DOI: 10.19327/j.cnki.zuaxb.1007-9734.2017.06.010

如何检验分组回归后的组间系数差异?

连玉君,廖俊平

(中山大学 岭南学院,广东 广州 510275)

摘 要:实证研究中,若检验在不同情况下一个变量对另一变量的影响程度,则需要对样本进行分组检验,进而比较两个子样本的系数差异。然而,单独比较子样本系数的显著性水平可能会存在偏差。为解决这一问题,本文介绍三种较为常见的组间系数差异检验方法的原理及Stata实现。其中,"Chow检验"通过引入交乘项的方式进行检验,该方法假定控制变量的系数不随组别发生变化,适用条件最为严格;"似无相关模型检验"则允许控制变量系数存在差异,且子样本扰动项相关,适用条件较为宽松;"费舍尔组合检验"基于自体抽样思想,通过不断抽样模拟总体特征,适用范围最为广泛。

关键词:组间系数差异;Chow检验;似无相关模型;费舍尔组合检验;企业金融

中图分类号:F275 文献标识码:A 文章编号:1007-9734(2017)06-0097-13

一、问题的提出

实证分析中,经常需要对比分析两个子样本组的系数是否存在差异。例如,在公司金融领域,研究薪酬激励是否有助于提升业绩时,模型设定为:

$$ROE_{it} = \alpha_i + Salary_{it-1} \cdot \beta + Controls_{it-1} \cdot \gamma + u_{it}$$

关注的重点是系数β。文献中经常把样本组分成"国有企业(SOE)"和"民营企业(PRI)"两个样本组,继而比较β_{SOE}和β_{PRI}是否存在差异。通常认为,民营企业的薪酬激励更有效果,即$\beta_{SOE} < \beta_{PRI}$。

如果两个样本组中的模型设定是相同的,则两组之间的系数大小是可以比较的,而且这种比较在多数实证分析中都是非常必要的。参见Cleary(1999),以及连玉君等(2010)。

下面使用Stata软件自带的一份数据文件(nlsw88.dta),说明此类问题的关键所在。这份数据包含了1988年采集的2 246个妇女的资料,包括:小时工资(wage),每周工作时数(hours),种族(race)等关键变量。

假如我们研究的问题是种族特征是否会影响妇女的工资决定机制,则我们最为关注的是

收稿日期:2017-10-09

作者简介:连玉君,男,河南淮阳人,副教授,经济学博士,研究方向为公司金融、金融计量。
廖俊平,男,教授,博士,研究方向为房地产经营管理、住房政策。

“白人组”和“黑人组”的工资决定因素是否存在差异。分析的重点集中于工龄(ttl_exp)和婚姻状况(married)这两个变量的系数在两组之间是否存在显著差异。

我们可以使用如下命令进行分组OLS回归,并呈现“白人组”和“黑人组”的系数估计值:

```
  sysuse "nlsw88.dta", clear
  gen agesq = age*age
*-分组虚拟变量
  drop if race==3
  gen black = 2.race
  tab black
*-删除缺漏值
  global xx "ttl_exp married south hours tenure age* i.industry"
  reg wage $xx i.race
  keep if e(sample)
*-分组回归
  global xx "ttl_exp married south hours tenure age* i.industry"
  reg wage $xx if black==0
  est store White
  reg wage $xx if black==1
  est store Black
 *-结果呈现和对比
  local m "White Black"
  esttab `m', mtitle(`m') b(%6.3f) nogap drop(*.industry) ///
              s(N r2_a) star(* 0.1 ** 0.05 *** 0.01)
```

执行上述命令后将呈现如下结果,如表1所示:

表1 “黑人组”和“白人组”系数对比

	(1) White	(2) Black
ttl_exp	0.251***	0.269***
	(6.47)	(4.77)
married	-0.737**	0.091
	(-2.31)	(0.23)
south	-0.813***	-2.041***
	(-2.71)	(-4.92)
hours	0.051***	0.037
	(3.81)	(1.39)
tenure	0.025	-0.004
	(0.77)	(-0.09)
age	0.042	0.995
	(0.03)	(0.54)
agesq	-0.001	-0.015
	(-0.09)	(-0.66)
_cons	3.333	-14.098
	(0.14)	(-0.39)
N	1615.000	572.000
r2_a	0.112	0.165

