第六届中国Stata用户大会

Stata中的标准误

陈强
山东大学经济学院
www.econometrics-stata.com

目录

1. 普通标准误
2. 异方差稳健标准误（HC0, HC1, HC2, HC3）
3. 刀切标准误（Jackknife）
4. 自助标准误（Bootstrap: pairs, residual, wild）
5. 异方差自相关稳健标准误（HAC）
6. 聚类稳健标准误（Cluster, two-way clustering, wild cluster bootstrap）

计量经济学的两大要点

- 估计量的一致性：$\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$
  （一切统计推断的前提）

- 方差估计量的一致性：$\widehat{Var(\hat{\boldsymbol{\beta}})} \xrightarrow{p} Var(\hat{\boldsymbol{\beta}})$
  （正确统计推断的保证）

  严格表述：

$$\widehat{\text{Avar}(\hat{\boldsymbol{\beta}})} \xrightarrow{p} \text{Avar}(\hat{\boldsymbol{\beta}}) \equiv \text{Var}\left[\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right]$$

线性回归模型

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i \quad (i = 1, \cdots, n)$$

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i \quad (i = 1, \cdots, n)$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_n' \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

OLS估计量

$$\hat{\boldsymbol{\beta}} \equiv (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

$$\hat{\boldsymbol{y}} \equiv \boldsymbol{X}\hat{\boldsymbol{\beta}} = \underbrace{\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'}_{\equiv \boldsymbol{P}}\boldsymbol{y} \equiv \boldsymbol{P}\boldsymbol{y}$$ (Projection Matrix)

$$0 \le h_i \equiv \boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_i \le 1$$ (Leverage for observation *i*)

若 $\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{I}$，则 $h_{ii} = \boldsymbol{x}_i'\boldsymbol{x}_i = \|\boldsymbol{x}_i\|^2$

OLS标准误的夹心估计量

- 抽样误差 (sampling error)

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{\beta}$$
$$= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) - \boldsymbol{\beta}$$
$$= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$$

- 夹心估计量 (Sandwich estimator)

$$\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) = \mathrm{Var}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \mid \boldsymbol{X}) = \mathrm{Var}\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon} \mid \boldsymbol{X}\right)$$
$$= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{Var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$
$$= (\boldsymbol{X}'\boldsymbol{X})^{-1}\underbrace{\boldsymbol{X}'\overbrace{\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X})}^{n\times n}\boldsymbol{X}}_{K\times K}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- 此公式无任何假定，永远成立

1、普通标准误

- 假设球型扰动项 (同方差、无自相关):

$$\mathrm{Var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \sigma^2 \boldsymbol{I}_n = \begin{pmatrix} \sigma^2 & & \\ & \ddots & \\ & & \sigma^2 \end{pmatrix}$$

$$\begin{aligned} \mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) &= (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' \mathrm{Var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) \boldsymbol{X} (\boldsymbol{X}'\boldsymbol{X})^{-1} \\ &= \sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{X} (\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1} \end{aligned}$$

普通标准误 (续)

- 对扰动项方差 σ^2 的无偏估计：

$$s^2 \equiv \frac{1}{n-K}\sum_{i=1}^{n} e_i^2 = \underbrace{\left(\frac{n}{n-K}\right)}_{\text{小样本校正}} \underbrace{\frac{1}{n}\sum_{i=1}^{n} e_i^2}_{\text{MLE}}$$

- 其中，$n-K$ 为自由度，故

$$\widehat{\text{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})} = s^2 (\boldsymbol{X}'\boldsymbol{X})^{-1} \quad \Rightarrow \quad \widehat{\text{Var}(\hat{\beta}_k)} = s^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}_{kk}$$

2、异方差稳健的标准误

- 若存在异方差，但无自相关，则

$$\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' \mathrm{Var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) \boldsymbol{X} (\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$= (\boldsymbol{X}'\boldsymbol{X})^{-1} \begin{pmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \end{pmatrix} \begin{pmatrix} \mathrm{E}(\varepsilon_1^2 \mid \boldsymbol{X}) & & \\ & \ddots & \\ & & \mathrm{E}(\varepsilon_n^2 \mid \boldsymbol{X}) \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_1' \\ \cdots \\ \boldsymbol{x}_n' \end{pmatrix} (\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$= (\boldsymbol{X}'\boldsymbol{X})^{-1} \left(\sum_{i=1}^{n} \mathrm{E}(\varepsilon_i^2 \mid \boldsymbol{x}_i) \boldsymbol{x}_i \boldsymbol{x}_i' \right) (\boldsymbol{X}'\boldsymbol{X})^{-1}$$

HC0与HC1稳健标准误

- 异方差稳健标准误 (Eicker, 1963,1967; Huber, 1967; White, 1980)

$$\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})}^{HC0} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\sum_{i=1}^{n} e_i^2 \boldsymbol{x}_i \boldsymbol{x}_i'\right)(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- 小样本校正的异方差稳健标准误（Stata的默认稳健标准误，选择项 robust）(Hinkley, 1977)

$$\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})}^{HC1} = \underbrace{\left(\frac{n}{n-K}\right)}_{\to 1}(\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\sum_{i=1}^{n} e_i^2 \boldsymbol{x}_i \boldsymbol{x}_i'\right)(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

HC2稳健标准误

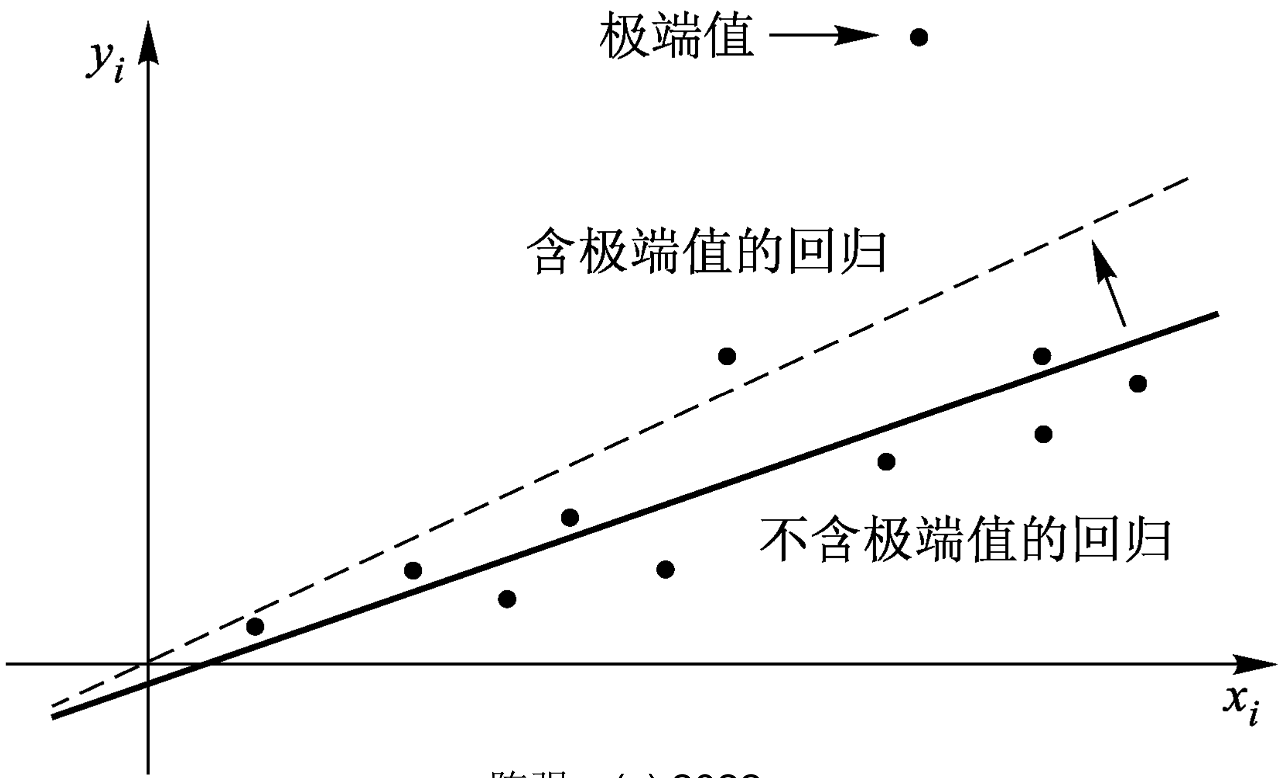
- 在同方差情况下，$\mathrm{E}(e_i^2)=(1-h_i)\sigma^2$， 故 $\mathrm{E}\left(\dfrac{e_i^2}{1-h_i}\right)=\sigma^2$
- 定义如下稳健标准误 (Horn, Horn and Duncan, 1975)：

$$\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}}\mid \boldsymbol{X})}^{HC2}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\sum_{i=1}^{n}\frac{e_i^2}{1-h_i}\boldsymbol{x}_i\boldsymbol{x}_i'\right)(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- 注：若 $\mathrm{E}(x_{ik}^2)<\infty$， 则 $\max\limits_{1\le i\le n} h_i \xrightarrow{p} 0$ 。
- Stata选择项：vce(hc2)

极端值的影响

- e_i为样本内拟合值，若存在极端值（outlier），则使用e_i估计扰动项 ε_i，可能有较大偏差



留一回归

- 使用“留一回归”（leave-one-out regression）进行样本外预测：

$$\hat{\boldsymbol{\beta}}_{(-i)} = \left(\mathbf{X}'_{(-i)}\mathbf{X}_{(-i)}\right)^{-1}\mathbf{X}'_{(-i)}\mathbf{y}_{(-i)} = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\tilde{e}_i$$

- 留一预测： $\tilde{y}_i = \mathbf{x}_i'\hat{\boldsymbol{\beta}}_{(-i)}$

- 留一残差： $\tilde{e}_i \equiv y_i - \tilde{y}_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_{(-i)} = \dfrac{e_i}{1-h_i}$

HC3稳健标准误

- 由于 $\tilde{e}_i = \dfrac{e_i}{1-h_i}$ ，故可定义如下稳健标准误：

$$\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})}^{HC3} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\sum_{i=1}^{n} \frac{e_i^2}{(1-h_i)^2}\boldsymbol{x}_i\boldsymbol{x}_i'\right)(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- Stata选择项：vec(hc3)

稳健标准误的大小比较

- 首先，$\widehat{\text{Var}(\hat{\boldsymbol{\beta}})}^{HC0} < \widehat{\text{Var}(\hat{\boldsymbol{\beta}})}^{HC1}$
- 这意味着 $\widehat{\text{Var}(\hat{\boldsymbol{\beta}})}^{HC1} - \widehat{\text{Var}(\hat{\boldsymbol{\beta}})}^{HC0}$ 为正定矩阵
- 其次，由于 $0 \le h_i \le 1$ ，故 $\dfrac{1}{(1-h_i)^2} \ge \dfrac{1}{1-h_i} \ge 1$
- 因此，$\widehat{\text{Var}(\hat{\boldsymbol{\beta}})}^{HC0} < \widehat{\text{Var}(\hat{\boldsymbol{\beta}})}^{HC2} < \widehat{\text{Var}(\hat{\boldsymbol{\beta}})}^{HC3}$

