:2 匹配
744
745
746  *-----
747  *-psmatch2 产生的变量
748

```

```

749     des _pscore-_pdif
750     *
751     *     _pscore      Propensity Score
752     *     _treated    Treatment assignment
753     *     _support    Common support
754     *     _weight     weight of matched controls 被用作配对样本的次数
755     *     _ln_wage     value of ln_wage of match(es), 配对公司的 Outcome
756     *     _id         Identifier (ID) // 自行生成的代号 1,2,3
757     *     _n1         ID of nearest neighbor nr. 1 // &
758     *     _nn         # matched neighbors
759     *     _pdif       abs(pscore - pscore[nearest neighbor])
760
761     global cc "id _treated _id _n1 _n2 _nn _pscore _pdif _weight ln_wage
       _ln_wage"
762     order $cc
763     sort _id
764     br $cc
765     sort _ps
766     br $cc
767
768
769     *-----
770     *-3.2.3 检验平行假设：匹配前后匹配变量的差异对比
771
772     *-具体的解释：Grilli_2011_ppt_PSM, pp.65-67
773     shellout "$path\PPT\Grilli_2011_ppt_PSM.pdf"
774
775     *-原则：
776     *
777     * (1) 匹配后各变量在 Treat-Control 组间应该不存在显著差异
778     *     Note：这个有时不容易满足
779     * (2) 标准化后的偏差应该小于 5%
780
781     pstest $x, both
782
783     *-标准化后的偏差(standardized bias)计算公式
784     *-Rosenbaum, P. R., D. B. Rubin, 1985,
785     *   Constructing a Control Group Using Multivariate Matched Sampling Methods
786     *   That Incorporate the Propensity Score,
787     *   American Statistician, 39 (1): 33-38.
788     shellout "$path\refs\Rosenbaum_1985_PSM.pdf" //pp.36, Table.2 Note
789
790
791     *-----
792     *-3.2.4 匹配前后的密度函数图
793
794     *-----
795     *-简化图形
796     twoway (kdensity _ps if _treat==1, legend(label(1 "Treat")))) ///
797            (kdensity _ps if _treat==0, legend(label(2 "Control"))), ///
798            xtitle(Pscore) title("Before Matching")
799
800     twoway (kdensity _ps if (_treat==1), legend(label(1 "Treat")))) ///
801            (kdensity _ps if (_treat==0&_wei!=.), legend(label(2 "Control"))),
802     ///
803            xtitle("Pscore") title("After Matching")
804
805     *-----完整的图形设定-----
806     *
807     *-匹配前的密度函数图
808     twoway                                     ///
809            (kdensity _ps if _treat==1, lp(solid) lw(*1.5)) ///
810            (kdensity _ps if _treat==0, lp(dash) lw(*1.5)), ///
811            title("Before Matching") ///
812            ylabel(,angle(0)) ///
813            xtitle("Propensity Score") xscale(titlegap(2)) ///
814            xlabel(0(0.2)1, format(%2.1f)) ///
            legend(label(1 "Treat") label(2 "Control") row(2)) ///

```

```

815             position(3) ring(0))                                     ///
816             scheme(slmono)    // 彩色图片 scheme(s2color)           ///
817
818             graph export "Figs\PSM_match_pre.wmf", /// 将图片输出到指定文件夹中
819             replace fontface("Times New Roman")
820
821             *-匹配后的密度函数图
822             twoway                                                     ///
823             (kdensity _ps if _treat==1,lp(solid) lw(*1.5))           ///
824             (kdensity _ps if _treat==0&_wei!=.,lp(dash) lw(*1.5)), ///
825             title("After Matching") ///
826             ytitle("Density") ylabel(,angle(0))                     ///
827             xtitle("Propensity Score") xscale(titlegap(2))          ///
828             xlabel(0(0.2)1, format(%2.1f))                          ///
829             legend(label(1 "Treat") label(2 "Control") row(2))      ///
830             position(3) ring(0))                                     ///
831             scheme(slmono)
832
833             graph export "Figs\PSM_match_post.wmf", ///
834             replace fontface("Times New Roman")
835
836             *-----
837
838
839             *-----
840             *-3.2.5 分组检验 ATT
841
842             local if "if collgrad==1" // 大学毕业
843             psmatch2 union $x `if', out(ln_wage) neighbor(1) ate qui
844
845             local if "if collgrad==0" // 非大学毕业
846             psmatch2 union $x `if', out(ln_wage) neighbor(1) ate qui nowarn
847
848
849             *-----
850             *-3.2.6 采用 Bootstrap 获取标准误
851
852             *-样本较大时会非常耗时，因此本例仅 BS 3 次
853             bootstrap r(att), reps(3):                               ///
854             psmatch2 union $x if collgrad==0, ///
855             out(ln_wage) neighbor(2)
856
857
858             *-----
859             *-3.2.7 其他匹配方法
860
861             *-----
862             *-半径匹配 Radius matching
863
864             psmatch2 union $x if collgrad==1, out(ln_wage) ate ///
865             radius caliper(0.0001)
866
867             *-Notes: (1)请耐心等待，可能需要较长的时间
868             *          (2)radius 是必须填写的选项
869             *          (3)caliper(0.0001) 由于设定搜索半径，即 abs(_pdif)<0.0001
870             *          显然，半径越大，Treat 的匹配对象就越多，需要 trade-off
871
872
873             *-----
874             *-核匹配 Kernel matching
875
876             psmatch2 union $x if collgrad==1, out(ln_wage) ate ///
877             kernel
878
879             *-Note: 可以通过 kerneltype(), 简写为 k() 来设定 kernel 的类型
880             *          很少用
881
882

```

```

883  *-----
884  *-3.2.8 敏感性分析
885
886      psmatch2 union $x, out(ln_wage)
887      rbounds ln_wage, gamma(1 (0.2) 2)
888      mhbounds ln_wage, gamma(1 (0.05) 2)
889
890
891
892
893
894
895  *=====
896  *-> 4 Lian et al. (2011) Stata 实现过程
897  *=====
898
899  *-----
900  * Lian, Y., Z. Su, Y. D. Gu, 2011,
901  * Evaluating the Effects of Equity Incentives Using PSM:
902  * Evidence from China,
903  * Frontiers of Business Research in China, 5(2): 266-290.
904  *-----
905
906      global path "`c(sysdir_personal)'PX_papers" //课程总文件夹
907      cd "$path\Lian_2011_PSM" //进入本讲目录
908
909  *-原始论文
910      shellout "$path\Refs\Lian_2011_PSM.pdf"
911
912
913  *-----
914  *-4.1 股权激励数据处理
915  *-----
916
917      clear all
918      set memory 100m
919
920      insheet using "Data\incentive_id.txt", clear
921
922      label var id "公司代码"
923      label var name "公司名称"
924      label var process "方案进度"
925      label var type "激励标的物"
926      label var shares_million "激励股份总数(万股/万份)"
927      label var share_ratio "激励总数占当时总股本比例(%)"
928      label var strikeprice "期权初始行权价"
929      label var duration "有效期(年)"
930      label var date "信息发布日期"
931
932      drop if id==600030 // 中信证券属于金融业
933      drop if process=="停止实施" // 美的电器取消了其股权激励方案
934      drop if id==2045&share_ratio==0.125 // 广州国光(2045)出现了两次
935      drop if id==600143&date==20050701 // 金发科技(600143)出现了两次,删除2005年的
936      drop if id==2014&share_ratio==. // 水新股份(2014)出现了两次,删除信息不全的
937
938      encode type, gen(jili_type) label(type_lab) //转换为数字
939      drop type
940
941      gen year = int(date/10000) //从日期中提取年份
942      gen jili = 1 //实施激励的公司虚拟变量
943
944      order id year
945      tsset id year
946
947      save incentive_id.dta, replace // 另存为 .dta 格式
948
949
950  *-----

```

```

951  *-4.2 财务资料处理
952  *-----
953
954  *-----
955  *-4.2.1 合并股权激励数据
956
957      use "$path\data\GTA2009.dta", clear // GTA 数据库, 这里仅包含2005年以后的数据
958
959      tsset id year
960
961      merge 1:1 id year using incentive_id.dta, nogen // 增加激励公司的标识
962
963      replace jili = 0 if jili ==.
964      replace jili=1 if L.jili==1
965      replace jili_type=L.jili_type if L.jili_type!=.
966
967      order id year process-jili
968      tsset id year
969
970  *-----
971  *-4.2.2 进一步筛选样本
972
973      drop if TA==.
974      drop if sicmen_str=="I" // 剔除金融业公司
975      replace stpt=0 if jili
976      drop if stpt>0 // 剔除 ST, PT 或退市的公司
977      drop if tagr>1&jili==0 // 这类公司可能发生了兼并收购行为
978      drop if t1>1 // 负债率大于1, 可能资不抵债
979      keep if year>=2005 // 2005 年以后的资料
980      drop if pyear>2007 // 删除2007年以后上市的公司
981                        // 实施激励的公司都是07年以前上市的
982
983  *-----
984  *-4.2.3 变量定义和处理
985
986  *-补充一些指标
987      gen gpay = ln(d2) //前三名高管薪酬总额
988      local lab_d2: var label d2
989      label var gpay "ln(`lab_d2')"
990      gen ac = b001210000/b001101000 //代理成本=管理费用/营业收入
991  *-激励公司的缺漏值 (gpay)
992      list id year gpay if id==600019
993      replace gpay = 14.38 if id==600019&year==2006
994      replace gpay = 14.63 if id==600019&year==2007
995      list id year gpay if id==600183
996      replace gpay = 15.48 if id==600183&year==2006
997  *-行业虚拟变量
998      tab sicmen if jili
999      drop if sicmen_str=="A" | sicmen_str=="F"
1000      cap drop _Isicmen_*
1001      xi i.sicmen
1002
1003  *-Logit 回归变量(配对变量)统计分析
1004      dropvars lev roe
1005      rename roe2 roe
1006      rename t1 lev
1007      rename ros_a prof
1008      rename d24 magstk
1009      rename y1001b heyi
1010      rename h5index hhi5
1011      rename top1 topone
1012      rename dum_my private
1013      rename dum_gy state
1014
1015  *-删除缺漏值
1016      local v "invst ac size lev tobin tang prof gpay magstk heyi hhi topone zindex"
1017      egen miss = rmiss(`v')
1018      drop if miss

```



```

1019
1020     *-winsor 主要变量
1021     local vv "ac tagr roe roa tobin prof"
1022     foreach v of varlist `vv'{
1023         local a: var lab `v'
1024         winsor `v', p(0.01) gen(`v'_x)
1025         drop `v'
1026         rename `v'_x `v'
1027         label var `v' "`a'"
1028     }
1029
1030
1031     *-----
1032     *-4.3 基本统计分析
1033     *-----
1034
1035     *-----
1036     *-4.3.1 基本统计量
1037
1038     *-----
1039     *-Table 1: Summary Statistics of Variables
1040     *-----
1041     tabstat roe roa ac invt tagr if year>=2005, ///
1042             s(mean sd min max N) f(%6.3f) c(s)
1043
1044     *-----
1045     *-Table 2: Summary Statistics for the Matching Variables
1046     *-----
1047     tabstat size lev tobin tang prof gpay magstk heyi ///
1048             hhi topone zindex state if year>=2005, ///
1049             s(mean sd min max N) f(%6.3f) c(s)
1050
1051     order id year jili roe roa ac invt tagr size lev ///
1052             tobin tang prof gpay magstk heyi hhi topone ///
1053             zindex state private _I*
1054
1055
1056     *-----
1057     *-4.3.2 Logit 分析
1058
1059     *-目的: (1) 确定匹配变量(主要);
1060     *        (2) 分析影响公司实施股权激励的因素(次要)
1061
1062     global xv "_Isicmen* state"
1063
1064     logit jili size lev tang prof tobin gpay heyi hhi5 $xv
1065     est store m1
1066     logit jili size lev          prof tobin gpay magstk hhi5 $xv
1067     est store m2
1068     logit jili size lev          prof tobin gpay topone $xv
1069     est store m3
1070     logit jili size lev          prof tobin gpay zindex $xv
1071     est store m4
1072     logit jili size lev          prof tobin gpay hhi5 $xv
1073     est store m5
1074
1075     /*
1076     *-附加 AUC 指标的结果 [论文中使用的, 有些耗时]
1077     dropvars ppr*
1078     logit jili size lev tang prof tobin gpay heyi hhi5 $xv `if'
1079     cap drop pr
1080     qui predict pr if e(sample)
1081     roctab jili pr          // 计算 AUC 值, 比较费时
1082     estadd scalar auc = r(area)
1083     est store m1
1084     logit jili size lev          prof tobin gpay magstk hhi5 $xv
1085     cap drop pr
1086     qui predict pr if e(sample)

```



```

1087         roctab jili pr
1088         estadd scalar auc = r(area)
1089         est store m2
1090         logit jili size lev          prof tobin gpay topone $xv
1091         cap drop pr
1092         qui predict pr if e(sample)
1093         roctab jili pr
1094         estadd scalar auc = r(area)
1095         est store m3
1096         logit jili size lev          prof tobin gpay  zindex $xv
1097         cap drop pr
1098         qui predict pr if e(sample)
1099         roctab jili pr
1100         estadd scalar auc = r(area)
1101         est store m4
1102         logit jili size lev          prof tobin gpay  hhi5 $xv
1103         cap drop pr
1104         qui predict pr if e(sample)
1105         roctab jili pr
1106         estadd scalar auc = r(area)
1107         est store m5
1108     */
1109
1110     *-----
1111     *-Table 3 : The Estimation Results of Logit Models
1112     *-----
1113     local mm "m1 m2 m3 m4 m5"
1114     local order "size lev tobin tang prof gpay magstk heyi hhi5 topone zindex
state"
1115     esttab `mm', mtitle(`mm') s(r2_p auc N) compress nogap ///
1116         star(* 0.1 ** 0.05 *** 0.01) b(%6.3f)          ///
1117         drop(_Isic*) order(`order')
1118
1119     *-统计样本数和公司数目
1120     tab jili year if e(sample)
1121
1122
1123
1124     *-----
1125     *-4.4    PSM 分析
1126     *-----
1127
1128     *-----
1129     *-4.4.1  匹配效果分析
1130
1131     *-基本设定
1132     global x "size lev  prof tobin gpay hhi5"
1133     global xv "_Isicmen* state"
1134     global if56 "if year>=2005"
1135     global if567 "if year>2004"
1136
1137     *-激励类型变量
1138     dropvars jili_gp jili_qq
1139     gen jili_gp = (jili_type==2)
1140     label var jili_gp "1=股票激励,0=其它"
1141     gen jili_qq = (jili_type==1)
1142     label var jili_qq "1=期权激励,0=其它"
1143
1144
1145     *-比较各变量匹配前后的情况
1146
1147     *-最近邻匹配
1148     cap drop u
1149     gen u = uniform()
1150     sort u
1151     psmatch2 jili $x $xv $if56, neighbor(3) logit
1152     order id year jili _pscore- _pdif
1153

```

```

1154      *-匹配前后匹配变量的差异对比 [Balancing Assumption]
1155      pstest $x, both    //论文中未报告
1156
1157      *-Comments:
1158      *-(1)匹配后[%bias]基本上都小于 5%，表明balance效果较好；
1159      *-(2)匹配后[t-test]结果表明，Treat-Control组的差异不显著，
1160      *      表明 balancing 假设得到满足；
1161      *-(3)最后三行的[LR]检验表明，匹配后已经无法根据 x 的特征区分 jili 与否了，
1162      *      从整体上表明 balancing 假设得到满足。
1163
1164
1165      *-----
1166      *-Figure 1: Kernel Density of the Incentive and Control Groups
1167      *-----
1168
1169      *-(a)before matching: 匹配前的密度函数图
1170      twoway (kdensity _ps if _treat==1,lp(solid) lw(*2.5))          ///
1171             (kdensity _ps if _treat==0,lp(dash) lw(*2.5)),          ///
1172             ytitle("Density")                                       ///
1173             ylabel(,angle(0))                                       ///
1174             xtitle("Propensity Score")                             ///
1175             xscale(titlegap(2))                                     ///
1176             xlabel(0(0.2)0.8, format(%2.1f))                       ///
1177             legend(label(1 "Incentive") label(2 "Control") row(2)   ///
1178                    position(3) ring(0))                             ///
1179             scheme(slmono)
1180
1181      graph export "Figs\kn01.wmf", replace fontface("Times New Roman")
1182
1183      *-(b)after matching: 匹配后的密度函数图
1184      twoway (kdensity _ps if _treat==1,lp(solid) lw(*2.5))          ///
1185             (kdensity _ps if _treat==0&_wei!=.,lp(dash) lw(*2.5)),  ///
1186             ytitle("Density") ylabel(,angle(0))                   ///
1187             xtitle("Propensity Score") xscale(titlegap(2))        ///
1188             xlabel(0(0.2)0.8, format(%2.1f))                       ///
1189             legend(label(1 "Incentive") label(2 "Control") row(2)   ///
1190                    position(3) ring(0))                             ///
1191             scheme(slmono)
1192
1193      graph export "Figs\kn02.wmf", replace fontface("Times New Roman")
1194
1195
1196      *-比较ROC曲线 面积接近于0.8表明达到了不错的匹配，参见Yang(2007,p.7)
1197      * 论文中未报告
1198
1199      *-1-模型拟合效果: AUC 越大越好
1200      roctab jili _pscore if _support==1,          ///
1201             ylabel(,angle(0)) graph              ///
1202             ytitle("敏感度(Sensitivity)")         ///
1203             xtitle("误判率(1-Specificity)")        ///
1204             xscale(titlegap(2))                   ///
1205             scheme(slmono) note(" ")              ///
1206             caption("ROC 曲线下的面积: AUC = 0.844")
1207      graph export "Figs\roc_fit.wmf", replace ///
1208             fontface("Times New Roman")
1209
1210      *-2-共同支撑假设检验: AUC 越接近于0.5越好
1211      roctab jili _pscore if _weight!=.,          ///
1212             ylabel(,angle(0)) graph              ///
1213             ytitle("敏感度(Sensitivity)")         ///
1214             xtitle("误判率(1-Specificity)")        ///
1215             xscale(titlegap(2))                   ///
1216             scheme(slmono) note(" ")              ///
1217             caption("ROC 曲线下的面积: AUC = 0.536")
1218      graph export "Figs\roc_common.wmf", replace ///
1219             fontface("Times New Roman")
1220

```

```

1221
1222 *-----
1223 *-4.4.2 -ATT- 比较
1224
1225 *-基本设定
1226     global x      "size lev  prof tobin gpay hhi5"
1227     global xv     "_Isicmen* state"
1228     global out    "outcome(roe roa ac invt tagr)"
1229     global if08   "if year>=2005"
1230
1231 *-----
1232 * Table 4: Comparison of ATTs (最近邻匹配, Full sample)
1233 *-----
1234     cap drop u
1235     gen u = uniform()
1236     sort u
1237     psmatch2 jili $x $xv $if08, neighbor(3) logit $out qui
1238     *-Note: 这的结果也出现在了 Table 5, Panel A 中
1239
1240
1241 *-----
1242 *-Table 5: 不同最终控制权的激励效果
1243 *-----
1244
1245 *-----
1246 *- 最近邻匹配 (Nearest neighbor matching)      Table 5: Panel A
1247 *-----
1248     cap drop u
1249     gen u = uniform()
1250     sort u
1251     *-全样本 [ Table 5 A: Full sample ]
1252     psmatch2 jili $x $xv $if08, neighbor(3) logit $out qui
1253     *-国有控股 [ Table 5 B: Stated owned ]
1254     local if "&gfinal==1"
1255     psmatch2 jili $x $xv $if08 `if', neighbor(3) logit $out qui
1256     *-民营控股 [ Table 5 C: Privately owned ]
1257     local if "&gfinal==2"
1258     psmatch2 jili $x $xv $if08 `if', neighbor(3) logit $out qui
1259
1260 *-----
1261 *- 半径匹配 (Radius matching)      Table 5: Panel B
1262 *-----
1263     cap drop u
1264     gen u = uniform()
1265     sort u
1266     *-全样本
1267     psmatch2 jili $x $xv $if08, radius caliper(0.001) logit $out qui
1268     *-国有控股
1269     local if "&gfinal==1"
1270     psmatch2 jili $x $xv $if08 `if', radius caliper(0.001) logit $out qui
1271     *-民营控股
1272     local if "&gfinal==2"
1273     psmatch2 jili $x $xv $if08 `if', radius caliper(0.001) logit $out qui
1274
1275 *-----
1276 *- 核匹配 (Kernel matching)      Table 5: Panel C
1277 *-----
1278
1279     *-样本总体
1280     psmatch2 jili $x $xv $if567, kernel logit $out qui
1281     *-国有控股
1282     local if "&gfinal==1"
1283     psmatch2 jili $x $xv $if567 `if', kernel logit $out qui
1284     *-民营控股
1285     local if "&gfinal==2"
1286     psmatch2 jili $x $xv $if567 `if', kernel logit $out qui
1287
1288 *-----

```

```

1289 *-----精简汇总结果-----begin-----
1290 *
1291 * 目的：把半径匹配、核匹配和分层匹配的结果精简至一张表格中
1292     global vv "roe roa ac invt tagr"
1293 *-----
1294 * 最近邻匹配
1295 *-----
1296     local rn: word count $vv
1297     mat A1 = J(`rn', 6, .)
1298 *-----全样本
1299     local j = 1
1300     global opt "neighbor(3) logit qui "
1301     foreach v of varlist $vv{
1302         qui psmatch2 jili $x $xv $if08`if', $opt out(`v')
1303         mat A1[`j++',1] = (`r(att_`v')', `r(seatt_`v')')
1304     }
1305 *-----国有控股
1306     local j = 1
1307     foreach v of varlist $vv{
1308         local if "&gfinal==1"
1309         qui psmatch2 jili $x $xv $if08`if', $opt out(`v')
1310         mat A1[`j++',3] = (`r(att_`v')', `r(seatt_`v')')
1311     }
1312 *-----民营控股
1313     local j = 1
1314     foreach v of varlist $vv{
1315         local if "&gfinal==2"
1316         qui psmatch2 jili $x $xv $if08`if', $opt out(`v')
1317         mat A1[`j++',5] = (`r(att_`v')', `r(seatt_`v')')
1318     }
1319 *-----
1320 * 半径匹配
1321 *-----
1322     local rn: word count $vv
1323     mat B1 = J(`rn', 6, .)
1324 * 全样本
1325     local j = 1
1326     global opt "radius caliper(0.001) logit"
1327     foreach v of varlist $vv{
1328         qui psmatch2 jili $x $xv $if08, $opt out(`v')
1329         mat B1[`j++',1] = (`r(att_`v')', `r(seatt_`v')')
1330     }
1331 * 国有控股
1332     local j = 1
1333     foreach v of varlist $vv{
1334         local if "&gfinal==1"
1335         qui psmatch2 jili $x $xv $if08`if', $opt out(`v')
1336         mat B1[`j++',3] = (`r(att_`v')', `r(seatt_`v')')
1337     }
1338 * 民营控股
1339     local j = 1
1340     foreach v of varlist $vv{
1341         local if "&gfinal==2"
1342         qui psmatch2 jili $x $xv $if08`if', $opt out(`v')
1343         mat B1[`j++',5] = (`r(att_`v')', `r(seatt_`v')')
1344     }
1345 *-----
1346 * 核匹配
1347 *-----
1348     local rn: word count $vv
1349     mat C1 = J(`rn', 6, .)
1350 *-----全样本
1351     local j = 1
1352     global opt "kernel logit"
1353     foreach v of varlist $vv{
1354         qui psmatch2 jili $x $xv $if567, $opt out(`v')
1355         mat C1[`j++',1] = (`r(att_`v')', `r(seatt_`v')')
1356     }

```

```

1357     *--国有控股
1358         local j = 1
1359         foreach v of varlist $vv{
1360             local if "&gfinal==1"
1361             qui psmatch2 jili $x $xv $if567`if', $opt out(`v')
1362             mat C1[`j++',3] = (`r(att_`v')', `r(seatt_`v')')
1363         }
1364     *--民营控股
1365         local j = 1
1366         foreach v of varlist $vv{
1367             local if "&gfinal==2"
1368             qui psmatch2 jili $x $xv $if567`if', $opt out(`v')
1369             mat C1[`j++',5] = (`r(att_`v')', `r(seatt_`v')')
1370         }
1371
1372     * 结果汇总
1373
1374     mat d = J(1,6,.)
1375     mat D = (A1 \ d \ B1 \ d \ C1)
1376     *mat list D
1377     cap drop v*
1378     cap drop t1 t2 t3
1379     cap drop s1 s2 s3
1380     cap drop att_*
1381     svmat D, names(v)
1382     gen t1 = v1/v2
1383     gen t2 = v3/v4
1384     gen t3 = v5/v6
1385     forvalues i = 1/3{
1386         gen str5 s`i' = "****" if abs(t`i')>=2.576
1387         replace s`i' = "***" if abs(t`i')>=1.960&abs(t`i')<2.576
1388         replace s`i' = "**" if abs(t`i')>=1.645&abs(t`i')<1.96
1389         replace s`i' = "." if t2==.
1390     }
1391     format v1 t1 v3 t2 v5 t3 %6.3f
1392     rename v1 att_all
1393     rename v3 att_gy
1394     rename v5 att_my
1395
1396     *--显示结果(变量形式)
1397     list att_a t1 s1 att_g t2 s2 att_m t3 s3 in 1/17, ///
1398         sep(0) noobs clean
1399
1400     *--显示结果(矩阵形式)
1401     mat d = J(1,9,.)
1402     cap drop xxx
1403     gen xxx = .
1404     mkmat att_a t1 xxx att_g t2 xxx att_m t3 xxx in 1/17, mat(DD)
1405     mat DD = (d \ DD)
1406     mat colname DD = att_all t sig att_gy t sig att_my t sig
1407     mat rowname DD = 最近邻匹配 $vv 半径匹配 $vv 核匹配 $vv
1408     mat list DD, format(%6.3f)
1409
1410     *-----精简汇总结果-----over-----
1411     *-----
1412
1413
1414
1415     *-----
1416     *--Table 6: 不同激励方式的激励效果
1417     *-----
1418
1419     * 最近邻匹配
1420
1421     local rn: word count $vv
1422     mat A3 = J(`rn', 4, .)
1423     global opt "neighbor(3) logit qui "
1424     *--股票

```

