应用微观计量经济学
Applied Microeconometrics

江　艇
中国人民大学经济学院

Last updated: March 18, 2019

Time:	周一晚 18:00-21:15
Venue:	明法 0502
Office:	明主 952
E-mail:	ting.jiang@ruc.edu.cn
Phone:	8250-0201
Grading:	期末闭卷考试

Lecture 2　OLS 回顾

2.1　OLS 的代数学

- 寻找能近似 y 的 $x_1(\equiv 1), x_2, \ldots, x_K$ 的最佳线性组合

$$\tilde{\beta}_1 + \tilde{\beta}_2 x_2 + \ldots + \tilde{\beta}_K x_K$$

$$\min_{\{\tilde{\beta}_1,\ldots,\tilde{\beta}_K\}} S = \sum_{i=1}^{n} \left(y_i - \tilde{\beta}_1 - \tilde{\beta}_2 x_{i2} - \ldots - \tilde{\beta}_K x_{iK} \right)^2$$

$$\frac{\partial S}{\partial \tilde{\beta}_1} = -2 \sum_{i=1}^{n} \left(y_i - \tilde{\beta}_1 - \tilde{\beta}_2 x_{i2} - \ldots - \tilde{\beta}_K x_{iK} \right) = 0$$

$$\frac{\partial S}{\partial \tilde{\beta}_2} = -2 \sum_{i=1}^{n} x_{i2} \left(y_i - \tilde{\beta}_1 - \tilde{\beta}_2 x_{i2} - \ldots - \tilde{\beta}_K x_{iK} \right) = 0$$

$$\vdots$$

$$\frac{\partial S}{\partial \tilde{\beta}_K} = -2 \sum_{i=1}^{n} x_{iK} \left(y_i - \tilde{\beta}_1 - \tilde{\beta}_2 x_{i2} - \ldots - \tilde{\beta}_K x_{iK} \right) = 0$$

- 向量表示

$$\min_{\tilde{\boldsymbol{\beta}}} S\left(\tilde{\boldsymbol{\beta}}\right) = \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i' \tilde{\boldsymbol{\beta}}\right)^2$$

$$\mathbf{x}_i = \left(1 \; x_{i2} \; \ldots \; x_{iK}\right)', \; \tilde{\boldsymbol{\beta}} = \left(\tilde{\beta}_1 \; \ldots \; \tilde{\beta}_K\right)'$$

$$\sum_{i=1}^{n} \mathbf{x}_i \left(y_i - \mathbf{x}_i' \tilde{\boldsymbol{\beta}}\right) = \mathbf{0}, \text{ 或 } \sum_{i=1}^{n} \mathbf{x}_i e_i = 0$$

$$\mathbf{b} = \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \sum_{i=1}^{n} \mathbf{x}_i y_i$$

- 几种特殊情形
  - 一元回归

$$y_i = b_0 + b_1 x_i + e_i$$

$$b_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\widehat{\mathsf{Cov}(x, y)}}{\widehat{\mathsf{Var}(x)}}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

  - 仅含截距项回归

$$y_i = b_0 + e_i \Rightarrow b_0 = \bar{y}$$

  - 不含截距项一元回归

$$y_i = b_1 x_i + e_i \Rightarrow b_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

  - 去均值后 (demeaned) 的一元回归

$$y_i - \bar{y} = b_1 (x_i - \bar{x}) + e_i$$

- 含截距项的虚拟变量一元回归

$$y_i = b_0 + b_1 D_i + e_i \Rightarrow b_0 = \bar{y}_0,\ b_1 = \bar{y}_1 - \bar{y}_0$$

- 不含截距项的虚拟变量回归

$$y_i = b_0 D_i^0 + b_1 D_i^1 + e_i \Rightarrow b_0 = \bar{y}_0,\ b_1 = \bar{y}_1$$

- 包含多组虚拟变量的回归，例：虚拟变量 $M_i = 1$ 表示 i 为男性（F 表示女性），H_i 表示 i 为高受教育水平（L 表示低受教育水平）。考虑如下的线性拟合

$$y_i = b_0 + b_1 M_i + b_2 H_i + e_i$$

此时是否仍然有类似 $b_1 = \bar{y}_M - \bar{y}_F$, $b_2 = \bar{y}_H - \bar{y}_L$ 的结论呢？一般情况下，答案是否定的。用 $\bar{y}$ 表示相应人群 y 之和，用 n 表示相应人群数量。

	H	L
M	$\bar{y}_{MH}, n_{MH}$	$\bar{y}_{ML}, n_{ML}$
F	$\bar{y}_{FH}, n_{FH}$	$\bar{y}_{FL}, n_{FL}$

考察这一线性拟合的正规方程组,

$$\sum_{i=1}^{n}(y_i - b_0 - b_1 M_i - b_2 H_i) = 0$$

$$\sum_{i=1}^{n} M_i(y_i - b_0 - b_1 M_i - b_2 H_i) = 0$$

$$\sum_{i=1}^{n} H_i(y_i - b_0 - b_1 M_i - b_2 H_i) = 0$$

可以等价写作

$$\sum_{i:M_i=1}(y_i - b_0 - b_1 - b_2 H_i) = 0$$
$$\sum_{i:F_i=1}(y_i - b_0 - b_2 H_i) = 0$$
$$\sum_{i:H_i=1}(y_i - b_0 - b_1 M_i - b_2) = 0$$
$$\sum_{i:L_i=1}(y_i - b_0 - b_1 M_i) = 0$$

进一步改写作

$$\begin{aligned}
(n_{MH} + n_{ML})b_0 + (n_{MH} + n_{ML})b_1 + n_{MH}b_2 &= n_{MH}\bar{y}_{MH} + n_{ML}\bar{y}_{ML} \\
(n_{FH} + n_{FL})b_0 + n_{FH}b_2 &= n_{FH}\bar{y}_{FH} + n_{FL}\bar{y}_{FL} \\
(n_{MH} + n_{FH})b_0 + n_{MH}b_1 + (n_{MH} + n_{FH})b_2 &= n_{MH}\bar{y}_{MH} + n_{FH}\bar{y}_{FH} \\
(n_{ML} + n_{FL})b_0 + n_{ML}b_1 &= n_{ML}\bar{y}_{ML} + n_{FL}\bar{y}_{FL}
\end{aligned}$$

易求得

$$b_1 = \frac{1}{1+\delta_1}(\bar{y}_{MH} - \bar{y}_{FH}) + \frac{\delta_1}{1+\delta_1}(\bar{y}_{ML} - \bar{y}_{FL}),\ \delta_1 \triangleq \frac{n_{MH}^{-1} + n_{FH}^{-1}}{n_{ML}^{-1} + n_{FL}^{-1}}$$

$$b_2 = \frac{1}{1+\delta_2}(\bar{y}_{MH} - \bar{y}_{ML}) + \frac{\delta_2}{1+\delta_2}(\bar{y}_{FH} - \bar{y}_{FL}),\ \delta_2 \triangleq \frac{n_{MH}^{-1} + n_{ML}^{-1}}{n_{FH}^{-1} + n_{FL}^{-1}}$$