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

可以看到,ttl_exp变量在[white组]和[black组]的系数分别为0.251和0.269,二者都在

1% 水平上显著异于零。问题在于：以上结果能解释为 0.269 比 0.251 大吗？从统计意义上来看，答案显然没有那么明确。

相对而言，若把注意力放在 married 这个变量上，或许更容易判断二者的差异是否显著。因为，married 在“白人组”和“黑人组”的系数分别为 $\hat{\beta}_{white} = -0.737$ 和 $\hat{\beta}_{black} = 0.091$——前者在 5% 水平上显著为负，而后者则不显著。然而，即使如此，我们仍然无法确信二者的差异在统计意义上是否显著。这是因为二者的 95% 置信区间（CI）尚存在重叠区域：$CI_{white} = [-1.36, -0.11]$，$CI_{black} = [0.69, 0.87]$。

下面，我们介绍三种检验组间系数差异的方法：其一，引入交叉项（Chow 检验）；其二，基于似无相关模型的检验方法（suest）；其三，费舍尔组合检验（Permutation test）。第一种方法是文献中普遍使用的，但假设条件最为严格，第三种方法主要以自抽样为基础，假设条件最为宽松。

二、检验方法

（一）引入交叉项

1. 基本思想

这是文献中最常用的方法，执行起来也最简单。以检验 ttl_exp 在两组之间的系数是否存在显著差异为例。引入一个虚拟变量 D_i，若某个妇女是黑人，则 $D_i = 1$，否则 $D_i = 0$。在如下命令中，black 变量即为这里的 D_i。模型设定为：

$$Wage_i = \alpha + \gamma \cdot D_i + \beta \cdot ttl\_exp_i + \delta(D_i \times ttl\_exp_i) + Controls_i \cdot \gamma + \varepsilon_i \quad (1)$$

这是最基本的包含虚拟变量，以及虚拟变量与一个连续变量交乘项的情形。

显然，对于“白人组”而言，$D_i = 0$，则（1）式可以写为：

$$Wage_i = \alpha + \beta \cdot ttl\_exp_i + Controls_i \cdot \gamma + \varepsilon_i \quad (1a)$$

对于“黑人组”，（1）式可以写为：

$$Wage_i = (\alpha + \gamma) + (\beta + \delta) \cdot ttl\_exp_i + Controls_i \cdot \gamma + \varepsilon_i \quad (1b)$$

由此可见，在（1）式中，参数 γ 和 δ 分别反映了黑人组相对于白人组的截距和斜率差异。此时，参数反映了 ttl_exp 这个变量在两个样本组中的系数差异。因此，检验 ttl_exp 在两组之间的系数是否存在显著差异就转变为 $H_0: \delta = 0$。

2. Stata 实现

估计（1b）式的 Stata 命令如下：

```
dropvars ttl_x_black marr_x_black
global xx "ttl_exp married south hours tenure age* i.industry" //Controls
gen ttl_x_black = ttl_exp*black  //交乘项
reg wage black ttl_x_black $xx   //全样本回归+交乘
```

为节省篇幅，仅列出最关键的结果如下：

```
reg wage black ttl_x_black $xx   //全样本回归+交乘
      Source |       SS       df       MS              Number of obs =    2187
-------------+------------------------------           F( 20,  2166) =   17.32
       Model |   10074.761    20  503.738052            Prob > F      =  0.0000
    Residual |  63000.2591  2166  29.0859922            R-squared     =  0.1379
-------------+------------------------------           Adj R-squared =  0.1299
       Total |  73075.0201  2186   33.428646            Root MSE      =  5.3931
------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |   -.818647   .8015272    -1.02   0.307     -2.39049    .7531957
ttl_x_black |  -.0181844   .0585517    -0.31   0.756    -.1330077    .0966389
     ttl_exp |   .2537358   .0351178     7.23   0.000     .1848676     .322604
     married |  -.4646649   .2530829    -1.84   0.066    -.9609756    .0316458
       south |  -1.127138   .2453123    -4.59   0.000     -1.60821   -.6460662
       hours |   .0516672   .0116886     4.42   0.000     .0287451    .0745894
      tenure |   .0198005   .0260971     0.76   0.448    -.0313775    .0709786
         age |   .1685498   1.035721     0.16   0.871    -1.862562    2.199661
       agesq |  -.0034668   .0131097    -0.26   0.791    -.0291756     .022242
             |                .....
       _cons |   .8448605   20.39973     0.04   0.967    -39.16022    40.84994
------------------------------------------------------------------------------
```

交乘项 [ttl_x_black] 的系数为 -0.01818，对应的 p-value 为 0.756，表明 [ttl_exp] 的系数在两组之间并不存在显著差异。我们也可以直接采用 Stata 的因子变量表达式，在不用预先生成交乘项的情况下得到完全相同的结果：

```
reg wage i.black ttl_exp i.black#c.ttl_exp $xx
```