```

1425     preserve
1426     drop if (jili==1&jili_type==1)
1427     local j = 1
1428     foreach v of varlist $vv{
1429         qui psmatch2 jili_gp $x $xv $if08, $opt out(`v')
1430         mat A3[`j++',1] = (`r(att_`v')', `r(seatt_`v')')
1431     }
1432     restore
1433 *--期权
1434     preserve
1435     drop if (jili==1&jili_type==2)
1436     local j = 1
1437     foreach v of varlist $vv{
1438         qui psmatch2 jili_qq $x $xv $if08, $opt out(`v')
1439         mat A3[`j++',3] = (`r(att_`v')', `r(seatt_`v')')
1440     }
1441     restore
1442
1443 *-----
1444 * 半径匹配
1445 *-----
1446     local rn: word count $vv
1447     mat B3 = J(`rn', 4, .)
1448     global opt "radius caliper(0.001) logit "
1449 *--股票
1450     preserve
1451     drop if (jili==1&jili_type==1)
1452     local j = 1
1453     foreach v of varlist $vv{
1454         qui psmatch2 jili_gp $x $xv $if08, $opt out(`v')
1455         mat B3[`j++',1] = (`r(att_`v')', `r(seatt_`v')')
1456     }
1457     restore
1458 *--期权
1459     preserve
1460     drop if (jili==1&jili_type==2)
1461     local j = 1
1462     foreach v of varlist $vv{
1463         qui psmatch2 jili_qq $x $xv $if08, $opt out(`v')
1464         mat B3[`j++',3] = (`r(att_`v')', `r(seatt_`v')')
1465     }
1466     restore
1467
1468 *-----
1469 * 核匹配
1470 *-----
1471     local rn: word count $vv
1472     mat C3 = J(`rn', 4, .)
1473     global opt "kernel logit "
1474 *--股票
1475     preserve
1476     drop if (jili==1&jili_type==1)
1477     local j = 1
1478     foreach v of varlist $vv{
1479         qui psmatch2 jili_gp $x $xv $if08, $opt out(`v')
1480         mat C3[`j++',1] = (`r(att_`v')', `r(seatt_`v')')
1481     }
1482     restore
1483 *--期权
1484     preserve
1485     drop if (jili==1&jili_type==2)
1486     local j = 1
1487     foreach v of varlist $vv{
1488         qui psmatch2 jili_qq $x $xv $if08, $opt out(`v')
1489         mat C3[`j++',3] = (`r(att_`v')', `r(seatt_`v')')
1490     }
1491     restore
1492

```



```

1493 *-----
1494 * 结果汇总
1495 *-----
1496 mat d = J(1,4,.)
1497 mat D = (d \ A3 \ d \ B3 \ d \ C3)
1498 *mat list D
1499 cap drop v*
1500 cap drop t1 t2
1501 cap drop s1 s2
1502 cap drop att_*
1503 svmat D, names(v)
1504 gen t1 = v1/v2
1505 gen t2 = v3/v4
1506 forvalues i = 1/2{
1507     gen str5 s`i' = "***" if abs(t`i')>=2.576
1508     replace s`i' = "***" if abs(t`i')>=1.960&abs(t`i')<2.576
1509     replace s`i' = "**" if abs(t`i')>=1.645&abs(t`i')<1.96
1510     replace s`i' = "." if t2==.
1511 }
1512 format v1 t1 v3 t2 %6.3f
1513 rename v1 att_gp
1514 rename v3 att_qq
1515
1516 * 显示结果(变量形式)
1517 list att_gp t1 s1 att_qq t2 s2 in 1/18, sep(0) noobs clean
1518
1519 * 显示结果(矩阵形式)
1520 cap drop xxx
1521 gen xxx = .
1522 mkmat att_gp t1 xxx att_qq t2 xxx in 1/18, mat(DD)
1523 mat colname DD = att_gp t sig att_qq t sig
1524 mat rowname DD = 最近邻匹配 $vv 半径匹配 $vv 核匹配 $vv
1525 mat list DD, format(%6.3f)
1526
1527
1528
1529 *-----
1530 *-Table 7: 股权集中度对激励效果的影响
1531 *-----
1532
1533 dropvars gtop med_topone
1534 bysort year: egen med_topone = mean(hhi5)
1535 gen gtop = hhi5>med_topone
1536 global rn: word count $vv
1537
1538 *-----
1539 * 最近邻匹配
1540 *-----
1541 mat A4 = J($rn, 4, .)
1542 global opt "neighbor(3) logit"
1543 *-股权集中
1544     local j = 1
1545     foreach v of varlist $vv{
1546         local if "&gtop==1"
1547         qui psmatch2 jili $x $xv $if08`if', $opt out(`v')
1548         mat A4[`j++',1] = (`r(att_`v)'), (`r(seatt_`v)')
1549     }
1550 *-股权分散
1551     local j = 1
1552     foreach v of varlist $vv{
1553         local if "&gtop==0"
1554         qui psmatch2 jili $x $xv $if08`if', $opt out(`v')
1555         mat A4[`j++',3] = (`r(att_`v)'), (`r(seatt_`v)')
1556     }
1557
1558 *-----
1559 * 半径匹配
1560 *-----

```

```

1561     mat B4 = J($rn, 4, .)
1562     global opt "radius caliper(0.001) logit"
1563 *-股权集中
1564     local j = 1
1565     foreach v of varlist $vv{
1566         local if "&gt;top==1"
1567         qui psmatch2 jili $x $xv $if08`if', $opt out(`v')
1568         mat B4[`j++',1] = (`r(att_`v')', `r(seatt_`v')')
1569     }
1570 *-股权分散
1571     local j = 1
1572     foreach v of varlist $vv{
1573         local if "&gt;top==0"
1574         qui psmatch2 jili $x $xv $if08`if', $opt out(`v')
1575         mat B4[`j++',3] = (`r(att_`v')', `r(seatt_`v')')
1576     }
1577
1578 *-----
1579 * 核匹配
1580 *-----
1581     mat C4 = J($rn, 4, .)
1582     global opt "kernel logit"
1583 *-股权集中
1584     local j = 1
1585     foreach v of varlist $vv{
1586         local if "&gt;top==1"
1587         qui psmatch2 jili $x $xv $if08`if', $opt out(`v')
1588         mat C4[`j++',1] = (`r(att_`v')', `r(seatt_`v')')
1589     }
1590 *-股权分散
1591     local j = 1
1592     foreach v of varlist $vv{
1593         local if "&gt;top==0"
1594         qui psmatch2 jili $x $xv $if08`if', $opt out(`v')
1595         mat C4[`j++',3] = (`r(att_`v')', `r(seatt_`v')')
1596     }
1597
1598 *-----
1599 * 结果汇总
1600 *-----
1601     mat d = J(1,4,..)
1602     mat D = (d \ A4 \ d \ B4 \ d \ C4)
1603     cap drop v*
1604     cap drop t1 t2
1605     cap drop s1 s2
1606     cap drop att_*
1607     svmat D, names(v)
1608     gen t1 = v1/v2
1609     gen t2 = v3/v4
1610     forvalues i = 1/2{
1611         gen str5 s`i' = "***" if abs(t`i')>=2.576
1612         replace s`i' = "***" if abs(t`i')>=1.960&abs(t`i')<2.576
1613         replace s`i' = "**" if abs(t`i')>=1.645&abs(t`i')<1.96
1614         replace s`i' = "." if t2==.
1615     }
1616     format v1 t1 v3 t2 %6.3f
1617     rename v1 att_jz
1618     rename v3 att_fs
1619
1620 * 显示结果(变量形式)
1621     list att_jz t1 s1 att_fs t2 s2 in 1/18, sep(0) noobs clean
1622
1623 * 显示结果(矩阵形式)
1624     cap drop xxx
1625     gen xxx = .
1626     mkmat att_jz t1 xxx att_fs t2 xxx in 1/18, mat(DD)
1627     mat colname DD = att_jz t sig att_fs t sig
1628     mat rowname DD = 最近邻匹配 $vv 半径匹配 $vv 核匹配 $vv

```

1629        `mat list` DD, format(%6.3f)

1630

1631

1632        \*-----over-----

1633

1634

```

1
2  *      -----
3  *      -----
4  *
5  *      Stata 学术论文专题
6  *
7  *      -----
8  *      -----
9
10
11  *      主讲人：连玉君 副教授
12  *
13  *      单 位：中山大学岭南学院金融系
14  *      电 邮：arlionn@163.com
15  *      博 客：http://blog.cnfol.com/arlionn
16  *      主 页：http://goo.gl/tRXba
17  *      微 博：http://weibo.com/arlionn
18
19
20  *-----
21  * 连玉君，钟经樊.
22  * 中国上市公司资本结构动态调整机制研究.
23  * 南方经济，2007,(1): 23-38.
24  *-----
25
26  *-主要目标:
27  * 了解公司财务领域中的一个重要的研究主题—资本结构；(他山之石,可以攻玉)
28  * 学习使用 Stata 处理数据和编写 NLS 程序；
29  * 交流一下与审稿人的沟通技巧。
30
31  *-主要方法
32  *-(1) Non-linear LS (NLS)    help nl
33  *-(2) HHI index
34  *-(3) Kruskal-Wallis test    help kwallis
35  *-(4) 数据处理：数据合并、离群值、循环语句等
36
37  *-PPT
38  shellout "`c(sysdir_personal)'PX_papers\PPT\连玉君_钟经樊_2007.ppt"
39
40  *-原始论文
41  shellout "`c(sysdir_personal)'PX_papers\Refs\连玉君_钟经樊_2007.pdf"
42
43
44
45  *-----
46  *-基本设定 (参见 Stata 初级视频: A1_Intro.do)
47
48  global path "`c(sysdir_personal)'PX_papers\Lian_NFJJ2007" //便于后续引用
49  display "`c(sysdir_personal)'" // 路径含义解释
50
51  cd "$path" // 进入课程目录
52  adopath + "$path\myprogs" // 自编程序
53
54  clear all
55  set memory 30m // 分配 30m 内存
56  set matsize 3000 // 矩阵运算的最大维度
57
58  capture log close // 关闭已经打开的日志文件
59  log using Lian_NFJJ2007_log, replace text // 定义日志文件
60
61  use "Lian_NFJJ2007_data.dta", clear // 调入样本数据
62
63
64  *=====
65  *- Part 1 : 样本初步筛选
66  *=====
67
68  *-1.1 排序 Variable 窗口中的变量

```

```

69     order code name year flowa_total fixa_total flowd asset ///
70         debt_total right_total main_income netprof          ///
71         stock dep sic* pyear both st* merge*
72
73     tsset code year      // 声明为 Panel Data
74
75
76     *-1.2 保留仅发行A股的公司、ST-PT 类公司、金融类公司
77     drop if bothab==1    // 删除同时发行 A,B 股的观察值
78     drop if st_b != .   // 删除 ST/PT 公司
79     labelbook sic_men    // 行业代码"数字-文字"对应表
80     drop if sic_men==9   // 删除金融类公司, 对应行业门类代码为 J
81
82
83     *-1.3 HHI Index  衡量行业竞争程度 (具体计算方法参见: 刘志彪等(2003,PP.64))
84
85     shellout "$path/刘志彪_2003_JJYJ.pdf"
86
87     *-----
88     *-HHI 计算方法
89
90     *- HHI = SUM(x_i/X)^2
91     *
92     *   x_i: 为企业 i 的销售额;
93     *   x: 为同一个行业内所有公司的总销售额;
94
95     *-1.3.1 Method 1
96     preserve
97     keep if year > 1997
98     forvalues t = 1998(1)2003{ //参见 A5_Prog, Section 5.5
99         egen sic_ssum_`t' = sum(main_income) if year==`t', by(sic_men)
100         //分母:每个行业-年度的销售额之和
101         replace sic_ssum_`t' = 0 if year != `t'
102         egen hhi_`t' = sum((main_income/sic_ssum_`t')^2) if year==`t', by(sic_men)
103         replace hhi_`t' = 0 if year != `t'
104     }
105     keep code year sic_men sic_ss* hhi*
106     *-汇总各个年度的计算结果
107     gen hhiyr = hhi_1998 +hhi_1999 +hhi_2000 +hhi_2001 +hhi_2002 +hhi_2003
108     keep sic_men year hhiyr
109     duplicates drop
110     egen hhi = mean(hhiyr), by(sic_men)
111     keep sic_men hhi
112     duplicates drop
113     rename hhi hhi_yr
114     sort sic_men
115     save hhi_yr.dta, replace // 另存一份数据, 随后合并到主数据中
116     restore
117
118     *-1.3.2 Method 2: 不区分年度
119     preserve
120     keep if year > 1997
121     egen sic_ssum = sum(main_income), by(sic_men)
122     egen hhi = sum((main_income/sic_ssum)^2), by(sic_men)
123         //同一行业中, 各个年度的 HHI 值相同
124     keep sic_men hhi
125     duplicates drop
126     rename hhi hhi_all
127     sort sic_men
128     save hhi_all, replace
129     restore
130
131
132     *-1.4 仅保留1998年1月1日以前上市的公司
133
134     keep if pyear<1998    // pyear 表示上市时间
135     drop if year<1998
136

```

```

137
138 *=====
139 *- Part 2 : 指标定义和进一步筛选
140 *=====
141
142 *-2.1 计算财务比率
143
144     tsset code year
145     gen t1      = debt_total/asset_total      // total leverage
146     gen t1_lag = L.t1
147     gen s1      = flowd_total/asset_total      // short-run leverage
148     gen l1      = longd_total/asset_total      // long-run leverage
149     gen fr      = flowa_total/flowd_total      // current ratio
150     gen size    = ln(asset_total)
151     gen ndts    = depreciate/asset_total      // 累计折旧/总资产
152     gen tang    = (fixa_total+stock)/asset_total // (固定资产+存货)/总资产
153     gen npr     = netprof/main_income         // 净利润率=净利润/主营业务收入
154     gen tshr    = trdshr/fullshr
155     gen sshr    = stateownshr/fullshr
156     gen tagr    = D.asset_total/L.asset_total
157
158     *-计算 Tobin's Q
159     gen napshr  = right_total/fullshr          // 每股净资产
160     gen shrmkv  = avyopr*trdshr + napshr*ntrdshr
161     gen mkv     = shrmkv + debt_total          // 分子：市场价值
162     gen mkv2    = avy_tmkv+ debt_total
163     gen tobin   = mkv/asset_total
164     gen tobin2  = mkv2/asset_total
165     gen mtl     = debt_total/mkv              // 市场负债率
166     gen mtl_lag= L.mtl
167
168     gen slr     = s1/t1
169     gen llr     = l1/t1
170
171 *-2.2 剔除奇异值和缺漏值
172
173     drop if tagr > 1 & tagr != .
174     drop if (t1>1 | t1<=0) | (napshr<0) | (abs(npr)>1)
175
176     local varlist t1 s1 l1 fr size ndts tang npr tshr tobin
177     foreach var of varlist `varlist' {
178         drop if `var' == .
179     }
180
181
182 *-2.3 处理为平行面板数据
183
184     xtbalance , range(1998 2003) /* Written by Arlion */
185
186
187 *-2.4 定义 age，序别化变量
188
189     _pctile byear, n(3)
190     return list          // 这是返回值
191     gen age = byear > r(r2)
192     replace age = 3 if byear <= r(r1)
193     replace age = 2 if age == 0
194
195 *-2.5 给新变量加标签
196
197     label var t1      "账面总负债率"
198     label var s1      "短期负债率"
199     label var t1      "长期负债率"
200     label var mtl     "市场总负债率"
201     label var fr      "流动比率"
202     label var size    "公司规模"
203     label var ndts    "非负债类税盾"
204     label var tang    "资产结构"

```



```

205     label var   npr      "净利润率"
206     label var   tshr     "流通股比例"
207     label var   sshr     "国家控股比例"
208     label var   age      "公司成熟度, 1 young 2 med 3 old"
209     label var   napshr   "每股净资产"
210     label var   shrmkv   "股票市场价值"
211     label var   mkv      "公司市场价值"
212     label var   tobin    "Tobin's Q 表示公司成长性"
213     label var   slr      "短期负债占总负债的比例"
214     label var   llr      "长期负债占总负债的比例"
215
216
217     *-----
218     *- 表 1: 样本描述性统计量
219     *-----
220
221     global xx "tl fr size ndts tang npr tshr age tobin"
222     des $xx
223     logout, save(Results/Tabel01_Var_stat) excel replace: ///
224     tabstat $xx, stat(mean sd min max) columns(s) format(%4.2f)
225
226     *-分行业统计主要指标 (没有呈现在正文中)
227     tabstat $xx, s(mean) by(sic_men) f(%4.2f)
228
229     *-分年度统计负债率 (没有呈现在正文中)
230     tabstat tl sl ll, s(mean) by(year) f(%4.2f)
231
232     *-分年度统计主要行业的负债率 (没有呈现在正文中)
233     tab sic_men
234     label list sic_men
235     format tl %4.2f // 设定显示格式, 这个必须提前执行
236     tab year sic_men , s(tl) mean // 整体统计
237     tabstat tl sl ll if sic_men==4, s(mean) by(year) f(%4.2f) //能源电力
238     tabstat tl sl ll if sic_men==6, s(mean) by(year) f(%4.2f) //交通运输
239     tabstat tl sl ll if sic_men==7, s(mean) by(year) f(%4.2f) //信息业
240     tabstat tl sl ll if sic_men==8, s(mean) by(year) f(%4.2f) //批发零售
241
242     *-2.6 产生年度虚拟变量
243
244     forvalues i = 1999(1)2003{
245         gen y`i' = 0
246         replace y`i' = 1 if year == `i'
247     }
248     *-更简洁的方法:
249     * tab year, gen(yr)
250     * drop yr1
251
252
253     *-2.7 产生行业虚拟变量, 按行业门类, 以综合类为基准
254
255     *-----
256     *- 表 2: 样本的行业分布及行业虚拟变量定义
257     *-----
258
259     tab sic_men if year==2003 // 表 2 第 1 列
260
261     *-2.7.1 方法1: 合并公司数目少于 10 家的行业门类
262
263     * 合并公司数目小于10家的行业到相应的行业中, A B L E
264     * C+B; J+E; M+A; D+L;
265     * 其余不变: G; F; H; D; 最终剩余8个行业门类
266
267     gen sicmen = sic_men
268     replace sic_men = 3 if (sic_men==3 | sic_men==2) /*C+B*/
269     replace sic_men = 10 if (sic_men==5 | sic_men==10) /*E+J*/
270     replace sic_men = 11 if (sic_men==11 | sic_men==12) /*K+L*/
271     replace sic_men = 13 if (sic_men==13 | sic_men==1) /*A+M*/
272

```

```

273   tab sic_men, gen(sic) // 定义行业虚拟变量
274   drop sic8
275
276   *-数字-文字 对应表
277   label define sic_men_lab ///
278       3 "制采" 4 "能源" 6 "交运" 7 "信息" ///
279       8 "批零" 10 "建房" 11 "社服" 13 "农综"
280   label value sic_men sic_men_lab
281
282
283
284   *-----
285   * 连玉君, 钟经樊.
286   * 中国上市公司资本结构动态调整机制研究. 04
287   * 南方经济, 2007,(1): 23-38.
288   *-----
289
290   *=====
291   *- Part 3 : 估计静态模型并获得初始值 (pp. 29)
292   *=====
293
294   *-3.1 估计静态模型作为对比 [p.29, Eq.(8)]
295
296   qui xtreg tl fr-tshr age tobin y1999-y2003 sic1-sic7, re
297   est store m_static
298
299
300   *-3.2 产生初始值
301
302   *-Step1:regress static model using OLS
303   reg tl fr-tshr age tobin y2000-y2003 sic1-sic7 // Eq.(5) 的初始值
304   est store m0_static
305
306   *-Step2: optimal TL, and Speed=delta
307   predict tl_hat // optimal leverage level
308   qui tsset
309   gen diff_star = tl_hat-tl_lag // the Right hand of Eq.(1)
310   gen diff_lag = tl-tl_lag // the Left hand of Eq.(1)
311   generate distance=abs(tl_hat-tl_lag) // the distance between TL and TL*
312   gen alpha=diff_lag/diff_star // p.29 第二步 delta=D.TL/D.TL*
313
314   *-Step3: Eq.(6) 的初始值
315   reg alpha tobin distance size y2000-y2003 sic1-sic7
316   est store alpha
317
318   *-结果简单对比
319   est table m0_static alpha, stats(r2 r2_a N) b(%6.4f) star(0.01 0.05 0.1)
320
321   * 这里得到的初始值应作为参数写入如下两个程序中, 需要手动操作
322   * dycs_xiao04.ado, 估计准动态模型; p.29, Eq.(9)
323   * dycs_NFJJ_final.ado, 估计异质性动态调整模型 p.29, Eq.(7)
324
325
326
327   *=====
328   *- Part 4 : 估计准动态模型 (肖作平,2004) (pp.29, Eq.(9))
329   *=====
330
331   cap drop tl_star_xiao
332   gen tl_star_xiao = 0
333   cap program drop _all
334
335   drop if tl_lag==.
336   run "myprogs\dycs_xiao04.ado" // 自编程序读入内存
337   doedit "myprogs\dycs_xiao04.ado" // 它存放于 myprogs 文件夹下, 参见 A1
338
339   nl dycs_xiao04 tl // 非线性估计程序的运行方式
340   est store m_xiao04

```

```

341     global rss_xiao = e(rss) // 残差平方和, 以便执行 F 检验: p.30, Eq.(10)
342
343
344
345     *=====
346     *- Part 5 : 估计动态模型 (本文)      (pp.29, Eq.(7))
347     *=====
348
349     *-5.1 产生两个代理变量用于非线性估计
350     cap drop alpha
351     gen tl_star = 0      // 记录最优负债率的估计值
352     gen alpha = 0       // 记录调整速度的估计值
353
354     *-5.2 估计非线性模型
355     drop if tl_lag==.
356     cap program drop _all
357
358     run      "myprogs\dycs_NFJJ_final.ado" // 将异质性动态调整模型读入内存
359     doedit   "myprogs\dycs_NFJJ_final.ado" // 查看程序代码
360     nl dycs_NFJJ_final tl      // 执行 NLS 估计
361     est store m_dynamic
362
363     global rss_dycs = e(rss) // 残差平方和
364     predict res, res        // 残差序列, 以备表3使用
365
366
367
368     *=====
369     *- Part 6 : 输出结果      (pp.29, Eq.(7))
370     *=====
371
372     *-6.1 假设检验
373
374     *-----
375     *-表 3: Panel A    残差的基本统计量
376     *-----
377     tabstat res if year==2003, stat(mean sd min med max sk k) format(%6.4f)
378
379     *-----
380     *-表 3: Panel B    残差正态分布假设检验
381     *-----
382     sktest res if year==2003
383
384     *-----
385     *-表 3: Panel C    残差一阶序列相关检验
386     *-----
387     qui tsset code year
388     xtserial res // Woodridge test for first-order serial correlation of Panel
data
389
390     *-----
391     *-pp.30, Eq.10 检验结果
392     *-检验调整系数是否为常数 参见Greene(2000,pp.439, Eq(10-29))
393     local n = _N
394     local n_K = `n' - 35
395     local F = (($rss_xiao - $rss_dycs)/14) / ($rss_dycs/`n_K') // Eq.(10)
396     dis "F(14,`n_K') = " in y `F'
397
398
399     *-6.2 呈现回归结果
400
401     *-----
402     *- 表 4: 静态模型、准动态模型和动态模型结果对比
403     *-----
404
405     local mm "m_static m_xiao04 m_dynamic"
406     esttab `mm', mtitle(`mm') ///
407     star(* 0.1 ** 0.05 *** 0.01) b(%6.4f) se(%6.4f) ///

```