究竟用哪个稳健标准误？

- 大样本中，HC0, HC1, HC2, HC3都是一致估计

- 有限样本中，MacKinnon and White (1985)的蒙特卡洛模拟显示：

  HC3 优于 HC2 优于 HC1 优于 HC0

- 但后续研究发现，HC2有时优于HC3

- 结论：推荐HC2与HC3

3、刀切法 (Jackknife)

- 用 $\hat{\beta}$ 估计 β，留一回归可得 $\left\{\hat{\beta}_{(-1)},\cdots,\hat{\beta}_{(-n)}\right\}$

- 可否用其样本方差估计Var($\hat{\beta}$)？

$$\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_{(-i)}-\hat{\beta}_{(\cdot)}\right)^2,\quad \text{其中}\hat{\beta}_{(\cdot)}\equiv\frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_{(-i)}$$

- 刀切法：$\widehat{\mathrm{Var}(\hat{\beta})}^{jackknife}=\frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\beta}_{(-i)}-\hat{\beta}_{(\cdot)}\right)^2$

- 刀切法效率不如自助法，也不适用于非光滑（如中位数）与非线性（如相关系数）的情形

4、自助标准误(bootstrap s.e.)

- 有时估计量的标准误没有解析表达式，可考虑使用“自助标准误”(bootstrap standard error)。
- 自助法是一种有放回的再抽样(resampling)，保持样本容量不变。对样本数据 $\{y_i, \mathbf{x}_i\}_{i=1}^{n}$ 成对再抽样
- 使用自助法得到原始样本的B个自助样本（比如B=1000），对每个样本进行估计，得到B个估计值 $\left\{\hat{\boldsymbol{\beta}}^{*1}, \cdots, \hat{\boldsymbol{\beta}}^{*B}\right\}$，并计算这B个估计值的标准差。

自助标准误一致性的例外

1. 统计量不是样本数据的光滑函数：倾向得分匹配 (Propensity Score Matching)

2. 由于“厚尾”(fat tail)，不存在高阶矩：IV (2SLS, GMM等) 不建议用自助标准误

3. 待估参数位于参数空间的边界上

4. 使用CV选择调节参数的估计量：Lasso

自助法的种类

- 成对自助法 (pairs bootstrap)
- 残差自助法 (residual bootstrap)
- 野自助法(wild bootstrap): Bootstrap DGP is often close to the true DGP
- 约束野自助法 (restricted wild bootstrap imposing the null hypothesis)

残差自助法(residual bootstrap)

- 对于回归模型 $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$，首先得到OLS残差 $\hat{\varepsilon}_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}$

- 对残差 $\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \cdots, \hat{\varepsilon}_n\}$ 使用自助法，得到残差自助样本 $\{\hat{\varepsilon}_1^*, \hat{\varepsilon}_2^*, \cdots, \hat{\varepsilon}_n^*\}$，计算 $y_i^* = \mathbf{x}_i'\hat{\boldsymbol{\beta}} + \hat{\varepsilon}_i^*$，得到自助样本 $\{(y_1^*, \boldsymbol{x}_1), \cdots, (y_n^*, \boldsymbol{x}_n)\}$，获得 $\hat{\boldsymbol{\beta}}^*$，重复B次，得到 $\{\hat{\boldsymbol{\beta}}^{*1}, \cdots, \hat{\boldsymbol{\beta}}^{*B}\}$

- 数据矩阵 $\mathbf{X}$ 始终不变 (conditioning on $\mathbf{X}$)

约束残差自助法 (restricted residual bootstrap)

- 假设想检验原假设 $H_0: \mathbf{R\beta} = \mathbf{r}$ ， 而约束OLS估计量为 $\tilde{\boldsymbol{\beta}}$ ，约束残差 $\tilde{\varepsilon}_i$ 。

- 对残差残差 $\{\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \cdots, \tilde{\varepsilon}_n\}$ 使用自助法，并使用约束估计量 $\tilde{\boldsymbol{\beta}}$ ，则更有效率：

$$y_i^* = \mathbf{x}_i' \tilde{\beta} + \tilde{\varepsilon}_i^*$$

- 缺点：(约束或无约束)残差自助法不适用于异方差的数据

野自助法(wild bootstrap)

- 对于回归模型 $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$，首先得到OLS残差 $\hat{\varepsilon}_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}$ 。对此残差进行随机扰动：
- 对个体 i ，生成随机变量 $v_i = \begin{cases} 1 & 以1/2概率 \\ -1 & 以1/2概率 \end{cases}$
- 生成“伪残差”(pseudo residual): $\hat{\varepsilon}_i^* = v_i \hat{\varepsilon}_i$
- 使用伪残差 $\left\{\hat{\varepsilon}_1^*, \hat{\varepsilon}_2^*, \cdots, \hat{\varepsilon}_n^*\right\}$ 进行残差自助法。
- 给定 $\hat{\varepsilon}_i$ ，则 $\mathrm{E}(\hat{\varepsilon}_i^*) = 0,\ \mathrm{Var}(\hat{\varepsilon}_i^*) = \hat{\varepsilon}_i^2$

Rademacher Distribution

- 辅助随机变量(auxiliary random variable, wild weight) 服从两点分布：

$$v_i = \begin{cases} 1 & \text{以}1/2\text{概率} \\ -1 & \text{以}1/2\text{概率} \end{cases}$$

- 优点：伪残差可保持原残差的一阶、二阶与四阶矩。缺点：三阶矩必为0

- 给定 $\hat{\varepsilon}_i$ ， $\mathrm{E}(\hat{\varepsilon}_i^{*3}) = \mathrm{E}(v_i^3 \hat{\varepsilon}_i^3) = 0$

$$\mathrm{E}(\hat{\varepsilon}_i^{*4}) = \mathrm{E}(v_i^4 \hat{\varepsilon}_i^4) = \hat{\varepsilon}_i^4$$

Mammen (1993) Two-point Distribution

- 辅助随机变量服从两点分布：

$$v_i = \begin{cases} \dfrac{1-\sqrt{5}}{2}, & \text{以}\dfrac{\sqrt{5}+1}{2\sqrt{5}}\text{的概率} \\ \dfrac{1+\sqrt{5}}{2}, & \text{以}\dfrac{\sqrt{5}-1}{2\sqrt{5}}\text{的概率} \end{cases}$$

- 优点：伪残差可保持原残差的一阶、二阶与三阶矩。缺点：四阶矩变为原来的2倍

Webb(2014) Six-point Distribution

- 辅助随机变量服从六点分布：

$$
v_i = \begin{cases} \sqrt{3/2} & \text{以}1/6\text{的概率} \\ \sqrt{2/2} & \text{以}1/6\text{的概率} \\ \sqrt{1/2} & \text{以}1/6\text{的概率} \\ -\sqrt{1/2} & \text{以}1/6\text{的概率} \\ -\sqrt{2/2} & \text{以}1/6\text{的概率} \\ -\sqrt{3/2} & \text{以}1/6\text{的概率} \end{cases}
$$

- 优点：可保持一阶与二阶矩，伪残差取值变化多
  缺点：三阶矩必为0，四阶矩变为原来的 $7/6$ 倍

Stata案例1：Nerlove (1963)

- use nerlove.dta,clear
- reg lntc lnq lnpl lnpk lnpf
- est sto homo

- reg lntc lnq lnpl lnpk lnpf,r
- est sto hc1

- reg lntc lnq lnpl lnpk lnpf,vce(hc2)
- sto hc2

- reg lntc lnq lnpl lnpk lnpf,vce(hc3)
- est sto hc3

Stata案例1 (续)

- `reg lntc lnq lnpl lnpk lnpf,vce(jack)`
- `est sto jack`

- `reg lntc lnq lnpl lnpk lnpf, vce(boot,reps(500) seed(1))`
- `est sto boot`

- `esttab homo hc1 hc2 hc3 jack boot, se mtitle`

	(1) homo	(2) hc1	(3) hc2	(4) hc3	(5) jack	(6) boot
lnq	0.721***	0.721***	0.721***	0.721***	0.721***	0.721***
	(0.0174)	(0.0325)	(0.0330)	(0.0340)	(0.0339)	(0.0316)
lnpl	0.456	0.456	0.456	0.456	0.456	0.456
	(0.300)	(0.260)	(0.263)	(0.270)	(0.269)	(0.262)
lnpk	-0.215	-0.215	-0.215	-0.215	-0.215	-0.215
	(0.340)	(0.323)	(0.327)	(0.337)	(0.336)	(0.322)
lnpf	0.426***	0.426***	0.426***	0.426***	0.426***	0.426***
	(0.100)	(0.0741)	(0.0746)	(0.0765)	(0.0763)	(0.0754)
_cons	-3.567*	-3.567*	-3.567*	-3.567*	-3.567*	-3.567*
	(1.779)	(1.718)	(1.741)	(1.795)	(1.789)	(1.675)
N	145	145	145	145	145	145

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

人为制造一个极端值

- list lnq if _n==145

	lnq
145.	9.724301

- replace lnq = lnq*3 if _n==145
- list lnq if _n==145

	lnq
145.	29.1729

Stata案例1 (续2)

- qui reg lntc lnq lnpl lnpk lnpf
- est sto homo_o

- qui reg lntc lnq lnpl lnpk lnpf,r
- est sto hc1_o

- qui reg lntc lnq lnpl lnpk lnpf,vce(hc2)
- est sto hc2_o

- qui reg lntc lnq lnpl lnpk lnpf,vce(hc3)
- est sto hc3_o

Stata案例1 (续3)

- qui reg lntc lnq lnpl lnpk lnpf,vce(jack)
- est sto jack_o

- qui reg lntc lnq lnpl lnpk lnpf, vce(boot,reps(500) seed(1))
- est sto boot_o

- esttab homo_o hc1_o hc2_o hc3_o jack_o boot_o, se mtitle

	(1)	(2)	(3)	(4)	(5)	(6)
	homo_o	hc1_o	hc2_o	hc3_o	jack_o	boot_o
lnq	0.424***	0.424**	0.424*	0.424	0.424	0.424*
	(0.0278)	(0.149)	(0.207)	(0.293)	(0.292)	(0.178)
lnpl	0.112	0.112	0.112	0.112	0.112	0.112
	(0.671)	(0.545)	(0.566)	(0.613)	(0.610)	(0.494)
lnpk	-0.363	-0.363	-0.363	-0.363	-0.363	-0.363
	(0.757)	(0.711)	(0.722)	(0.750)	(0.748)	(0.662)
lnpf	0.224	0.224	0.224	0.224	0.224	0.224
	(0.223)	(0.182)	(0.209)	(0.258)	(0.257)	(0.180)
_cons	-0.0347	-0.0347	-0.0347	-0.0347	-0.0347	-0.0347
	(3.950)	(4.108)	(4.492)	(5.229)	(5.209)	(3.999)
N	145	145	145	145	145	145