注意到,

$$b_1 \ \neq \bar{y}_M - \bar{y}_F = \frac{n_{MH}\bar{y}_{MH} + n_{ML}\bar{y}_{ML}}{n_{MH} + n_{ML}} - \frac{n_{FH}\bar{y}_{FH} + n_{FL}\bar{y}_{FL}}{n_{FH} + n_{FL}}$$

$$b_2 \ \neq \bar{y}_H - \bar{y}_L = \frac{n_{MH}\bar{y}_{MH} + n_{FH}\bar{y}_{FH}}{n_{MH} + n_{FH}} - \frac{n_{ML}\bar{y}_{ML} + n_{FL}\bar{y}_{FL}}{n_{ML} + n_{FL}}$$

除非 $n_{MH} = n_{ML} = n_{FH} = n_{FL}$。

一个更灵活的模型是加入 $M_i \times H_i$,

$$y_i = b_0 + b_1 M_i + b_2 H_i + b_3 M_i \times H_i + e_i$$

易证,

$$\begin{aligned}
b_0 &= \bar{y}_{FL} \\
b_1 &= \bar{y}_{ML} - \bar{y}_{FL} \\
b_2 &= \bar{y}_{FH} - \bar{y}_{FL} \\
b_3 &= (\bar{y}_{MH} - \bar{y}_{FH}) - (\bar{y}_{ML} - \bar{y}_{FL})
\end{aligned}$$

- Frisch-Waugh-Lovell 定理
  - 考察如下多元回归，

$$y_i = b_1 + b_2 x_{i2} + \cdots + b_k x_{ik} + \cdots + b_K x_{iK} + e$$

  - 如果我们只关心系数 b_k，一种等价做法是，y_i 和 x_{ik} 分别对其它解释变量进行回归，保留残差，

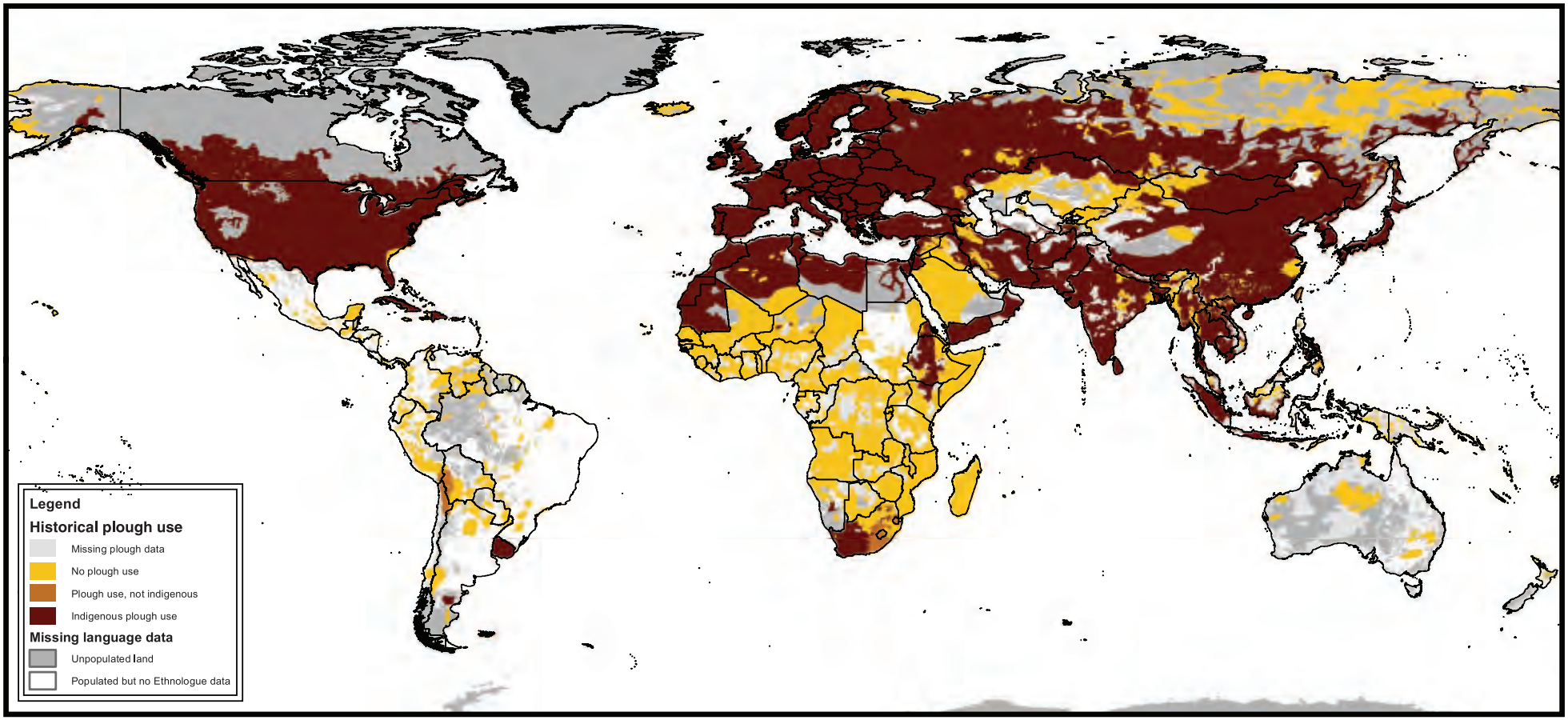
$$y_i = c_1 + c_2 x_{i2} + \cdots + c_{k-1} x_{i,k-1} + c_{k+1} x_{i,k+1} + \cdots + c_K x_{iK} + \tilde{y}_i$$
$$x_{ik} = d_1 + d_2 x_{i2} + \cdots + d_{k-1} x_{i,k-1} + d_{k+1} x_{i,k+1} + \cdots + d_K x_{iK} + \tilde{x}_{ik}$$