```

408         scalar(N r2_o r2_a) wide nogaps order(_cons) replace
409
410     *-输出到 Excel
411     local mm "m_static m_xiao04 m_dynamic"
412     esttab `mm' using Results/Table04_reg.csv, mtitle(`mm') ///
413         star(* 0.1 ** 0.05 *** 0.01) b(%6.4f) se(%6.4f) ///
414         scalar(N r2_a) wide nogaps order(_cons) replace
415
416     *-6.3 统计调整速度和最优比率
417
418     count if tl_star<0
419     count if tl_star>1
420     replace tl_star=0.01 if tl_star<0
421     replace tl_star=0.99 if tl_star>1
422     count if alpha<0
423     count if alpha>1
424
425     gen opratio = tl_star/tl    // 最优比率
426
427     *-将公司按规模分成 5 类，产生变量 size_id 用以标示
428     centile size, centile(20(20)100)
429     gen size_id = 0
430     replace size_id=1 if size<=r(c_1)
431     replace size_id=2 if size>r(c_1)&size<=r(c_2)
432     replace size_id=3 if size>r(c_2)&size<=r(c_3)
433     replace size_id=4 if size>r(c_3)&size<=r(c_4)
434     replace size_id=5 if size>r(c_4)&size<=r(c_5)
435
436     *-----
437     *-表 5：调整速度和最优比率的分类统计描述
438     *-----
439
440     *-Panel A: by Year
441     tabstat alpha opratio, stat(mean med) format(%6.3f) by(year)
442
443     *-Panel B: by Industry
444     tabstat alpha opratio, stat(mean med) format(%6.3f) by(sic_men)
445
446     *-Panel C: by Firm Size
447     tabstat alpha opratio, stat(mean med) format(%6.3f) by(size_id)
448
449
450     *-6.4 从债务期限结构和竞争强度角度解释调整速度的行业差异
451
452     *-估算样本内行业竞争强度
453
454     preserve
455     egen sic_ssum = sum(main_income), by(sic_men)
456     egen hhi = sum((main_income/sic_ssum)^2), by(sic_men)
457     keep sic_men hhi
458     duplicates drop
459     rename hhi hhi_samp
460     sort sic_men
461     save hhi_samp, replace
462     restore
463
464     merge m:1 sic_men using hhi_yr.dta, nogen    // 分年度计算的 HHI
465     merge m:1 sic_men using hhi_all.dta, nogen  // 样本筛选前计算出的 HHI
466     merge m:1 sic_men using hhi_samp.dta, nogen // 样本内公司计算出的 HHI
467     drop if tl == .
468
469     *-----
470     *-表 6：不同行业的财务特征和竞争强度
471     *-----
472
473     tabstat alpha opratio tl tang npr hhi_all hhi_samp hhi_yr ///
474         if tl!=. , stat(mean) by(sic_men) format(%6.3f)
475

```

```

476
477 *-----
478 *-表 7: 调整速度的行业、规模和年度差异检验
479 *-----
480
481 *-Kruskal-Wallis test by sic_men and size [see ]
482
483 help kwallis
484
485 *local rvars alpha opratio // 被检验变量
486 local rvars "alpha" // 正文中只检验了 alpha
487 local gvars "sic_men size_id" // 分组变量
488 foreach Rvar of local rvars{
489     foreach Gvar of local gvars{
490         mat kw_men = J(4,5,0)
491         forvalues i=1999(1)2003{
492             qui kwallis `Rvar' if year == `i', by(`Gvar')
493             local df = r(df)
494             local chi2 = r(chi2)
495             local pvalue = chi2tail(`df',`chi2')
496             mat kw_men[1,`i'-1998] = `i'
497             mat kw_men[2,`i'-1998] = r(chi2)
498             mat kw_men[3,`i'-1998] = r(df)
499             mat kw_men[4,`i'-1998] = `pvalue'
500         }
501         dis _n in g "The K-W test for " in y "`Rvar'" in g " by " in w "`Gvar':"
502         mat rownames kw_men = year kw-value k p-value
503         mat list kw_men
504     }
505 }
506
507 *-K-W test by year
508
509 kwallis alpha , by(year)
510
511 *===== Over =====
512
513 capture log close // 关闭日志文件
514 exit // 退出 Stata
515
516
517
518
519
520
521
522 *=====
523 *===== Appendix A: 自编程序代码 =====
524 *=====
525
526
527 *-A1-----nldycs_xiao04.ado-----
528 *
529 *! Author: Lian Yu-jun
530 *! 准动态模型 (9) 估计程序
531
532 program define nldycs_xiao04
533 version 8.0
534 if "`1'"=="?" {
535
536     #delimit ;
537     global S_1 "a_fr a_sz a_nd a_tg a_np a_tr a_ag a_tb
538               a_y0 a_y1 a_y2 a_y3
539               a_s1 a_s2 a_s3 a_s4 a_s5 a_s6 a_s7
540               a_0
541               alpha";
542     #delimit cr
543

```

```

544     global S_2 "Estimaion for semi dynamic capital structure model"
545
546     global a_fr = -.03951422
547     global a_sz = .01915718
548     global a_nd = -.25094612
549     global a_tg = -.04223138
550     global a_np = -.31237543
551     global a_tr = -.0329844
552     global a_ag = .02264043
553     global a_tb = -.03447056
554
555     global a_y0 = .0191199
556     global a_y1 = .02061439
557     global a_y2 = .01355629
558     global a_y3 = .02546505
559     global a_s1 = -.03569658
560     global a_s2 = -.02559723
561     global a_s3 = -.02511563
562     global a_s4 = .00103758
563     global a_s5 = -.04008194
564     global a_s6 = .03730375
565     global a_s7 = -.07317321
566     global a_0 = .21209572
567
568     global alpha = 0.4
569     exit
570 }
571
572 #delimit ;
573     replace tl_star_xiao = $a_fr*fr + $a_sz*size + $a_nd*ndts + $a_tg*tang + $a_np*npr
574                        + $a_tr*tshr + $a_ag*age + $a_tb*tobin
575                        + $a_y0*y2000 + $a_y1*y2001 + $a_y2*y2002 + $a_y3*y2003
576                        + $a_s1*sic1 + $a_s2*sic2 + $a_s3*sic3
577                        + $a_s4*sic4 + $a_s5*sic5 + $a_s6*sic6 + $a_s6*sic7
578                        + $a_0;
579 #delimit cr
580
581     if tl_star_xiao<0 {
582         replace tl_star_xiao = 0
583     }
584     else if tl_star_xiao>1 {
585         replace tl_star_xiao = 1
586     }
587
588     replace `1'=(1-$alpha)*tl_lag + $alpha*tl_star_xiao
589
590 end
591
592 *-----
593
594
595 *-A2-----nldycs_NFJJ_final.ado-----
596 *
597 *! Author: Lian Yu-jun
598 *! 动态模型 (7) 估计程序
599 *! First Draft: 2004.08.20
600 *! This Version: 2006.09.12
601
602 program define nldycs_NFJJ_final
603     version 8.0
604     if "`1'"=="?" {
605         #delimit ;
606         global S_1 "a_fr a_sz a_nd a_tg a_np a_tr a_ag a_tb
607                    a_y0 a_y1 a_y2 a_y3
608                    a_s1 a_s2 a_s3 a_s4 a_s5 a_s6 a_s7
609                    a_0
610                    b_sz b_ds b_tb
611                    b_y0 b_y1 b_y2 b_y3"

```

```

612          b_s1 b_s2 b_s3 b_s4 b_s5  b_s6 b_s7
613          b_0";
614      #delimit cr
615
616      global S_2 "Estimaion for dynamic capital structure"
617
618      global a_fr = -.03951422
619      global a_sz = .01915718
620      global a_nd = -.25094612
621      global a_tg = -.04223138
622      global a_np = -.31237543
623      global a_tr = -.0329844
624      global a_ag = .02264043
625      global a_tb = -.03447056
626
627      global a_y0 = .0191199
628      global a_y1 = .02061439
629      global a_y2 = .01355629
630      global a_y3 = .02546505
631      global a_s1 = -.03569658
632      global a_s2 = -.02559723
633      global a_s3 = -.02511563
634      global a_s4 = .00103758
635      global a_s5 = -.04008194
636      global a_s6 = .03730375
637      global a_s7 = -.07317321
638      global a_0  = .21209572
639
640      global b_tb = -3.9267479
641      global b_ds = -30.374831
642      global b_sz = -.89620473
643
644      global b_y0 = 1.6372949
645      global b_y1 = .76052199
646      global b_y2 = 7.3130424
647      global b_y3 = 1.9304668
648      global b_s1 = -1.9559097
649      global b_s2 = -2.3695235
650      global b_s3 = -.86199012
651      global b_s4 = 26.957073
652      global b_s5 = -1.3170207
653      global b_s6 = -2.5345035
654      global b_s7 = -2.1545516
655      global b_0 = 28.487197
656      exit
657  }
658
659  #delimit ;
660      replace tl_star = $a_fr*fr +$a_sz*size +$a_nd*ndts +$a_tg*tang +$a_np*npr
661                      +$a_tr*tshr +$a_ag*age +$a_tb*tobin
662                      +$a_y0*y2000 +$a_y1*y2001 +$a_y2*y2002
663                      +$a_y3*y2003
664                      +$a_s1*sic1 +$a_s2*sic2 +$a_s3*sic3
665                      +$a_s4*sic4 +$a_s5*sic5 +$a_s6*sic6 +$a_s6*sic7
666                      +$a_0;
667      replace alpha = $b_tb*tobin +$b_ds*abs(tl_star-tl_lag) +$b_sz*size
668                      +$b_y0*y2000 +$b_y1*y2001 +$b_y2*y2002 +$b_y3*y2003
669                      +$b_s1*sic1 +$b_s2*sic2 +$b_s3*sic3
670                      +$b_s4*sic4 +$b_s5*sic5 +$b_s6*sic6 +$b_s6*sic7
671                      +$b_0;
672  #delimit cr
673
674      if tl_star<0 {
675          replace tl_star=0
676      }
677      else if tl_star>1 {
678          replace tl_star=1
679      }

```



---

```

679     replace `1' = alpha*tl_star + (1-alpha)*tl_lag  // pp.29, Eq.(7)
680
681 end
682
683 *-----
684
685
686
687
688
689
690
691
692 *-----
693 * 连玉君, 钟经樊.
694 *    中国上市公司资本结构动态调整机制研究.    05
695 *    南方经济, 2007,(1): 23-38.
696 *-----
697
698
699 *=====
700 *===== Appendix B: 外审意见及答复 =====
701 *=====
702 *
703 *-外审意见:
704     shellout "$path\审稿意见.doc"
705
706 *-答复及修改说明:
707     shellout "$path\答辩书1.doc"
708     shellout "$path\答辩书2.doc"
709
710
711
712
713

```

```

1
2  *      -----
3  *      -----
4  *
5  *      Stata 学术论文专题
6  *
7  *      -----
8  *      -----
9
10
11  *      主讲人：连玉君 副教授
12  *
13  *      单 位：中山大学岭南学院金融系
14  *      电 邮：arlionn@163.com
15  *      博 客：http://blog.cnfol.com/arlionn
16  *      主 页：http://goo.gl/tRXba
17  *      微 博：http://weibo.com/arlionn
18
19
20  *-----
21  * Love, I., L. Zicchino, 2006,
22  *   Financial development and dynamic investment behavior:
23  *   Evidence from panel VAR,
24  *   Quarterly Review of Economics and Finance, 46(2):190-210.
25  *-----
26
27  global path "`c(sysdir_personal)'PX_papers" //课程总文件夹
28  adopath + "$path\adofiles" //自编程序(这条命令一定要执行)
29
30  *-原始论文
31  shellout "$path\Refs\Love_2006_pvar.pdf"
32
33  *-主要方法
34  *
35  *- PVAR: Panel VAR      || help pvar; help pvar2; help xtvar
36  *- PVAR 滞后阶数筛选    || help pvar2
37  *- 面板 Granger 检验    || help pvar2
38  *- PVAR IRF 和 FEVD     || help pvar2
39  *
40  *- Note: 有关 VAR 的介绍, 参见stata高级视频教程 B6_TimeS.do
41
42
43  *-----
44  *-与本专题中的关联论文:
45  *
46  *   Fazzari et al. (1988,BPEA) || Fazzari_1988.do
47  *   Cleary (1999,JF)          || Cleary_1999.do
48  *   Opler et al. (1999,JFE)   || Opler_1999.do
49
50
51  *-----
52  *-其他几篇重要的文献:
53  *
54
55  *-Hayashi, F., 1982, (q model: marginal q and average q)
56  *   Tobin's Marginal q and Average q: A Neoclassical Interpretation,
57  *   Econometrica, 50(1): 213-224.
58  *   shellout "$path\Refs\Hayashi_1982.pdf"
59
60  *-Love, I., 2003, (Introduction to Euler equation)
61  *   Financial Development and Financing Constraints:
62  *   International Evidence from the Structural Investment Model,
63  *   Review of Financial Studies, 16(3): 765-791.
64  *   shellout "$path\Refs\Love_2003_RFS.pdf"
65
66  *-Whited, T., 1992, (Euler equation)
67  *   Debt, Liquidity Constraints, and Corporate Investment:
68  *   Evidence from Panel Data,

```

```

69 *   Journal of Finance, 47(4): 1425-1460.
70   shellout "$path\Refs\Whited_1992_JF.pdf"
71
72 *-Gilchrist, S., C. Himmelberg, 1995, (Fundamental Q: FQ)
73 *   Evidence on the role of cash flow for investment,
74 *   Journal of Monetary Economics, 36(3): 541-572.
75   shellout "$path\Refs\Gilchrist_1995_FQ_PVAR.pdf"
76
77 *-McLean, R. D., T. Zhang, M. Zhao, 2012,
78 *   Why does the law matter?
79 *   Investor protection and its effects on investment, finance, and growth,
80 *   Journal of Finance, 67(1): 313-350.
81   shellout "$path\Refs\McLean_2012_JF.pdf"
82
83
84
85
86 *-----
87 *-> 1 Panel VAR 模型简介
88 *-----
89   global path "`c(sysdir_personal)'PX_papers" // 课程总文件夹
90   cd "$path\Love_2006" // 进入本讲目录
91   adopath + "$path\adofiles" // 自编程序
92
93   help pvar2 // PVAR 的估计, IRF, FEVD, 滞后阶数选择、因果检验等
94   which pvar2
95
96   help xtvar // PVAR 的估计和冲击反应
97   which xtvar
98   help xtvarsoc // PVAR 最佳滞后阶数的筛选
99   which xtvarsoc
100  help xtvarstable // PVAR 稳定性检验
101  which xtvarstable
102  help xtvargranger // PVAR Granger 因果检验
103  which xtvargranger
104
105
106 *-----
107 *-1.1 Panel VAR 简介
108 *-----
109
110 *-VAR 简介
111   shellout "$path\refs\连玉君_chp11_TS.pdf" //时间序列基础
112   shellout "$path\Refs\连玉君_chp15_VAR.pdf" //书稿, 较全面, 勿外传!
113   *-Stata高级视频教程: B6_TimeS.do
114
115 *-PVAR 简介
116   shellout "$path\Refs\连玉君_PhD_chp4_Panel_VAR.pdf" // 博士论文第四章
117   shellout "$path\Refs\Holtz_1988_PVAR.pdf" // Holtz 论文
118
119 *-PVAR 应用
120   shellout "$path\Refs\Love_2006_pvar.pdf" // I.Love 的论文
121
122   *-在 Google Scholar 中能搜到很多采用 Love 程序的论文
123   *-搜索关键词: "pvar Love"
124
125
126
127 *-----
128 *-1.2 pvar 命令: LOVE 的原程序
129 *-----
130
131   cd "$path\Love_2006"
132   use grunfeld.dta, clear
133   rename company id // 截面变量的名称必须为 id
134   xtset id year
135
136   helm invest mvalue kstock // 前向差分, 去除个体相应

```

```

137
138     pvar kstock invest mvalue, lag(2) gmm monte 50 "decomp 30"
139
140     pvar kstock invest mvalue, lag(2) gmm monte 500 list_mon ///
141         "decomp 30" impulse list_imp
142
143     mat dir
144
145     *-重新绘制冲击反应图
146     preserve
147         use imp_50.dta, clear
148         twoway line kstock* s if order==1
149     restore
150
151
152 *-----
153 *-1.3 pvar2命令: 连玉君更新后的程序
154 *-----
155
156 *-----
157 *-1.3.1 主要改进
158
159     help pvar2    // see Section: `What is new?'
160
161     *-(1)  tsset
162     *-(2)  helm 内化
163     *-(3)  timeeffect 内化
164     *-(4)  irf(#)  irfformat()
165     *-(5)  decomp(#)
166     *-(6)  AIC, BIC, HQIC (滞后阶数选择依据)
167     *      估计结果可以用 est store 存储, 并进而用 esttab 输出
168     *-(7)  nograph, saving() 选项
169     *      设定 irf() 选项后, MC 后产生的 CI 数据会自动存储在 irf_data.dta 文件中
170     *      若附加 nograph 选项, 则只采用 MC 获得 IRF 数据, 不输出图形
171     *-(8)  Granger 因果检验
172
173     *-高级应用: 可以自行修改改程序(建议另存后修改)
174     doedit "$path\adofiles\pvar2.ado"  //打开 ado 文件
175     doedit "$path\adofiles\pvar2.hlp"  //打开 帮助文件
176
177
178
179 *-----
180 *-1.3.2 PVAR 模型的 GMM 估计
181
182     cd "$path\Love_2006"
183     use grunfeld.dta, clear
184     xtset company year
185
186     pvar2 kstock invest mvalue          // 默认为 PVAR(1) 模型
187
188     pvar2 kstock invest mvalue, lag(2)
189
190     *-屏幕呈现方式
191     esttab, b(%6.3f) star(* 0.1 ** 0.05 *** 0.01) ///
192         s(N AIC BIC HQIC)      nogap replace          //基本列表方式
193
194     esttab, b(%6.3f) star(* 0.1 ** 0.05 *** 0.01) ///
195         s(N AIC BIC HQIC) wide nogap replace          //横向列表
196
197     esttab, b(%6.3f) star(* 0.1 ** 0.05 *** 0.01) ///
198         s(N AIC BIC HQIC) wide onecell staraux replace //横向列表(接排)
199
200
201 *-----
202 *-1.3.3 PVAR 滞后阶数的选择
203
204     pvar2 kstock invest mvalue, lag(4) soc

```

```

205     pvar2 kstock invest mvalue, lag(5) soc
206
207     *-判断规则:
208     * (1) 选择 AIC, BIC 或 HQIC 值最小的模型;
209     * (2) 但三者不一致时, BIC/HQIC 倾向于选择比较精简的模型,
210     *     AIC 倾向于选择比较"丰满"的模型,
211     *     通常, BIC/HQIC 优于 AIC.
212     * (3) 有时也不能完全依赖上述准则, 需要做一些人为判断
213     * (4) 计算公式: pp.56 Eq.(4-31)-Eq.(4-33)
214     shellout "$path\Refs\连玉君_PhD_chp4_Panel_VAR.pdf" // 博士论文第四章
215
216     *-引用方法:
217     * 连玉君. 中国上市公司投资效率研究. 北京: 经济管理出版社, 2009: pp.80.
218
219     *-基本原理
220     pvar2 kstock invest mvalue, lag(1)
221     est store L1
222     pvar2 kstock invest mvalue, lag(2)
223     est store L2
224     pvar2 kstock invest mvalue, lag(3)
225     est store L3
226     pvar2 kstock invest mvalue, lag(4)
227     est store L4
228     *-呈现结果
229     global m "L1 L2 L3 L4"
230     esttab $m, mtitle($m) star(* 0.1 ** 0.05 *** 0.01) ///
231     s(N AIC BIC HQIC) b(%6.3f) nogap
232
233
234     *-----
235     *-1.3.4 PVAR 冲击反应函数 (IRFs)
236
237     *-What is IRF?
238     shellout "$path\refs\连玉君_chp11_TS.pdf" //时间序列基础
239     shellout "$path\Refs\连玉君_chp15_VAR.pdf" //参见第 15.2.4 小节
240     shellout "$path\PPT\IRFs.pdf"
241     * An IRF measures the effect of a shock to an endogenous variable
242     * on itself or on another endogenous variable
243
244     *-参考资料:
245     * Lutkepohl (2005, 51-63)
246     * Lütkepohl, H. New Introduction to Multiple Time Series Analysis. Springer,
247     2005.
248     * Hamilton (1994, 318-323)
249     * Hamilton, J. Time Series Analysis. Princeton: Princeton University Press,
250     1994.
251
252     *-pvar2 命令
253     pvar2 kstock invest mvalue, lag(2) irf(6)
254
255     *-Notes:
256     *
257     *-(1) 向前 10 期 IRFs
258     *     期数的选择决定于你的资料结构, 年\月\日?
259     pvar2 kstock invest mvalue, lag(2) irf(10)
260
261     *-(2) 默认 MC 200次, 通过 reps() 选项可以更改 MC 次数, e.g.
262     *     报告论文最终结果时, 以 500-1000 次为宜, 视样本数而定;
263     *     设定 reps() 时必须同时设定 irf()
264     pvar2 kstock invest mvalue, lag(2) irf(10) reps(500)
265
266     *-(3) 种子值(seed)的作用
267
268     *-A. 不设种子值的效果
269     *-----请执行2次,对比结果-----begin----
270     qui pvar2 kstock invest mvalue, reps(100) irf(6) nograph
271     preserve // IRF 结果
272     use IRF_data, clear

```

```

271         list s order kstock_5 kstock kstock_95 in 1/5
272         restore
273     *-----over-----
274
275     *-B. 设种子值的效果
276     *-----begin-----
277         qui pvar2 kstock invest mvalue, reps(10) irf(6) nograph seed(13579)
278         preserve // IRF 结果
279         use IRF_data, clear
280         list s order kstock_5 kstock kstock_95 in 1/5
281         restore
282     *-----over-----
283     *-结论：发表前一定要设定种子值，以保证结果的可重现
284
285     *-c. 也可以预先设定种子值（两条命令一起执行）
286     *-----begin-----
287         set seed 13579
288         qui pvar2 kstock invest mvalue, reps(10) irf(6) nograph
289         preserve // IRF 结果
290         use IRF_data, clear
291         list s order kstock_5 kstock kstock_95 in 1/5
292         restore
293     *-----over-----
294
295     *-(4) 默认自动输出 .gph 图片，可以进一步输出为 .wmf 格式，以便贴入 word
296         pvar2 kstock invest mvalue, lag(2) irf(6)
297         graph export Figure.wmf, replace
298
299     *-(5) 图形纵轴的刻度标签
300         pvar2 kstock invest mvalue, lag(2) irf(6) irfformat(%4.0f)
301
302     *-(6) 绘制黑白图片：更改图片模板
303         set scheme slmono //设定 reps(100)，只是为了节省时间
304         pvar2 kstock invest mvalue, lag(2) irf(6)
305         graph export Figure_black.wmf, replace
306         *-Note: 其他模板
307         help scheme
308         set scheme s2color //改回默认值(彩色图片)
309
310     *-(7) 关于 IRF 图形的进一步说明
311
312     *-更改图形的显示模式为黑白图片
313         graph use gr1_2.gph, scheme(slmono) // IRF of v1 to v2
314         graph export fig_kstock_invest.wmf, replace // 另存图片
315
316     *-这个模式的效果也不错
317         graph use gr1_3.gph, scheme(slcolor)
318         graph export fig_kstock_mvalue.wmf, replace
319
320     *-利用内存中的数据绘制 IRFs 图形
321         xtset company year
322         pvar2 kstock invest mvalue, lag(2) reps(500) irf(10) nograph
323         dir *.dta // 当前工作路径下存储的数据文件列表
324         preserve // why?
325         use irf_data.dta, clear //此前自动存储的 IRF 数据
326         twoway line invest_5 invest invest_95 s if varname=="kstock", ///
327             title(IRFs of kstock to invest) ///
328             xtitle(Steps) ylabel(,angle(0)) ///
329             yline(0, lpattern(dash)) ///
330             scheme(slmono) legend(off)
331         graph export fig_kstock_invest.wmf, replace
332         restore
333
334     *-重新合并多个图片
335         graph combine gr1_1.gph gr1_2.gph gr1_3.gph, ///
336             title("IRFs of kstock to other variables") ///
337             rows(1) ycommon colfirst imargin(small) ///
338             scheme(s2color)

```

```

339
340
341 *-----
342 *-1.3.5 PVAR 预测误差之方差分解 (FEVD)
343
344 *-含义: Forecast Error Variance Decomposition, FEVD
345 *      对内生变量预测的误差变异数中, 每一个正交单位冲击的贡献比例。
346
347 *-详见:
348   shellout "$path\Refs\连玉君_chp15_VAR.pdf"
349   *-Section 15.1.6; Section 15.2.5
350
351   use grunfeld.dta, clear
352   xtset company year
353   pvar2 kstock invest mvalue, lag(2) irf(5) nograph decomp(5)
354   *-Note: 附加 nograph 选项有助于节省时间
355
356   pvar2 kstock invest mvalue, lag(2) irf(10) nograph decomp(10) //更改: steps=10
357
358 *-----
359 *-1.3.6 Granger 因果检验
360
361   pvar2 kstock invest mvalue, lag(3) granger
362
363   *-检验过程解析: invest 是否为 kstock 的 Granger 因?
364   local Eq "h_kstock" // 方程名称
365   local v "h_invest"
366   test [`Eq']L.`v' = [`Eq']L2.`v' = [`Eq']L3.`v' = 0
367
368   *-Notes:
369   * 执行 Granger 因果检验时, 可以不必拘泥于 xtvarsoc 选出的滞后阶数;
370   * 因为 Granger 因果检验的目的在于检验 x 滞后项对 y 的联合影响
371
372 *-----
373 *-1.3.7 pvar2.ado 的返回值
374
375   macro dir // 暂元
376   matrix dir // 矩阵
377   dir *.dta // 数据文件 IRFs
378   dir gr*.gph // 图形文件 IRFs
379
380
381
382
383 *-----
384 *-1.4 xtvar 命令: 连玉君编写的程序
385 *-----
386
387 *-----
388 *-参考资料对应表
389
390 *-1.4.1 PVAR 的估计 连玉君_博士论文 Section 4.4.1
391 *-1.4.2 最佳滞后阶数的选择 连玉君_博士论文 Section 4.4.2
392 *-1.4.3 稳定性检验 连玉君_博士论文 Section 4.4.3
393   shellout "$path\Refs\连玉君_PhD_chp4_Panel_VAR.pdf"
394
395 *-1.4.4 Granger 因果检验 连玉君_chp15_VAR.pdf Section 15.1.4
396 *-1.4.5 冲击反应 连玉君_chp15_VAR.pdf Section 15.2.4
397   shellout "$path\Refs\连玉君_chp15_VAR.pdf"
398
399
400 *-----
401 *-1.4.1 PVAR 的估计
402
403   cd "$path\Love_2006"
404   use grunfeld.dta, clear
405   xtset company year
406

```