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Stata案例2：Griliches (1976)

- use grilic.dta,clear
- ivregress 2sls lw s expr tenure rns smsa (iq=med kww mrt)
- est sto iv_homo
- ivregress 2sls lw s expr tenure rns smsa (iq=med kww mrt),r
- est sto iv_hc
- ivregress 2sls lw s expr tenure rns smsa (iq=med kww mrt),vce(jack)
- est sto iv_jack
- ivregress 2sls lw s expr tenure rns smsa (iq=med kww mrt),vce(boot,reps(500) seed(1))
- est sto iv_boot
- esttab iv_homo iv_hc iv_jack iv_boot, se mtitle

	(1)	(2)	(3)	(4)
	iv_homo	iv_hc	iv_jack	iv_boot
iq	0.00413	0.00413	0.00413	0.00413
	(0.00525)	(0.00552)	(0.00753)	(0.00815)
s	0.0902***	0.0902***	0.0902***	0.0902***
	(0.0168)	(0.0170)	(0.0229)	(0.0246)
expr	0.0397***	0.0397***	0.0397***	0.0397***
	(0.00657)	(0.00683)	(0.00718)	(0.00768)
tenure	0.0338***	0.0338***	0.0338***	0.0338***
	(0.00799)	(0.00814)	(0.00873)	(0.00866)
rns	-0.0721*	-0.0721*	-0.0721*	-0.0721*
	(0.0324)	(0.0323)	(0.0360)	(0.0366)
smsa	0.136***	0.136***	0.136***	0.136***
	(0.0282)	(0.0283)	(0.0290)	(0.0306)
_cons	3.841***	3.841***	3.841***	3.841***
	(0.344)	(0.370)	(0.494)	(0.532)
N	758	758	758	758

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Stata案例2 (续): GMM

- ivregress gmm lw s expr tenure rns smsa (iq=med kww mrt),vce(unadjusted)
- est sto gmm_homo
- ivregress gmm lw s expr tenure rns smsa (iq=med kww mrt),r
- est sto gmm_hc
- ivregress gmm lw s expr tenure rns smsa (iq=med kww mrt),vce(jack)
- est sto gmm_jack
- ivregress gmm lw s expr tenure rns smsa (iq=med kww mrt),vce(boot,reps(500) seed(1))
- est sto gmm_boot
- esttab gmm_homo gmm_hc gmm_jack gmm_boot, se mtitle

	(1)	(2)	(3)	(4)
	gmm_homo	gmm_hc	gmm_jack	gmm_boot
iq	0.00334	0.00334	0.00413	0.00413
	(0.00552)	(0.00552)	(0.00753)	(0.00815)
s	0.0969***	0.0969***	0.0902***	0.0902***
	(0.0169)	(0.0169)	(0.0229)	(0.0246)
expr	0.0412***	0.0412***	0.0397***	0.0397***
	(0.00680)	(0.00680)	(0.00718)	(0.00768)
tenure	0.0359***	0.0359***	0.0338***	0.0338***
	(0.00812)	(0.00812)	(0.00873)	(0.00866)
rns	-0.0789*	-0.0789*	-0.0721*	-0.0721*
	(0.0322)	(0.0322)	(0.0360)	(0.0366)
smsa	0.130***	0.130***	0.136***	0.136***
	(0.0283)	(0.0283)	(0.0290)	(0.0306)
_cons	3.825***	3.825***	3.841***	3.841***
	(0.370)	(0.370)	(0.494)	(0.532)
N	758	758	758	758

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

5.异方差自相关稳健的标准误

- 如果扰动项存在自相关 (常见于时间序列数据)，仍可用OLS来估计回归系数，但应使用“异方差自相关稳健的标准误”(Heteroskedasticity and Autocorrelation Consistent Standard Error，简记HAC)，即在存在异方差与自相关的情况下也成立的稳健标准误。

- 称为“Newey-West估计法”(Newey and West, 1987)，它只改变标准误的估计值，并不改变回归系数的估计值。也称“Newey-West标准误”。

时间序列的线性回归模型

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_K x_{tK} + \varepsilon_i \quad (t = 1, \cdots, n)$$

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta} + \varepsilon_t \quad (t = 1, \cdots, n)$$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{y}$$

$$\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}) \boldsymbol{X} (\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- 能否去掉期望，仍用 $\mathbf{e}$ 替代 $\boldsymbol{\varepsilon}$ ？但根据OLS的正交性：

$$\underbrace{\mathbf{X}'\mathbf{e}}_{=\mathbf{0}}\underbrace{\mathbf{e}'\mathbf{X}}_{=\mathbf{0}'} = \mathbf{0}$$

自协方差矩阵

- 抽样误差 $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(\frac{\boldsymbol{X}'\boldsymbol{X}}{n}\right)^{-1} \frac{1}{\sqrt{n}} (\boldsymbol{x}_1 \quad \cdots \quad \boldsymbol{x}_T) \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{pmatrix} = \left(\frac{\boldsymbol{X}'\boldsymbol{X}}{n}\right)^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \boldsymbol{x}_t \varepsilon_t$$

- 考虑 $\boldsymbol{g}_t = \boldsymbol{x}_t \varepsilon_t$ 滞后 j 阶的自协方差矩阵：

$$\boldsymbol{\Gamma}(j) \equiv \mathrm{Cov}(\boldsymbol{g}_t, \boldsymbol{g}_{t-j}) = \mathrm{Cov}(\boldsymbol{x}_t \varepsilon_t, \boldsymbol{x}_{t-j} \varepsilon_{t-j}) = \mathrm{E}(\varepsilon_t \varepsilon_{t-j} \boldsymbol{x}_t \boldsymbol{x}'_{t-j}) \neq \boldsymbol{\Gamma}(-j)$$

- 滞后 -j 阶（前移j 阶）的自协方差矩阵：

$$\boldsymbol{\Gamma}(-j) \equiv \mathrm{Cov}(\boldsymbol{g}_t, \boldsymbol{g}_{t+j}) = \mathrm{Cov}(\boldsymbol{x}_t \varepsilon_t, \boldsymbol{x}_{t+j} \varepsilon_{t+j}) = \mathrm{E}(\varepsilon_t \varepsilon_{t+j} \boldsymbol{x}_t \boldsymbol{x}'_{t+j})$$

$$\mathrm{E}(\varepsilon_t \varepsilon_{t-j} \boldsymbol{x}_t \boldsymbol{x}'_{t-j}) = \boldsymbol{\Gamma}(j) = \boldsymbol{\Gamma}(-j)' = \mathrm{E}(\varepsilon_{t+j} \varepsilon_t \boldsymbol{x}_{t+j} \boldsymbol{x}'_t)$$

长期方差矩阵

- $\boldsymbol{g}_t = \boldsymbol{x}_t \varepsilon_t$ 的长期方差矩阵(long-run variance matrix)

$$\mathrm{Var}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{x}_t\varepsilon_t\right) = \frac{1}{n}\mathrm{Var}\left(\boldsymbol{g}_1 + \boldsymbol{g}_2 + \cdots + \boldsymbol{g}_n\right)$$

$$= \frac{1}{n}\mathrm{Cov}\left(\boldsymbol{g}_1 + \boldsymbol{g}_2 + \cdots + \boldsymbol{g}_n,\ \boldsymbol{g}_1 + \boldsymbol{g}_2 + \cdots + \boldsymbol{g}_n\right)$$

$$= \frac{1}{n}\left[\boldsymbol{\Gamma}(-(n-1)) + \cdots + (n-1)\boldsymbol{\Gamma}(-1) + n\boldsymbol{\Gamma}(0) + (n-1)\boldsymbol{\Gamma}(1) + \cdots + \boldsymbol{\Gamma}(n-1)\right]$$

$$= \frac{1}{n}\boldsymbol{\Gamma}(-(n-1)) + \cdots + \frac{n-1}{n}\boldsymbol{\Gamma}(-1) + \boldsymbol{\Gamma}(0) + \frac{n-1}{n}\boldsymbol{\Gamma}(1) + \cdots + \frac{1}{n}\boldsymbol{\Gamma}(n-1)$$

$$= \sum_{j=-(n-1)}^{n-1}\left(1 - \frac{|j|}{n}\right)\boldsymbol{\Gamma}(j) \xrightarrow{p} \sum_{j=-\infty}^{\infty}\boldsymbol{\Gamma}(j) < \infty \quad \text{(by assumption)}$$

估计长期方差矩阵

- 天真的估计量　　　　注：$\boldsymbol{\Gamma}(j)=\mathrm{E}(\varepsilon_t\varepsilon_{t-j}\boldsymbol{x}_t\boldsymbol{x}'_{t-j})$

$$\hat{\boldsymbol{V}}=\sum_{j=-(n-1)}^{n-1}\hat{\boldsymbol{\Gamma}}(j)$$

$$=\hat{\boldsymbol{\Gamma}}(-(n-1))+\cdots+\hat{\boldsymbol{\Gamma}}(-1)+\hat{\boldsymbol{\Gamma}}(0)+\hat{\boldsymbol{\Gamma}}(1)+\cdots+\hat{\boldsymbol{\Gamma}}(n-1)$$

- 当 $j=0$ 时，$\hat{\boldsymbol{\Gamma}}(0)=\frac{1}{n}\sum_{t=1}^{n}e_t^2\boldsymbol{x}_t\boldsymbol{x}'_t$

- 当 $j=1,\cdots,n-1$ 时，$\hat{\boldsymbol{\Gamma}}(j)=\frac{1}{n}\sum_{t=j+1}^{n}e_te_{t-j}\boldsymbol{x}_t\boldsymbol{x}'_{t-j}$