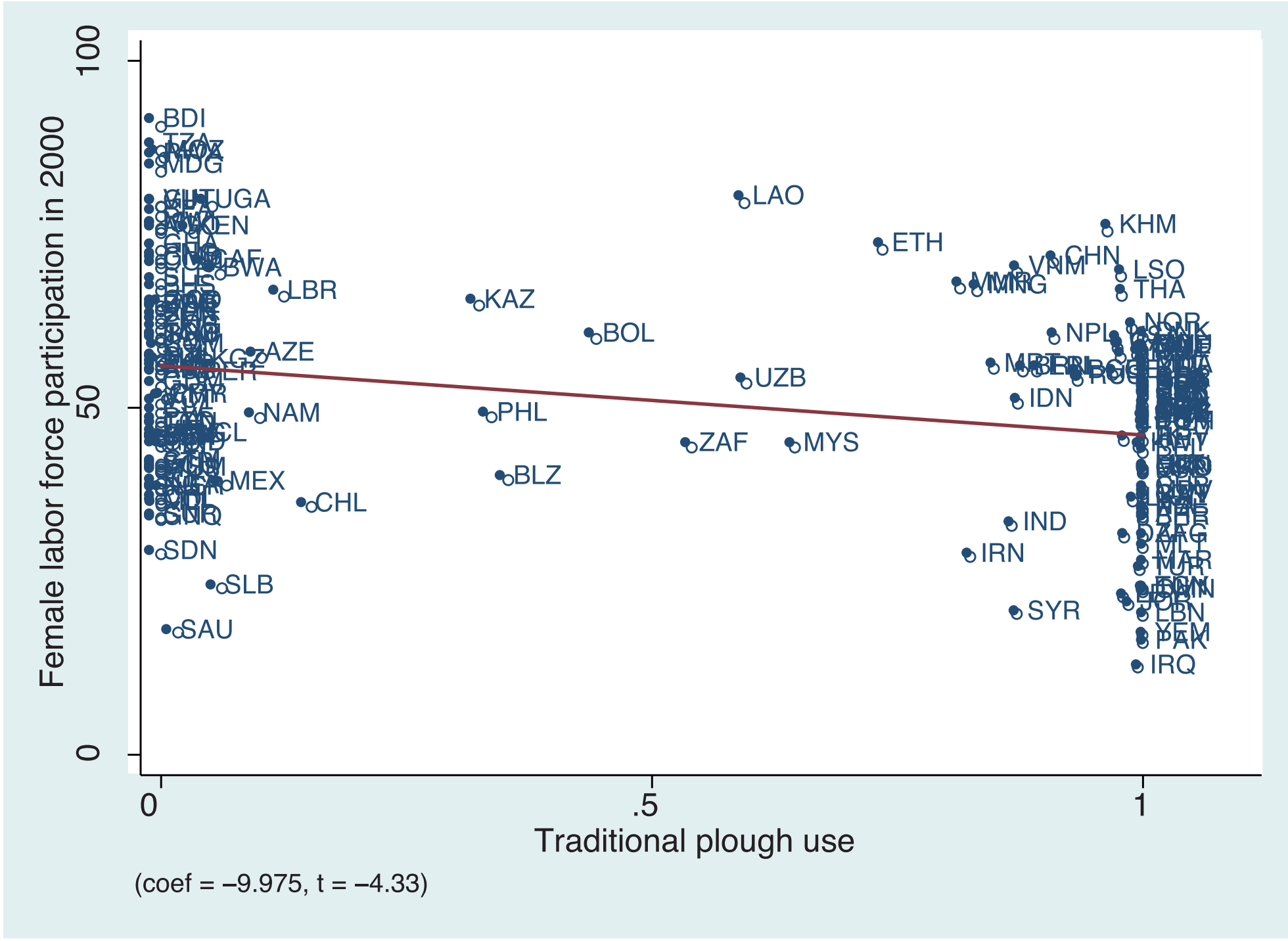
则

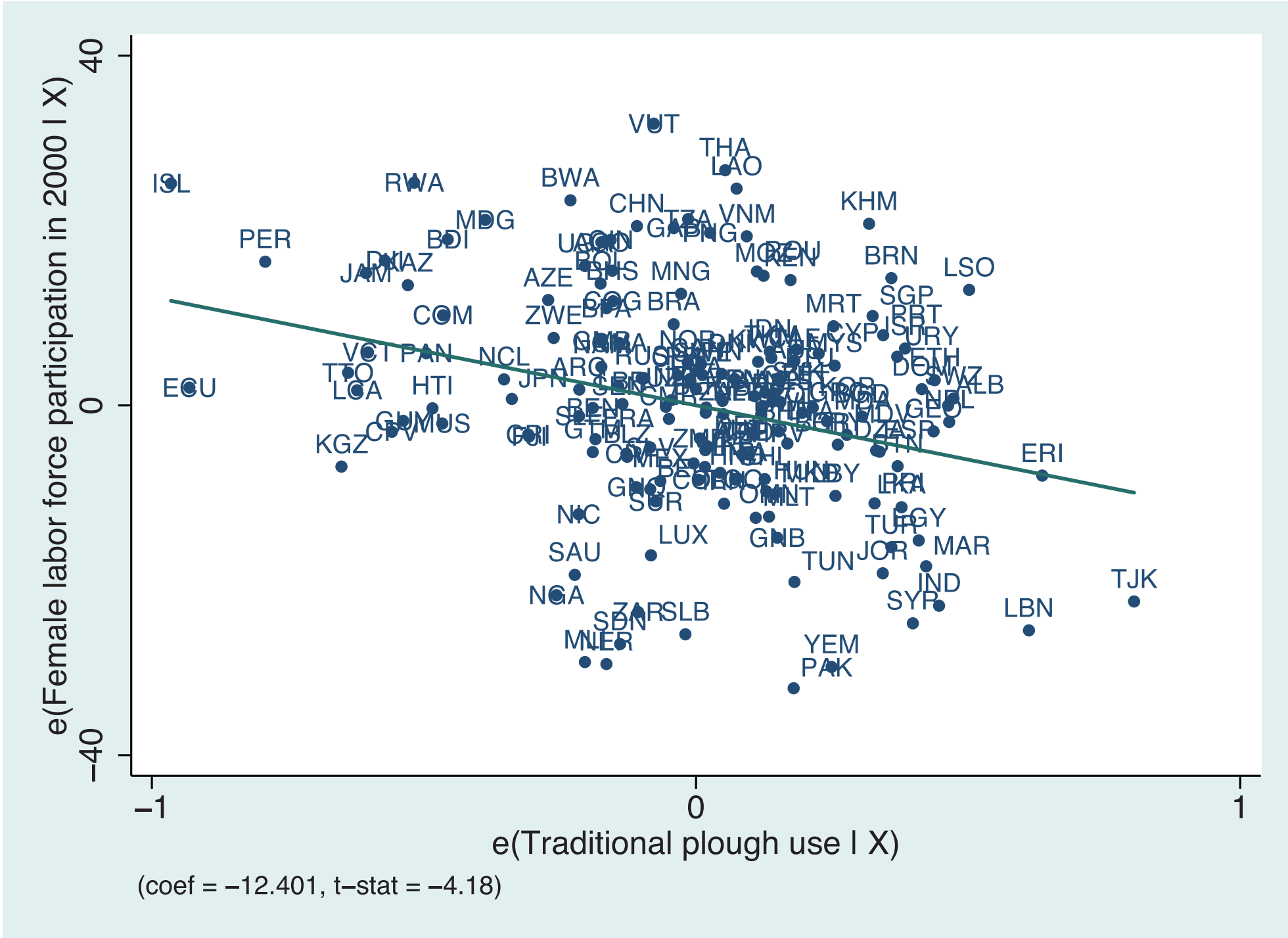
$$b_k = \frac{\widehat{\text{Cov}}\left(y, \tilde{x}_k\right)}{\widehat{\text{Var}}\left(\tilde{x}_k\right)} = \frac{\widehat{\text{Cov}}\left(\tilde{y}, \tilde{x}_k\right)}{\widehat{\text{Var}}\left(\tilde{x}_k\right)}$$