```

407     xtvar kstock invest mvalue, var(2)
408
409     *-把 mvalue 设定为外生变量
410     xtvar kstock invest, var(2) ex(mvalue)
411
412
413     *-----
414     *-1.4.2 最佳滞后阶数的选择 (基于 AIC, SBIC, HQIC 准则)
415
416     xtvarsoc kstock invest mvalue, maxlag(4)
417
418
419     *-----
420     *-1.4.3 稳定性检验
421
422     *-s1- 估计 PVAR 模型
423     xtvar kstock invest mvalue, var(1)
424
425     *-s2- 执行稳定性检验
426     xtvarstable
427     xtvarstable, graph           // 图形呈现
428     xtvarstable, graph dlabel    // 距离
429     xtvarstable, graph modlabel  // 模
430
431
432     *-----
433     *-1.4.4 Granger 因果检验
434
435     xtvar kstock invest mvalue, var(3)
436     xtvargranger
437     xtvar kstock invest mvalue, var(1)
438     xtvargranger
439
440
441     *-----
442     *-1.4.5 冲击反应
443
444     xtvar kstock invest mvalue, var(1)
445     xtvar kstock invest mvalue, irfs bs reps(50)
446
447     *-Note: 目前我们还没有很好地解决 IRF 绘图的问题,
448
449
450
451
452
453
454     *-----
455     *-> 2 Love and Zicchino (2006) 的实证分析过程
456     *-----
457
458     shellout "$path\Refs\Love_2006_pvar.pdf" // I.Love 的论文
459
460     *-----
461     *-2.1 样本初筛
462
463     cd "$path\Love_2006"
464
465     clear all
466     set memory 80m // 设定 Stata 占用的内存空间
467     use "$path\data\GTA2008_my.dta", clear //调入数据
468
469     keep id year tobin size tl TA TD ll sgr tagr ///
470     div_cash div_yes sicda sicmen ///
471     b001101000 c001000000 a001212000 ///
472     a001101000 c002006000 a001100000 ///
473     a002100000 a001101000 a001123000 ///
474     b002000000 a003000000 province

```

```

475
476     rename a001212000 netpeq //期初固定资产净值
477     rename c002006000 capex  //购建固定资产、无形资产和其他长期资产支付的现金
478     rename c001000000 cflow  //经营活动产生的现金流净额
479     rename b001101000 sales  //营业收入
480
481     keep if year>1997          //1998年以后采用现金流量表
482
483
484 *-----
485 *-2.2 定义变量
486
487     gen STD = netpeq-capex //分母
488     gen IKB = capex/STD
489     gen SKB = sales/STD
490     gen CFKB = cflow/STD
491     rename tobin TOBINQ
492     tsset id year
493     order id year SKB CFKB IKB TOBINQ
494
495
496 *-----
497 *-2.3 样本再次筛选
498
499 *-----
500 *-M&A (兼并收购)
501     des tagr
502     sum tagr, detail
503     histogram tagr
504     count if tagr>1.5
505     drop if (tagr>1.5&tagr!=.)
506     *-Note: 这个完全是主观判断了, 这个处理对结果有何影响?
507
508 *-----
509 *-处理为平行面板
510     xtbalance, range(1999 2007) miss(*KB TOBINQ) //市场化指数数据: 1999-2007
511     xtides
512
513 *----- pp.209, Appendix A
514 *-Trim 处理
515     foreach v of varlist IKB SKB CFKB{
516         qui sum `v', detail
517         replace `v'=. if `v'<r(p5)
518         replace `v'=. if `v'>r(p95)
519     }
520
521 /*
522 *-----
523 *-winsor 处理 || 会有很严重的离群值
524     foreach v of varlist IKB SKB CFKB{
525         rename `v' `v'_a
526         winsor `v'_a, gen(`v') p(0.01)
527         drop `v'_a
528     }
529 */
530
531
532 *-----
533 *-2.4 合并市场化进程指数数据
534
535 *-----
536 *-重新定义省名
537
538     cap safedrop prov
539     gen str6 prov = "."
540     replace prov = "安徽" if province==1
541     replace prov = "北京" if province==2
542     replace prov = "福建" if province==3|province==31 //31=厦门

```

```

543     replace prov = "甘肃"    if province==4
544     replace prov = "广东"    if province==5|province==27 //27=深圳
545     replace prov = "广西"    if province==6|province==7 // 7=广西
546     *replace prov = "桂林"    if province==7
547     replace prov = "贵州"    if province==8
548     replace prov = "海南"    if province==9
549     replace prov = "河北"    if province==10
550     replace prov = "河南"    if province==11
551     replace prov = "黑龙江"  if province==12
552     replace prov = "湖北"    if province==13
553     replace prov = "湖南"    if province==14
554     replace prov = "吉林"    if province==15
555     replace prov = "江苏"    if province==16|province==19 //19=南京
556     replace prov = "江西"    if province==17
557     replace prov = "辽宁"    if province==18
558     *replace prov = "南京"    if province==19
559     replace prov = "内蒙古"  if province==20
560     replace prov = "宁夏"    if province==21
561     replace prov = "青海"    if province==22
562     replace prov = "山东"    if province==23
563     replace prov = "山西"    if province==24
564     replace prov = "陕西"    if province==25
565     replace prov = "上海"    if province==26
566     *replace prov = "深圳"    if province==27
567     replace prov = "四川"    if province==28
568     replace prov = "天津"    if province==29
569     replace prov = "西藏"    if province==30
570     *replace prov = "厦门"    if province==31
571     replace prov = "新疆"    if province==32
572     replace prov = "云南"    if province==33
573     replace prov = "浙江"    if province==34
574     replace prov = "重庆"    if province==35
575
576     *-----
577     *-合并数据
578
579     merge n:1 prov year using "$path\Data\China_MI.dta"
580
581     drop if prov=="."|prov=="西藏" //删除没有省份归属数据的观察值
582
583
584     *-----
585     *-2.5 Love 文中的结果
586     *-----
587
588     *-定义 High-Low 虚拟变量
589
590     dropvars *FD*
591     gen FD = s4a // 4a. 金融业市场化程度
592     bysort prov: egen FD_p50 = median(FD) //各省的中位数,去除年度差异
593     qui sum FD_p50, detail
594     gen High_FD = FD_p50>r(p50) //FD大于样本中位数的省份定义为1, 否则为0
595
596
597     *-----
598     *-表 1 pp.196
599     *-----
600
601     *-Column 2-3
602     tab prov if High_FD==1
603     tab prov if High_FD==0
604     *-Column 4
605     panels id: tab prov if High_FD==1
606     panels id: tab prov if High_FD==0
607     *-Column 5
608     tabstat FD if High_FD==1, s(mean) by(prov) f(%6.3f)
609     tabstat FD if High_FD==0, s(mean) by(prov) f(%6.3f)
610

```

```

611 *-----
612 *-表 3      pp.198
613 *-----
614     bysort High_FD:  ///
615         tabstat SKB CFKB IKB TOBINQ, s(mean sd min q max) c(s) f(%4.2f)
616
617
618 *-模型设定：滞后阶数的选取
619
620     pvar2 SKB CFKB IKB, lag(4) soc           // Full sample
621     pvar2 SKB CFKB IKB if High_FD==1, lag(4) soc // High
622     pvar2 SKB CFKB IKB if High_FD==0, lag(4) soc // Low
623
624 *-----
625 *-表 4      3-variables PVAR model
626 *-----
627
628 *-区分金融发展水平
629     pvar2 SKB CFKB IKB if High_FD==1, lag(1)
630     est store High
631     pvar2 SKB CFKB IKB if High_FD==0, lag(1)
632     est store Low
633
634     local m "Low High"
635     esttab `m', mtitle(`m') star(* 0.1 ** 0.05 *** 0.01) ///
636         s(N AIC BIC HQIC) b(%6.3f) wide nogap compress replace
637
638 *-----
639 *-表 5      4-variables PVAR model
640 *-----
641     xtset id year
642     pvar2 SKB CFKB IKB TOBINQ, lag(1)
643     est store All4
644     pvar2 SKB CFKB IKB TOBINQ if High_FD==1, lag(1)
645     est store High4
646     pvar2 SKB CFKB IKB TOBINQ if High_FD==0, lag(1)
647     est store Low4
648
649     local m "All4 High4 Low4"
650     esttab `m', mtitle(`m') star(* 0.1 ** 0.05 *** 0.01) ///
651         s(N AIC BIC HQIC) b(%6.3f) wide nogap compress replace
652
653 *-----
654 *-图 1      Low-FD  IRF 3-variables PVAR model || 冲击反应函数
655 *-----
656     xtset id year
657     * 正式投稿前一定要设定种子值
658     pvar2 SKB CFKB IKB if High_FD==0, reps(1000) irf(6) seed(13579)
659     graph export Fig01_3IRF_low.wmf, replace //保存图片，可直接贴入 word
660
661     *-Note: 变量的顺序很重要(see pp.194-195)
662     *
663     *(1) 所谓"正交冲击反应函数"是指经过 Cholesky 分解后的干扰项
664     *      Cholesky 分解与变量的先后顺序有关
665     *(2) 排序依据：外生性越强的变量越先出现
666     *      SKB[t] --> IKB[t] --> SKB[t+1]
667     *      CFKB[t]--> IKB[t] --> CFKB[t+1]
668     *      so, SKB and CFKB 应该放在 IKB 之前;
669     *      TOBINQ[t-1] --> SKB[t], CFKB[t], IKB[t]
670     *      so, TOBINQ 应该放在最后
671     *
672     * 白话版规则：分析一下"连锁反应"的先后顺序，进而按此顺序放置变量即可
673
674
675
676 *-----
677 *-图 2      High-FD  IRF 3-variables PVAR model || 冲击反应函数
678 *-----

```

```

679     xtset id year
680     pvar2 SKB CFKB IKB if High_FD==1, reps(1000) irf(6) seed(13579)
681     graph export Fig01_3IRF_high.wmf, replace //保存图片, 可直接贴入 word
682
683
684 *-----
685 *-图 3      IRF-difference High v.s. Low
686 *-----
687 *-计算依据: pp.195 倒数第9行(合理性值得怀疑)
688 *-Assumption: 两个样本独立
689
690     global option "reps(1000) irf(6) seed(13579) nograph"
691     xtset id year
692     pvar2 SKB CFKB IKB if High_FD==0, saving(IRF3_low) $option
693     pvar2 SKB CFKB IKB if High_FD==1, saving(IRF3_high) $option
694
695     preserve
696     pvar2_irf_diff "IRF3_low" "IRF3_high" %4.2f
697     graph export Fig03_3IRF_diff.wmf, replace
698     restore
699
700
701 *-----
702 *-图 4      Low-FD  IRF 4-variables PVAR model || 冲击反应函数
703 *-----
704     xtset id year
705     pvar2 SKB CFKB IKB TOBINQ if High_FD==0, reps(1000) irf(6) seed(13579)
706     graph export Fig04_4IRF_low.wmf, replace
707
708
709 *-----
710 *-图 5      High-FD  IRF 4-variables PVAR model || 冲击反应函数
711 *-----
712     xtset id year
713     pvar2 SKB CFKB IKB TOBINQ if High_FD==1, reps(1000) irf(6) seed(13579)
714     graph export Fig05_4IRF_high.wmf, replace
715
716
717 *-----
718 *-图 6      IRF-difference High v.s. Low (4-variables)
719 *-----
720 *-计算依据: pp.195 倒数第9行(合理性值得怀疑)
721 *-Assumption: 两个样本独立
722
723     global option "reps(1000) irf(6) seed(13579) nograph"
724     xtset id year
725     pvar2 SKB CFKB IKB TOBINQ if High_FD==0, saving(IRF4_low) $option
726     pvar2 SKB CFKB IKB TOBINQ if High_FD==1, saving(IRF4_high) $option
727
728     preserve
729     pvar2_irf_diff "IRF4_low" "IRF4_high" %4.2f
730     graph export Fig06_4IRF_diff.wmf, replace
731     restore
732
733
734 *-----
735 *-表 6      方差分解 3-variables PVAR model
736 *-----
737
738 *-Panel A: Low FD    3 variables
739
740     pvar2 SKB CFKB IKB if High_FD==0, ///
741           irf(10) nograph decomp(10) //10 periods ahead
742
743     *-Notes:
744     * (1) 结果释义
745     *      最后三行就是论文中报告的结果
746     * (2) bug

```

```

747      *      就目前这个版本而言，首次计算 FEVD 时，还需设定 irf(10) nograph
748
749      *-Love 程序的做法
750      preserve
751          xtset id year
752          helm SKB CFKB IKB    // 前向差分，去除个体相应
753          pvar SKB CFKB IKB , lag(1) gmm "decomp 10"
754      restore
755
756
757      *-Panel B: High FD    3 variables
758          pvar2 SKB CFKB IKB if High_FD==1,    ///
759              irf(10) nograph decomp(10)
760
761      *-Panel C: Low FD    4 variables
762          pvar2 SKB CFKB IKB TOBINQ if High_FD==0,    ///
763              irf(10) nograph decomp(10)
764
765      *-Panel D: High FD    4 variables
766          pvar2 SKB CFKB IKB TOBINQ if High_FD==1,    ///
767              irf(10) nograph decomp(10)
768
769
770
771
772
773      *=====
774      *=====
775      *
776      * Appendix: 中国市场化进程指数 (1999-2007) 处理过程
777      *
778      *=====
779      *=====
780
781      *-----
782      *-Notes:
783      *-----
784      *-(1) 相关文件存放路径
785      *      D:\stata11\ado\personal\PX_papers\Data\mkt_index
786      *-(2) 调用方法
787      *      cd D:\stata11\ado\personal\PX_papers\Data\mkt_index //数据所在文件夹
788      *      use "China_MI.dta", clear
789      *-(3) 致谢：这些数据 and 主要处理过程由人民大学的黄继承博士提供，特此致谢！
790      *      姜付秀,黄继承,2011.
791      *      市场化进程与资本结构动态调整. 管理世界,(3):124-134.
792      *-----
793
794
795      clear all      //让STATA 处于初始状态，清除所有使用过的痕迹
796
797      global path "`c(sysdir_personal)'PX_papers"    //课程总目录
798      cd "$path\Data\mkt_index"
799
800      *-----
801      *-1- 将原始的 csv 文件另存为 Stata 格式的 .dta 文件
802      *-----
803      global f1 "1 1a 1b 1c 1d 1e 2 2a 2b 2c 3 3a 3b"
804      global f2 "4 4a 4a1 4a2 4b 4c 4d 5 5a 5a1 5a2 5b 5c 5c1 5c2 5d"
805      foreach file in 0 $f1 $f2 {
806          insheet using "Excel_data/\`file'.csv", clear //导入原始 csv 文件
807          drop in 1
808          rename v1 prov
809          local i=2
810          local j=1997
811          while `i'<=12{
812              rename v`i' x`j'
813              local ++i
814              local ++j

```



```

815     }
816     reshape long x, i(prov) j(year)
817     set obs 341
818     gen id = _n
819     order id
820     rename x s`file'
821     save "Stata_data/`file'.dta", replace //另存为 Stata 格式的数据文件
822 }
823
824 *-----
825 *-2- 将上述 .dta 文件合并为一份
826 *-----
827 use "Stata_data/0.dta", clear
828 foreach f in $f1 $f2{
829     merge 1:1 id using "Stata_data/`f'.dta" //依次合并
830     drop _merge //删除合并时自动产生的标示变量
831 }
832
833 *-----
834 *-3- 重新定义省份名称和地区虚拟变量
835 *-----
836 drop id
837 encode prov, gen(id) lab(id_lab) //省份名称数字化
838 labelbook
839 #delimit ;
840 gen str6 region = "西部";
841 replace region="东部"
842     if prov=="北京" | prov=="天津" | prov=="河北" |
843     prov=="上海" | prov=="江苏" | prov=="浙江" |
844     prov=="福建" | prov=="山东" | prov=="广东" |
845     prov=="海南" ;
846 replace region="中部"
847     if prov=="山西" | prov=="安徽" | prov=="江西" |
848     prov=="河南" | prov=="湖北" | prov=="湖南" ;
849 replace region="东北"
850     if prov=="辽宁" | prov=="吉林" | prov=="黑龙江";
851 #delimit cr
852
853 *-----
854 *-3- 变量标签
855 *-----
856 label var id "省份代码"
857 label var prov "省份名称"
858 label var year "年份"
859 label var region "地区"
860 label var s0 "0.总得分"
861 label var s1 "1.政府与市场关系"
862 label var s1a "1a.市场分配经济资源比重"
863 label var s1b "1b.减轻农民税费负担"
864 label var s1c "1c.减少政府对企业干预"
865 label var s1d "1d.减轻企业税外负担"
866 label var s1e "1e.缩小政府规模"
867 label var s2 "2.非国有经济发展"
868 label var s2a "2a.在工业销售收入中所占比重"
869 label var s2b "2b.固定资产总投资中所占比重"
870 label var s2c "2c.占城镇总就业人数的比例"
871 label var s3 "3.产品市场发育程度"
872 label var s3a "3a.价格由市场决定程度"
873 label var s3b "3b.减少商品市场地方保护"
874 label var s4 "4.要素市场发育程度"
875 label var s4a "4a.金融业市场化程度"
876 label var s4a1 "4a1.金融业竞争"
877 label var s4a2 "4a2.信贷资金分配市场化"
878 label var s4b "4b.引进外资程度"
879 label var s4c "4c.劳动力流动性"
880 label var s4d "4d.技术成果市场化"
881 label var s5 "5.中介组织发育和法律制度环境"
882 label var s5a "5a.中介组织发育"

```



```

883     label var s5a1 "5a1. 律师、会计师等市场中介组织服务条件"
884     label var s5a2 "5a2. 行业协会对企业帮助程度"
885     label var s5b "5b. 对生产者合法权益保护"
886     label var s5c "5c. 知识产权保护"
887     label var s5c1 "5c1. 三项专利申请受理量除以科技人员数"
888     label var s5c2 "5c2. 三项专利申请批准数除以科技人员数"
889     label var s5d "5d. 消费者权益保护"
890
891     *-----
892     *-3- 生成指数排序(各省的得分排序)
893     *-----
894     foreach newvar of varlist s0-s5d{
895         gsort year -`newvar'
896         set obs 341
897         gen `newvar'_rank = mod(_n,31)/`bysort year:gen `newvar'_rank=_n*/
898     }
899     label var s0_rank "排名0. 总得分"
900     label var s1_rank "排名1. 政府与市场关系"
901     label var s1a_rank "排名1a. 市场分配经济资源比重"
902     label var s1b_rank "排名b. 减轻农民税费负担"
903     label var s1c_rank "排名1c. 减少政府对企业干预"
904     label var s1d_rank "排名1d. 减轻企业税外负担"
905     label var s1e_rank "排名1e. 缩小政府规模"
906     label var s2_rank "排名2. 非国有经济发展"
907     label var s2a_rank "排名2a. 在工业销售收入中所占比重"
908     label var s2b_rank "排名2b. 固定资产投资中所占比重"
909     label var s2c_rank "排名2c. 占城镇总就业人数的比例"
910     label var s3_rank "排名3. 产品市场发育程度"
911     label var s3a_rank "排名3a. 价格由市场决定程度"
912     label var s3b_rank "排名3b. 减少商品市场地方保护"
913     label var s4_rank "排名4. 要素市场发育程度"
914     label var s4a_rank "排名4a. 金融业市场化程度"
915     label var s4a1_rank "排名4a1. 金融业竞争"
916     label var s4a2_rank "排名4a2. 信贷资金分配市场化"
917     label var s4b_rank "排名4b. 引进外资程度"
918     label var s4c_rank "排名4c. 劳动力流动性"
919     label var s4d_rank "排名4d. 技术成果市场化"
920     label var s5_rank "排名5. 中介组织发育和法律制度环境"
921     label var s5a_rank "排名5a. 中介组织发育"
922     label var s5a1_rank "排名5a1. 律师、会计师等市场中介组织服务条件"
923     label var s5a2_rank "排名5a2. 行业协会对企业帮助程度"
924     label var s5b_rank "排名5b. 对生产者合法权益保护"
925     label var s5c_rank "排名5c. 知识产权保护"
926     label var s5c1_rank "排名5c1. 三项专利申请受理量除以科技人员数"
927     label var s5c2_rank "排名5c2. 三项专利申请批准数除以科技人员数"
928     label var s5d_rank "排名5d. 消费者权益保护"
929
930     *-----
931     *-4- 删除部分观察值
932     *-----
933     drop if prov=="西藏"
934     drop if year==1997|year==1998
935
936     *-----
937     *-5- 保存数据
938     *-----
939     order id year prov region
940     tsset id year // 定义为 Panel Data
941     label data "China Market Index: 1999-2007" // 数据标签
942
943     save "China_MI.dta", replace
944
945     use "China_MI.dta", clear // 调用方法
946
947

```

```

1
2  *      -----
3  *      -----
4  *
5  *      Stata 学术论文专题
6  *
7  *      -----
8  *      -----
9
10
11  *      主讲人：连玉君 副教授
12  *
13  *      单 位：中山大学岭南学院金融系
14  *      电 邮：arlionn@163.com
15  *      博 客：http://blog.cnfol.com/arlionn
16  *      主 页：http://goo.gl/tRXba
17  *      微 博：http://weibo.com/arlionn
18
19
20  *-----
21  *      卢洪友，连玉君，卢盛峰，
22  *      "中国医疗服务市场中的信息不对称程度测算"，
23  *      《经济研究》，2011(4)：94-106.
24  *-----
25  *--Ref: Kumbhakar and Christopher (2009,JPA)
26  *      Kumb_2009_SFA.do
27
28
29  *--主要方法
30  *-- OLS
31  *-- Two-tier Stochastic Frontier Model
32  *      help SFA2tier
33
34  *--原始论文
35  cd "`c(sysdir_personal)'"PX_papers"
36  shellout "Refs\连玉君_2011_JJYJ_Twotier_SFA.pdf" // 两条命令一起执行
37
38
39
40
41
42  *-----
43  *--1- 数据处理和基本统计量
44  *-----
45
46  cd "`c(sysdir_personal)'"PX_papers\Lu_Lian_JJYJ2011" // 进入课程目录
47  adopath + "`c(sysdir_personal)'"PX_papers\adofiles" // 自编程序
48
49  *--调入数据
50  clear all
51  use "Data\Lian2011_JJYJ_data.dta", clear
52
53  *--增加变量标签
54  label var unit "样本编号"
55  label var year "观察年度"
56  label var price "Medical Price"
57  label var age "年龄"
58  label var symptoms "生病状况，数值越大表示越严重"
59  label var sex "性别"
60  label var urban "城乡分布"
61  label var married "婚姻状况(1=未婚,0=已婚)"
62  label var education "受教育程度(升序)"
63  label var insurance "是否有医疗保险"
64  label var job "是否有固定工作"
65  label var endurance "忍耐力强弱, 1=强,0=弱"
66  label var liaoning "辽宁=1"
67  label var hlj "黑龙江=1"
68  label var jiangsu "江苏=1"

```

```

69      label var shandong  "山东=1"
70      label var henan     "河南=1"
71      label var hubei     "湖北=1"
72      label var hunan     "湖南=1"
73
74      *-数字-文字对应表
75
76      label define edu_lab 0 "文盲" 1 "小学" 2 "初中" 3 "高中" ///
77                          4 "中职" 5 "大专及本科" 6 "研究生以上"
78      label value education edu_lab
79
80      label define urban_lab 1 "城市" 0 "农村"
81      label value urban urban_lab
82
83      label define sex_lab 1 "男" 0 "女"
84      label value sex sex_lab
85
86      label define insur_lab 1 "有医保" 0 "无医保"
87      label value insurance insur_lab
88
89      *-离群值和特殊值的处理
90      *replace price=1.01 if price==1 // 防止 lnprice=0
91      *replace lnprice = ln(price)
92      local winsor = 0 //在此处设置是否需要 winsor 处理, 1=需要
93      if `winsor'{
94          rename price price0
95          winsor price0, gen(price) p(0.05) // new
96          replace lnprice = ln(price)
97      }
98
99      local trim = 0 //在此设置是否需要"截尾"处理, 1=需要
100     if `trim'{
101         qui sum price, d
102         drop if price<r(p5) | price>r(p95)
103     }
104
105     *-产生新变量
106     gen lnage = ln(age)
107     gen age2 = age^2
108
109     *-教育层次虚拟变量
110     tab education // 直接放入教育年限变量并不好, 故转换成定性变量
111     recode education (min/1=1) (1/4=2) (4/6=3), gen(g_edu)
112     label define g_edu_lab 1 "文盲及小学" 2 "初中和高中" 3 "大学以上"
113     label value g_edu g_edu_lab
114     tab g_edu, gen(dum_edu)
115     drop dum_edu1
116     label var dum_edu2 "初中高中"
117     label var dum_edu3 "大学以上"
118
119     *-年龄段虚拟变量
120     tab age
121     dropvars g_age dum_age*
122     recode age (min/35=1) (35/45=2) (45/60=3) (60/max=4), gen(g_age)
123     tab g_age, gen(dum_age)
124     drop dum_age1
125     label var dum_age2 "age:(35-45]"
126     label var dum_age3 "age:(45-60]"
127     label var dum_age4 "age:>60"
128
129     *-消费紧迫性
130     tab symp, gen(dum_sym)
131     drop dum_sym1
132     label var dum_sym2 "symp=2, 病情一般"
133     label var dum_sym2 "symp=3, 病情相当严重"
134
135
136     *-----

```