估计长期方差矩阵 (2)

$$\hat{\boldsymbol{\Gamma}}(-j) = [\hat{\boldsymbol{\Gamma}}(-j)]' = \left(\frac{1}{n} \sum_{t=j+1}^{n} e_t e_{t-j} \boldsymbol{x}_t \boldsymbol{x}_{t-j}' \right)' = \frac{1}{n} \sum_{t=j+1}^{n} e_t e_{t-j} \boldsymbol{x}_{t-j} \boldsymbol{x}_t'$$

$$\begin{aligned}
\hat{\boldsymbol{V}} &= \sum_{j=-(n-1)}^{n-1} \hat{\boldsymbol{\Gamma}}(j) \\
&= \hat{\boldsymbol{\Gamma}}(-(n-1)) + \cdots + \hat{\boldsymbol{\Gamma}}(-1) + \hat{\boldsymbol{\Gamma}}(0) + \hat{\boldsymbol{\Gamma}}(1) + \cdots + \hat{\boldsymbol{\Gamma}}(n-1) \\
&= [\hat{\boldsymbol{\Gamma}}(n-1)]' + \cdots + [\hat{\boldsymbol{\Gamma}}(1)]' + \hat{\boldsymbol{\Gamma}}(0) + \hat{\boldsymbol{\Gamma}}(1) + \cdots + \hat{\boldsymbol{\Gamma}}(n-1) \\
&= \frac{1}{n} \sum_{t=1}^{n} e_t^2 \boldsymbol{x}_t \boldsymbol{x}_t' + \frac{1}{n} \sum_{j=1}^{n-1} \sum_{t=j+1}^{n} e_t e_{t-j} \left(\boldsymbol{x}_t \boldsymbol{x}_{t-j}' + \boldsymbol{x}_{t-j} \boldsymbol{x}_t' \right)
\end{aligned}$$

HAC标准误

- HAC标准误的核心是估计各阶自相关系数，并以此校正标准误（表达式较复杂）。

- 如果样本容量为n，则(n -1)阶自相关系数只有一个观测值，无法准确估计。

- 解决方法：确定截断参数(truncation parameter) $p = n^{1/4}$ 或 $p = 0.75n^{1/3}$，再取整数。

截断求和 (truncated sum)

- 设定截断参数，求部分和：

$$\hat{V} = \sum_{j=-p}^{p} \hat{\Gamma}(j)$$

$$= \frac{1}{n}\sum_{t=1}^{n} e_t^2 \boldsymbol{x}_t \boldsymbol{x}_t' + \frac{1}{n}\sum_{j=1}^{p}\sum_{t=j+1}^{n} e_t e_{t-j}\left(\boldsymbol{x}_t \boldsymbol{x}_{t-j}' + \boldsymbol{x}_{t-j}\boldsymbol{x}_t'\right) \xrightarrow{p} \boldsymbol{V}$$

- 但 $\hat{\mathbf{V}}$ 未必半正定(positive semidefinite)。解决方法：加权求和，以kernel function为权重

Newey-West标准误

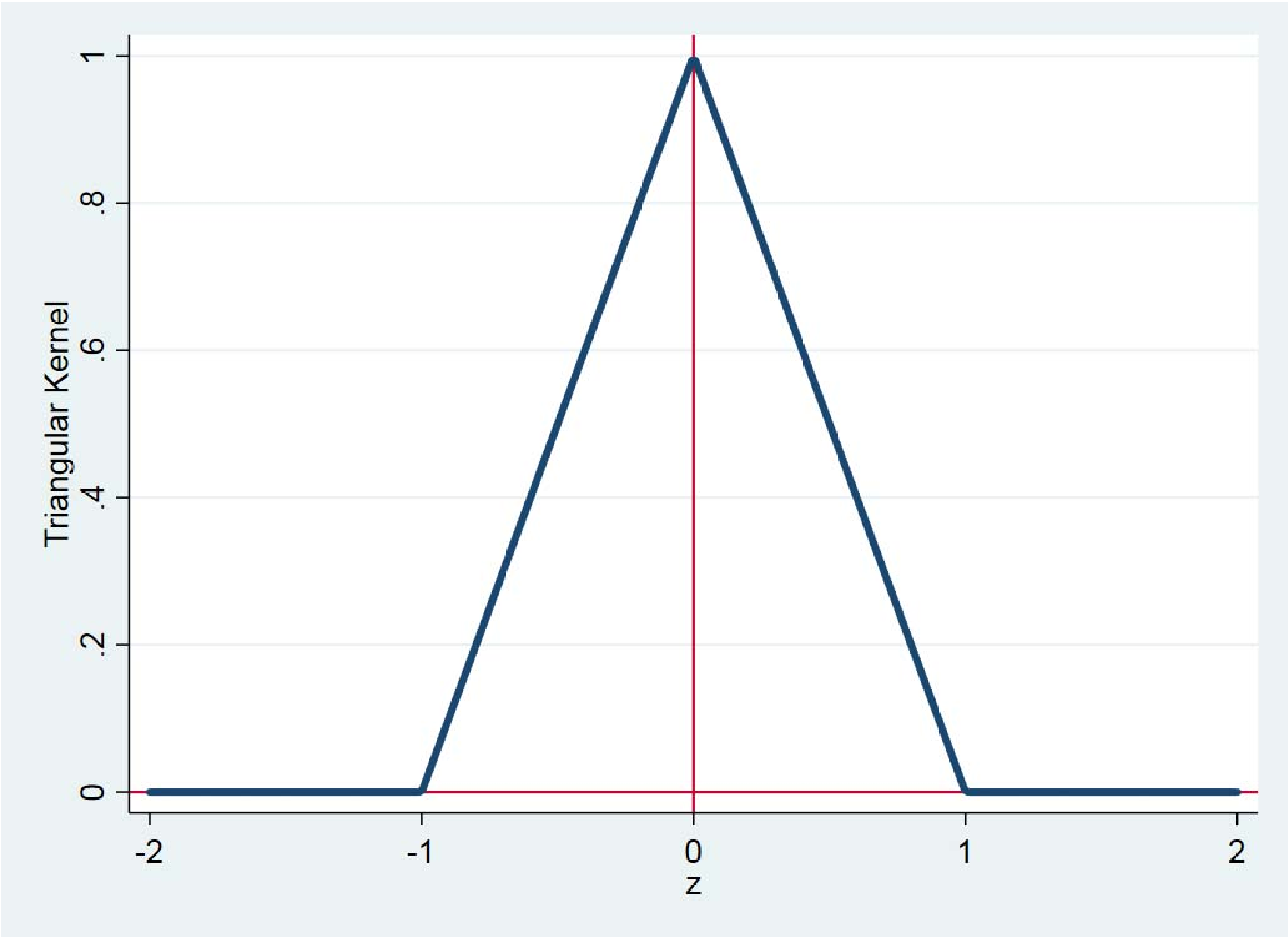
- j的绝对值越小，权重越大：$\hat{V}=\sum_{j=-p}^{p}k\left(\frac{j}{p+1}\right)\hat{\Gamma}(j)$
- $k(\cdot)$为核函数(kernel function)。Newey and West (1987)使用Bartlett kernel (triangular kernel)：

$$k(z)=(1-|z|)\mathbf{1}(|z|\le 1)$$

- 故

$$k\left(\frac{j}{p+1}\right)=\begin{cases}1-\frac{|j|}{p+1} & \text{if } -p\le j\le p\\ 0 & \text{if } |j|\ge p+1\end{cases}$$

三角核 (Triangular Kernel)



Stata的Newey-West标准误

- Stata所用的Newey-West标准误(小样本校正)：

$$\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}})}^{HAC} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\frac{n}{n-K}\right)\left[\sum_{t=1}^{n} e_t^2 \boldsymbol{x}_t\boldsymbol{x}_t' + \sum_{j=1}^{p}\left(1-\frac{j}{p+1}\right)\sum_{t=j+1}^{n} e_t e_{t-j}\left(\boldsymbol{x}_t\boldsymbol{x}_{t-j}' + \boldsymbol{x}_{t-j}\boldsymbol{x}_t'\right)\right](\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$= (\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\frac{n}{n-K}\right)\sum_{t=1}^{n} e_t^2 \boldsymbol{x}_t\boldsymbol{x}_t'(\boldsymbol{X}'\boldsymbol{X})^{-1} +$$

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\frac{n}{n-K}\right)\left[\sum_{j=1}^{p}\left(1-\frac{j}{p+1}\right)\sum_{t=j+1}^{n} e_t e_{t-j}\left(\boldsymbol{x}_t\boldsymbol{x}_{t-j}' + \boldsymbol{x}_{t-j}\boldsymbol{x}_t'\right)\right](\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$= \widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}})}^{HC1} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\frac{n}{n-K}\right)\left[\sum_{j=1}^{p}\left(1-\frac{j}{p+1}\right)\sum_{t=j+1}^{n} e_t e_{t-j}\left(\boldsymbol{x}_t\boldsymbol{x}_{t-j}' + \boldsymbol{x}_{t-j}\boldsymbol{x}_t'\right)\right](\boldsymbol{X}'\boldsymbol{X})^{-1}$$

- 若$p = 0$，则HAC标准误还原为HC1的稳健标准误
  故 `newey,lag(0)` 等价于 `reg,r`

自回归模型是否需要用HAC标准误？

- 考虑AR(1)模型： $y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$

- 一致估计要求$\{\varepsilon_t\}$无自相关，即$\mathrm{Cov}(\varepsilon_t, \varepsilon_s) = 0,\ \forall\, t \neq s$

- 由于y_{t-1}依赖于$\{\varepsilon_{t-1}, \cdots, \varepsilon_1\}$，而扰动项$\varepsilon_t$与$\{\varepsilon_{t-1}, \cdots, \varepsilon_1\}$不相关，故$y_{t-1}$与$\varepsilon_t$不相关，OLS一致。

- 如果$\{\varepsilon_t\}$有自相关，应考虑高阶自回归AR(*p*)，直至扰动项无自相关。故AR与ADL（自回归分布滞后）模型都不用HAC标准误。