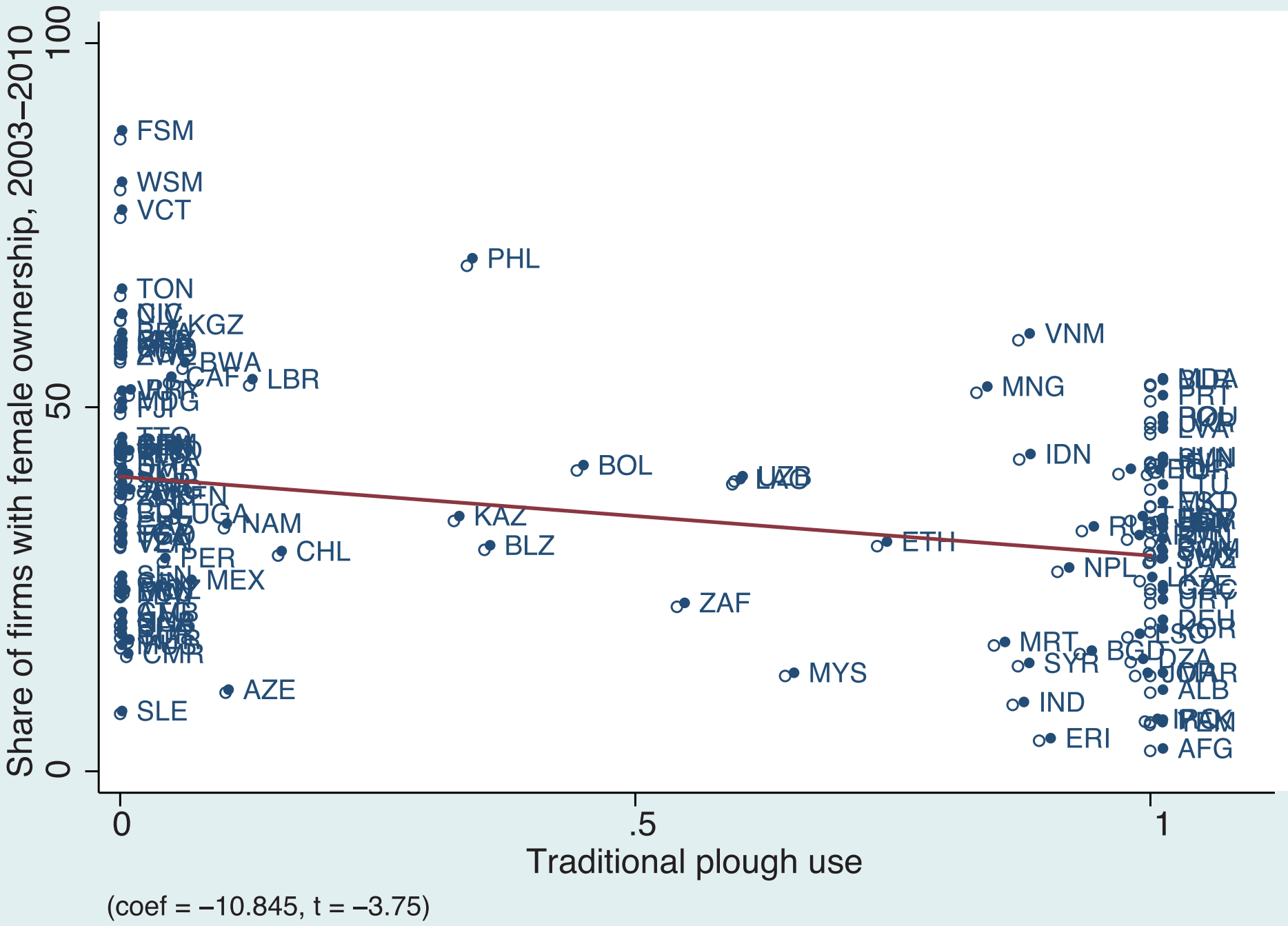
其中后者还能给出正确的标准误。

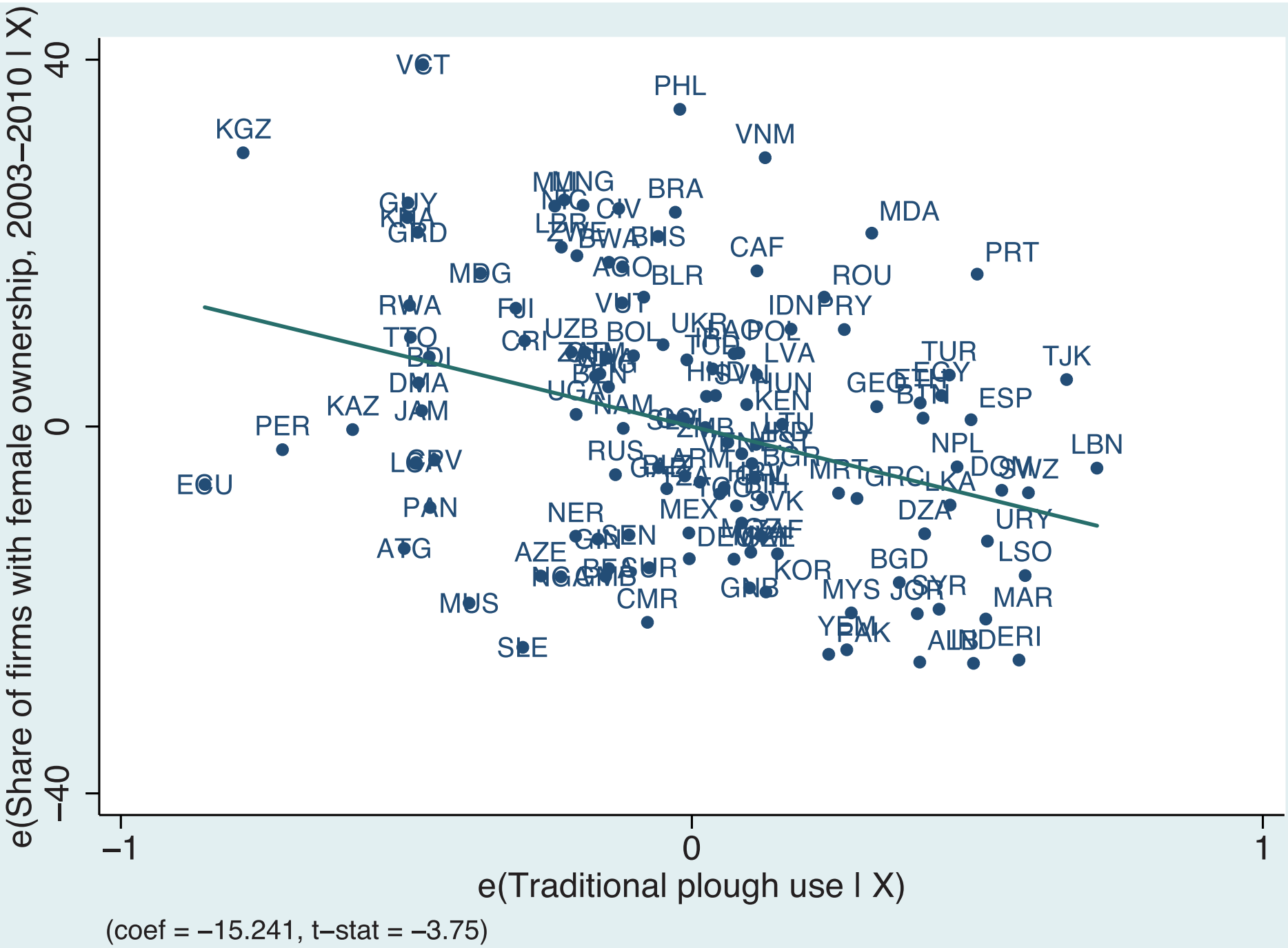
示例 1. 女性与犁具 (Alesina *et al*, 2013, *QJE*).