```

137  *--表 1: 样本分布状况
138  *-----
139
140  *--变量列表
141  des
142  des2  // new command: 2012-03-21
143
144  *--省份名称
145  gen province = .
146  local i = 1
147  foreach v of varlist liaoning-guangxi{
148      replace province=`i' if `v'==1
149      local i = `i'+1
150  }
151  replace province=9 if province==. //贵州
152  label define province_lab 1 "辽宁" 2 "黑龙江" 3 "江苏" ///
153      4 "山东" 5 "河南" 6 "湖北" 7 "湖南" 8 "广西" 9 "贵州"
154  label value province province_lab
155
156  *--表 1: 地区分布状况(上半部分)
157  tabulate province
158  tab2way province urban, cellpct nofreq
159  tab2way province sex, cellpct nofreq
160  tab2way province insurance, cellpct nofreq
161
162  *--表 1: 年度分布状况(下半部分)
163  tabulate year
164  tab2way year urban, cellpct nofreq
165  tab2way year sex, cellpct nofreq
166  tab2way year insurance, cellpct nofreq
167
168
169  *-----
170  *--表 2: 变量的统计性描述
171  *-----
172
173  logout, save("Results\Table02") excel replace:  ///
174  tabstat price lnprice age sym urban  ///
175  sex married education insur job endur,  ///
176  s(N mean sd min p25 p50 p75 max) f(%6.3f) c(s)
177
178
179  *-----
180  *--2- 模型估计和筛选
181  *-----
182
183  *-----
184  *--表 3: 议价能力效应模型估计
185  *-----
186  *--注意: 有些模型有些时候会不收敛, 多执行几次即可
187
188  global year "year1991-year2006" // 加入时间虚拟变量
189  *global year "" // 不加时间虚拟变量
190
191  *-----
192  *--OLS (模型 1)
193  reg lnprice lnage symp urban sex married education
194  est store OLS
195  dropvars e
196  predict e, res // OLS 残差
197  sum e, de
198
199  *-----
200  *--模型 2: MLE
201  *--Test sig_w = sig_u = 0
202  *-- MLE of OLS model
203  constraint define 1 [sigma_u]_cons = 0 //约束条件: 不加常数项
204  constraint define 2 [sigma_w]_cons = 0

```

```

205     local xx "lnage symp urban sex married education"
206     ml model lf SFA2tier_KP05_extend_ll (lnprice: lnprice = `xx') ///
207         (sigma_v:) (sigma_u:) (sigma_w:), constraint(1 2)
208     qui ml check
209     qui ml search
210     ml max
211     est store m0
212
213     *-----
214     *-Two-tier SFA
215
216     *-模型 3
217     cap program drop SFA2tier
218     SFA2tier lnprice lnage symp urban sex married education, search
219     version 9
220     ml max, difficult //可能不收敛, 多执行几次即可(五行命令同时执行)
221     est store m1
222
223     *-模型 4
224     SFA2tier lnprice lnage symp urban sex married education ///
225         job endurance insur, search
226     version 9
227     ml max, difficult //可能不收敛, 多执行几次即可(五行命令同时执行)
228     est store m2
229
230     *-模型 5
231     SFA2tier lnprice lnage symp urban education ///
232         job endurance insur, search
233     version 9 //可能不收敛, 多执行几次即可(五行命令同时执行)
234     ml max, difficult tolerance(0.001)
235     est store m3
236
237     *-模型 6
238     SFA2tier lnprice lnage symp urban education ///
239         job endurance insur ///
240         liaoning hlj jiangsu shandong henan hubei hunan guangxi ///
241         , search
242     version 9
243     ml max, difficult
244     est store m4
245
246     *-模型 7
247     SFA2tier lnprice lnage symp urban education ///
248         job endurance insur $year ///
249         liaoning hlj jiangsu shandong henan hubei hunan guangxi ///
250         , search
251     version 9
252     ml max, difficult
253     est store m5
254
255
256     *-----输出结果-----
257     *
258     *-Table 3, 议价能力效应模型估计结果
259
260     *-屏幕呈现
261     local mm "OLS m0 m1 m2 m3 m4 m5"
262     esttab `mm', mtitle(`mm') b(%6.3f) t(%6.3f) ///
263         nogaps star(* 0.1 ** 0.05 *** 0.01) ///
264         s(N r2_a ll) sfmt(%8.2f) compress
265
266     *-输出到Excel
267     local mm "OLS m0 m1 m2 m3 m4 m5"
268     esttab `mm' using Results\Table03.csv, ///
269         mtitle(`mm') b(%6.3f) t(%6.3f) compress ///
270         nogaps star(* 0.1 ** 0.05 *** 0.01) replace ///
271         s(N r2_a ll) sfmt(%6.0f %4.3f %12.2f)
272

```

```

273
274 *-----
275 *-表 3 倒数第二和第三行 检验结果
276 *
277 * LR test based on the null model,
278 * i.e. Model(7), with sigma_u = sigma_w = 0
279
280 *-示意范例: m0 v.s. m1
281     lrtest m0 m1
282     return list
283
284 *-检验所有模型, 并将结果输出到矩阵中
285     local model m1 m2 m3 m4 m5
286     local k : word count `model'
287     mat LR1 = J(2, `k', 0)
288     local j = 1
289     foreach md of local model{
290         qui lrtest m0 `md'
291         mat LR1[1, `j++'] = (r(chi2) \ r(p))
292     }
293     mat colnames LR1 = `model'
294     mat rownames LR1 = chi2 p-value
295     dis in g "LR test based on Model(0)"
296     mat list LR1, format(%6.3f)
297 *-----
298
299 *-----
300 *-配对检验, 论文中未呈现此结果
301 *
302 * LR test based on the null model,
303 * i.e. Model(2), with sigma_u = sigma_w = 0
304     local model m0 m1 m2 m3 m4 m5
305     local k : word count `model'
306     mat LR2 = J(`=(`k'-1)*2', `k', .)
307     local mm "`model'"
308     forvalues i=1/`k'{
309         gettoken v mm: mm
310         local ii = `i'*2-1
311         local j = `i'+1
312         foreach subm of local mm{
313             qui lrtest `v' `subm'
314             mat LR2[`ii', `j++'] = (r(chi2) \ r(p))
315         }
316     }
317     mat colnames LR2 = `model'
318     mat rownames LR2 = chi2 p-value chi2 p chi2 p chi2 p chi2 p
319     dis in y "LR test for model selection: "
320     mat list LR2, format(%6.3f)
321 *-----
322
323
324 *-----
325 *-3- 方差分解和效率估算
326 *-----
327
328 *-后续分析以表 3 中的模型 7 为基础
329
330 *-----
331 *-表 4: 方差分解 (议价能力因素的医疗价格效应分析)
332
333     SFA2tier lnprice lnage symp urban education    ///
334             job endurance insur year1991-year2006  ///
335             liaoning hlj jiangsu shandong henan hubei hunan guangxi ///
336             , search
337     version 9
338     ml max, difficult
339
340     SFA2tier_sigs    // 连玉君自编程序

```



```

341
342
343 *-----
344 *-估计单边效应(效率值)
345
346     dropvars *_hat* *_diff* pat doc doc_pat
347     SFA2tier_eff // 连玉君自编程序
348     replace uw_diff_exp = -uw_diff_exp //医生-病人
349
350     foreach v of varlist u_hat w_hat uw_diff *_exp{
351         replace `v' = `v'*100 // 转换为百分比形式
352     }
353
354 *-----
355 *-表 5: 议价中医生和患者获得的总剩余 (估算效率)
356
357     logout, save("Results\Table05") excel replace: ///
358         tabstat w_hat_exp u_hat_exp uw_diff_exp, ///
359         s(mean sd p25 p50 p75) f(%6.2f) c(s) save
360
361
362 *-----
363 *-表 6: 分年度统计
364
365 *-详细结果
366     bysort year: tabstat w_hat_exp u_hat_exp uw_diff_exp, ///
367         s(mean sd p25 p50 p75) f(%6.2f) c(s)
368
369 *-论文呈现结果: 净剩余
370     logout, save("Results\Table06") excel replace: ///
371         tabstat uw_diff_exp, by(year) ///
372         s(mean sd p25 p50 p75) f(%6.2f) c(s)
373
374
375 *-----
376 *-表 7: 城市 v.s. 农村
377     logout, save("Results\Table07") excel replace: ///
378         bysort urban: tabstat w_hat_exp u_hat_exp uw_diff_exp, ///
379         s(mean sd p25 p50 p75) f(%6.2f) c(s)
380
381 *-----
382 *-表 8: 学历因素
383     *-g_edu
384     logout, save("Results\Table08") excel replace: ///
385         bysort g_edu: tabstat w_hat_exp u_hat_exp uw_diff_exp, ///
386         s(mean sd p25 p50 p75) f(%6.2f) c(s)
387
388
389 *-----
390 *-其它结果: 论文中未呈现
391
392 *-sex
393     bysort sex: tabstat w_hat_exp u_hat_exp uw_diff_exp, ///
394         s(mean sd p25 p50 p75) f(%6.2f) c(s)
395
396 *-married
397     bysort married: tabstat w_hat_exp u_hat_exp uw_diff_exp, ///
398         s(mean sd p25 p50 p75) f(%6.2f) c(s)
399     ttest uw_diff_exp, by(married)
400
401 *-insurance
402     bysort insurance: tabstat w_hat_exp u_hat_exp uw_diff_exp, ///
403         s(mean sd p25 p50 p75) f(%6.2f) c(s)
404     ttest uw_diff_exp, by(insurance)
405
406 *-job
407     bysort job: tabstat w_hat_exp u_hat_exp uw_diff_exp, ///
408         s(mean sd p25 p50 p75) f(%6.2f) c(s)

```



```

409      ttest uw_diff_exp, by(job)
410
411      *-g_age
412      bysort g_age: tabstat w_hat_exp u_hat_exp uw_diff_exp, ///
413          s(mean sd p25 p50 p75) f(%6.2f) c(s)
414
415
416      *-----
417      *-直方图
418      *-----
419
420      *set scheme s2color // 彩色图片
421      set scheme slmono // 黑白图片
422
423      *-图 1: 医生获得剩余的频数分布
424      histogram w_hat_exp, percent ///
425          title(Percent, place(10) size(*0.7)) ///
426          xtitle("Surplus extracted by doctors (%)") ///
427          xscale(titlegap(3) outergap(-2)) ///
428          ytitle("") ylabel(,angle(0))
429      *-输出图片
430      graph export "Results\Fig01_w_his.wmf", replace
431
432      *-图 2: 患者获得剩余的频数分布
433      histogram u_hat_exp, percent ///
434          title(Percent, place(10) size(*0.7)) ///
435          xtitle("Surplus extracted by patients (%)") ///
436          xscale(titlegap(3) outergap(-2)) ///
437          ytitle("") ylabel(,angle(0))
438      *-输出图片
439      graph export "Results\Fig02_u_his.wmf", replace
440
441      *-图 3: 净剩余的频数分布
442      histogram uw_diff_exp, percent ///
443          title(Percent, place(10) size(*0.7)) ///
444          xtitle("Net Surplus (%)") ///
445          xscale(titlegap(3) outergap(-2)) ///
446          ytitle("") ylabel(,angle(0))
447      *-输出图片
448      graph export "Results\Fig03_uw_his.wmf", replace
449
450
451      *-----
452      *-其它应用
453      *-----
454
455      *-[1] 投资效率分析
456      *      Lian, Y., C.-F. Chung, 2008,
457      *      Are Chinese Listed Firms Over-Investing?,
458      *      Available at SSRN: http://ssrn.com/abstract=1296462.
459      *-[2] 房地产价格分析
460      *-[3] 银行贷款利率
461
462
463      *-----
464      *-合作和修改经验
465      *-----
466
467      *-初稿和批注
468      shellout "$path\Lu_Lian_JJYJ2011\comments\20100502_初稿.doc"
469
470      *-修改建议
471      shellout "$path\Lu_Lian_JJYJ2011\comments\comments_0503.doc"
472

```

```

1
2  *      -----
3  *      -----
4  *
5  *      Stata 学术论文专题
6  *
7  *      -----
8  *      -----
9
10
11  *      主讲人：连玉君 副教授
12  *
13  *      单 位：中山大学岭南学院金融系
14  *      电 邮：arlionn@163.com
15  *      博 客：http://blog.cnfol.com/arlionn
16  *      主 页：http://goo.gl/tRXba
17  *      微 博：http://weibo.com/arlionn
18
19
20  *-----
21  * -Opler,1999,JFE-
22  *   Opler, T., L. Pinkowitz, R. Stulz, R. Williamson, 1999,
23  *   The determinants and implications of corporate cash holdings,
24  *   Journal of Financial Economics, 52(1): 3-46.
25  *-----
26  * -相关阅读：Flannery and Rangan (2006, JFE)  || Flannery_2006.do
27  *              Faulkender and Wang (2006, JF)  || Faulkender_2006.do
28
29  * -主要方法
30  *   OLS: regress
31  *   Fama-MacBeth (1973) regression: xtfrm
32  *   FE (Fixed Effect Model): xtreg, fe
33  *   Cross-sectional regression: xtreg, be
34  *   分组回归: statsby (一个非常强大的命令)
35  *   基本统计分析和 t-test: tabstat, ttest
36  *   分位数 (Quartiles): quantiles, pctlile
37  *   Markov 状态转移矩阵: xttrans
38
39
40  * -设定课程存储路径
41  *   global path "`c(sysdir_personal)'PX_papers" //请先执行该命令
42
43  *   cd "$path\Flannery_2006" // 进入课程目录
44
45  * -原始论文
46  *   shellout "$path\Refs\Opler_1999.pdf"
47
48  * -PPT
49  *   shellout "$path\PPT\Opler_1999.ppt"
50
51
52
53  *=====
54  * 样本筛选
55  *=====
56
57  *   clear all
58  *   set memory 80m // 设定 Stata 占用的内存空间
59
60  *   use "$path\data\GTA2008_my.dta", clear //调入数据
61
62  *   keep if year>1997 // 1998年以后采用现金流量表
63
64  * -剔除金融类企业和公用事业企业
65  *   drop if sicmen_str == "I"
66  *   drop if sicmen_str == "F"
67  *   drop if b001101000<0 //营业收入<0
68

```

```

69
70 *=====
71 * 指标构建
72 *=====
73
74 *-----
75 *-生成新变量
76 *-----
77 dropvars cash cflow t1 ll size
78 gen NetTA = TA - a001101000 // 净资产=TA-货币资金
79 gen cash = a001101000/NetTA // 货币资金/净资产
80 gen trcash= cash
81 replace trcash=1 if cash>1 // 大于1的替换成1
82 gen cflow= c001000000/NetTA // 经营活动产生的现金流净额/净资产
83 gen capExp = c002006000/NetTA
84 // 购建固定资产、无形资产和其他长期资产支付的现金/净资产
85 gen t1 = TD/TA // 负债合计/净资产
86 gen ll = a002206000/NetTA // 长期负债合计/净资产
87 gen size = ln(TA) // ln(总资产)
88 bysort sicda year: egen indsigma = sd(cflow) // updated
89 tsset id year
90 gen salegr = D.b001101000/L.b001101000 //营业收入增长率
91
92 gen NetWC = (a001100000-a002100000-a001101000)/NetTA
93 //[(营运资本=流动资产-流动负债)-货币资金]/净资产
94 replace div_cash=0 if div_cash==. // 参考 p.16 R&D 缺漏值的处理方法
95 gen payout = div_cash*nshra/NetTA // 现金股利(无奈之举)
96 gen deficit = payout + capExp + D.NetWC + ll - cflow
97
98 *-----
99 *-winsor 主要变量
100 *-----
101 local x1 "cash trcash capExp cflow t1 size"
102 local x2 "tobin indsigma NetWC payout salegr deficit"
103 foreach v of varlist `x1' `x2'{
104     rename `v' `v'_a
105     winsor `v'_a, gen(`v') p(0.01)
106     drop `v'_a
107 }
108
109 *-负债率需要特别处理
110 count if t1>1
111 sum t1 if t1>1, detail
112 replace t1=0.99 if t1>1 //特殊缩尾处理
113 *replace t1=. if t1>1 //删除处理
114
115
116 *-定义为平行面板(酌情选择, Opler, pp.14, 未作此处理)
117 * xtbalance, range(1999 2008)
118
119 *-----
120 *-删除缺漏值
121 *-----
122 local x1 "cash trcash capExp cflow t1 size tobin"
123 local x2 "indsigma NetWC payout div_cash"
124 qui reg `x1' `x2'
125 keep if e(sample) //一种巧妙地删除缺漏值的方法, 无需记命令
126 //also see: help missing()
127
128 *-----
129 *-增加变量标签
130 *-----
131 order id year sicda sicmen trcash cash capExp cflow salegr t1 size ///
132 tobin indsigma NetWC deficit payout TA salegr div_cash div_yes
133
134 label var cflow "现金流量"
135 label var cash "现金持有比率"
136 label var trcash "大于1截断的cash"

```

```

137     label var capExp      "投资支出率"
138     label var tl          "总负债率"
139     label var size        "公司规模_Ln总资产"
140     label var tobin        "Tobin's Q"
141     label var indsigma     "行业现金流波动sigma"
142     label var NetWC        "营运资本"
143     label var payout       "股利支付"
144     label var salegr        "主营业务成长率"
145     label var deficit      "财务赤字:Pecking order proxy"
146
147     keep id year sicda sicmen trcash cash capExp cflow salegr tl size ///
148           tobin indsigma NetWC deficit payout TA salegr div_cash div_yes
149
150     save "Cash_Opler_1999_JF.dta", replace //存一份小数据，方便日后调用
151
152
153
154     *=====
155     * Table 1 : Description of variables
156     *=====
157
158     clear all
159     use "Cash_Opler_1999_JF.dta", clear
160
161     local v "cash trcash size tobin cflow NetWC capExp payout indsigma tl"
162     logout, save(Table01) excel replace: ///
163         tabstat `v', stat(mean min p25 p50 p75 max N) f(%9.3f) c(s)
164
165
166
167     *=====
168     * Fig 2 : Mean and Median cash-to-assets across firm sizes
169     *=====
170
171     *-----
172     *-mean
173     bysort year: egen TA_mean = mean(TA) //各年度的样本均值
174     bysort year: egen avgcash_small = mean(cash) if TA<TA_mean
175     bysort year: egen avgcash_large = mean(cash) if TA>TA_mean
176
177     *global scheme "s2mono" //黑白图片
178     global scheme "s2color" //彩色图片
179     twoway (area avgcash_small year) ///
180           (function y=0.15, range(1999 2008) lp(dash) lc(white) lw(*1.5)), ///
181           xlabel(1999(1)2008, grid) ///
182           xscale(titlegap(2)) ///
183           ylabel(0.1(0.05)0.25, format(%4.2f) angle(0)) ///
184           ylabel(现金持有比率) ///
185           scheme($scheme) legend(off) ///
186           title("Panel A: Firms with Total Assets less than year mean") ///
187           xtitle(Year) ytitle(Cash-asset ratio) ///
188           saving(Fig1_a, replace)
189
190     twoway (area avgcash_large year) ///
191           (function y=0.15, range(1999 2008) lp(dash) lc(white) lw(*1.5)), ///
192           xlabel(1999(1)2008, grid) ///
193           xscale(titlegap(2)) ///
194           ylabel(0.1(0.05)0.25, format(%4.2f) angle(0)) ///
195           ylabel(现金持有比率) ///
196           scheme($scheme) legend(off) ///
197           title("Panel B: Firms with Total Assets greater than year mean") ///
198           xtitle(Year) ytitle(Cash-asset ratio) ///
199           saving(Fig1_b, replace)
200
201     graph combine Fig1_a.gph Fig1_b.gph, ///
202           cols(1) scheme(s1mono) note(mean)
203
204     *-更为直观的呈现方式

```

```

205 twoway connect avgcash_small avgcash_large year, sort
206
207 *-完整版本
208 twoway connect avgcash_small avgcash_large year, sort ///
209 xlabel(1999(1)2008, grid) ///
210 xscale(titlegap(2)) ///
211 ylabel(0.15(0.05)0.25, format(%4.2f) angle(0)) ///
212 scheme(s2mono) legend(off) ///
213 xtitle(Year) ///
214 ytitle(Cash-asset ratio) ///
215 saving(Fig1_b, replace)
216
217 *-----
218 *-Median
219
220 bysort year: egen medcash_small = median(cash) if TA<3*10^9
221 bysort year: egen medcash_large = median(cash) if TA>3*10^9
222
223 twoway (area medcash_small year) ///
224 (function y=0.1, range(1999 2008) lp(dash) lc(white) lw(*1.5)), ///
225 xlabel(1999(1)2008, grid) ///
226 xscale(titlegap(2)) ///
227 ylabel(0.05(0.05)0.2, format(%4.2f) angle(0)) ///
228 ytitle(现金持有比率) ///
229 scheme(slmono) legend(off) ///
230 title("Panel B: Firms with Total Assets less than 3*10^9") ///
231 xtitle(Year) ytitle(Cash-asset ratio) ///
232 saving(Fig1_amed, replace)
233
234 twoway (area medcash_large year) ///
235 (function y=0.1, range(1999 2008) lp(dash) lc(white) lw(*1.5)), ///
236 xlabel(1999(1)2008, grid) ///
237 xscale(titlegap(2)) ///
238 ylabel(0.05(0.05)0.2, format(%4.2f) angle(0)) ///
239 ytitle(现金持有比率) ///
240 scheme(slmono) legend(off) ///
241 title("Panel B: Firms with Total Assets greater than 3*10^9") ///
242 xtitle(Year) ytitle(Cash-asset ratio) ///
243 saving(Fig1_bmed, replace)
244
245 graph combine Fig1_amed.gph Fig1_bmed.gph, ///
246 cols(1) scheme(slmono) note(Median) // 合并图片
247
248 graph export Fig02.wmf, replace // 输出图片
249
250
251
252 *=====
253 * Fig 3 help statsby
254 *=====
255
256 *-Distribution of Coefficients on Lagged Change in Cash/Assets.
257 * from the "firm-wise regression
258
259 *-The model:
260 *
261 *  $D.Cash[t] = a + b \cdot D.Cash[t-1] + e[t]$ 
262 *-or
263 *  $X[t] - X[t-1] = a + b(X[t-1] - X[t-2]) + e[t]$ 
264 *
265 *
266 * 若存在均值回复, 则 b 应显著为负
267
268 qui tsset
269 bysort id: egen idN = count(id) //每家公司中的观察值的个数
270 statsby _b[L.D.cash], by(id) saving(Fig3data, replace): ///
271 reg D.cash L.D.cash if (idN>=5)
272 *-Note: 红色 "x" 表示观察值不足的公司(至少要3个)

```

```

273  *--绘图
274  preserve
275  use "Fig3data.dta", clear
276  sum _stat_1, detail
277  histogram _stat_1 //winsor前的图形
278  winsor _stat_1, gen(beta) p(0.005)
279  histogram beta, fraction scheme(slmono) ///
280  ylabel(0(0.02)0.12,angle(0) format(%4.2f)) ///
281  xlabel(-2(0.5)2,format(%4.1f)) ///
282  xtitle("Regression Model Coeffieicients ({&beta})")
283  graph export Fig03.wmf, replace // 保存图片,可以插入 word
284  tabstat beta, s(mean min q max N) f(%9.3f) c(s)
285  restore
286
287  *--样本均值
288  qui tsset
289  reg D.cash L.D.cash
290  panels id if e(sample) //统计参与回归的公司
291
292  *--Comments:
293  * (1) 由于样本数很小, Fig.3 里得到的估计系数低于真实值;
294  * (2) 上述结果表明存在"均值回复", 即公司不会任由 Cash 过高或过低
295
296
297
298
299  *=====
300  * Table 2 : Time series analysis of liquid asset holdings
301  *=====
302
303  *--The model:
304
305  * 设定依据:
306  * Shyam-Sunder and Myers(1999, pp.226, Eq.3, and pp.230, Table2)
307  *
308  * Opler et al.(1999, pp.21)
309  *
310  *  $Cash[t+1]-Cash[t] = a + b*(Cash[t]-Cash\_Target[t]) + e[t]$ 
311  *--or
312  *  $X[t+1]-X[t] = a + b(X[t]-X^*[t]) + e[t]$ 
313  *
314  *--系数含义:
315  *
316  *  $X[t]-X^*[t]>0$  表示 公司在第 t 年度超额持有现金
317  * 若存在调整行为, 则下一年度应该减少现金持有, 即,
318  *  $X[t+1]-X[t]<0$ 
319  *
320  * 这意味着  $b<0$ 
321
322  *--Table 2: 注释部分的描述(pp.20), 但表2中呈现的估计结果并不是这个模型
323  *
324  *  $Cash[t]-Cash[t-1] = a + b*(Cash\_Target[t]-Cash[t-1]) + e[t]$ 
325  *
326  * 参见 Flannery and Rangan(2006,JFE)
327
328  *--(1) Mean target MV(4)
329  qui tsset
330  dropvars cash_mv4 D_cash_mv4
331  gen cash_mv4 = (L4.cash+L3.cash+L2.cash+L.cash)/4
332  gen D_cash_mv4 = cash - cash_mv4
333  reg D.F.cash D_cash_mv4, robust
334  est store tab2_1
335
336  *--(2) target: size indsigma 的拟合值, OLS 估计
337  dropvars sizesigma D_sizesigma
338  gen sizesigma = . // 目标 cash ratio
339  forvalues t = 1999/2008{
340  qui reg cash size indsigma if year==`t' //第t年回归

```