或如下更为简洁的方式（详情参见 help fvvarlist）：

```
reg wage i.black ttl_exp i.black#c.ttl_exp $xx
*-或
reg wage i.black##c.ttl_exp $xx
```

然而，在上述检验过程中，我们无意识中施加了一个非常严格的假设条件：只允许变量 [ttl_exp] 的系数在两组之间存在差异，而其他控制变量（如 married，south，hours 等）的系数则不随组别发生变化。这显然是一个非常严格的假设。因为，从表 1 中的结果来看，married，south，hours 等变量在两组之间的差异都比较明显。

我们可以放松上述假设，允许 married，south，hours 等变量的系数存在组间差异。这主要通过引入虚拟变量 black 与上述变量的交乘项来实现（模型 2）：

```
reg wage i.black i.black#(c.ttl_exp i.married i.south c.hours) $xx  //Model 2
```

在这种相对灵活的设定下，[ttl_exp] 的系数为 -0.016147，相应的 p-value = 0.787，依然不显著。

当然，我们也可以采用更为灵活的方式：允许所有的变量在两组之间都存在系数差异 1（模型 3）：

```
global xx "c.ttl_exp married south c.hours c.tenure c.(age*) i.industry"
reg wage i.black##($xx)                                          //Model 3
```

这其实就是大名鼎鼎的 Chow test（邹检验），可以用 chowtestP 命令快捷地完成。

3. 小结

引入交乘项来检验某个或某几个变量的系数是否存在组间差异，只需在普通线性回归中加

入交乘项即可，但需要注意这一方法背后隐含的假设条件（为了便于说明，重新将（1）式列出）：

$$Wage_i = \alpha + \gamma \cdot D_i + \beta \cdot ttl\_exp_i + \delta(D_i \times ttl\_exp_i) + Controls_i \cdot \gamma + \varepsilon_i \quad (1)$$

A1：所有控制变量的系数在两组之间无差异，即 $\gamma_{white} = \gamma_{black}$；

A2：两组的干扰项具有相同的分布（因为估计时是将两组样本混合在一起进行估计的），即 $e_{white} \sim N(0,\sigma^2)$，且 $e_{black} \sim (0,\sigma^2)$，换言之，$\sigma^2_{white} = \sigma^2_{black}$。

因此，当其他变量的系数在两组之间也存在明显差异（A1 不满足），或存在异方差（A2 不满足）时，上述检验方法得到的结果都存在问题。

4. 解决办法

对于 A1，实际操作过程中，可以通过引入更多的交项来放松 A1，如上文提到的“模型 2”或“模型 3”。

对于 A2，则可以在上述回归分析过程中加入 vce（robust）选项，以便允许干扰项存在异方差；或加入 vce（cluster varname）以便得到聚类调整后的稳健型标准误。

上述范例中，是以基于截面数据的 OLS 回归为例的，但这一方法也适用于其他命令，如针对面板数据的 xtreg，针对离散数据的 logit，probit 等。

（二）SUEST（基于似无相关模型 SUR 的检验）

1. 基本思想

顾名思义，所谓的似无相关模型（seemingly unrelated regression）其实就是表面上看起来没有关系，但实质上有关系的两个模型。这听起来有点匪夷所思。这种“实质上”的关系其实是假设白人组和黑人组的干扰项彼此相关。为了表述方便，将白人和黑人组的模型简写如下：

白人组：$y_{1i} = x_{1i}\beta_1 + e_{1i}, i = 1,2,\cdots,N_1$　　(2a)

黑人组：$y_{2j} = x_{2j}\beta_2 + e_{2j}, j = 1,2,\cdots,N_2$　　(2b)

若假设 $corr(e_1, e_2) = 0$，则我们可以分别对白人组和黑人组进行 OLS 估计。然而，虽然白人和黑人种族不同，但所处的社会和法律环境，面临的劳动法规都有诸多相似之处，使得二者的干扰项可能相关，即 $corr(e_1, e_2) \neq 0$。此时，对两个样本组执行联合估计（GLS）会更有效率（详见 Greene（2012），292 - 304）。执行完 SUR 估计后，对两组之间的系数差异进行检验。