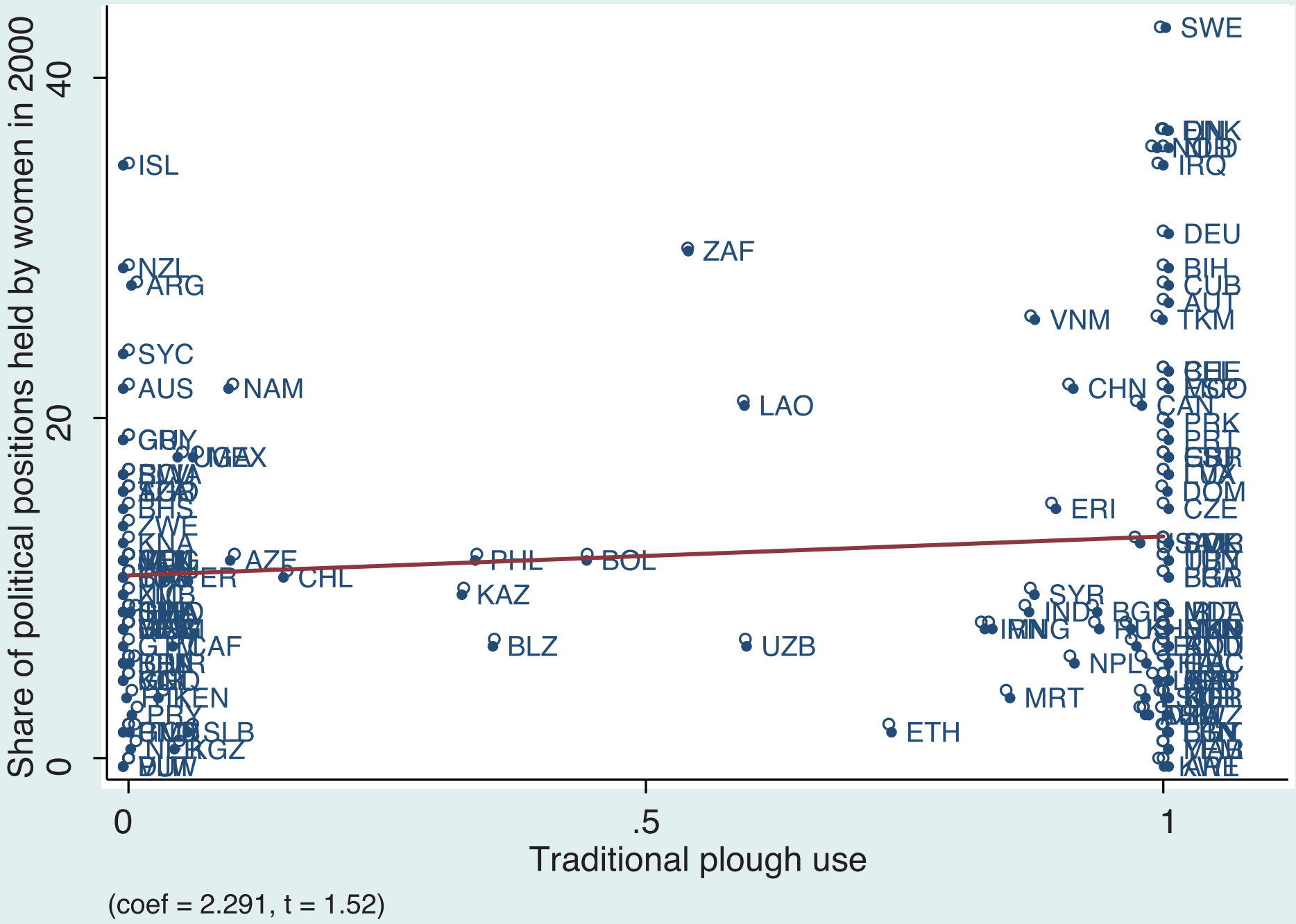

Legend
Historical plough use
Missing plough data
No plough use
Plough use, not indigenous
Indigenous plough use
Missing language data
Unpopulated land
Populated but no Ethnologue data