```

341     predict cash_star_t if e(sample) //第t年拟合值
342     replace sizesigma = cash_star_t if year==`t'
343     drop cash_star_t
344 }
345 gen D_sizesigma = cash - sizesigma
346 reg D.F.cash D_sizesigma, robust
347 est store tab2_2
348
349 *-(3) target: size indsigma 的拟合值, Fama-MacBench,1973 估计
350 dropvars fmtarget D_fmtarg
351 xtfmb cash size indsigma
352 predict fmtarget, xb
353 gen D_fmtarg = cash - fmtarget
354 reg D.F.cash D_fmtarg, robust
355 est store tab2_3
356
357 *-(4) test pecking order
358 *pecking order: 采用前面定义的 deficit 变量
359 sum deficit
360 reg D.F.cash deficit, robust
361 est store tab2_4
362
363 *-(5) Model (1) + Pecking order
364 reg D.F.cash D_cash_mv4 deficit, robust
365 est store tab2_5
366
367 *-(6) Model (2) + Pecking order
368 reg D.F.cash D_sizesigma deficit, robust
369 est store tab2_6
370
371 *-(7) Model (3) + Pecking order
372 reg D.F.cash D_fmtarg deficit, robust
373 est store tab2_7
374
375 *-(8) Model (5) + [Pecking order]x[above target dummy]
376 dropvars dum_mean_above po_mean_above
377 gen dum_mean_above = (D_cash_mv4>0) //above target dummy
378 gen po_mean_above = deficit*dum_mean_above
379 reg D.F.cash D_cash_mv4 deficit po_mean_above, robust
380 est store tab2_8
381
382 *-(9) Model (6) + [Pecking order]x[above target dummy]
383 dropvars dum_size_above po_size_above
384 gen dum_size_above = (D_sizesigma>0)
385 gen po_size_above = deficit*dum_size_above
386 reg D.F.cash D_sizesigma deficit po_size_above, robust
387 est store tab2_9
388
389 *-(10) Model (7) + [Pecking order]x[above target dummy]
390 dropvars dum_sop_above po_sop_above
391 gen dum_sop_above = (D_fmtarg>0)
392 gen po_sop_above = deficit*dum_sop_above
393 reg D.F.cash D_fmtarg deficit po_sop_above, robust
394 est store tab2_10
395
396
397 *--结果汇总
398
399 label var D_cash_mv4 "Mean target"
400 label var D_sizesigma "Size&sigma target"
401 label var D_fmtarg "Sophist target"
402 label var deficit "Pecking order"
403 label var po_mean_above "Peck_x_above_tar_Mean"
404 label var po_size_above "Peck_x_above_tar_Size"
405 label var po_sop_above "Peck_x_above_tar_Soph"
406
407 *--屏幕显示
408 local beta "beta" //执行时选中此行, 报告 Beta 系数

```



```

409      *
410      local m1      "tab2_1 tab2_2 tab2_3 tab2_4 tab2_5"
411      global order  "_cons D_cash_mv4 D_sizesigma D_fmtarg deficit"
412      esttab `m1', mtitle(`m1') b(%6.3f) nogaps compress ///
413              star(* 0.1 ** 0.05 *** 0.01) `beta'      ///
414              stats(N r2_a) order($order) label
415
416      local m2      "tab2_6 tab2_7 tab2_8 tab2_9 tab2_10"
417      esttab `m2', mtitle(`m2') b(%6.3f) nogaps compress ///
418              star(* 0.1 ** 0.05 *** 0.01) `beta'      ///
419              stats(N r2_a) order($order) label
420
421      *-输出到 Excel
422      local beta     "beta"    //执行时选中此行, 报告 Beta 系数
423      local label    "label"   //执行时选中此行, 显示变量标签而非变量名称
424      local m1       "tab2_1 tab2_2 tab2_3 tab2_4 tab2_5"
425      local m2       "tab2_6 tab2_7 tab2_8 tab2_9 tab2_10"
426      esttab `m1' `m2' using tab2.csv, mtitle(`m1' `m2') compress ///
427              b(%6.3f) star(* 0.1 ** 0.05 *** 0.01) replace nogaps ///
428              scalar(N r2_a) order($order) `label' `beta'
429
430      *-Comments:
431      *
432      *(1) Opler(1999,Tab2) 中得到的调整系数为负, 如何解释?
433      *(2) 文中使用了多种 Target Cash 设定方法, why? any implications?
434
435
436
437      *=====
438      *   Table 3       : Firm characteristics by cash/assets quartiles
439      *=====
440
441      *-Goals & Methods:
442      *
443      *(1) Cash/NetTA 的决定因素分析: 单因素分析
444      *(2) 每个年度, 基于 Q1, Q2, Q3 分成四组, t-test
445      *(3) 列表方法
446
447      *-----
448      * 纵向列表
449      *-----
450      cap drop dumch
451      bysort year: quantiles cash, gen(dumch) n(4) //按年度分组
452      *-第一行的结果
453      tabstat cash, by(dumch) s(min max) f(%4.3f) nototal save
454      mat q1R = r(Stat1)
455      mat q2R = r(Stat2)
456      mat q3R = r(Stat3)
457      mat q4R = r(Stat4)
458      *-其他行
459      local vars "cash size tobin cflow NetWC capExp payout indsigma t1"
460      local nv = wordcount("`vars'")*2+2 //确定结果矩阵的行数
461      mat aa = J(`nv', 5, .) //存储结果的矩阵
462      local i = 3
463      foreach var of varlist `vars' {
464          forvalues j=1(1)4 {
465              qui sum `var' if dumch==`j',d
466              mat aa[`i',`j'] = (r(mean) \ r(p50))
467          }
468          preserve
469              qui drop if dumch==2|dumch==3
470              qui ttest `var', by(dumch)
471              mat aa[`i',5] = (r(t) \ r(p))
472          restore
473          local i = `i'+2
474      }
475      mat aa[1,1] = (q1R, q2R, q3R, q4R) // 插入前两行的取值范围
476      mat rownames aa = Min Max ///

```

```

477     cash [p50] size [p50] tobin [p50] cflow [p50] NetWC [p50] ///
478     capExp [p50] payout [p50] indsigma [p50] t1 [p50]
479     mat colnames aa = Q1th Q2th Q3th Q4th t-value(p)
480
481     *-屏幕呈现结果
482     mat list aa, format(%6.3f) noheader
483
484     *-输出到Excel表格中
485     logout, save(Tab03) excel fix replace: ///
486     mat list aa, format(%6.3f) noheader
487
488
489
490     *-----
491     * 横向列表 (self-reading)
492     *-----
493     cap drop dumch
494     bysort year: quantiles cash, gen(dumch) n(4)
495     local vars "cash size tobin cflow NetWC capExp payout indsigma t1"
496     local nv = wordcount("`vars'") + 1
497     mat aa = J(`nv', 10, 0)
498     local i = 2
499     foreach var of varlist `vars' {
500         forvalues j=1(1)4 {
501             qui sum `var' if dumch==`j', d
502             local z = (`j'-1)*2 + 1
503             mat aa[`i', `z'] = (r(mean) , r(p50))
504         }
505         preserve
506         qui drop if dumch==2 | dumch==3
507         qui ttest `var', by(dumch)
508         mat aa[`i', 9] = (r(t) , r(p))
509         restore
510         local i = `i'+1
511     }
512     qui sum cash
513     local min = r(min)
514     local max = r(max)
515     _pctile cash , n(4)
516     mat aa[1,1] = (`min', r(r1), r(r1), r(r2), r(r2), r(r3), r(r3), `max', ., .)
517     mat rownames aa = rang `vars'
518     mat colnames aa = Q1th p50 Q2th p50 Q3th p50 Q4th p50 t-value p-value
519     mat list aa, format(%6.3f)
520
521     *-Note: 这里有个 bug, 请找出来并修改
522
523
524
525     *=====
526     * Table 4 : Regressions predicting firm liquidity levels
527     *=====
528
529     *-Methods:
530     *
531     *-Times-series cross-sectional regression: Pooled OLS (help regress)
532     *-其他方法前面已经介绍
533
534     cap drop lncash
535     gen lncash = ln(cash)
536
537     *-(1) Fama-MacBeth Model
538     xtfrm lncash tobin size cflow NetWC capExp t1 indsigma div_yes
539     est store m4_1
540
541     *-(2) Pooled OLS + year dummies
542     dropvars yeardummy*
543     qui tab year, gen(yeardummy)
544     reg lncash tobin size cflow NetWC capExp t1 indsigma div_yes ///

```

```

545             yeardummy*, nocons
546     est store m4_2
547
548     *-(3) Pooled OLS + year dummies + industry dummies
549     dropvars sicdummy*
550     qui tab sicmen, gen(sicdummy)
551     drop sicdummy1
552     reg lncash tobin size cflow NetWC capExp tl indsigma div_yes ///
553             yeardummy* sicdummy*, nocons
554     est store m4_3
555
556     *-(4) cross-sectional regression: between effect model
557     xtreg lncash tobin size cflow NetWC capExp tl indsigma div_yes ///
558             , be
559     est store m4_4
560
561     *-(5) Fixed effects regression
562     xtreg lncash tobin size cflow NetWC capExp tl indsigma div_yes ///
563             ,fe
564     est store m4_5
565
566     *-输出结果
567     local save "using Tab04.csv" //执行时选中此行, 则输出 Excel 表格
568     local mm "m4_1 m4_2 m4_3 m4_4 m4_5"
569     local order "order(_cons)" //变量显示顺序
570     local drop "drop(year* sic*)" //无需显示的变量
571     esttab `mm' `save', scalar(N r2_a r2_w r2_b) b(%6.3f) t(%6.2f) ///
572             star(* 0.1 ** 0.05 *** 0.01) nogap replace `order' `drop' ///
573             mtitles(Fama-MacBeth Year Year&industry Cross-Section FE)
574
575     *-----
576     *-Comments and Discussions:
577     *
578     *-(0) add [label] option to present vairiable labels
579     *-(1) why use different specifications?
580     *-(2) in Col.4, Size is signifciantly Negative,
581     *      in Col.5, Size is signifciantly Positive, Why? Implications?
582     *-(3) use Cash instead of ln(Cash) as dependent vairiable, try
583
584
585
586     *=====
587     * Table 5      Panel A      [reduced form model]
588     *=====
589     *
590     *-Modified regressions predicting firm liquidity levels
591     *
592     *-Goals: Firm may choose leverage, cash holdings, and investment
593     *         policy simultaneosly, which would make our estimates in
594     *         Table 4 inconsistent.
595     *
596     *-Methods: drop Leverage, investment proxies: tl capExp
597
598     global xx "tobin size cflow NetWC indsigma" //tl, capExp is dropped
599
600     *-(1)
601     xtfmb lncash $xx
602     est store m5_1
603
604     *-(2)
605     dropvars yeardummy*
606     qui tab year, gen(yeardummy)
607     reg lncash $xx yeardummy*, nocons
608     est store m5_2
609
610     *-(3)
611     dropvars sicdummy*
612     qui tab(sicmen), gen(sicdummy)

```

```

613     drop sicdummy1
614     reg   lncash $xx yeardummy* sicdummy*, nocons
615     est store m5_3
616
617     *-(4)
618     xtreg lncash $xx, be
619     est store m5_4
620
621     *-(5)
622     xtreg lncash $xx, fe
623     est store m5_5
624
625     *-----
626     *-输出结果
627     local s "using Tab05_A.csv"
628     local m "m5_1 m5_2 m5_3 m5_4 m5_5"
629     esttab `m' `s', scalar(N r2_a r2_w) replace          ///
630             star(* 0.1 ** 0.05 *** 0.01) nogap b(%6.3f) ///
631             t(%6.2f) drop(*year* sic*) order(_cons)      ///
632             mtitles(FM Year Year&industry Cross-Section FE)
633
634     *-----
635     *-Comments:
636     *
637     *-掩耳盗铃?  IV-GMM should be used. How?
638
639
640     *=====
641     *  Table 5      Panel B      [add D.cash in Table 4]
642     *=====
643     *
644     *-Note: some of the cash holdings are transitory,
645     *  because a firm might have raised funds that it is waiting to spend,
646     *  or the firm has raised funds simply because it is away from its target
647     *  holdings.
648     *  To allow for the existence of transitory cash holdings,
649     *  we add next year's change in cash holdings as an explanatory variable.
650
651     qui tsset
652     cap drop FDcash
653     gen FDcash = F.cash - cash //留待后用,所以用 F.
654     global xx "tobin size cflow NetWC capExp tl indsigma div_yes FDcash"
655
656     *-(1)
657     xtfmb lncash $xx
658     est store m5_b1
659
660     *-(2)
661     dropvars yeardummy*
662     qui tab year, gen(yeardummy)
663     reg   lncash $xx yeardummy*, nocons
664     est store m5_b2
665
666     *-(3)
667     dropvars sicdummy*
668     qui tab(sicmen), gen(sicdummy)
669     drop sicdummy1
670     reg   lncash $xx yeardummy* sicdummy*, nocons
671     est store m5_b3
672
673     *-(4)
674     xtreg lncash $xx, be
675     est store m5_b4
676
677     *-(5)
678     xtreg lncash $xx, fe
679     est store m5_b5

```

```

680 *-----
681 *-输出结果
682     local s "using Tab05_B.csv"
683     local m "m5_b1 m5_b2 m5_b3 m5_b4 m5_b5"
684     esttab `m' `s', scalar(N r2_a r2_w) replace ///
685             star(* 0.1 ** 0.05 *** 0.01) nogap b(%6.3f) ///
686             t(%6.2f) drop(*year* sic*) order(_cons) ///
687             mtitles(FM Year Year&industry Cross-Section FE)
688
689
690
691 *=====
692 *   table6           omitted: no data
693 *=====
694
695
696
697 *=====
698 * Table 7
699 *=====
700
701 *-Spending patterns based on market-to-book ratio
702 *
703
704     global xx "tobin size cflow NetWC capExp tl indsigma div_yes"
705     cap drop normal
706     cap drop excess
707     cap drop Lexcess
708     gen normal = .
709     gen excess = . // excess cash
710     forvalues t = 1999/2008{
711         qui reg cash $xx if year==`t'
712         cap drop norm_fit
713         predict norm_fit if e(sample)
714         replace normal = norm_fit if year==`t'
715         cap drop res
716         qui predict res, res
717         replace excess = res if year==`t'
718     }
719
720 *-定义 excess cash 的四个分组 (分年度进行)
721     dropvars g_ex Lexcess
722     qui tsset
723     gen Lexcess = L.excess if L.excess>0 // 只分析 L.excess>0 的样本
724     bysort year: quantiles Lexcess, gen(g_ex) n(4)
725     cap label drop g_ex_lbl
726     label define g_ex_lbl 1 "First" 2 "Second" 3 "Third" 4 "Fourth"
727     label value g_ex g_ex_lbl
728
729 *-定义 M/B 分组 Low=tobin<p25 High=tobin>p75
730     cap drop g_tobin
731     local if "if g_ex!="
732     bysort year: quantiles tobin `if', gen(g_tobin) n(4)
733     replace g_tobin=. if inlist(g_tobin, 2,3)
734     replace g_tobin= -g_tobin //为了先呈现 High M/B
735 *-定义数字-文字对应表
736     cap label drop g_tobin_lbl
737     label define g_tobin_lbl -1 "Low M/B" -4 "High M/B"
738     label values g_tobin g_tobin_lbl
739
740     label var g_ex "Quartiles of previous year excess cash"
741     label var g_tobin "Market-to-book ratio"
742
743 *-----Table 7 (Panel A)-----
744
745     global vvv "capExp" // Capital expenditures
746
747     *-t-test (bottom)

```

```

748     mat bottom_T = J(2,4,..)
749     forvalues i=1/4{
750         qui ttest $vvv if g_ex==`i', by(g_tobin)
751         mat bottom_T[1, `i'] = (r(t) \ r(p))
752     }
753     *-t-test (right)
754     mat right_T = J(4,1,..)
755     local j=-4
756     forvalues i=1(2)3{
757         qui ttest $vvv if g_tobin==`j' & (g_ex==1 | g_ex==4), by(g_ex)
758         mat right_T[`i', 1] = (r(t) \ r(p))
759         local j=`j'+3
760     }
761     *-汇总结果
762     mat T = J(6,5,..)
763     mat T[5,1] = bottom_T
764     mat T[1,5] = right_T
765     mat colnames T = First Second Third Fourth "t-value(p-value)"
766     mat rownames T = "High M/B" "-" "Low M/B" "-" "t-statistic" "(p-value)"
767     mat list T, format(%6.3f) title("$vvv")
768
769     *-basic statistics
770     format $vvv %6.4f
771     tab g_tobin g_ex, sum($vvv) nostandard
772     * see: help tabulate_summarize
773     *-----
774
775
776     *-----Table 7 (Panel B)-----
777     *-Sorry, no data available to perform this analysis
778     *-----
779
780
781     *-----Table 7 (Panel C)-----
782
783     global vvv "div_cash"    // Payment to shareholders
784
785     *-t-test (bottom)
786     mat bottom_T = J(2,4,..)
787     forvalues i=1/4{
788         qui ttest $vvv if g_ex==`i', by(g_tobin)
789         mat bottom_T[1, `i'] = (r(t) \ r(p))
790     }
791     *-t-test (right)
792     mat right_T = J(4,1,..)
793     local j=-4
794     forvalues i=1(2)3{
795         qui ttest $vvv if g_tobin==`j' & (g_ex==1 | g_ex==4), by(g_ex)
796         mat right_T[`i', 1] = (r(t) \ r(p))
797         local j=`j'+3
798     }
799     *-汇总结果
800     mat T = J(6,5,..)
801     mat T[5,1] = bottom_T
802     mat T[1,5] = right_T
803     mat colnames T = First Second Third Fourth "t-value(p-value)"
804     mat rownames T = "High M/B" "-" "Low M/B" "-" "t-statistic" "(p-value)"
805     mat list T, format(%6.3f) title("$vvv")
806
807     *-basic statistics
808     format $vvv %6.4f
809     tab g_tobin g_ex, sum($vvv) nostandard
810     *-----
811
812
813
814     *-----Table 7 (Panel D)-----
815

```



```

816     global vvv "cflow"    // Operating cash flows
817
818     *-t-test (bottom)
819     mat bottom_T = J(2,4,.)
820     forvalues i=1/4{
821         qui ttest $vvv if g_ex==`i', by(g_tobin)
822         mat bottom_T[1, `i'] = (r(t) \ r(p))
823     }
824     *-t-test (right)
825     mat right_T = J(4,1,.)
826     local j=-4
827     forvalues i=1(2)3{
828         qui ttest $vvv if g_tobin==`j' & (g_ex==1 | g_ex==4), by(g_ex)
829         mat right_T[`i', 1] = (r(t) \ r(p))
830         local j=`j'+3
831     }
832     *-汇总结果
833     mat T = J(6,5,.)
834     mat T[5,1] = bottom_T
835     mat T[1,5] = right_T
836     mat colnames T = First Second Third Fourth "t-value(p-value)"
837     mat rownames T = "High M/B" "-" "Low M/B" "-" "t-statistic" "(p-value)"
838     mat list T, format(%6.3f) title("$vvv")
839
840     *-basic statistics
841     format $vvv %6.4f
842     tab g_tobin g_ex, sum($vvv) nostandard
843     *-----
844
845
846
847
848     *=====
849     * Table 8 : Determinants of capital expenditures
850     *=====
851
852     *-Goals:
853     * add excess cash to traditional investment equations.
854     * test the precautionary propensity to save cash
855
856     dropvars posx*
857     gen posx = (excess>0)
858     tsset id year
859     gen posx_x_Lnormal = posx*L.normal
860     gen posx_x_Lexcess = posx*L.excess
861     dropvars L1_normal L2_normal L3_normal
862     gen L1_normal = L1.normal
863     gen L2_normal = L2.normal
864     gen L3_normal = L3.normal
865     dropvars L1_excess L2_excess L3_excess
866     gen L1_excess = L1.excess
867     gen L2_excess = L2.excess
868     gen L3_excess = L3.excess
869
870     #delimit ;
871     global vlist1 "capExp
872                 cflow salegr tobin L1_normal L2_normal L3_normal
873                 posx posx_x_Lnormal" ;
874     #delimit cr
875     global vlist2 "L1_excess L2_excess L3_excess posx_x_Lexcess"
876
877     *-(1)
878     reg $vlist1, robust
879     est store m_1
880
881     *-(2)
882     reg $vlist1 $vlist2, robust
883     est store m_2

```

```

884
885 *-(3)
886     xtreg $vlist1, robust fe
887     est store m_3
888
889 *-(4)
890     xtreg $vlist1 $vlist2, robust fe
891     est store m_4
892
893 *-----
894 *-右侧: Industry-adjusted results
895 preserve
896     bysort sicda: center $vlist1 $vlist2, replace //去除行业均值
897     /*
898         foreach v of varlist $vlist1 $vlist2{
899             bysort sicda: egen `v'_m = mean(`v')
900             replace `v' = `v' - `v'_m
901             drop `v'_m
902         }
903     */
904
905 *-(5)
906     reg $vlist1, robust
907     est store m_5
908
909 *-(6)
910     reg $vlist1 $vlist2, robust
911     est store m_6
912
913 *-(7)
914     areg $vlist1, absorb(id) vce(robust)
915     est store m_7
916
917 *-(8)
918     areg $vlist1 $vlist2, absorb(id) vce(robust)
919     est store m_8
920
921 restore
922 *-----
923
924     esttab m_1 m_2 m_3 m_4,          ///
925         scalar(r2 N F)  nogap        ///
926         star(* 0.1 ** 0.05 *** 0.01) ///
927         b(%6.3f) t(%6.2f)           ///
928         mtitles(OLS OLS FE FE)       ///
929         title(Raw regression results:)
930
931     esttab m_5 m_6 m_7 m_8,          ///
932         scalar(r2 N F)  nogap        ///
933         star(* 0.1 ** 0.05 *** 0.01) ///
934         b(%6.3f) t(%6.2f)           ///
935         mtitles(OLS OLS FE FE)       ///
936         title(Industry-adjusted results:)
937
938 *-输出到 Excel表格中
939     local mm "m_1 m_2 m_3 m_4 m_5 m_6 m_7 m_8"
940     esttab `mm' using Tab08.csv, scalar(r2 N F) nogap ///
941         star(* 0.1 ** 0.05 *** 0.01)           ///
942         b(%6.3f) t(%6.2f) compress replace     ///
943         mtitles(OLS OLS FE FE OLS_indadj OLS_indadj FE_indadj FE_indadj)
944
945 *-Comments:
946 *
947 *   see Opler(1999, pp.35)
948 *   (1) greater excess cash leads firms to invest more,
949 *       whether they have good investment opportunities or not.
950 *   (2) Overall, the results suggest that the propensity to spend positive excess
951     cash

```

```

951 *      is small.
952 * (3) Table 8 also provides no evidence to support the view that it takes time
953 *      for excess cash to affect investment,
954 *      since most of the effect seems to take place within one year.
955 *      That is, only L1_excess is significant
956 * (4) 为何 Col.1-4 的结果与 Col.5-8 的如此相似?
957
958
959
960
961 *=====
962 * Table 9
963 *=====
964
965 // 我认为应该允许公司的 excess 为负
966 dropvars g_ex
967 bysort year: quantiles excess, gen(g_ex) n(4)
968 cap label define g_ex_lbl 1 "Q1" 2 "Q2" 3 "Q3" 4 "Q4"
969 label value g_ex g_ex_lbl
970 label list g_ex_lbl
971
972 dropvars Year*
973 qui tsset
974 by id: gen Year0 = g_ex      if _n<=_N // Year 0
975 by id: gen Year1 = g_ex[_n+1] if _n<_N // Year 1
976 by id: gen Year2 = g_ex[_n+2] if _n<_N // Year 2
977 by id: gen Year3 = g_ex[_n+3] if _n<_N // Year 3
978 by id: gen Year4 = g_ex[_n+4] if _n<_N // Year 4
979 by id: gen Year5 = g_ex[_n+5] if _n<_N // Year 5
980
981 order id year g_ex Year*
982 browse id year g_ex Year*
983
984 *-----
985 *-Panel A: highest quartile of excess cash in year 0
986 *-----
987 mat Num = J(6,4,.)
988 forvalues t = 0/5{
989     forvalues q = 4(-1)1{
990         qui count if (Year`t'==`q' & Year0==4)
991         local r = `t'+1
992         local c = -`q'+5
993         mat Num[`r', `c'] = r(N)
994     }
995 }
996 global cns "Q4 Q3 Q2 Q1"
997 global rns "Year0 Year1 Year2 Year3 Year4 Year5"
998 mat rownames Num = $rns
999 mat colnames Num = $cns
1000 *-呈现公司状态转移的【数目】
1001 mat list Num, title(Distribution of Firms (Numbers))
1002
1003 mat det = J(4,1,1)
1004 mat list Num
1005 mat list det
1006 mat b = Num*det
1007 mat list b
1008 mat inv_b = vecdiag(inv(diag(b)))' //
1009 mat list inv_b
1010 mat one = J(1,4,1)
1011 mat inv_b = inv_b#one
1012 mat list inv_b
1013 mat prob = hadamard(Num, inv_b)*100 //直乘
1014 mat rownames prob = $rns
1015 mat colnames prob = $cns
1016 *-呈现公司状态转移的【概率】
1017 mat list prob, format(%6.1f) title(Transition Probabilities)
1018

```