从上面的原理介绍，可以看出，基于 SUR 估计进行组间系数差异检验时，假设条件比第一种方法要宽松一些：

其一，在估计过程中，并未预先限定白人组和黑人组中各个变量的系数相同，因此在（2）式中，我们分别用 β_1 和 β_2 表示白人组和黑人组各个变量的系数向量；

其二，两个组的干扰项可以有不同的分布，即 e_{1i} 和 e_{2j} 可以不同，即 $e_{1i} \sim N(0,\sigma_1^2)$，$e_{2j} \sim N(0,\sigma_2^2)$，且允许二者的干扰项相关，$corr(e_1, e_2) \neq 0$。

2. Stata 实现方法

在 Stata 中执行上述检验的步骤为：（1）分别针对白人组和黑人组进行估计（不限于 OLS 估计，可以执行 logit，tobit 等估计），存储估计结果；（2）使用 suest 命令执行 SUR 估计；（3）使用 test 命令检验组间系数差异。范例如下：

```
*-Step1: 分别针对两个样本组执行估计
  reg wage $xx if black==0
  est store w  //white
  reg wage $xx if black==1
  est store b  //black
*-Step 2: SUR
  suest w b
*-Step 3: 检验系数差异
  test [w_mean]ttl_exp = [b_mean]ttl_exp
  test [w_mean]married = [b_mean]married
  test [w_mean]south = [b_mean]south
```

Step 2 的结果如下(为便于阅读,部分变量的系数未呈现):

```
.       suest w b

Simultaneous results for w, b
                                                  Number of obs   =       2187
------------------------------------------------------------------------------
             |    Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
w_mean       |
      ttl_exp |   .2505707   .0362302     6.92   0.000     .1795609    .3215805
      married |  -.7367238   .3486479    -2.11   0.035    -1.420061   -.0533865
        south |  -.8125914   .2892359    -2.81   0.005    -1.379483   -.2456994
        hours |   .0507118   .0126268     4.02   0.000     .0259637    .0754599
       tenure |   .0246063   .0289939     0.85   0.396    -.0322207    .0814333
          age |   .0415616   1.107902     0.04   0.970    -2.129887    2.213011
        agesq |  -.0014454   .0138965    -0.10   0.917     -.028682    .0257911
              |                     ... ...
        _cons |   3.333308   22.03159     0.15   0.880    -39.84781    46.51442
-------------+----------------------------------------------------------------
w_lnvar      |
        _cons |     3.4561   .0971137    35.59   0.000     3.265761     3.64644
-------------+----------------------------------------------------------------
b_mean       |
      ttl_exp |   .2686185   .0511831     5.25   0.000     .1683015    .3689355
      married |   .0913607   .4050685     0.23   0.822    -.7025589    .8852804
        south |  -2.040701   .4252889    -4.80   0.000    -2.874252    -1.20715
        hours |   .0367536   .0229076     1.60   0.109    -.0081444    .0816516
       tenure |   -.003914   .0450917    -0.09   0.931    -.0922921    .0844641
          age |   .9952064   1.814351     0.55   0.583    -2.560856    4.551269
        agesq |  -.0153824   .0229045    -0.67   0.502    -.0602744    .0295096
        _cons |  -14.09831   35.40206    -0.40   0.690    -83.48508    55.28846
              |                     ... ...
-------------+----------------------------------------------------------------
b_lnvar      |
        _cons |   3.077588   .2090237    14.72   0.000     2.667909    3.487267
------------------------------------------------------------------------------
```

对上述命令和结果的简要解释如下:

其一,白人组和黑人组估计结果分别存储于 w 和 b 两个临时性文件中。

其二,执行 suest w b 命令时,白人组和黑人组的被视为两个方程,分别对应于(2a)和(2b)式。Stata 会自动将两个方程对应的样本联合起来,采用 GLS 执行似无相关估计(SUR)。

由于 SUR 属于多方程模型,因此需要指定每个方程的名称,在下面呈现的回归结果中,[w_mean] 和 [b_mean] 分别是白人组和黑人组各自对应的方程名称。因此,[w_mean] ttl_exp 表示白人组方程中 ttl_exp 变量的系数,而 [b_mean] ttl_exp 则表示黑人组中 ttl_exp 变量的系数。

执行组间系数差异检验的结果如下(Step 3):

```
*-Step 3: 检验系数差异
test [w_mean]ttl_exp = [b_mean]ttl_exp
(1)  [w_mean]ttl_exp - [b_mean]ttl_exp = 0
          chi2(  1) =    0.08
          Prob > chi2 =    0.7735
test [w_mean]married = [b_mean]married
 (2)  [w_mean]married - [b_mean]married = 0
          chi2(  1) =    2.40
          Prob > chi2 =    0.1213
test [w_mean]south = [b_mean]south
 (3)  [w_mean]south - [b_mean]south = 0
          chi2(  1) =    5.70
          Prob > chi2 =    0.0169
```