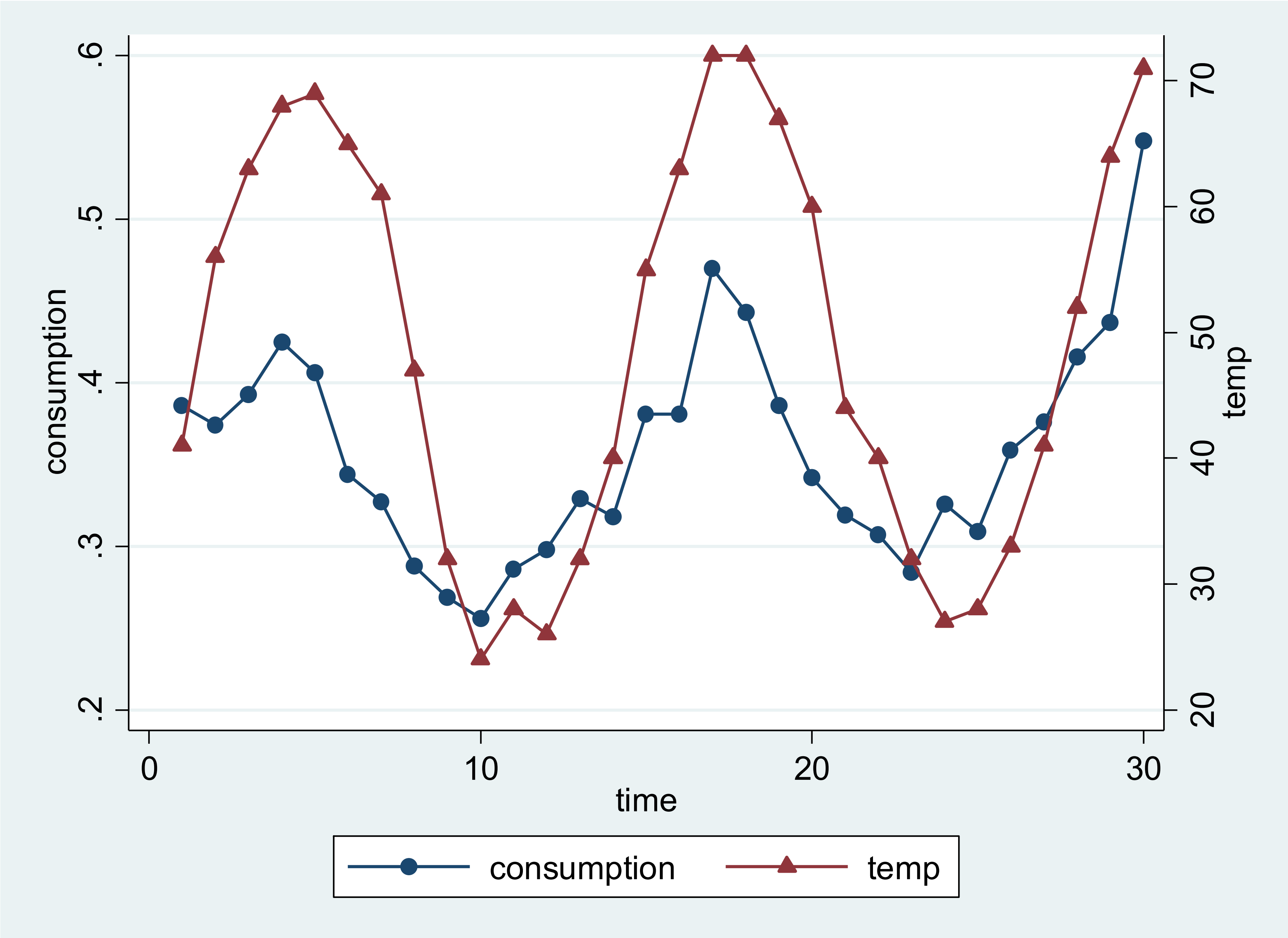
Stata案例3：冰淇淋的需求函数

- Hildreth and Lu(1960)对冰淇淋需求函数的经典研究。数据集icecream.dta包含了下列变量的30个月度时间序列数据：

- consumption(人均冰淇淋消费量)，income(平均家庭收入)，price(冰淇淋价格)，temp(平均华氏气温)，time(时间)。

时间变量与时间趋势图

- use icecream.dta,clear
- tsset time （宣布时间变量为time）
- twoway connect consumption time,msymbol(circle) yaxis(1) || connect temp time, msymbol(triangle) yaxis(2)
- 其中，“connect”表示将观测点用线连接起来，选择项“msymbol(circle)”与“msymbol(triangle)”分别表示点的“图标”(marker symbol)分别为圆圈与三角形；选择项“yaxis(1)”与“yaxis(2)”指定使用不同的纵坐标。


.6
.5
.4
.3
.2
consumption
70
60
50
40
30
20
temp
0
10
20
30
time
consumption
temp

OLS回归

- reg consumption temp price income

Source	SS	df	MS
Model	.090250523	3	.030083508
Residual	.035272835	26	.001356647
Total	.125523358	29	.004328392

Number of obs =	30
F(3, 26) =	22.17
Prob > F =	0.0000
R-squared =	0.7190
Adj R-squared =	0.6866
Root MSE =	.03683

consumption	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
temp	.0034584	.0004455	7.76	0.000	.0025426	.0043743
price	-1.044413	.834357	-1.25	0.222	-2.759458	.6706322
income	.0033078	.0011714	2.82	0.009	.0008999	.0057156
_cons	.1973149	.2702161	0.73	0.472	-.3581223	.752752

自相关检验

- estat bgodfrey

Breusch-Godfrey LM test for autocorrelation

lags(*p*)	chi2	df	Prob > chi2
1	4.237	1	0.0396

H0: no serial correlation

Newey-West回归

- 由于扰动项存在自相关，故应使用异方差自相关稳健的标准误。由于n^(1/4)=30^(1/4)=2.34，故取截断参数为p = 3：

- newey consumption temp price income,lag(3)

```
Regression with Newey-West standard errors          Number of obs  =        30
maximum lag: 3                                      F(  3,      26)  =     27.63
                                                    Prob > F         =    0.0000
```

consumption	Coef.	Newey-West Std. Err.	t	P>|t|	[95% Conf.	Interval]
temp	.0034584	.0004002	8.64	0.000	.0026357	.0042811
price	-1.044413	.9772494	-1.07	0.295	-3.053178	.9643518
income	.0033078	.0013278	2.49	0.019	.0005783	.0060372
_cons	.1973149	.3378109	0.58	0.564	-.4970655	.8916952

6、聚类数据 (Cluster Data)

- 如果样本可分为不同的“聚类”(clusters)，在同一聚类里的观测值互相相关，而不同聚类之间的观测值不相关，称为“聚类样本”(clustered sample)。

- 例：考察全国各地的电力企业。同一省的电力企业可能受到相同省政策的影响而自相关，但不同省之间的企业不相关。此时，“省”(province)被称为“聚类变量”(cluster variable)。

- 例：面板数据（每位个体不同时期的观测值构成聚类）

聚类稳健标准误

- 如果将观测值按聚类的归属顺序排列，则扰动项的协方差矩阵为“块对角”(block diagonal)。
- 仍可用OLS来估计系数，但需使用“聚类稳健标准误”(cluster robust standard error)，表达式更复杂些。聚类稳健标准误也是异方差稳健的（未假设同方差）。
- 使用聚类稳健标准误的前提：聚类数目 (面板数据的*n*)很大，聚类中的个体数 (面板数据的*T*) 无所谓(Hansen, 2007允许*T*趋无穷，即*n*大*T*也大的大面板)；农户数据以很多村为聚类，每个村有很多农户

以双下标表示聚类数据

- 以 $(y_{ig}, \mathbf{x}_{ig})$ 表示第 $g \in \{1, \cdots, G\}$ 个聚类中的第 $i \in \{1, \cdots, n_g\}$ 位个体，其中 $n = n_1 + \cdots + n_G$
- 第 g 个聚类：

$$\boldsymbol{y}_g \equiv \begin{pmatrix} y_{1g} \\ \vdots \\ y_{n_g g} \end{pmatrix}, \quad \boldsymbol{X}_g \equiv \begin{pmatrix} \boldsymbol{x}'_{1g} \\ \vdots \\ \boldsymbol{x}'_{n_g g} \end{pmatrix}_{n_g \times K}, \quad \boldsymbol{\varepsilon}_g \equiv \begin{pmatrix} \varepsilon_{1g} \\ \vdots \\ \varepsilon_{n_g g} \end{pmatrix}$$

$$\boldsymbol{y}_g = \boldsymbol{X}_g \boldsymbol{\beta} + \boldsymbol{\varepsilon}_g \quad \Rightarrow \quad \boldsymbol{y} \equiv \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_G \end{pmatrix} = \underbrace{\begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_G \end{pmatrix}}_{X} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_G \end{pmatrix} \equiv \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

$$\begin{aligned}
\mathrm{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}) &= (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}' \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}) \boldsymbol{X} (\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1} \begin{pmatrix} \boldsymbol{X}_1' & \cdots & \boldsymbol{X}_G' \end{pmatrix} \mathrm{E}\left[\begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_G \end{pmatrix} \begin{pmatrix} \boldsymbol{\varepsilon}_1' & \cdots & \boldsymbol{\varepsilon}_G' \end{pmatrix} \middle| \boldsymbol{X} \right] \begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_G \end{pmatrix} (\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1} \begin{pmatrix} \boldsymbol{X}_1' & \cdots & \boldsymbol{X}_G' \end{pmatrix} \mathrm{E}\left[\begin{pmatrix} \boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1' & & \\ & \ddots & \\ & & \boldsymbol{\varepsilon}_G\boldsymbol{\varepsilon}_G' \end{pmatrix} \middle| \boldsymbol{X} \right] \begin{pmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_G \end{pmatrix} (\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1} \left[\sum_{g=1}^{G} \boldsymbol{X}_g' \, \mathrm{E}\left(\boldsymbol{\varepsilon}_g \boldsymbol{\varepsilon}_g' \middle| \boldsymbol{X}_g \right) \boldsymbol{X}_g \right] (\boldsymbol{X}'\boldsymbol{X})^{-1}
\end{aligned}$$

$$\widehat{\operatorname{Var}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})}$$

$$= \frac{n-1}{n-K} \cdot \frac{G}{G-1} (\boldsymbol{X}'\boldsymbol{X})^{-1} \left(\sum_{g=1}^{G} \boldsymbol{X}_g' \boldsymbol{e}_g \boldsymbol{e}_g' \boldsymbol{X}_g \right) (\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$= \frac{n-1}{n-K} \cdot \frac{G}{G-1} (\boldsymbol{X}'\boldsymbol{X})^{-1} \left(\sum_{g=1}^{G} \sum_{i=1}^{n_g} \sum_{j=1}^{n_g} e_{ig} e_{jg} \boldsymbol{x}_{ig} \boldsymbol{x}_{jg}' \right) (\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$= \frac{n-1}{n-K} \cdot \frac{G}{G-1} (\boldsymbol{X}'\boldsymbol{X})^{-1} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} e_i e_j \boldsymbol{x}_i \boldsymbol{x}_j' \boldsymbol{1}(i, j \text{ in same cluster}) \right) (\boldsymbol{X}'\boldsymbol{X})^{-1}$$

其中， $\boldsymbol{X}_g \equiv \begin{pmatrix} \boldsymbol{x}_{1g}' \\ \vdots \\ \boldsymbol{x}_{n_g g}' \end{pmatrix}_{n_g \times K}$ ， $\boldsymbol{e}_g \equiv \begin{pmatrix} e_{1g} \\ \vdots \\ e_{n_g g} \end{pmatrix}$

注：若$G = n$ (每位个体自成聚类)，则等价于HC1

Stata案例4: Duflo et al. (2011)

Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya†

By Esther Duflo, Pascaline Dupas, and Michael Kremer*

To the extent that students benefit from high-achieving peers, tracking will help strong students and hurt weak ones. However, all students may benefit if tracking allows teachers to better tailor their instruction level. Lower-achieving pupils are particularly likely to benefit from tracking when teachers have incentives to teach to the top of the distribution. We propose a simple model nesting these effects and test its implications in a randomized tracking experiment conducted with 121 primary schools in Kenya. While the direct effect of high-achieving peers is positive, tracking benefited lower-achieving pupils indirectly by allowing teachers to teach to their level. (*JEL* I21, J45, O15)

Stata案例4 (续)

- `use DDK2011.dta,clear`
- `reg stdR_totalscore tracking`
- `est sto homo`
- `reg stdR_totalscore tracking,r`
- `est sto hc`
- `reg stdR_totalscore tracking, cluster(schoolid)`
- `est sto cluster`
- `esttab homo hc cluster,se mtitle`

- Note：Data downloaded from AER, replicates Table 2 Column 1 of Duflo et al. (2011)。变量tracking在聚类中无变化，故无法控制聚类固定效应

```
. reg stdR_totalscore tracking, cluster(schoolid)

Linear regression                               Number of obs     =      5,795
                                                F(1, 120)         =       3.20
                                                Prob > F          =     0.0763
                                                R-squared         =     0.0048
                                                Root MSE          =     1.0089

                                 (Std. err. adjusted for 121 clusters in schoolid)
```

stdR_total~e	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
tracking	.1396412	.0781031	1.79	0.076	-.0149975	.2942799
_cons	-4.28e-08	.0550039	-0.00	1.000	-.1089039	.1089038

Cluster-robust SE Often (Much) Larger

	(1)	(2)	(3)
	homo	hc	cluster
tracking	0.140***	0.140***	0.140
	(0.0265)	(0.0265)	(0.0781)
_cons	-4.28e-08	-4.28e-08	-4.28e-08
	(0.0190)	(0.0189)	(0.0550)
N	5795	5795	5795

Standard errors in parentheses

* p<0.05, ** p<0.01, *** p<0.001

如何选择聚类变量

- 例：学生 -- 班级 -- 学校 -- 学区 -- 城市
- 例：农户 -- 村庄 -- 乡镇 -- 县 -- 地级市 -- 省
- 聚类到底层：聚类数目多，方差小；但可能忽略聚类之间的相关性，导致偏差
- 聚类到高层：偏差小；但聚类数目少，方差大

如何选择聚类变量 (续)

- 经验规则 #1 (Cameron and Miller, 2015)：只要聚类数目足够多，则聚类到最高层

- 经验规则 #2 (Cameron and Miller, 2015)：从底层向高层聚类，直到标准误基本不变

- 经验规则 #3 (Angrist and Pischke, 2008)：保守做法：聚类到标准误最大的那一层

需要多少聚类数目

- Angrist and Pischke (2009, sec 8.2.3) are skeptical about the reliability of clustered errors when the number of clusters is less than 42.
- Bertrand, Duflo and Mullainathan (QJE 2004, Table VIII) present evidence that as few as 20 clusters may be sufficient.
- 非平衡面板(每个聚类包含不同的个体数)需要更多的聚类数目

双向聚类 (Two-way Clustering)

- 企业面板：同时以firm与year作为聚类变量

$$\boldsymbol{y}_{gh} = \boldsymbol{X}_{gh}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{gh}, \quad g = 1, \cdots, G; h = 1, \cdots, H$$

$$\widehat{\boldsymbol{X}' \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X})\boldsymbol{X}} = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} e_i e_j \boldsymbol{x}_i \boldsymbol{x}_j' \boldsymbol{1}(i, j \text{ share any of the two clusters}) \right)$$

$$= \left(\sum_{i=1}^{n} \sum_{j=1}^{n} e_i e_j \boldsymbol{x}_i \boldsymbol{x}_j' \boldsymbol{1}(i, j \text{ in same cluster } g) \right)$$

$$+ \left(\sum_{i=1}^{n} \sum_{j=1}^{n} e_i e_j \boldsymbol{x}_i \boldsymbol{x}_j' \boldsymbol{1}(i, j \text{ in same cluster } h) \right)$$

$$- \left(\sum_{i=1}^{n} \sum_{j=1}^{n} e_i e_j \boldsymbol{x}_i \boldsymbol{x}_j' \boldsymbol{1}(i, j \text{ in both clusters } g \text{ and } h) \right)$$

双向聚类 (续)

$$\widehat{\mathrm{Var}_{twoway}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})} = \widehat{\mathrm{Var}_{G}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})} + \widehat{\mathrm{Var}_{H}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})} - \widehat{\mathrm{Var}_{G\cap H}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})}$$

- $\widehat{\mathrm{Var}_{G}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})}$：以*G*为聚类变量的聚类协方差矩阵
- $\widehat{\mathrm{Var}_{H}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})}$：以*H*为聚类变量的聚类协方差矩阵
- $\widehat{\mathrm{Var}_{G\cap H}(\hat{\boldsymbol{\beta}} \mid \boldsymbol{X})}$：以*GH*为聚类变量的聚类协方差矩阵
- 要求 $\min(G, H) \to \infty$ 。
- 可推广至“多向聚类 ” (multi-way clustering)

Robust Inference With Multiway Clustering

JBES, 2011

A. Colin CAMERON
Department of Economics, University of California–Davis, One Shields Ave, Davis, CA

Jonah B. GELBACH
Yale Law School, 127 Wall Street, New Haven, CT 06511

Douglas L. MILLER
Department of Economics, University of California–Davis, One Shields Ave, Davis, CA 95616
(*dlmiller@ucdavis.edu*)

In this article we propose a variance estimator for the OLS estimator as well as for nonlinear estimators such as logit, probit, and GMM. This variance estimator enables cluster-robust inference when there is two-way or multiway clustering that is nonnested. The variance estimator extends the standard cluster-robust variance estimator or sandwich estimator for one-way clustering (e.g., Liang and Zeger 1986; Arellano 1987) and relies on similar relatively weak distributional assumptions. Our method is easily implemented in statistical packages, such as Stata and SAS, that already offer cluster-robust standard errors when there is one-way clustering. The method is demonstrated by a Monte Carlo analysis for a two-way random effects model; a Monte Carlo analysis of a placebo law that extends the state–year effects example of Bertrand, Duflo, and Mullainathan (2004) to two dimensions; and by application to studies in the empirical literature where two-way clustering is present.

双向聚类稳健标准误

- 双向聚类稳健标准误：下载命令cgmreg

http://cameron.econ.ucdavis.edu/research/cgmreg.ado

1. 复制此网页的文本文件；
2. 粘贴至Stata的Do Editor；
3. 存为ado文件(扩展名为ado)；
4. 存入Stata的plus文件夹 (可用命令sysdir查找)

```
   STATA:  C:\Program Files\Stata17\
    BASE:  C:\Program Files\Stata17\ado\base\
    SITE:  C:\Program Files\Stata17\ado\site\
    PLUS:  C:\Users\DELL\ado\plus\
PERSONAL:  C:\Users\DELL\ado\personal\
OLDPLACE:  c:\ado\
```

多向聚类稳健标准误

- cgmreg只能估计双向聚类稳健标准误
- 多向聚类稳健标准误：
  ssc install clus_nway
- clus_nway建立在cgmreg的基础上，前者允许多向聚类，但使用*t*分布计算*p*值

Stata案例5：双向聚类

- webuse nlsw88, clear
- reg wage tenure ttl_exp collgrad, cluster(industry)

```
Linear regression                               Number of obs     =      2,217
                                                F(3, 11)          =      56.79
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1255
                                                Root MSE          =      5.402

                             (Std. err. adjusted for 12 clusters in industry)
```

wage	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
tenure	.0304488	.0282654	1.08	0.304	-.0317629	.0926605
ttl_exp	.2729597	.0469742	5.81	0.000	.1695701	.3763492
collgrad	3.255491	.3967566	8.21	0.000	2.382235	4.128746
_cons	3.419859	.2853613	11.98	0.000	2.791783	4.047935

Stata案例5 (续)

- reg wage tenure ttl_exp collgrad, cluster(age)

```
Linear regression                               Number of obs     =      2,231
                                                F(3, 12)          =     111.16
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1270
                                                Root MSE          =     5.3895

                                   (Std. err. adjusted for 13 clusters in age)
```

wage	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
tenure	.0321168	.0202474	1.59	0.139	-.0119984	.076232
ttl_exp	.273864	.0333345	8.22	0.000	.2012343	.3464938
collgrad	3.251781	.2816042	11.55	0.000	2.638218	3.865344
_cons	3.389638	.3316122	10.22	0.000	2.667117	4.112159

Stata案例5 (续2)

- cgmreg wage tenure ttl_exp collgrad, cluster(industry age)

```
Note: +/- means the corresponding matrix is added/subtracted

Calculating cov part for variables:  industry (+)
Calculating cov part for variables:  industry age (-)
Calculating cov part for variables:  age (+)

                                                Number of obs      =     2217
                                                Num clusvars       =     2
                                                Num combinations   =     3

                                                G(industry)        =     12
                                                G(age)             =     13
```

wage	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
tenure	.0304488	.0278042	1.10	0.273	-.0240465	.0849441
ttl_exp	.2729597	.0495781	5.51	0.000	.1757884	.3701309
collgrad	3.255491	.3613499	9.01	0.000	2.547258	3.963723
_cons	3.419859	.3528189	9.69	0.000	2.728347	4.111371

Stata案例5 (续3)

- clus_nway reg wage tenure ttl_exp collgrad, cluster(industry age)

```
                                    Number of obs       =    2217
                                    Num clusvars        =    2
                                    Num combinations    =    3

                                    G(age) =      13
                                    G(industry) =      12
Dependent variable: wage
                              (Std. err. adjusted for 12 clusters in __000006)
```

wage	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
tenure	.0304488	.0278042	1.10	0.297	-.0307479	.0916456
ttl_exp	.2729597	.0495781	5.51	0.000	.1638391	.3820803
collgrad	3.255491	.3613499	9.01	0.000	2.460165	4.050816
_cons	3.419859	.3528189	9.69	0.000	2.64331	4.196408

聚类数目太少(few clusters)怎么办？

- Use leave-one-out residuals $\tilde{e}_i$ (Bell and McCaffrey, 2002)，类似于HC2与HC3
- Leave-one-cluster-out jackknife standard errors
- Cluster bootstrap
- Wild cluster bootstrap (推荐)

Wild Cluster Bootstrap

- Wild cluster bootstrap unrestricted (WCU)

$$\boldsymbol{y}_g^* = \boldsymbol{X}_g \hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}_g^*, \quad \boldsymbol{\varepsilon}_g^* = v_g^* \boldsymbol{e}_g$$

- Wild cluster bootstrap restricted (WCR)

$$\boldsymbol{y}_g^* = \boldsymbol{X}_g \tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\varepsilon}}_g^*, \quad \tilde{\boldsymbol{\varepsilon}}_g^* = v_g^* \tilde{\boldsymbol{e}}_g$$

- $\tilde{\boldsymbol{\beta}}$ 与 $\tilde{\boldsymbol{e}}_g$分别为H_0下的约束估计量与约束残差

The Stata Journal (2019)
19, Number 1, pp. 4–60 DOI: 10.1177/1536867X19830877

Fast and wild: Bootstrap inference in Stata using boottest

David Roodman
Open Philanthropy Project
San Francisco, CA
david.roodman@openphilanthropy.org

James G. MacKinnon
Queen's University
Kingston, Canada
jgm@econ.queensu.ca

Morten Ørregaard Nielsen
Queen's University
Kingston, Canada
and CREATES, Aarhus University, Denmark
mon@econ.queensu.ca

Matthew D. Webb
Carleton University
Ottawa, Canada
matt.webb@carleton.ca

Abstract. The wild bootstrap was originally developed for regression models with heteroskedasticity of unknown form. Over the past 30 years, it has been extended to models estimated by instrumental variables and maximum likelihood and to ones where the error terms are (perhaps multiway) clustered. Like bootstrap methods in general, the wild bootstrap is especially useful when conventional inference methods are unreliable because large-sample assumptions do not hold.