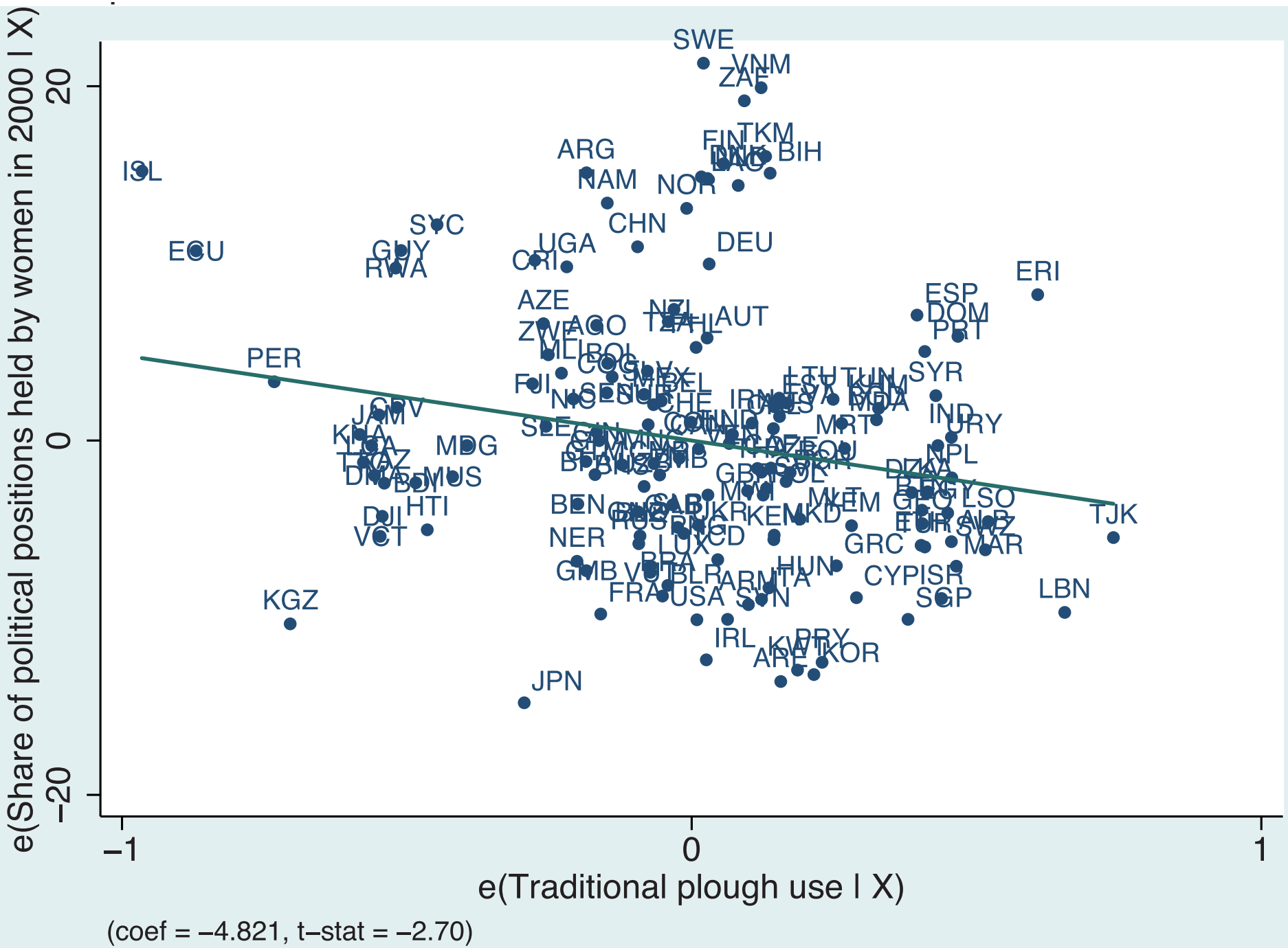

Female labor force participation in 2000
100
50
0
0
.5
1
Traditional plough use
(coef = –9.975, t = –4.33)
BDI
MDG
UGA
BWA
LBR
AZE
NAM
MEX
CHL
SDN
SLB
SAU
LAO
KAZ
BOL
UZB
PHL
ZAF
MYS
BLZ
ETH
KHM
CHN
VNM
LSO
THA
NPL
IDN
IND
IRN
SYR
LBN
PAK
IRQ


e(Female labor force participation in 2000 | X)
40
0
−40
−1
0
1
e(Traditional plough use | X)
(coef = −12.401, t−stat = −4.18)
VUT
THA
LAO
ISL
RWA
BWA
CHN
KHM
VNM
MDG
PER
BDI
BRN
LSO
JAM
AZE
MNG
SGP
MRT
COM
ZWE
BRA
VCT
PAN
NCL
TTO
LCA
HTI
ECU
JPN
KGZ
ERI
NIC
LUX
SAU
GNB
TUN
JOR
MAR
IND
TJK
SYP
LBN
NGA
SLB
SDN
YEM
PAK




e(Share of firms with female ownership, 2003–2010 | X)
40
0
−40
−1
0
1
e(Traditional plough use | X)
VCT
PHL
KGZ
VNM
BRA
MDA
CAF
PRT
MDG
BLR
ROU
RWA
FJI
IDN
PRY
UZB
BOL
TUR
TJK
LVA
DMA
KAZ
JAM
ESP
PER
NPL
LBN
RUS
ECU
PAN
NER
MEX
DZA
URY
ATG
AZE
LSO
BGD
KOR
MYS
MAR
CMR
MUS
SLE
ALB
ERI
(coef = −15.241, t−stat = −3.75)


Share of political positions held by women in 2000
0
20
40
0
.5
1
Traditional plough use
(coef = 2.291, t = 1.52)
SWE
ISL
ZAF
DEU
NZL
ARG
BIH
CUB
AUT
VNM
TKM
SYC
AUS
NAM
CHN
LAO
ERI
DOM
CZE
AZE
PHL
BOL
CHL
KAZ
SYR
BLZ
UZB
NPL
MRT
ETH


e(Share of political positions held by women in 2000 | X)
20
0
−20
−1
0
1
e(Traditional plough use | X)
(coef = −4.821, t−stat = −2.70)
SWE
VNM
ZAF
TKM
ARG
BIH
ISL
NAM
NOR
CHN
SYC
UGA
DEU
ECU
GUY
CRI
RWA
ERI
ESP
AZE
AUT
DOM
PRT
PER
SYR
IND
URY
MDG
MRT
NPL
MUS
HTI
LSO
TJK
VCT
NER
GRC
MAR
KGZ
HUN
CYP
ISR
LBN
SGP
IRL
PRY
KOR
JPN

- 从 FWL 定理看控制虚拟变量的作用
  - We need enough **variation** in an explanatory variable to obtain its coefficient (otherwise it is collinear with the intercept). Consider, for example, the return-to-education equation,

$$\texttt{wage}_i = \beta_0 + \beta_1 \texttt{edu}_i + e_i$$

    we need individuals of all levels of education in our sample to estimate β_1.
  - Now augment the RHS with gender controls,

$$\texttt{wage}_i = \beta_0 + \beta_1 \texttt{edu}_i + \gamma \texttt{male}_i + e_i$$

    It is equivalent to

$$\widetilde{\texttt{wage}}_i = \beta_0 + \beta_1 \widetilde{\texttt{edu}}_i + \varepsilon_i$$

    where $\widetilde{\texttt{wage}}_i$ and $\widetilde{\texttt{edu}}_i$ are the residuals of regressing $\texttt{wage}_i$ and $\texttt{edu}_i$, respectively, on $\texttt{male}_i$ (and an intercept).

– It is easy to see that

$$\widetilde{\texttt{wage}}_i = \begin{cases} \texttt{wage}_i - \overline{\texttt{wage}}_m & \text{if } \texttt{male}_i = 1 \\ \texttt{wage}_i - \overline{\texttt{wage}}_f & \text{if } \texttt{male}_i = 0 \end{cases}$$

$$\widetilde{\texttt{edu}}_i = \begin{cases} \texttt{edu}_i - \overline{\texttt{edu}}_m & \text{if } \texttt{male}_i = 1 \\ \texttt{edu}_i - \overline{\texttt{edu}}_f & \text{if } \texttt{male}_i = 0 \end{cases}$$

so the **working source of variation** is **within-gender-group variation** (in levels of education). We call the practice of including $\texttt{male}_i$ on the RHS controlling for **gender fixed effects**.

– Similarly, if we augment the RHS with regional dummies,

$$\texttt{wage}_i = \beta_0 + \beta_1 \texttt{edu}_i + \sum_{r=2}^{R} \delta^r \texttt{region}_i^r + e_i$$

where

$$\texttt{region}_i^r = \begin{cases} 1 & \text{if } i \text{ is from region } r \\ 0 & \text{otherwise} \end{cases}$$

it is equivalent to

$$\widetilde{\texttt{wage}}_i = \beta_0 + \beta_1 \widetilde{\texttt{edu}}_i + \varepsilon_i$$

where $\widetilde{\texttt{wage}}_i$ and $\widetilde{\texttt{edu}}_i$ are the residuals of regressing $\texttt{wage}_i$ and $\texttt{edu}_i$, respectively, on all the regional dummies (and an intercept). They are nothing but **within-region demeaned** variables, i.e., the original variables subtracting **region-specific means**. After controlling for **regional fixed effects**, the working source of variation is now **within-region variation**.