此时，ttl_exp 在两组之间的系数差异仍然不显著，这与采用第一种方法得到的结论是一致的。在我们测试的三个变量中，只有 south 的系数在两组之间存在显著差异，对应的p－value为0.0169。

3. 使用 bdiff 命令

上述过程可以使用 bdiff 命令快捷地加以实现，结果的输出方式也更为清晰[②]。举例如下:

```
preserve
  drop if industry==2  // 白人组中没有处于 Mining (industry=2) 的观察值
  tab industry, gen(d)  //手动生成行业虚拟变量
  local dumind "d2 d3 d4 d5 d6 d7 d8 d9 d10 d11" //行业虚拟变量
  global xx "c.ttl_exp married south c.hours c.tenure c.age c.agesq `dumind'"
  bdiff, group(black) model(reg wage $xx) surtest
restore
```

输出结果如下:

Variables	b0-b1	Chi2	p-value
ttl_exp	-0.017	0.07	0.788
married	-0.814	2.32	0.128
south	1.238	5.80	0.016
hours	0.014	0.28	0.597
tenure	0.030	0.32	0.571
age	-1.027	0.23	0.629
agesq	0.015	0.31	0.578
d2	-2.732	1.63	0.202
d3	-1.355	1.45	0.228
d4	-2.708	2.23	0.135
d5	-1.227	1.00	0.317
d6	0.087	0.00	0.950
d7	-0.534	0.07	0.785
d8	-1.316	1.26	0.261
d9	0.346	0.06	0.807
d10	-1.105	0.94	0.333
d11	-1.689	1.81	0.179
_cons	18.770	0.20	0.652

两点说明: 其一,使用 suest 时,允许两个样本组的解释变量个数不同。但由于一些技术上的问题尚未解决,bdiff 命令要求两个样本组中的解释变量个数相同。在上例中,白人组在 Mining 行业的观察值个数为零(输入 tab industry black 命令可以查看) ,导致我们加入行业虚拟变量时,白人组只有 10 个行业虚拟变量,而黑人组则有 11 个行业虚拟变量。为此,在上述命令中,我使用 drop if industry = =2 命令删除了 Mining 行业的观察值。其二,目前,bdiff 命令还不能很好地支持因子变量的写法(help fvvarlist) ,因此上例中的行业虚拟变量不能通配符方式写成 d* ,而必须写成原始模样: d2 d3 d4 d5 d6 d7 d8 d9 d10 d11。

4. 面板数据的处理方法

suest 不支持 xtreg 命令,因此无法直接将该方法直接应用于面板数据模型,如 FE 或 RE。此时,可以预先手动去除个体效应,继而对变换后的数据执行 OLS 估计,步骤如下: 第一步,去除个体效应。对于固定效应模型而言,可以使用 center 或 xtdata 命令去除个体效应; 对于随机效应模型而言,可以使用 xtdata 命令去除个体效应。第二步,按照截面数据的方法对处理后的数据进行分组估计,并执行 suest 估计和组间系数检验。

5. 小结

相对于方法 1(引入交乘项) ,基于 SUR 的方法更为灵活一些。在上例中,白人组和黑人组的被解释变量相同(均为 wage) ,此时方法 1 和方法 2 都能用。有些情况下,两个组中的被解释变量不同,此时方法 1 不再适用,而方法 2 则可以。

对于面板数据而言,可以预先使用 center 或 xtdata 命令去除个体效应,变换后的数据可以视为截面数据,使用 regress 命令进行估计即可。

为了便于呈现结果,可以使用 estadd 命令将上述检验结果(chi2 值或 p 值) 加入内存,进而使用 esttab 命令列示出来。可以参考 help bdiff 中的类似范例。

(三) 费舍尔组合检验 (Fisher's Permutation test)

1. 基本思想

这一方法最早见于 Efron 和 Tibshirani(1993) (Section 15.2,pp. 202) 。Cleary(1999) ,连玉君等(2008) , 连玉君等(2010) 等都使用了该方法进行组间系数差异检验。

同样以上述"白人组"和"黑人组"为例,假设两组的样本数分别为 n_1 和 n_2,模型设定如下:

$$y = x\beta_1 + e_1 \quad (3a)$$

$$y = x\beta_2 + e_2 \quad (3b)$$

将二者的系数差异定义为 $d = \beta_1 - \beta_2$,检验的原假设为: H_0 : $d = 0$。

我们仍然关注 ttl_exp 变量在两组之间的系数差异。以表 1 中的结果为例,可以看到,ttl_exp 变量在 [white 组] 和 [black 组] 的系数估计值分别为 0.251($\hat{\beta}_1$) 和 0.269($\hat{\beta}_2$) ,因此,实际观察到的系数差异为 $\hat{d}_0 = \hat{\beta}_1 - \hat{\beta}_2 = 0.251 - 0.269 = -0.018$。