Cluster-robust *t* statistic

For concreteness, we suppose that our objective is to test the hypothesis that $\beta_j = 0$. Then the WCU and WCR algorithms are as follows:

1. Regress $\mathbf{y}$ on $\mathbf{X}$ to obtain $\widehat{\boldsymbol{\beta}}$, $\hat{\mathbf{u}}$, and $\widehat{\mathbf{V}}$ as given in (3), (4), and (12).

2. Calculate the usual cluster–robust t statistic for the hypothesis that $\beta_j = 0$,

$$t_j = \frac{\widehat{\beta}_j}{\sqrt{\widehat{V}_{jj}}} \tag{16}$$

   where $\widehat{V}_{jj}$ is the jth diagonal element of $\widehat{\mathbf{V}}$.

3. For the WCU bootstrap, set $\ddot{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}$ and $\ddot{\mathbf{u}} = \hat{\mathbf{u}}$. For the WCR bootstrap, regress $\mathbf{y}$ on $\mathbf{X}$ subject to the restriction that $\beta_j = 0$ to obtain $\widetilde{\boldsymbol{\beta}}$ and $\tilde{\mathbf{u}}$, and set $\ddot{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}$ and $\ddot{\mathbf{u}} = \tilde{\mathbf{u}}$.

Percentile-*t* Bootstrap

4. For each of B bootstrap replications:

   a. Using (15), generate a new set of bootstrap error terms $\mathbf{u}^{*b}$ and dependent variables $\mathbf{y}^{*b}$.

   b. By analogy with step 1, regress $\mathbf{y}^{*b}$ on $\mathbf{X}$ to obtain $\widehat{\boldsymbol{\beta}}^{*b}$ and $\widehat{\mathbf{u}}^{*b}$, and use the latter in (12) to compute the bootstrap CRVE $\widehat{\mathbf{V}}^{*b}$.

   c. Calculate the bootstrap t statistic for the null hypothesis that $\beta_j^{*b} = \ddot{\beta}_j$ as

$$t_j^{*b} = \frac{\widehat{\beta}_j^{*b} - \ddot{\beta}_j}{\sqrt{\widehat{V}_{jj}^{*b}}} \tag{17}$$

   where $\widehat{V}_{jj}^{*b}$ is the jth diagonal element of $\widehat{\mathbf{V}}^{*b}$. For the WCR bootstrap, the numerator of (17) can also be written as $\widehat{\beta}_j^{*b}$ because $\ddot{\beta}_j = 0$.

5. For a one-tailed test, use the distribution of all the t_j^{*b} to compute the lower-tail or upper-tail p-value,

usual cluster-robust t-statistic

$$\hat{P}_{\mathrm{L}}^* = \frac{1}{B}\sum_{b=1}^{B}\mathbb{I}(t_j^{*b} < t_j) \quad \text{or} \quad \hat{P}_{\mathrm{U}}^* = \frac{1}{B}\sum_{b=1}^{B}\mathbb{I}(t_j^{*b} > t_j) \tag{18}$$

where $\mathbb{I}(\cdot)$ is the indicator function. For a two-tailed test, compute either the symmetric or the equal-tail p-value,

$$\hat{P}_{\mathrm{S}}^* = \frac{1}{B}\sum_{b=1}^{B}\mathbb{I}\big(|t_j^{*b}| > |t_j|\big) \quad \text{or} \quad \hat{P}_{\mathrm{ET}}^* = 2\min(\hat{P}_{\mathrm{L}}^*, \hat{P}_{\mathrm{U}}^*) \tag{19}$$

The former is appropriate if the distribution of t_j is symmetric around a mean of zero. In that case, if B is not very small, $\hat{P}_{\mathrm{S}}^*$ and $\hat{P}_{\mathrm{ET}}^*$ will be similar. If the

- 命令boottest默认使用常见的“对称*p*值”(symmetric *p*-value)

We make a few observations on this algorithm. First, whether the WCR or WCU bootstrap is used in step 3, the t statistic calculated in step 4c tests a hypothesis, $\beta_j = \ddot{\beta}_j$, that is true by construction in the bootstrap samples because $\ddot{\boldsymbol{\beta}}$ is used to construct the samples in step 4a. Because the bootstrap distribution is generated by testing a null hypothesis on samples from a DGP for which the null is correct, the resulting bootstrap distribution should mimic the distribution of the sample test statistic, t_j, when the null hypothesis of interest, $\beta_j = 0$, is also correct.

Second, because the small-sample correction factor m defined in (13) and used in the CRVE (12) affects both t_j and the t_j^{*b} proportionally, the choice of m does not affect any of the bootstrap p-values.

Third, the algorithm does not produce standard errors (which is why `boottest` does not attempt to compute them). Instead, inference is based on p-values and confidence sets; the latter are discussed below in section 3.5. One could compute the standard deviation of the bootstrap distribution of $\widehat{\boldsymbol{\beta}}^{*b}$ and then use it for inference in several ways. However, this approach relies heavily on the asymptotic normality of $\widehat{\boldsymbol{\beta}}$ in a case where large-sample theory may not apply. Higher-order asymptotic theory for the bootstrap (Davison and Hinkley 1997, chap. 5) predicts that this approach should not perform as well as the algorithm discussed above, and Monte Carlo simulations in CGM (2008) confirm this prediction.

3.2 Imposing the null on the bootstrap DGP

The bootstrap algorithm defined above lets the wild bootstrap DGP (15) either impose the restriction being tested (WCR) or not (WCU). Usually, it is better to impose the restriction. For any bootstrap DGP like (15) that depends on estimated parameters, those parameters are estimated more efficiently when restricted estimates are used (Davidson and MacKinnon 1999). Intuitively, because inference involves estimating the probabilities of obtaining certain results under the assumption that the null is true, inference is improved by using bootstrap datasets in which the null in fact holds. Simulation evidence on this issue is presented in, among many others, Davidson and MacKinnon (1999) and Djogbenou, MacKinnon, and Nielsen (2018).

For this reason, `boottest` uses the restricted estimates $\tilde{\beta}$ and restricted residuals $\tilde{\mathbf{u}}$ by default.[4] Nevertheless, `boottest` does allow the use of unrestricted estimates. This can be useful because WCU makes it possible to invert a hypothesis test to calculate confidence intervals for all parameters using just one set of bootstrap samples, whereas WCR requires constructing many sets of bootstrap samples to do so; see Davidson and MacKinnon (2004, sec. 5.3) and section 3.5. Thus, if the computational cost of WCR is prohibitive (although it rarely is with `boottest`), WCU is a practical alternative.

4.2 Wild-bootstrapping the multiway CRVE

The wild bootstrap and the multiway CRVE appear to have been combined first in the package `cgmwildboot` (Caskey 2010). Bringing them together raises several practical issues, in addition to those discussed in the context of the one-way wild cluster bootstrap.

The first practical issue that arises in wild-bootstrapping the multiway CRVE is what to do when some of the bootstrap variance matrices, $\widehat{\mathbf{V}}^{*b}$, are not positive definite. In their simulations, MacKinnon, Nielsen, and Webb (2017) apply the CGM (2011) correction (27) to these, too. In contrast, `boottest` merely omits instances where the test statistic is degenerate from the bootstrap distribution and decrements the value of B accordingly.[7] Like the CGM (2011) approach, deleting infeasible statistics from the bootstrap distribution is atheoretical. However, reassuringly, rerunning the Monte Carlo experiments of MacKinnon, Nielsen, and Webb (2017) with the CGM (2011) correction disabled suggests that the change has little effect in the cases examined in those experiments.[8]

The second practical issue is that, in contrast with the one-way case, the choice of small-sample correction factors now affects results. `boottest` applies CGM's (2011) proposed values for m_G, m_H, and m_{GH} in (26) also to each bootstrap sample for the reasons discussed above. Because each component of (26) is scaled by a different factor, the scaling affects the actual and bootstrap CRVEs differently. Although the impact is likely minor in most cases, it might not be when at least one of G and H is very small. An alternative would be to set all three factors to $\max(m_G, m_H)$, as `ivreg2` does (or would, if it were bootstrapped). If this were done for both the actual and bootstrap CRVEs, it would be equivalent to using no small-sample correction at all.

The third issue is that the elegant symmetry of the multiway CRVE formula does not carry over naturally to the wild bootstrap. The wild cluster bootstrap is designed to preserve the pattern of correlations within each cluster for one-way clustering, but it cannot preserve the correlations in two or more dimensions at once. Therefore, we must

7 The boottest package

The `boottest` package performs wild bootstrap tests of linear hypotheses. It is compatible with Stata versions back to 11.0, but it runs faster in Stata versions 13.0 and later because they include the Mata `panelsum()` function. The syntax is modeled on that of Stata's built-in command for Wald testing, `test`. Like `test`, but unlike other Stata implementations of the wild bootstrap, `boottest` is a postestimation command. It determines the context for inference from the current estimation results.

The following three commands implement the extended example in section 5:

```
webuse nlsw88, clear
regress wage tenure ttl_exp collgrad, cluster(industry)
boottest tenure
```

Here, by default, `boottest` generates 999 wild cluster bootstrap samples using the Rademacher distribution, with the null hypothesis imposed. It reports the t statistic from the Wald test and its bootstrapped p-value (by default, symmetric). It then automatically inverts the test, as described in section 3.5, and reports the bounds of the confidence set for the default level of confidence, which is normally 95%. Finally, it plots the "confidence curve" underlying this calculation, that is, the bootstrap p-value for the hypothesis $\beta_{\texttt{tenure}} = c$ as a function of c. It marks the points where the curve crosses 0.05, which are the limits of the confidence set.