2.2 OLS 的有限样本理论

- 考察总体线性模型

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$$

- 在随机抽样前提下，有限样本理论的关键假定为扰动项均值独立假定（也称为严格外生性假定）。

Assumption FS.1: $\mathsf{E}\left(\varepsilon_i|\mathbf{x}_i\right) = \mathsf{E}(\varepsilon_i) = 0$

- 对系数的解释

$$\mathsf{E}\left(y_i|\mathbf{x}_i\right) = \mathbf{x}_i'\boldsymbol{\beta}, \ \frac{\partial \mathsf{E}\left(y_i|\mathbf{x}_i\right)}{\partial x_{ik}} = \beta_k$$

- 用 OLS 估计系数，可以把 OLS 看作一种矩估计方法。

$$\mathsf{E}(\mathbf{x}_i\varepsilon_i) = 0$$

$$\mathsf{E}\left(\mathbf{x}_i(y_i - \mathbf{x}_i'\boldsymbol{\beta})\right) = 0$$

$$\boldsymbol{\beta} = \left(\mathsf{E}(\mathbf{x}_i\mathbf{x}_i')\right)^{-1}\mathsf{E}(\mathbf{x}_i y_i)$$

Naturally, we can estimate population moments (expectation of function of random vectors) with their sample analogue (sample mean).

$$\mathbf{b} = \left(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\sum_{i=1}^{n}\mathbf{x}_i y_i\right)$$

- 在关键假定**FS.1**下，系数的 OLS 估计量是无偏的。

$$\mathsf{E}\left(\mathbf{b}|\mathbf{X}\right) = \boldsymbol{\beta}$$

- 如何理解系数具有因果含义？关键在于如何看待总体线性模型。
  - 部分教科书认为，总体线性模型是在对给定 $\mathbf{x}_i$ 时 y_i 的条件期望进行建模，这种认识是错误的。
  - 如果我们定义

$$\begin{aligned} y_i &\equiv \mathsf{E}\left(y_i|\mathbf{x}_i\right) + \left[y_i - \mathsf{E}\left(y_i|\mathbf{x}_i\right)\right] \\ &\triangleq \quad \mathbf{x}_i'\boldsymbol{\beta} \quad + \qquad \varepsilon_i \end{aligned}$$

    那么 $\mathbf{x}_i$ 的严格外生性假定天然成立

$$\mathsf{E}\left(\varepsilon_i|\mathbf{x}_i\right) = 0$$

  - 换言之，不论 y_i 和 $\mathbf{x}_i$ 为何，我们总能用 OLS 方法得到 $\mathsf{E}\left(y_i|\mathbf{x}_i\right)$ 的无偏估计（对于非线性条件期望，OLS 方法总能得到其最佳线性近似的无偏估计）。此时 ε_i 仅仅是期望残差 (expectational residual)，$\boldsymbol{\beta}$ 不具有因果含义。

– 严格外生性假定并不是总体线性模型的固有特征。ε_i 应该被定义成所有其它直接影响 y_i 的因素，严格外生性假定有可能不成立，

$$\mathsf{E}\left(\varepsilon_i|\mathbf{x}_i\right) \neq 0$$

此时称总体线性模型为结构模型，称 ε_i 为结构性误差，称 $\boldsymbol{\beta}$ 为结构参数或因果参数。

– 我们需要区分预测性问题和因果性问题。

▷ **预测性问题**：若我们观测到 $\mathbf{x}_i$ 的值为 $\mathbf{x}_0$，我们预期 y_i 的值是多少？

▷ **因果性问题**：若我们干预并设定 $\mathbf{x}_i$ 的值为 $\mathbf{x}_0$，我们预期 y_i 的值是多少？

– 条件期望旨在回答预测性问题，而结构模型旨在回答因果性问题，结构模型中的 $\mathbf{x}_0'\boldsymbol{\beta}$ 一般不等于条件期望 $E(y_i|\mathbf{x}_i = \mathbf{x}_0)$，而是等于我们干预并设定 $\mathbf{x}_i$ 的值为 $\mathbf{x}_0$ 时 y_i 的期望值。Pearl 称之为“基于干预的期望”，并将其表示成 $E(y_i|do(\mathbf{x}_0))$，尽管这一表示法并不常用。

- 此时我们说 $\boldsymbol{\beta}$ 刻画了其它因素保持不变时 y_i 与 $\mathbf{x}_i$ 之间的因果关系（这里保持不变的其它因素既包含结构模型中的控制变量，也包含结构性误差）。不论严格外生性假定是否成立，$\boldsymbol{\beta}$ 始终具有因果含义：

$$\boldsymbol{\beta} = \frac{\partial}{\partial \mathbf{x}_i} E(y_i | do(\mathbf{x}_i))$$

- 我们总能得到条件期望 $E(y_i|\mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\alpha}$ 的无偏估计，但只有当严格外生性假定成立时（例如随机实验），对条件期望的 OLS 估计才是对干预性期望的无偏估计 ($\boldsymbol{\alpha} = \boldsymbol{\beta}$)。而当严格外生性假定不成立时，我们需要另外寻找 $\mathbf{x}_i$ 的外生变动来估计干预性期望，因此，考察具体研究情境中严格外生性假定的经济学含义是否合理才显得至关重要。

– For example, suppose the structural model is

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i = 1 + 2x_i + \varepsilon_i$$

$$\begin{pmatrix} x \\ \varepsilon \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix} \right)$$

β_1 is the structural parameter that depicts the causal effect of x on y.
In a Monte Carlo exercise we will show that the OLS estimator of β_1 is biased. The reason is of course $E(\varepsilon_i|x_i) \neq 0$.
If instead we model

$$\begin{aligned} y_i &= E(y_i|x_i) + [y_i - E(y_i|x_i)] \\ &= \tilde{\beta}_0 + \tilde{\beta}_1 x_i + \tilde{\varepsilon}_i \\ &= 1 + 2.5x_i + \tilde{\varepsilon}_i \end{aligned}$$

where we use the fact that

$$E(\varepsilon_i|x_i) = E(\varepsilon_i) + \rho_{x\varepsilon}\frac{\sigma_\varepsilon}{\sigma_x}[x_i - E(x_i)]$$

Since $E(\tilde{\varepsilon}|x) = 0$ by construction, the OLS estimator of $\tilde{\beta}_1$ is always unbiased.

- 扰动项条件同方差假定，据此计算 OLS 估计量的条件方差。

Assumption FS.2 $\quad \mathrm{E}\left(\varepsilon_i^2|\mathbf{x}_i\right)=\sigma^2$

$$\mathrm{Var}\left(\mathbf{b}|\mathbf{X}\right)=\sigma^2\left(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}$$

一元回归中，

$$\mathrm{Var}\left(b_1|\mathbf{X}\right)=\frac{\sigma^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}$$

条件方差的估计

$$\widehat{\mathrm{Var}\left(\mathbf{b}|\mathbf{X}\right)}=s^2\left(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right)^{-1},\ s^2\triangleq\frac{\sum_{i=1}^{n}e_i^2}{n-K}$$

等价地,

$$\widehat{\text{Var}}\left(b_k|\mathbf{X}\right) = \frac{s^2}{(n-1)\left(1-R_{c,k}^2\right)\widehat{\text{Var}}\left(x_k\right)}$$

where $R_{c,k}^2$ is the centered-R^2 in the regression of x_k on all other variables.

- The larger the sample size,
- The greater the variation in x_k,
- The less the correlation of x_k with the other variables,
- The better the overall fit of the regression,

the lower the variance of b_k will be.