```

1019  *--合并上述两个矩阵
1020      mat s = J(2,1,1)
1021      mat PA = prob#s
1022      mat list PA
1023      forvalues i = 2(2)12{
1024          local j = `i'/2
1025          mat PA[`i',1] = Num[`j', ....]
1026      }
1027      mat rownames PA = $rns
1028      mat colnames PA = $cns
1029      mat list PA, format(%6.1f) title(Table 9, Panel A)
1030
1031
1032  *--进一步美化
1033      dropvars artile1-artile4 uartile4-uartile1 Quartile4-Quartile1 Y_Year
1034      svmat PA, names(artile)
1035      renvars artile1-artile4 / uartile4 uartile3 uartile2 uartile1
1036      foreach v of varlist uartile4-uartile1{
1037          qui gen Q`v' = string(`v')
1038      }
1039
1040      gen Y_Year = " " in 1/12
1041      forvalues j=2(2)12{
1042          local jj = `j'-1
1043          qui replace Y_Year = "Year `='j'/2-1'" in `jj'
1044          foreach v of varlist Quartile4-Quartile1{
1045              qui replace `v' = substr(`v',1,5) in `jj'
1046              qui replace `v' = "["+"`v'+"]" in `j'
1047          }
1048      }
1049
1050      logout, save(Table09_A) excel replace:      ///
1051      list Y_Year Quartile4-Quartile1 in 1/12, ///
1052      clean noobs abbreviate(10)
1053
1054
1055  *-----
1056  *--Panel B: Lowest quartile of excess cash in year 0
1057  *-----
1058      mat Num_B = J(6,4,.)
1059      forvalues t = 0/5{
1060          forvalues q = 1(1)4{
1061              qui count if (Year`t'==`q' & Year0==1)
1062              local r = `t'+1
1063              local c = -`q'+5
1064              mat Num_B[`r', `c'] = r(N)
1065          }
1066      }
1067      global cns "Q4 Q3 Q2 Q1"
1068      global rns "Year0 Year1 Year2 Year3 Year4 Year5"
1069      mat rownames Num_B = $rns
1070      mat colnames Num_B = $cns
1071      *--呈现公司状态转移的【数目】
1072      mat list Num_B, title(Distribution of Firms (Num_Bbers))
1073
1074      mat det = J(4,1,1)
1075      mat b = Num_B*det
1076      mat list b
1077      mat inv_b = vecdiag(inv(diag(b)))'  //
1078      mat list inv_b
1079      mat one = J(1,4,1)
1080      mat inv_b = inv_b#one
1081      mat list inv_b
1082      mat prob_B = hadamard(Num_B, inv_b)*100
1083      mat rownames prob_B = $rns
1084      mat colnames prob_B = $cns
1085      *--呈现公司状态转移的【概率】
1086      mat list prob_B, format(%6.1f) title(Transition probabilities)

```

```

1087
1088     *-合并上述两个矩阵
1089     mat s = J(2,1,1)
1090     mat PB = prob_B#s
1091     mat list PB
1092     forvalues i = 2(2)12{
1093         local j = `i'/2
1094         mat PB[`i',1] = Num_B[`j', ....]
1095     }
1096     mat rownames PB = $rns
1097     mat colnames PB = $cns
1098     mat list PB, format(%6.1f) title(Table 9, Panel B)
1099
1100     *-进一步美化
1101     dropvars artile1-artile4 uartile4-uartile1 Quartile4-Quartile1 Y_Year
1102     svmat PB, names(artile)
1103     renvars artile1-artile4 / uartile4 uartile3 uartile2 uartile1
1104     foreach v of varlist uartile4-uartile1{
1105         qui gen Q`v' = string(`v')
1106     }
1107
1108     gen Y_Year = " " in 1/12
1109     forvalues j=2(2)12{
1110         local jj = `j'-1
1111         qui replace Y_Year = "Year `='j'/2-1'" in `jj'
1112         foreach v of varlist Quartile4-Quartile1{
1113             qui replace `v' = substr(`v',1,5) in `jj'
1114             qui replace `v' = "["+"`v'+" in `j'
1115         }
1116     }
1117
1118     logout, save(Table09_B) excel replace:    ///
1119     list Y_Year Quartile4-Quartile1 in 1/12, ///
1120         clean noobs abbreviate(10)
1121
1122
1123
1124     *=====
1125     * Table 10&11
1126     *=====
1127
1128     *-do it by your self (Ref to Table 9)
1129
1130

```

```

1
2  *      -----
3  *      -----
4  *
5  *      Stata 学术论文专题
6  *
7  *      -----
8  *      -----
9
10
11  *
12  *      主讲人：连玉君 副教授
13  *
14  *
15  *      单 位：中山大学岭南学院金融系
16  *      电 邮：arlionn@163.com
17  *      博 客：http://blog.cnfol.com/arlionn
18  *      主 页：http://goo.gl/tRXba
19  *      微 博：http://weibo.com/arlionn
20
21
22  *-----
23  * 叶德珠,连玉君,黄有光,李东辉.
24  *   "消费文化、认知偏差与消费行为偏差".
25  *   经济研究, 2012(2): 80-92.
26  *-----
27
28  *-原始论文
29  shellout "`c(sysdir_personal)'PX_papers\Refs\连玉君_2012_消费文化.pdf"
30
31  cd "`c(sysdir_personal)'PX_papers\Ye_Lian_JJYJ2012" // 进入课程目录
32  *cd "D:\stata11\ado\personal\PX_papers\Ye_Lian_JJYJ2012"
33
34
35  *-----
36  *-数据处理
37  *-----
38
39  use T_consum_sex_data.dta, clear // 调入数据
40
41  order id year
42  tsset id year // 声明为面板
43
44  *-----
45  *-变量标签 (多余的变量我未删除)
46
47  label var id "国家名称"
48  label var domsaving "国内储蓄率"
49  label var consume "最终消费率"
50  label var netsaving "净国民储蓄占gdp比例"
51  label var srate "一年期储蓄利率"
52  label var gdpper "人均收入"
53  label var ruralpop "郊区人口"
54  label var edu "大学入学率"
55  label var infla "通胀率"
56  label var ss "社会保障"
57  label var ss2 " "
58  label var insur8 "保险"
59  label var fd1 "金融发展"
60  label var sex1 "性生活指数"
61  label var sex2 "最早性生活的年龄"
62  label var sex3 "进行性教育的年龄"
63  label var sex4 "进行不安全性行为的比例"
64  label var sex5 "每年性生活次数"
65  label var rategap "长短期利率差"
66  label var tour2 "旅游支出占比"
67  label var tour3 " "
68  label var tour1 " "

```

```

69     label var      tour4
70     rename sex1    sex_index    // 部分变量更名
71     rename sex2    sex_age
72     rename sex3    sex_EDU_age
73     rename sex4    sex_unsafe
74     rename sex5    sex_feq
75     label var      bankloan     "银行部门提供的国内信贷除以GDP"
76     label var      privateloan  "对私营部门提供的国内信贷除以GDP"
77     label var      ss_new       "社会保障支出占GDP比例"
78     label var      old_dep      "老年人负担系数"
79     label var      young_dep    "儿童负担系数"
80
81
82     *-改变部分变量的单位, 以便估计系数显示在正常范围内(0.00-10.00)
83
84     local w1 "domsaving netsaving srate edu infla ss ss2"
85     local w2 "insur8 fd1 sex_unsafe ratemap old young"
86     foreach v of varlist `w1' `w2' {
87         replace `v' = `v'/100
88     }
89     replace bankloan = bankloan/10000
90     replace privateloan = privateloan/10000
91
92     clonevar sexfeq0 = sex_feq    // 备份一个变量, 绘图时使用 (see Figure 1)
93
94     replace sex_feq = sex_feq/360
95     label var sex_feq "每年性生活次数/360"
96
97     *-Note: 对变量执行放大或缩小处理, 并不影响统计推断的结论,
98     *        只是影响系数估计值的大小而已.
99
100
101     *-儒家文化虚拟变量
102
103     labelbook id
104     gen rujia = inlist(id,7,17,21,25,26,28,34,39,44,48)
105     *-中国、香港、印度尼西亚、日本、韩国、马来西亚、菲律宾、新加坡、泰国和越南
106
107
108     *-处理离群值, winsor 处理
109
110     *-原始变量的分布特征
111     des srate infla ss
112     sum srate infla ss
113
114     local vv "srate infla ss*"    // 需要 winsor 的变量都可以加入此处
115     foreach v of varlist `vv' {
116         local a: var label `v'
117         if ("`v'"=="infla"|"`v'"=="srate"){
118             winsor `v', p(0.02) gen(`v'_x)
119         }
120         else{
121             winsor `v', p(0.01) gen(`v'_x)
122         }
123         drop `v'
124         rename `v'_x `v'
125         label var `v' "`a'"
126     }
127
128     *-winsor 前的分布特征
129     des bankloan privateloan
130     sum bankloan privateloan
131     histogram bankloan
132     *-winsor 处理
133     winsor bankloan, gen(bank_w) p(0.05) high
134     winsor privateloan, gen(private_w) p(0.05) high
135     *-winsor 后的分布特征
136     sum bank_w private_w

```



```

137     histogram bank_w
138
139     *-新变量
140     gen real_i0 = srate - infla           // 用实质利率会更好一些
141     winsor real_i0, gen(real_i) p(0.01)
142     gen depend = young_dep + old_dep      // 二者之间相关系数过高，所以合并起来
143
144     *-处理部分缺漏值
145     bysort id: egen av_insor8 = mean(insur8) // 缺漏值用每个国家的样本均值代替
146     replace insur8 = av_insor8 if insur8==.
147
148
149     *=====
150     *-Table 1 : 基本统计量
151     *=====
152
153     global x "gdpper edu real_i rategap ss depend bank_w"
154     reg consume $x
155     ereturn list           // 返回值
156     gen yes = e(sample)    // 只统计参与回归分析的观察值
157     global v1 "netsaving domsaving consume"
158     global v2 "gdpper edu real_i rategap ss depend bank_w"
159     global v3 "rujia sex_feq sex_unsafe sex_age"
160     global v $v1 $v2 $v3
161
162     *-输出到 Excel
163     logout, save(Results/Var_statistic) excel replace: ///
164         tabstat $v if yes==1,           ///
165         s(mean p50 sd min max N)       ///
166         format(%6.3f) column(statis)
167
168
169     *-----
170     *-变量的相关系数矩阵 (not reported in the paper)
171     *-----
172
173     format $v %6.2f
174     logout, save(Results/Var_corr) excel replace: ///
175         pwcorr $v if yes, star(.05)
176
177     *-----
178     *-基本统计分析 (not reported in the paper)
179     *-----
180
181     *-分别统计rujia和非rujia地区的消费率，性生活指数数据，做基本的统计分析
182
183     *-分儒家和非儒家两组统计
184     tabstat sex* *sav* consume, by(rujia) f(%4.2f)
185
186     *-分年度统计
187     foreach t of numlist 1978 1980(5)2005 2007{
188         display in w _n _n "Year: " `t'
189         tabstat sex* *sav* if year==`t', ///
190             by(rujia) format(%4.1f) nottotal
191     }
192
193
194
195     *=====
196     *-Table 2 : 推论1检验结果：自我控制认知偏差对最终消费率的影响
197     *=====
198
199     dropvars yr* fix*
200
201     *-国家个体效应虚拟变量
202     tab id, gen(fix)
203     drop fix1
204

```

```

205  *-年度效应虚拟变量
206      tab year, gen(yr)
207      drop yr1
208
209  *-设定被解释变量和解释变量
210      global y  "consume"      // 被解释变量—最终消费率 (文中使用了这个)
211      global x  "gdpper real_i ratgap ss depend bank_w"
212
213  *-column (1)
214      reg $y $x
215      est store c1              // OLS
216
217  *-column (2)
218      reg $y $x fix*
219      est store c2              // FE
220      ereturn list              // 呈现返回值, 用于随后输出结果
221  *-检验个体效应              (p.86, 脚注 1)
222      lrtest c1 c2
223      estadd scalar FeChi2 = r(chi2)
224      estadd scalar FeChi2p= r(p)
225      ereturn list              // 呈现返回值
226
227  *-column (3)
228      reg $y $x fix* yr*
229      est store c3              // FE+Year
230  *-检验年度效应              (p.86, 脚注 2)
231      lrtest c2 c3
232      estadd scalar YrChi2 = r(chi2)
233      estadd scalar YrChi2p= r(p)
234
235  *-column (4)
236      global z "rujia"
237      reg $y $x $z
238      est store c4              // OLS+rujia
239
240  *-column (5)
241      global z "sex_feq"
242      reg $y $x $z
243      est store c5              // OLS+sex_feq
244
245  *-用第 (5) 栏的样本重新估计第 (4) 栏的模型 (p.87, 脚注 2)
246      reg $y $x rujia if e(sample)
247      dis "adj-R2 = " e(r2_a)
248
249  *-column (6)
250      global z "sex_unsafe"
251      reg $y $x $z
252      est store c6
253
254  *-column (7)
255      global z "sex_age"
256      reg $y $x $z
257      est store c7
258
259  *-column (8)
260      global z "rujia sex_feq"
261      reg $y $x $z
262      est store c8
263
264
265  *-呈现结果
266      global mm "c1 c2 c3 c4 c5 c6 c7 c8"
267      dis "Dependent = " in y "$y"
268      esttab $mm, b(%6.3f) star(* 0.1 ** 0.05 *** 0.01) ///
269      scalar(r2 r2_a F N FeChi2 FeChi2p YrChi2 YrChi2p) ///
270      compress nogaps          ///
271      order($x) drop(*fix* yr*) ///
272      mtitle(OLS FE FE+Year OLS OLS OLS OLS OLS)

```

```

273
274     *-输出到 Excel
275     esttab $mm using Results/Table02_$y.csv, replace    ///
276         b(%6.3f) star(* 0.1 ** 0.05 *** 0.01)    ///
277         scalar(r2 r2_a F N FeChi2 FeChi2p YrChi2 YrChi2p)    ///
278         nogaps order($x) drop(*fix* yr*)    ///
279         mtitle(OLS FE FE+Year OLS OLS OLS OLS OLS)
280
281
282     *-----
283     *-稳健性检验      (not reported in the final version of the paper)
284     *-----
285
286     *-分别采用 domsaving 和 netsaving 做为被解释变量
287
288     *-----
289     *-domsaving
290
291     global y  "domsaving"  // 被解释变量 1 国内储蓄率
292     *global x  "gdpper real_i rategap ss depend bank_w"
293
294     reg $y $x
295     est store rd0
296
297     reg $y $x fix*
298     est store rd1
299
300     global z "rujia"
301     reg $y $x $z
302     est store rd2
303
304     global z "sex_feq"
305     reg $y $x $z
306     est store rd3
307
308     *-----
309     *-netsaving
310
311     global y  "netsaving"  // 被解释变量2 净储蓄率
312
313     reg $y $x
314     est store rn0
315
316     reg $y $x fix*
317     est store rn1
318
319     global z "rujia"
320     reg $y $x $z
321     est store rn2
322
323     global z "sex_feq"
324     reg $y $x $z
325     est store rn3
326
327     *-呈现结果
328     global mm "rd0 rd1 rd2 rd3  rn0 rn1 rn2 rn3"
329     esttab $mm, b(%6.3f) star(* 0.1 ** 0.05 *** 0.01)    ///
330         scalar(N r2_a) compress nogaps    ///
331         order($x) drop(*fix*)    ///
332         mtitle(OLS FE domsave OLS OLS FE netsave OLS)
333
334     *-输出到 Excel
335     esttab $mm using Results/Table02_Robust.csv, replace    ///
336         b(%6.3f) star(* 0.1 ** 0.05 *** 0.01)    ///
337         scalar(N r2_a) nogaps    ///
338         order($x) drop(*fix*)    ///
339         mtitle(OLS FE domsave OLS OLS FE netsave OLS)
340

```

```

341
342
343 *=====
344 *-Table 3 : 推论2的检验结果：异常消费与理性消费者占比
345 *=====
346
347     global y  "domsaving"  // 被解释变量1 国内储蓄率
348     * global y  "netsaving"  // 被解释变量2 净储蓄率
349     * global y  "consume"    // 被解释变量3 最终消费率
350
351 *-----
352 *-第一步：计算离差（异常消费和异常性行为）
353 *-----
354
355     *-y 的离差：全样本 OLS 回归法
356
357     global x  "gdpper real_i ss depend bank_w"
358     dropvars Ex_$y Exdum abs_Ex_$y
359     foreach v of varlist consume domsaving netsaving{
360         qui reg `v' $x
361         qui predict Ex_`v' if e(sample), res
362         qui gen abs_Ex_`v' = abs(Ex_`v') // Ex_$y 的绝对值
363     }
364
365     gen Exdum = (Ex_$y>0)
366     replace Exdum=. if Ex_$y==.
367
368     tabstat sex* *sav* consume, by(Exdum) f(%4.2f)
369     tabstat sex* *sav* consume, by(rujia) f(%4.2f)
370
371
372     *-sex_index 的离差
373
374     dropvars av_sex* Ex_sex*
375     foreach v of varlist sex_feq sex_age sex_unsafe{
376         bysort year: egen av_`v' = mean(`v')
377         gen Ex_`v' = abs(`v' - av_`v')
378     }
379
380
381 *-----
382 *-第二步：回归分析（表 3）
383 *-----
384
385     dropvars Ex_sexfeq_x_edu
386     gen Ex_sexfeq_x_edu = Ex_sex_feq*edu
387
388     local y "consume"
389     reg abs_Ex_`y' Ex_sex_feq
390     est store a1
391     reg abs_Ex_`y' edu
392     est store a2
393     reg abs_Ex_`y' edu Ex_sex_feq
394     est store a3
395
396     local y "domsaving"
397     reg abs_Ex_`y' Ex_sex_feq
398     est store b1
399     reg abs_Ex_`y' edu
400     est store b2
401     reg abs_Ex_`y' edu Ex_sex_feq
402     est store b3
403
404     local y "netsaving"
405     reg abs_Ex_`y' Ex_sex_feq
406     est store c1
407     reg abs_Ex_`y' edu
408     est store c2

```

```

409     reg abs_Ex_`y' edu Ex_sex_feq
410     est store c3
411
412     *-呈现结果
413
414     *-Panel A
415     local mm "a1 a2 a3"
416     esttab `mm', mtitle("Ex_Consume") b(%6.3f) nogaps ///
417         star(* 0.1 ** 0.05 *** 0.01) scalar(r2 r2_a N)
418
419     *-Panel B
420     local mm "b1 b2 b3"
421     esttab `mm', mtitle("Ex_Domsaving") b(%6.3f) nogaps ///
422         star(* 0.1 ** 0.05 *** 0.01) scalar(r2 r2_a N)
423
424     *-Panel C
425     local mm "c1 c2 c3"
426     esttab `mm', mtitle("Ex_Netsaving") b(%6.3f) nogaps ///
427         star(* 0.1 ** 0.05 *** 0.01) scalar(r2 r2_a N)
428
429
430     *-输出到 Excel
431     local mm "a1 a2 a3"
432     esttab `mm' using Results/Table03_Ex.csv, ///
433         star(* 0.1 ** 0.05 *** 0.01) b(%6.3f) ///
434         scalar(N r2_a) wide nogaps replace
435
436     local mm "b1 b2 b3"
437     esttab `mm' using Results/Table03_Ex.csv, ///
438         star(* 0.1 ** 0.05 *** 0.01) b(%6.3f) ///
439         scalar(N r2_a) wide nogaps append // 追加结果到上一张表中
440
441     local mm "c1 c2 c3"
442     esttab `mm' using Results/Table03_Ex.csv, ///
443         star(* 0.1 ** 0.05 *** 0.01) b(%6.3f) ///
444         scalar(N r2_a) wide nogaps append // 追加结果到上一张表中
445
446
447     *=====
448     *-Figure 00 Discounting function Eq.(1) (not reported)
449     *=====
450
451     *-----Fig.00-----begin-----
452     preserve
453     clear
454     set obs 10
455     local delta = 0.8
456     gen t = _n - 1
457     gen dt = `delta'^t
458     gen sum = sum(dt)
459     local b = 0.4 // beta<1
460     gen yn1 = 1
461     replace yn1 = `b'*`delta'^t if t>0
462     local b = 1 // beta=1
463     gen ye1 = 1
464     replace ye1 = `b'*`delta'^t if t>0
465     local b = 1.5 // beta>1
466     gen yg1 = 1
467     replace yg1 = `b'*`delta'^t if t>0
468     #delimit ;
469     twoway (line yn1 t,lw(*1.9))
470         (line ye1 t,lw(*1.9))
471         (line yg1 t,lw(*1.9))
472         (dropline yg1 t if t<9, lstyle(grid) msymbol(none))
473     ,
474     scheme(slmono)
475     ytitle("贴" "现" "函" "数", place(top))
476     orientation(horizontal) size(*1.3)

```

```

477         xtitle("{it:t}", size(*1.5) place(right))
478         ylabel(1, angle(0) labsize(*1.3))
479         xlabel(1(1)8, labsize(*1.3))
480         yscale(range(0 1.4) titlegap(-4))
481         xscale(titlegap(-4) outergap(-3))
482         plotregion(style(none) margin(0))
483         legend(off)
484         text(0.99 2.5 "{it:{&beta}} >1", size(*1.3))
485         text(0.68 2.5 "{it:{&beta}} =1", size(*1.3))
486         text(0.31 2.5 "{it:{&beta}} <1", size(*1.3))
487     ;
488     #delimit cr
489     graph export Results/Fig00_discount.wmf, replace // 保存图片
490     restore
491     *-----Fig.00-----over-----
492
493
494     *=====
495     *-Figure 01 图1 消费偏差与每年性生活次数的散点图（1998-2007）
496     *=====
497
498     *-部分国家名称简写
499     global y "consume"
500     label define id ///
501         2 "AU" /// // Austalia
502         3 "AT" /// // Austia
503         4 "BE" /// // Belgium
504         7 "China" ///
505         10 "Crech" ///
506         11 "DK" /// // Denmark
507         13 "FI" /// // Finland
508         15 "DE" /// // Germany
509         40 "ZA" /// // South Africa
510         41 "ES" /// // Spain
511         46 "UK" /// // United Kingdom
512         47 "USA", modify
513
514     dropvars pos
515     gen pos = 3
516     replace pos = 2 if inlist(id,2,21,39,46,47)
517     replace pos = 4 if inlist(id,6,13,41,45)
518     replace pos = 6 if inlist(id,35,44)
519     replace pos = 9 if inlist(id,15,16,23,24,48)
520     replace pos = 10 if inlist(id,11,19,36)
521     replace pos = 12 if inlist(id,40)
522
523     dropvars av_Exy av_sexfeq
524     bysort id: egen av_Exy = mean(Ex_$y)
525     bysort id: egen av_sexfeq = mean(sexfeq0)
526     #delimit ;
527     twoway (scatter av_Exy av_sexfeq
528         if rujia==1,
529             mcolor(black) msymbol(0) mlabel(id) mlabvp(pos)
530         )
531         (scatter av_Exy av_sexfeq
532             if rujia==0&Ex_$y<=0.2,
533             mcolor(black*0.6) msymbol(Th) mlabel(id) mlabvp(pos) mlabsz(size(small))
534         )
535     ,
536     ytitle("与正常消费比率的偏离", size(*0.9))
537     yscale(titlegap(1) outergap(0))
538     ylabel(, angle(0) format(%3.1f))
539     xtitle("每年性生活次数", size(*0.9))
540     xscale(titlegap(3) outergap(-2))
541     yline(0, lp(dash) lc(black*0.4))
542     legend(label(1 "儒家") label(2 "非儒家") size(*0.8) pos(5) ring(0))
543     scheme(slmono) ;
544     #delimit cr

```

```
545      graph export Results/Fig02_Ex_sex.wmf, replace // 保存图片
546
547
548      *-----Over-----
549
550
551      *-----
552      *-附录：外审意见及修改说明
553      *-----
554
555      shellout "外审意见及修改说明.pdf"
556
557
558      *=====
559      *-2011-06-06 更新说明
560
561      *-增加了几个新变量
562      *-重新设定 Normal Consumption 模型
563      *-重新估计 Excess Consumption 的决定因素
564
565      *-2011-06-23 更新说明
566      *-将 edu 从 Normal Regression 中去掉
567      *-残差的绝对值衡量异常消费
568
569      *-2011-12-13
570      *-作为稳健性检验，可以采用旅游支出作为过度消费的替代指标。
571      *=====
572
573
```