显然,d 是一个统计量,若预先知道其分布特征,便可通过分析 $\hat{d}$ 在 d 的分布中的相对位置来判断我们实际观察到 $\hat{d}_0 = -0.018$ 的概率。若概率很小,则表明 H_0 : $d = 0$ 是小概率事件,此时拒绝原假设,反之则无法拒绝原假设。

例如，若假设 d 服从标准正态分布，即 d ~ N(0,1)，则基于实际观察到的 $\hat{d}_0 = 0.018$，我们很容易得出结论：无法拒绝原假设，即两组之间的 ttl_exp 的系数不存在显著差异。p - value 很容易计算（可以通过查表得到）：

```
. dis normal(-0.018)    //单尾检验
.49281943
```

然而，在实际应用过程中，我们并不知道 d 的分布特征。此时，可以对现有样本进行重新抽样，以得到经验样本（empirical sample），进而利用经验样本构造出组间系数差异统计量 d 的经验分布（empirical distribution），从而最终得到经验 p 值（empirical p - value）。

下面先通过一个小例子说明“经验 p 值”和“经验分布”的概念，进而介绍使用组合检验获得“经验 p 值”的流程。

2. 经验 p 值（empirical p - value）

在这个小例子中，我们先随机生成一个服从标准正态分布的随机数 d，共有 10 000 个观察值。这些观察值是通过模拟产生的。如果这些观察值构成的样本是通过从原始样本（原始样本是从母体中一次随机抽样，称为“抽样样本，sample”）中二次抽样得到的，则称为“经验样本（empirical sample）”。

然后，我们数一下在这 10 000 个随机数中，有多少个是大于（我们实际观察到的数值），命令为 count if d < -0.018。一共有 4 963 个观察值大于，由此可得，经验 p 值 = 4963/10000 = 0.4963。这与采用 normal( ) 函数得到的结果（0.4928）非常接近。

```
preserve
    clear
    set obs 10000
    set seed 1357
    gen d = rnormal()  // d~N(0,1) 服从标准正态分布的随机数
    sum d, detail
    count if d<-0.018
    dis "Empirical p-value = "  4963/10000
restore
```

3. 经验样本（empirical sample）

上例中，我们假设 d 服从标准正态分布，从而可以通过 monte carlo 模拟的方式产生 10 000 个观察值，这事实上是构造了一个经验样本。但多数情况下，我们并不知道 d 的分布特征，此时无法使用 monte carlo 模拟。然而，若假设抽样样本（sample）是从母体（population）中随机抽取的，则可以通过抽样样本中二次抽样得到经验样本（empirical sample），这些经验样本也可以视为对母体的随机抽样。

若 H_0 抽样过程中为无放回抽样（sampling with no replacement），则相应的检验方法称为“组合检验（permutation test）”；若抽样过程中为有放回抽样（也称为可重复抽样，sampling with replacement），则对应的检验方法称为“基于 Bootstrap 的检验”。

4. 费舍尔组合检验的步骤

若 H_0 是正确的，则对于任何一个妇女而言（不论她是白人还是黑人），其 x 对 y 的边际影响都是相同的。因此，我们可以将白人组和黑人组的观察值混合起来，从中随机抽取 n_1 个观察值，并将其视为“白人组”，剩下的 n_2 个观察值可以视为“黑人组”。

Step 0：分别针对白人组和黑人组估计模型（3a）和（3b），得到系数估计值 $\hat{\beta}_1$ 和 $\hat{\beta}_2$，以及二者的系数差异 $\hat{d}_0 = \hat{\beta}_1 - \hat{\beta}_2$；

Step 1：将白人组和黑人组的样本混合起来，得到 $n_1 + n_2$ 个观察值构成的样本 S；

Step 2：获得经验样本——从 S 中随机抽取（无放回）n_1 个观察值，将其视为“白人组”（记为 Sw），剩下的 n_2 个观察值可以视为“黑人组”（记为 Sb）；

Step 3：分别针对经验样本 Sw 和 Sb，估计模型（3a）和（3b），得到 $\hat{\beta}_1^{S_1}$ 和 $\hat{\beta}_2^{S_1}$（上标 S_1 表示利用第一笔经验样本得到的估计值），以及二者的差异 $d^{S_1} = \hat{\beta}_1^{S_1} - \hat{\beta}_2^{S_1}$；