Stata案例6

- `webuse nlsw88, clear`

- `regress wage tenure ttl_exp collgrad, cluster(industry)`

- 聚类数目仅为12

Linear regression		
	Number of obs =	2,217
	F(3, 11) =	56.79
	Prob > F =	0.0000
	R-squared =	0.1255
	Root MSE =	5.402

(Std. err. adjusted for 12 clusters in industry)

wage	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
tenure	.0304488	.0282654	1.08	0.304	-.0317629	.0926605
ttl_exp	.2729597	.0469742	5.81	0.000	.1695701	.3763492
collgrad	3.255491	.3967566	8.21	0.000	2.382235	4.128746
_cons	3.419859	.2853613	11.98	0.000	2.791783	4.047935

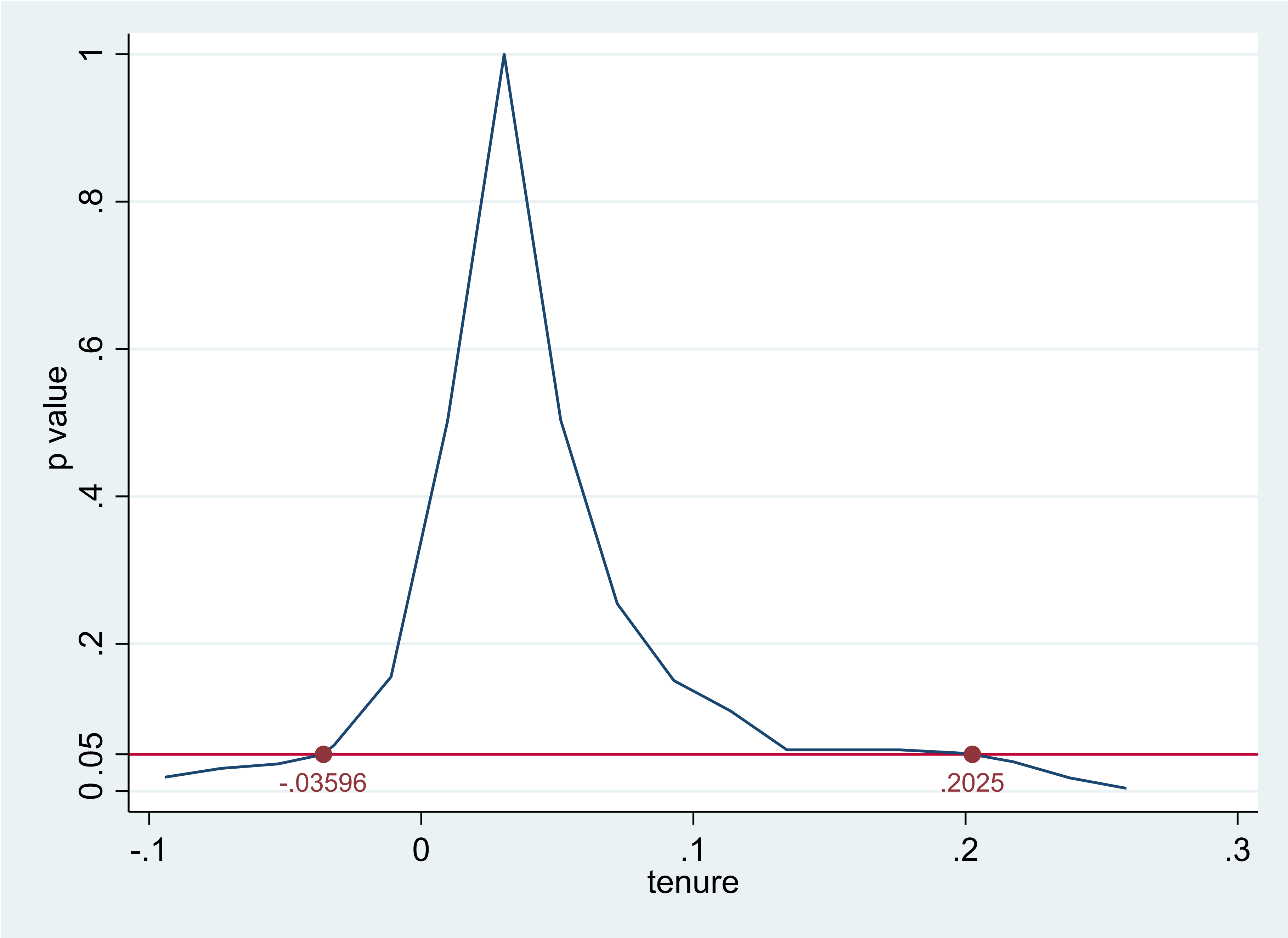
Wild Bootstrapping *t*-statistic

- boottest tenure,seed(1)

```
Wild bootstrap-t, null imposed, 999 replications, Wald test, bootstrap clusteri
> ng by industry, Rademacher weights:
  tenure

                           t(11) =     1.0772
                        Prob>|t| =     0.2993

95% confidence set for null hypothesis expression: [-.03596, .2025]
```


p value
1
.8
.6
.4
.2
0.05
0
-.03596
.2025
-.1
0
.1
.2
.3
tenure

Unrestricted with Webb Distribution

- boottest tenure,seed(1) weighttype(webb) nonull nograph

```
Wild bootstrap-t, null not imposed, 999 replications, Wald test, bootstrap clus
> tering by industry, Webb weights:
  tenure

                        t(11) =     1.0772
                     Prob>|t| =     0.3163

95% confidence set for null hypothesis expression: [-.02067, .08157]
```

Two-way Wild Clustering

- `boottest tenure,seed(1) cluster(industry age)`

```
Wild bootstrap-t, null imposed, 999 replications, Wald test, clustering by indu
> stry age, bootstrap clustering by industry age, Rademacher weights:
  tenure

                          t(11) =     1.0951
                       Prob>|t| =     0.3539

95% confidence set for null hypothesis expression: [-.05359, .1122]
```

Joint Test

- boottest tenure ttl_exp, seed(1)

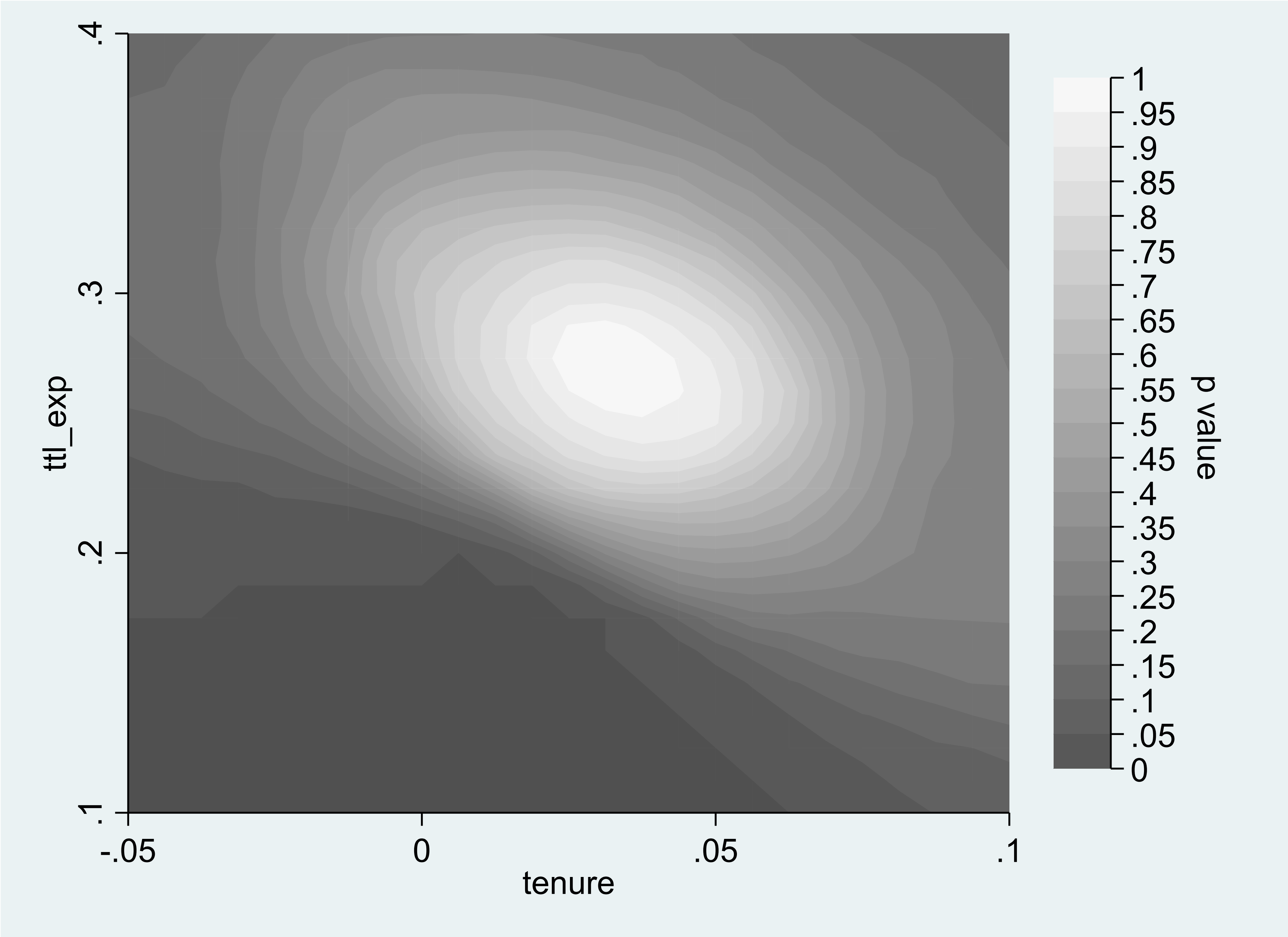
```
Overriding estimator's cluster/robust settings with cluster(industry age)
(bootcluster(industry age) assumed)

Warning: 41 replications returned an infeasible test statistic and were deleted
>  from the bootstrap distribution.

Wild bootstrap-t, null imposed, 999 replications, Wald test, clustering by indu
> stry age, bootstrap clustering by industry age, Rademacher weights:
  tenure

                           t(11) =     1.0951
                        Prob>|t| =     0.3539

95% confidence set for null hypothesis expression: [-.05359, .1122]
```


ttl_exp
.1
.2
.3
.4
-.05
0
.05
.1
tenure
p value
0
.05
.1
.15
.2
.25
.3
.35
.4
.45
.5
.55
.6
.65
.7
.75
.8
.85
.9
.95
1

第六届中国Stata用户大会

Stata中的标准误

THANKS!

陈强
山东大学经济学院
www.econometrics-stata.com