```

1
2
3  *=====
4  *=====
5  *
6  *   Appendix: Introduction to Logit Models
7  *
8  *=====
9  *=====
10
11 *-本专题中，如下论文使用了该方法
12
13 *-----
14 * -Chang&Wong,2009,JCF -
15 *   Chang, E. C., Wong, S. M. L., 2009.
16 *   Governance with multiple objectives:
17 *   Evidence from top executive turnover in china.
18 *   Journal of Corporate Finance, 15 (2): 230-244.
19 *-----
20 *-原始论文
21   shellout "`c(sysdir_personal)'PX_papers\Refs\Chang_2009_JCF.pdf"
22
23 *-----
24 * -PSM-
25 *   Lian, Y., Z. Su, Y. D. Gu, 2011,
26 *   Evaluating the Effects of Equity Incentives Using PSM:
27 *   Evidence from China,
28 *   Frontiers of Business Research in China, 5(2):266-290.
29 *-----
30   shellout "`c(sysdir_personal)'PX_papers\Refs\Lian_2011_PSM.pdf"
31
32
33 *-参考资料
34   global path "`c(sysdir_personal)'PX_papers" //课程总文件夹
35   shellout "$path\Refs\Lian_15_Logit.pdf" // 我的讲义
36   shellout "$path\Refs\Long_2001_Logit.pdf" // 极力推荐
37
38
39 *-本讲路径
40   global path "`c(sysdir_personal)'PX_papers"
41   cd "$path\z01_Logit"
42
43
44 *-----
45 *-> 1.1 离散选择模型简介
46 *-----
47
48 *-数据特征:
49
50   *-{0/1} 变量: Logit, Probit
51
52   *-e.g. 是否参与扶贫计划; 是否获得政府资助; 是否安装减排装置
53   *       是否读博士; 是否外出务工; 是否违约等
54
55   *-类别变量: Multinomial Logit (MLogit)
56
57   *-e.g. 融资方式的选择; 职业的选择;
58
59   *-序别变量: Ordered Logit (OLogit)
60
61   *-e.g. 选举倾向; 幸福指数; 顾客满意度; 融资序位;
62   *       专利数量; 论文发表数量; 购房意向;
63   *       CEO 离职次数; 会计舞弊次数; 子女个数;
64
65   *-多层嵌套选择: Nested Logit (NLogit)
66
67   *-e.g. 交通工具的选择; 交叉上市; 多元化经营; 融资方式的选择;
68   *       外出务工的选择行为; 移民行为;

```

```

69
70     *-低值聚集特征:
71
72     *- Poisson Model (Poisson)
73     *- Zero-inflated Poisson regression (zip)
74     *- Negative binomial regression (nbreg)
75     *- Zero-inflated negative binomial regression (zinb)
76
77
78
79 *-----
80 *-> 1.2 二元 logit 模型
81 *-----
82
83
84     *      { 1  if 进口车
85     *  y_i = {
86     *      { 0  if 国产车
87
88     *  假设 y_i 是随机变量 Y_i 的观察值,
89     *  Pr(Y_i=1)= p_i, Pr(Y_i=0) = 1-p_i
90
91     *  E(Y_i) = 1*p + 0*(1-p) = p_i    // 期望值取值有限 [0,1]
92     *  Var(Y_i) = p_i*(1-p_i)          // 异方差
93
94     *-Modeling idea:
95     *
96     *  E(Y_i) = p_i = x1*b1 + x2*b2 + ... = X*b
97
98
99 *-----
100 *-1.2.1 线性概率模型
101
102     sysuse auto, clear
103     reg foreign price wei len
104     predict y_hat
105     histogram y_hat    // 直方图
106     kdensity y_hat    // 密度函数图
107     kdensity y_hat, normal
108
109     *-Limits?
110
111     *-Solution: G(p_i) << [-oo, +oo]
112
113
114 *-----
115 *-1.2.2 Logit 变换
116
117 *-----
118     *-odds 变换      p/(1-p): 胜算比
119
120     local a = 0.6          // try: 0.4, 0.5, 0.6,
121     local odd_a = `a'/(1-`a')
122     dis `odd_a'
123
124     *-Note: odds 变换可以使      [0,1) --> [0,+oo)
125
126 *-----
127     *-Logit 变换      logit(p) = log(odds) = log[p/(1-p)]
128     local a = 0.6          // try: 0.4, 0.5, 0.6
129     local odd_a = `a'/(1-`a')
130     local ln_odd = ln(`odd_a')
131     dis `ln_odd'
132
133     *-Note: ln(odds) 变换可以使 [0,1) --> [-oo,+oo)
134
135
136 *-----

```

```

137      *-图示
138
139      clear
140      set obs 10000
141      egen p = seq(), from(0)
142      replace p = p/10000
143      gen odd_p = p/(1-p)
144      gen ln_odd_p = ln(odd_p)
145
146      scatter p      odd_p      // [0,1) --> [0,+oo)
147      scatter p ln_odd_p      // [0,1) --> [-oo,+oo)
148
149      histogram ln_odd_p, normal
150      sum ln_odd_p, detail
151
152
153      *-----
154      *-Implication:
155
156      * ln_odd_p ~ N(0, sigma^2)
157
158      * logit(p_i) = x_i*b
159
160      *-经过 Logit 变换, 若将 logit(p_i) 视为被解释变量, 则可以转换为线性模型
161
162      *      p_i
163      * ln ----- = xb ==> p_i = ----- = E(Y_i)
164      *      1-p_i      1+exp(xb)
165
166
167      *-----
168      *-1.2.3 估计方法: MLE
169
170      *-似然函数: 基于二项分布可以写出
171
172      sysuse auto, clear
173      logit foreign wei mpg
174
175      *-weight 的系数含义
176      *
177      * -1- 重量越大, 越不可能是进口车
178      *
179      * -2- weight 增加一个单位, logit(p_i) 增加 -0.0039 个单位
180      *
181      * -3- weight 增加一个单位, odds 增加 exp(-0.0039) 个单位
182      *      因为 b[weight] = D.logit(p_i) = D.ln(odds)
183      *      所以 D.odds = exp(b[weight])
184
185      logit foreign wei mpg
186      dis exp(-.00390672)
187
188      logit foreign wei mpg, or // odds ratio
189      dis ln(.9961009 )
190
191
192      *-进一步的解释
193      *      p_i
194      * ln ----- = xb ==> p_i = ----- = E(Y_i)
195      *      1-p_i      1+exp(xb)
196
197      sum foreign
198      reg foreign // 常数项的含义?
199      logit foreign
200      dis exp(-.8602013)/(1+exp(-.8602013))
201      *-或
202      dis exp(_b[_cons])/(1+exp(_b[_cons]))
203
204      *-系数的呈现方式

```

```

205         logit foreign wei mpg, nolog
206         listcoef, help
207         listcoef, help percent
208
209
210     *-----
211     *-1.2.4 假设检验
212
213         *-Wald test
214         test weight mpg
215
216         *-LR test
217         qui logit foreign
218         est store logit0
219         qui logit foreign wei mpg
220         est store logitfull
221         lrtest logitfull logit0
222
223
224     *-----
225     *-1.2.5 拟合优度
226
227         logit foreign wei mpg price
228         fitstat
229
230         *-模型的比较
231         qui logit foreign wei mpg
232         qui fitstat, saving(m1)
233         qui logit foreign wei mpg price
234         fitstat, using(m1)
235
236
237     *-----
238     *-1.2.6 拟合值和预测概率
239
240         *-----
241         *           p_i                exp(xb)
242         *   ln ----- = xb ==>  p_i = ----- = E(Y_i)
243         *           1-p_i              1+exp(b)
244         *-----
245         *-将 Logit 模型的估计值 b_hat 带入上式即可得到概率值
246
247         sysuse auto, clear
248         logit foreign wei mpg price
249         predict pr
250         sum pr
251         label var pr "Logit: Pr(foreign)"
252         dotplot pr, ylabel(0(.2)1)
253
254         gen Yes = pr>0.5
255         gen str1 Error = ""
256         replace Error="*" if (foreign==0&pr>0.5)|(foreign==1&pr<0.5)
257         sort Error
258         browse foreign pr Yes Error
259
260         *-样本外预测
261         sysuse auto, clear
262         gen u = uniform()
263         gen S = (u<=0.6)
264         logit foreign wei mpg price if S==1 // 随机抽取 60% 的样本
265         predict pr // 利用上面的参数进行全样本预测
266
267         *-误判情况
268         gen Yes = pr>0.5
269         gen str1 Error = ""
270         replace Error="*" if (foreign==0&pr>0.5)|(foreign==1&pr<0.5)
271         sort Error
272         browse foreign pr Yes Error if S==1 // 样本内误判

```

```

273      browse foreign pr Yes Error if S==0 // 样本外误判
274
275
276
277  *-----
278  *-> 1.3 多元 logit 模型      mlogit
279  *-----
280
281  *-----
282  *-1.3.1 简介
283
284      *- Y = 1,2,3,4... 但没有大小关系(类别变量)
285
286      *- 建模思路: 两两配对, 形成多个二元 Logit 模型; 需要施加约束条件
287
288  *-----
289  *-1.3.2 估计
290
291      use nomocc2.dta, clear
292      *label define oc 1 "佣人" 2 "蓝领" 3 "手工艺者" 4 "白领" 5 "专业人士"
293      *label value occ oc
294      des /*变量的基本信息*/
295      sum /*变量的基本统计量*/
296
297      mlogit occ ed exper white
298
299  *-与二元 Logit 结果对比      1:5
300      gen occ2 = occ
301      replace occ2 = . if occ2>=2&occ2<=4 // 去掉 2,3,4
302      replace occ2 = 0 if occ2==5
303      tab occ2, miss
304      logit occ2 ed exper white, nolog
305
306  *-基准组的设定
307      mlogit occ ed exper white, base(2) nolog
308
309  *-结果的解释
310      qui mlogit occ ed exper white
311      listcoef white
312      listcoef, pvalue(0.05) // 仅列出显著的配对
313
314  *-拟合优度
315      qui mlogit occ ed exper white
316      fitstat
317
318
319  *-----
320  *-1.3.3 假设检验
321
322  *-----
323  *-1.3.3.1 检验系数的显著性
324
325      *-LR test
326      *-单变量
327      qui mlogit occ ed exper white
328      est store mFull
329      qui mlogit occ ed exper
330      est store m0
331      lrtest mFull m0
332      *-多变量
333      qui mlogit occ ed exper white
334      est store mFull
335      qui mlogit occ ed
336      est store m0
337      lrtest mFull m0
338
339      *-Wald test
340      qui mlogit occ ed exper white

```

```
341         test white
342
343     *-mlogtest 命令
344     qui mlogit occ ed exper white
345     mlogtest white, lr // 指定变量名称
346
347     qui mlogit occ ed exper white
348     mlogtest, lr wald // 针对所有变量检验
349
350
351 *-----
352 *-1.3.3.2 检验组间差异
353
354     *-Wald test
355     qui mlogit occ ed exper white
356     test [Menial]
357
358     qui mlogit occ ed exper white
359     test [Menial=BlueCol]
360
361     qui mlogit occ ed exper white
362     mlogtest, combine
363
364     *-LR test
365     qui mlogit occ ed exper white
366     est store mFull
367
368     constraint define 88 [Menial]
369
370     mlogit occ ed exper white, constraint(88) base(3) nolog
371     est store m0
372
373     lrtest mFull m0
374
375     qui mlogit occ ed exper white
376     mlogtest, lrcomb
377
378
379 *-----
380 *-1.3.3.3 检验 IIA
381
382     *-Hausman test of IIA
383
384     qui mlogit occ ed exper white, base(5)
385     est store mFull
386     qui mlogit occ ed exper white if occ!=1
387     est store mRist
388     hausman mRist mFull, alleq constant
389
390     qui mlogit occ ed exper white
391     mlogtest, hausman base
392
393     *-Small-Hsiao test of IIA
394
395     qui mlogit occ ed exper white
396     mlogtest, smhsiao base
397
398     qui mlogit occ ed exper white
399     set seed 13579
400     mlogtest, smhsiao
401
402     qui mlogit occ ed exper white
403     set seed 1357911
404     mlogtest, smhsiao
405
406     *-Q: IIA 无法满足时如何处理?
407     *-A: Maybe, Nested-Logit, see Section 5.4
408
```

```
409
410      *-----
411      *-1.3.4 拟合值和预测概率
412
413      *-predict 命令: 全样本
414      use nomocc2.dta, clear
415      qui mlogit occ ed exper white
416      predict prM prB prC prW prP
417      des pr*
418      sum pr*
419
```



```

1
2  *      -----
3  *      -----
4  *
5  *      Stata 学术论文专题
6  *
7  *      -----
8  *      -----
9
10
11  *
12  *      主讲人：连玉君 副教授
13  *
14  *
15  *      单 位：中山大学岭南学院金融系
16  *      电 邮：arlionn@163.com
17  *      博 客：http://blog.cnfol.com/arlion
18  *      主 页：http://toran.cn/arlion
19  *      微 博：http://weibo.com/arlionn
20
21
22
23
24  *      =====
25  *      =====
26  *
27  *      Appendix 2: 事件研究法(Event Study)
28  *
29  *      =====
30  *      =====
31
32  *-Note: 这个专题本不在“Stata学术论文专题”的制作计划中，
33  *      由于部分学员需要这部分资料，故以附录的形式呈现于此。
34
35
36  *-本讲路径(执行后续命令之前，请先执行如下两条命令)
37  global path "`c(sysdir_personal)'PX_papers"
38  cd "$path\z02_Event_Study"
39
40  *-重点参考：
41  view browse "http://dss.wikidot.com/event-study"
42
43  *-有关本讲中数据的来源与初步处理方法，请参见该文档末尾的说明
44
45
46
47  *-----
48  *-> 2.1 定义事件日
49  *-----
50
51  use eventda, clear
52  drop comnam
53  tsset company_id date
54
55  sort company_id date
56  by company_id: gen datenum=_n // 定义新的日期变量，取值为 1,2,3,...
57  by company_id: gen target=datenum if date==event_date // 标定事件日
58  egen td = min(target), by(company_id) // 扩充target的观察值
59  gen dif = datenum-td // 以事件日0为基准定义日期
60
61  list comp date event_date datenum target td dif in 1/20
62
63  *For calendar days:
64  * gen difc=date-event_date
65
66
67
68

```

```

69
70 *-----
71 *-> 2.2 定义事件窗口和估计窗口
72 *-----
73
74 *-----
75 *-2.2.1 定义事件窗口 [-2, 2], 即时间前后各两天
76
77     by company_id: gen event_window=1 if (dif>=-2 & dif<=2)
78     egen count_event_obs=count(event_window), by(company_id)
79
80 *-----
81 *-2.2.2 定义估计窗口 [-60, -30]
82
83     by company_id: gen estimation_window = 1 if (dif<-30 & dif>=-60)
84     egen count_est_obs = count(estimation_window), by(company_id)
85
86     * 如果 count_est_obs<30, 该公司将被删除, 因为没有足够的样本来估计正常回报率
87
88 *-----
89 *-2.2.3 用 0 和 1 区分事件窗口和非事件窗口
90
91     list comp date target dif *window in 1/30, sep(0)
92     list comp date target dif *window in 290/360, sep(0)
93     replace event_window = 0 if event_window==.
94     replace estimation_window = 0 if estimation_window==.
95     list comp date target dif *window in 1/30, sep(0)
96     list comp date target dif *window in 290/360, sep(0)
97
98 *-----
99 *-2.2.4 删除样本数不足的公司
100
101     *-事件窗口观察值不足5天的公司
102
103     tab company_id if count_event_obs<5
104     xtides if count_event_obs<5, p(0) // 计算此类公司数目
105     drop if count_event_obs<5
106
107     *-估计窗口观察值不足30的公司
108
109     tab company_id if count_est_obs<30
110     xtides if count_est_obs<30, p(0) // 计算此类公司数目
111     drop if count_est_obs<30
112
113     *-剩余公司数目统计
114
115     xtides, p(0)
116
117
118 *-----
119 *-> 2.3 估计正常回报率
120 *-----
121
122     cap drop predicted_return
123     gen predicted_return=. // 用于存放正常回报率的变量
124     cap drop id
125     egen id=group(company_id) // 重新定义公司代码为 1,2,3...,便于执行循环程序
126     qui tab id
127     local N = r(r) // 记录公司数目
128
129     forvalues i=1(1)`N' {
130         *-在估计窗口内估计市场模型
131         qui reg ret market_return if (id==`i' & estimation_window==1)
132         *-得到全样本范围内的正常回报率, 即 market_return 的全样本拟合值
133         predict p if id==`i'
134         *-在事件窗口内计算正常回报率
135         replace predicted_return = p if (id==`i' & event_window==1)
136         drop p // 把相关结果记录到了predicted_return变量中, 变量p已经无用

```

```

137     }
138
139
140
141 *-----
142 *-> 2.4 估计超额回报率和累积超额回报率
143 *-----
144
145 *--计算超额回报率 Abnormal Returns
146
147     gen abnormal_return = ret - predicted_return if event_window==1
148
149
150 *--计算累积超额回报率 CAR_id: sum of each firm (简单加总)
151
152     sort id date
153     by id: egen CAR_id = sum(abnormal_return) // egen
154     list id date dif abnormal_return CAR_id if (event_window==1 & id==1)
155
156
157 *--计算累积超额回报率 CAR_date: sum according distance from event
    day(按时间累积加总)
158
159     cap drop CAR_date // gen
160     bysort comp: gen CAR_date = sum(abnormal_return) if event_window==1
161     list id date dif abnormal_return CAR_date if (event_window==1 & id ==1)
162
163
164
165 *-----
166 *-> 2.5 假设检验
167 *-----
168
169 *-----
170 *--2.5.1 检验个别公司的累积超常收益率是否显著
171
172     cap drop ar_sd
173     cap drop test_id
174     cap drop sig
175     cap drop pvalue
176
177 *--计算 t 值
178
179     sort id date
180     by id: egen ar_sd = sd(abnormal_return) // 标准差
181     gen test_id =(1/sqrt(5))*(CAR_id /ar_sd) // t值,注意调整自由度
182
183 *--计算 p 值
184
185     gen pvalue = normal(test_id)
186     replace pvalue = 1 - pvalue if test_id>0
187     replace pvalue = pvalue*2 // 双尾检定
188
189 *--标明显著水平
190
191     gen sig = "*" if pvalue <= 0.10
192     replace sig = "***" if pvalue <= 0.05
193     replace sig = "****" if pvalue <= 0.01
194
195 *--列示结果
196
197     gsort -sig
198     list company_id CAR_id test_id pvalue sig if dif==0&sig!=" "
199
200     gsort -sig
201     list company_id CAR_id test_id sig if dif==1 & sig!=" "
202     list company_id CAR_id test_id sig if dif==2 & sig!=" "
203     list company_id CAR_id test_id sig if dif==0 & sig!=" "

```

```

204 list company_id CAR_id test_id sig if dif==1 & sig!=" "
205 list company_id CAR_id test_id sig if dif==2 & sig!=" "
206
207
208 *Note: this test uses the sample standard deviation.
209 *      A less conservative alternative is to
210 *      use the population standard deviation.
211 *      To derive this from the sample standard deviation produced by Stata,
212 *      multiply ar_sd by the square root of n-1/n;
213 *      in our example, by the square root of 4/5.
214
215
216
217 *-----
218 *-2.5.2 样本总体的累积超常回报率是否显著?
219
220 *-2.5.2.1 事件窗内的简单加总回报率
221
222 reg CAR_id, robust // 方法一: 回归法
223 ttest CAR_id=0 // 方法二: 传统的t检验
224
225 *-2.5.2.2 事件窗内的按时间累积加总回报率
226
227 reg CAR_date if dif== -2, robust noheader
228 reg CAR_date if dif== -1, robust noheader
229 reg CAR_date if dif== 0, robust noheader
230 reg CAR_date if dif== 1, robust noheader
231 reg CAR_date if dif== 2, robust noheader
232
233 *-结论: 不存在事前信息泄露问题
234
235 *-a- 采用列表的方式来呈现
236
237 local j=1
238 forvalues i = -2(1)2{
239     dis in g "date = " in y `i'
240     reg CAR_date if dif==`i', robust noheader
241     est store r`j++'
242 }
243 esttab r1 r2 r3 r4 r5, mtitle(-2 -1 0 1 2)
244
245 *-b- 采用矩阵的方式来呈现
246
247 mat A = J(5,5,0)
248 forvalues i = -2(1)2{
249     qui reg CAR_date if dif==`i', robust
250     local b = _b[_cons] // 系数
251     mat V = e(V)
252     local se= sqrt(V[1,1]) // 标准误
253     local t = `b'/'se' // t值
254     local pvalue = normal(`t')
255     if `t'>0{
256         local pvalue = 1-`pvalue'
257     }
258     local pvalue = `pvalue'*2
259     mat A[`=i'+3',1] = (`i',`b',`se',`t',`pvalue')
260 }
261 mat colnames A = date coef se t pvalue
262 mat list A
263
264 *-c- 采用图形方式来呈现: 绘图显示逐日累加 累积超常回报率
265
266 preserve
267 keep id date dif event_window CAR_date
268 keep if event_window == 1
269 cap drop CAR_t
270 keep if event_window == 1
271 sort dif

```

```

272     by dif: egen CAR_t = mean(CAR_date)
273     keep dif CAR_t
274     duplicates drop
275     twoway connect CAR_t dif
276     restore
277
278     *-d- 虚拟变量法
279
280     * 参见 Cowan(2005, Evnetus User's Guide, p.78)
281     * Karafiath(1988, Financial Review, 23(3), 351-358.)
282
283     * 定义事件窗口虚拟变量
284
285     cap drop dum*
286     tab dif if abs(dif)<4&dif!=3, gen(dum)
287     drop dum1
288
289     *-OLS
290
291     reg ret market_return dum*, robust
292
293     *-考虑截面异方差
294
295     qui tsset company_id dif
296     xtglm ret market_return dum*, p(het)
297
298     *-同时考虑截面异方差和截面相关
299
300     xtglm ret market_return dum*, p(c)
301
302
303     *-----
304     *-2.5.4 符号检定法
305     *-----
306
307     *-统计量:  $J3 = [N^+ / N - 0.5] * (2 * \sqrt{N}) \sim N(0, 1)$ 
308
309     *-参见: Campbell, J., A. Lo, A. MacKinlay.
310     *         The Econometrics of Financial Markets[M].
311     *         Cambridge Univ Press, 1997.
312
313     *-整体检验
314
315     preserve
316     keep if event_window == 1
317     qui count if abnormal_return > 0
318     local Np = r(N)
319     local N = _N
320     local J3 = (`Np' / `N' - 0.5) * 2 * sqrt(`N')
321     local pvalue = normal(`J3')
322     if `J3' > 0 {
323         local pvalue = 1 - `pvalue'
324     }
325     local pvalue = `pvalue' * 2
326     dis in g " J3" _col(8) " = " in y %6.4f `J3' _n ///
327         in g "p-value = " in y %6.4f `pvalue'
328     restore
329
330
331     *-逐日检验
332
333     mat A = J(5, 5, 0)
334     forvalues i = -2(1)2 {
335         qui count if (dif==`i' & abnormal_return>0 & event_window==1)
336         local Np = r(N) // 超常回报率为正的公司数目
337         qui count if (dif==`i' & abnormal_return!=. & event_window==1)
338         local N = r(N) // 公司总数
339         local J3 = (`Np' / `N' - 0.5) * 2 * sqrt(`N')

```

```

340     local pvalue = normal(`J3')
341     if `J3' > 0{
342         local pvalue = 1- `pvalue'
343     }
344     local pvalue = `pvalue'*2
345     mat A[`=i'+3',1] = (`i', `J3', `pvalue', `Np', `N')
346 }
347 mat colnames A = date J3 p-value N+ N
348 mat list A
349
350
351
352 *==== A simulation ====
353 * to test the time expenditure
354 clear
355 set memory 200m
356 set seed 12345
357 set obs 1000
358 gen t=_n
359 postutil clear
360 tempfile temp
361 postfile v id e1 e2 e3 using `temp'
362 set rmsg on
363
364 qui{
365     gen car=sum(uniform()+.01)
366     forv id=1/130 {
367         tempfile t`id'
368         gen event=uniform(>.9999
369         replace event=0 if sum(event)>1
370         su t if event==1, meanonly
371         gen et=t-r(mean)
372         save `t`id''
373         drop event et
374     }
375 }
376
377 drop _all
378 qui{
379     forv id=1/130 {
380         use `t`id''
381         qui reg car et if inrange(et,-230,10)
382         predict resid if e(sample), resid
383         forv i=1/3 {
384             su resid if et==`i'-2, meanonly
385             local e`i'=r(mean)
386         }
387         post v (`id') (`e1') (`e2') (`e3')
388         drop _all
389     }
390 }
391
392 postclose v
393
394
395
396 *-----
397 *-> 2.6 Appendix: 数据处理说明
398 *-----
399
400 *-1- 本例中的数据来源于Princeton大学网站
401
402 view browse ///
403     "http://dss.princeton.edu/usingdata/stata/analysis/eventstudy.html"
404 view browse ///
405     "http://dss.princeton.edu/usingdata/stata/analysis/eventstudydataprep.html"
406
407 *-2- 下载数据（建议采用Ftp下载，否则容易中断）

```

```
408
409 * 该网站提供了两份数据，一份记录了事件日，另一份记录了个股单日回报率
410 * use http://dss.princeton.edu/sampleData/eventdates.dta /* about 11k */
411 * save D:\stata10\ado\personal\Net_course\C4_Event_study\eventdates.dta
412 * use http://dss.princeton.edu/sampleData/stockdata.dta, clear /* about 90m */
413 * save D:\stata10\ado\personal\Net_course\C4_Event_study\stockdata
414
415 *-3- 合并数据：在正式分析之前，我们需要把两份数据合并起来，
416 * 并保留与事件相关的数据
417                                     /*
418     cd D:\stata10\ado\personal\Net_course\C4_Event_study
419     clear
420     set memory 120m
421     use eventdates.dta, clear
422     drop targetname
423     sort company_id
424     save eventdate_cy, replace
425
426     use D:\stata10\ado\personal\Net_course\C4_Event_study\stockdata, clear
427     merge company_id using eventdate_cy
428     drop if _merge != 3
429     save eventda, replace
430                                     */
431 * 随后的分析采用 eventda.dta 文件即可
432 *=====
433
```