Step 4：获得统计量 d 的经验分布——将 Step 2 和 Step 3 重复执行 K 次（如 K = 1 000），则可以得到 $\{d^{S_1}, d^{S_2}, \cdots, d^{S_{1000}}\}$，亦可简记为 $d^{S_j}(j = 1, 2, \cdots, K)$；

Step 5：计算经验 p 值，$\hat{p} = \frac{\#\{d^{S_j} > \hat{d}_0\}}{K}$，其中 $\{d^{S_j} > \hat{d}_0\}$ 表示 Step 4 中得到的 K 个 d^{S_j} 中大于我们实际观测到 $\hat{d}_0$ 的个数。若 $\hat{p} < 0.05$，则可以在 5% 水平上拒绝原假设，表明两组的系数差异是显著的。

需要说明的是，由于 d^{sj} 的分布未必是对称的，因此，$\hat{p} = 0.03$ 与 $\hat{p} = 0.97$ 都可以视为在 5% 水平上拒绝原假设的证据。因为，前者意味着 $\hat{d}_0$ 在 1 000 个 d^{sj} 中属于非常大的数值，而后者意味着它是非常小的数值。无论如何，在原假设 H_0 下观察到 d_0 都是小概率事件，也就意味着原假设是不合理的。

此外，该方法并不局限于普通的线性回归模型（regress 命令），可以应用于各种模型的估计命令，如 xtreg，xtabond，logit，ivregress 等。

5. Stata 实现

上述过程可以使用 bdiff 命令来实现。③先使用一个简单的例子，不考虑行业虚拟变量：

```
*-数据处理
  sysuse "nlsw88.dta", clear
  gen agesq = age*age
  drop if race==3
  gen black = 2.race
  global xx "ttl_exp married south hours tenure age agesq"
*-检验
  bdiff, group(black) model(reg wage $xx) reps(1000) detail
```

选项 group( ) 中填写用于区分组别的类别变量（若有多个组，可以预先删除不参与比较的

组，类似于上面的 drop if race = =3 命令)；

选项 model( ) 用于设定回归模型，即上面提到的模型(3a) 或(3b)；

选项 reps(#) 用于设定抽样次数，即上文提到的 K，通常设定 1 000 ~ 5 000 次即可；

附加 detail 选项，可以进一步列表呈现两组的实际估计系数 $\hat{\beta}_1$ 和 $\hat{\beta}_2$。

上述过程大约用时 13 秒，结果如下：

```
-Permutaion (1000 times)- Test of Group (black 0 v.s 1) coeficients difference

   Variables |      b0-b1     Freq     p-value
-------------+------------------------------
     ttl_exp |       0.007      490       0.490
     married |      -0.824      920       0.080
       south |       1.411        5       0.005
       hours |       0.010      344       0.344
      tenure |      -0.006      512       0.488
         age |      -1.579      751       0.249
       agesq |       0.022      218       0.218
       _cons |      28.051      267       0.267
-------------------------------------------
```

可以看到，ttl_exp 的经验 p 值为 0. 49，表明白人组和黑人组的 ttl_exp 系数不存在显著差异; married 变量的 p 值为 0. 08，我们可以在 10% 水平上拒绝原假设。

若需在模型中加入虚拟变量，处理过程会稍微复杂一些。需要手动生成行业虚拟变量，并保证两个样本组中参与回归的行业虚拟变量个数相同。此外，书写命令时，不能使用通配符。(后续版本的 bdiff 命令会使用 fvunab 命令解决这些 bugs) 。

```
*-数据预处理(可以忽略)
  sysuse "nlsw88.dta", clear
  gen agesq = age*age
*-分组虚拟变量
  drop if race==3
  gen black = 2.race
*-删除缺漏值
  global xx "ttl_exp married south hours tenure age* i.industry"
  qui reg wage $xx i.race
  keep if e(sample)
*-生成行业虚拟变量
  drop if industry==2  // 白人组中没有处于 Mining (industry=2) 的观察值
  tab industry, gen(d)  //手动生成行业虚拟变量
  local dumind "d2 d3 d4 d5 d6 d7 d8 d9 d10 d11" //行业虚拟变量
  global xx "c.ttl_exp married south c.hours c.tenure c.age c.agesq `dumind'"
*-permutation test
  bdiff, group(black) model(reg wage $xx) reps(1000) detail
```

若原始数据为面板数据，通常会采用 xtreg，xtabond 等考虑个体效应的方法进行估计。抽样过程必须考虑面板数据的特征。在执行 bdiff 命令之前，只需设定 xtset id year，声明数据为面板数据格式，则抽样时便会以 id(公司或省份代码) 为单位，以保持 id 内部的时序特征。

```
*-数据预处理
  webuse "nlswork.dta", clear
  xtset id year                    //声明为面板数据，否则视为截面数据
  gen agesq = age*age
  drop if race==3
  gen black = 2.race
*-检验
  global x "ttl_exp hours tenure south age agesq"
  local m "xtreg ln_wage $x, fe" //模型设定
  bdiff, group(black) model(`m') reps(1000) bs first detail
```

结果如下:

```
-Bootstrap (1000 times)- Test of Group (black 0 v.s 1) coeficients difference

   Variables |      b0-b1    Freq     p-value
-------------+------------------------------
     ttl_exp |       0.011      47       0.047
       hours |       0.003      58       0.058
      tenure |       0.004     116       0.116
       south |       0.093      14       0.014
         age |      -0.022     984       0.016
       agesq |       0.000      45       0.045
       _cons |       0.302      22       0.022
-------------------------------------------------
Ho: b0(ttl_exp) = b1(ttl_exp)
Observed difference =  0.011
Empirical p-value =  0.047
```

解释和说明:

bdiff **命令中设定** bs **选项,则随机抽样为可重复抽样**(bootstrap) ;

first **选项便于将组间系数差异检验结果保存在内存中,方便后续使用** esttab **合并到回归结果表格中。具体使用方法参见** help bdiff。

由于抽样过程具有随机性,因此每次检验的结果都有微小差异。在投稿之前,可以附加 seed( ) **选项,以保证检验结果的可复制性。**

其他选项和使用方法参阅 help bdiff **的帮助文件。**

三、小 结

方法 1(**加入交乘项**) **在多数模型中都可以使用,但要注意其背后的假设条件是比较严格的。若在混合回归中,只引入关心的变量**(ttl_exp) **与分组变量**(black) **的交乘项**(ttl_exp * black) ,**则相当于假设其他控制变量在两组之间不存在系数差异。相对保守的处理方法是: 在混合估计时,引入所有变量与分组变量的交乘项,同时附加** vce(robust) **选项,以克服异方差的影响。**

方法 2(**基于** SUR **模型的检验**) **执行起来也比较方便,假设条件也比较宽松: 允许两组中所有变量的系数都存在差异,也允许两组的干扰项具有不同的分布,且彼此相关。局限在于,有些命令无法使用** suest **执行联合估计,如几乎所有针对面板数据的命令都不支持**(xtreg,xtabond **等**) 。

方法 3（组合检验）是三种方法中假设条件最为宽松的，只要求原始样本是从母体中随机抽取的（看似简单，但很难检验），而对于两个样本组中干扰项的分布，以及衡量组间系数差异的统计量 $d = \beta_1 - \beta_2$ 的分布也未做任何限制。事实上，在获取经验 p 值的过程中，我们采用的是“就地取材”“管中窥豹”的思路，并未假设 d 的分布函数；另一个好处在于可以适用于各种命令，如 regress，xtreg，logit，ivregress 等，而 suest 则只能应用于部分命令。

方法无优劣。无论选择哪种方法，都要预先审视一下是否符合这些检验方法的假设条件。

注　释：

①注意：所有离散变量前都要加前缀 i.，否则将被视为连续变量进行处理（对于取值为 0/1 的虚拟变量，可以省略前缀 i.）；连续变量则需加前缀 c.

②在 stata 命令窗口中输入 ssc install bdiff，replace 可以下载最新版命令包，进而输入 help bdiff 查看帮助文件.

③在命令窗口中输入 ssc install bdiff，replace 可以自动安装该命令。帮助文件中提供了多个范例.

参考文献：

[1] Cleary S. The relationship between firm investment and financial status [J]. Journal of Finance，1999，54（2）：673－692.

[2] 连玉君，彭方平，苏　治. 融资约束与流动性管理行为 [J]. 金融研究，2010，（10）：158－171.

[3] 连玉君，苏　治，丁志国. 现金—现金流敏感性能检验融资约束假说吗？ [J]. 统计研究，2008，（10）：92－99.

[4] Willian H Greene. Econometric Analysis [M]. New York：Prentice Hall，2012.

[5] Efron B R Tibshirani. An introduction to the bootstrap [M]. New York：Chapmann & Hall，1993.

责任编校：张　静，罗　红