- 扰动项正态分布假定，据此进行假说检验。

Assumption FS.3 $\quad \varepsilon_i|\mathbf{x}_i \sim N\left(0, \sigma^2\right)$

- OLS 估计量的准确分布

$$\mathbf{b}|\mathbf{X} \sim N\left(\boldsymbol{\beta}, \sigma^2\left(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\right)$$

- 单个回归系数显著性 t 检验

$$H_0 : \beta_k = \bar{\beta}_k$$

t 统计量

$$t_k \triangleq \frac{b_k - \bar{\beta}_k}{\widehat{\mathsf{SE}\left(b_k\right)}} = \frac{b_k - \bar{\beta}_k}{\sqrt{s^2\left[\left(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\right]_{kk}}} \sim t(n-K)$$

- t 检验的决策规则
  - 当 t 统计量落入拒绝域则拒绝原假说。

$$\text{Prob}\left(-t_{\alpha/2}(n-K) < t_k < t_{\alpha/2}(n-K)\right) = 1-\alpha$$

  - 当 p 值小于显著性水平 α 则拒绝原假说。

$$p = \text{Prob}\left(t > |t_k|\right) \times 2$$

  - 当置信区间不包含 $\bar{\beta}_k$ 则拒绝原假说。

$$b_k - \widehat{\text{SE}(b_k)} t_{\alpha/2}(n-K) < \beta_k < b_k + \widehat{\text{SE}(b_k)} t_{\alpha/2}(n-K)$$

- 单个线性约束的 t 检验

$$H_0 : r_1\beta_1 + \ldots + r_K\beta_K = \mathbf{r}'\boldsymbol{\beta} = q$$

t 统计量

$$t_{\text{stat}} \triangleq \frac{\mathbf{r}'\mathbf{b} - q}{\widehat{\text{SE}\left(\mathbf{r}'\mathbf{b}\right)}} = \frac{\mathbf{r}'\mathbf{b} - q}{\sqrt{s^2\mathbf{r}'\left(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\mathbf{r}}} \sim t(n-K)$$

- 多个线性约束的 F 检验

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{q}$$

F 统计量

$$F_{\text{stat}} = (\mathbf{Rb} - \mathbf{q})' \left(s^2 \mathbf{R} \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \mathbf{R}' \right)^{-1} (\mathbf{Rb} - \mathbf{q}) / \dim(\mathbf{q})$$

$$\sim F(\dim(\mathbf{q}), n - K)$$

其中 $\dim(\mathbf{q})$ 表示待检验线性约束的个数。

- F 检验的决策规则
  - 当 F 统计量大于临界值则拒绝原假说。

$$\text{Prob}(F < F_\alpha(\dim(\mathbf{q}), n - K)) = 1 - \alpha$$

  - 当 p 值小于显著性水平 α 则拒绝原假说。

$$p = \text{Prob}(F > F_{\text{stat}})$$

2.3 OLS 的大样本理论

- 我们希望放松有限样本理论的正态扰动项假定，OLS 估计量仍然具有合意性质。
- 大数定律 (Law of Large Numbers)。$\{x_i\}$ i.i.d., $\mathsf{E}(x_i) = \mu$, then

$$\overline{x}_n \longrightarrow_{\mathsf{p}} \mathsf{E}(\overline{x}_n) = \mu$$

  弱版本的大数定律允许 $\{x_i\}$ 具有时序或横截面的相关性。
- 中心极限定理 (Central Limit Theorem). $\{x_i\}$ i.i.d., $\mathsf{E}(x_i) = \mu$, $\mathsf{Var}(x_i) = \Sigma$, then

$$\overline{x}_n \to_{\mathsf{d}} N\left(\mathsf{E}(\overline{x}_n),\ \mathsf{Var}(\overline{x}_n)\right) = N\left(\mu,\ \Sigma/n\right)$$

  弱版本的中心极限定理允许 $\{x_i\}$ 具有时序或横截面相关性，此时 $\mathsf{Var}(\overline{x}_n)$ 中不仅包含 x_i 的方差项，还包含其自协方差项。

- 在同分布和弱持续 (weak persistence) 的抽样前提下，大样本理论的关键假定为扰动项与解释变量正交（也称为前定解释变量假定）。

$$\textbf{Assumption LS.1:} \quad \mathsf{E}(\mathbf{x}_i \varepsilon_i) = \mathsf{E}\left[\mathbf{x}_i \left(y_i - \mathbf{x}_i' \boldsymbol{\beta}\right)\right] = \mathbf{0}$$

- 在关键假定**LS.1**下，系数的 OLS 估计量是一致的。

$$\begin{aligned}
\mathbf{b} &= \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i y_i\right) \\
&= \boldsymbol{\beta} + \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \varepsilon_i\right) \\
&\to_{\mathrm{p}} \boldsymbol{\beta}
\end{aligned}$$

在一元回归中，

$$
\begin{aligned}
b_1 = \frac{\widehat{\mathrm{Cov}(y_i, x_i)}}{\widehat{\mathrm{Var}(x_i)}} &\to_p \frac{\mathrm{Cov}(y_i, x_i)}{\mathrm{Var}(x_i)} \\
&= \frac{\mathrm{Cov}(\beta_0 + \beta_1 x_i + \varepsilon_i, x_i)}{\mathrm{Var}(x_i)} \\
&= \beta_1 + \frac{\mathrm{Cov}(\varepsilon_i, x_i)}{\mathrm{Var}(x_i)}
\end{aligned}
$$

- 当 $\mathbf{x}_i\varepsilon_i$ 不相关时，系数的 OLS 估计量渐进服从正态分布。

$$
\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\varepsilon_i \to_{\mathrm{d}} N\left(\mathbf{0},\ \frac{\mathrm{Var}\left(\mathbf{x}_i\varepsilon_i\right)}{n}\right) = N\left(\mathbf{0},\ \frac{\mathrm{E}\left(\varepsilon_i^2\mathbf{x}_i\mathbf{x}_i'\right)}{n}\right)
$$

$$
\mathbf{b} \to_{\mathrm{d}} N\left(\boldsymbol{\beta},\ \frac{1}{n}\left(\mathrm{E}\left(\mathbf{x}_i\mathbf{x}_i'\right)\right)^{-1}\mathrm{E}\left(\varepsilon_i^2\mathbf{x}_i\mathbf{x}_i'\right)\left(\mathrm{E}\left(\mathbf{x}_i\mathbf{x}_i'\right)\right)^{-1}\right)
$$

此时系数估计标准误称为**异方差稳健标准误**。

- 当 $\mathbf{x}_i\varepsilon_i$ 聚类相关时，系数的 OLS 估计量渐进服从正态分布。此时系数估计量的渐进方差中应包含同一类内不同个体之间 $\mathbf{x}_i\varepsilon_i$ 的协方差。其具体形式取决于对 $\mathbf{x}_i\varepsilon_i$ 聚类相关结构的假定。此时系数估计标准误称为**聚类稳健标准误**。
  - 控制类固定效应可以消除相当一部分（但有时不是全部）聚类相关性。
  - 聚类层级的选择涉及稳健性和估计偏误之间的权衡。合适的聚类层级依研究情境和数据特征而异。一个起码的经验法则是，当解释变量的数据层级高于被解释变量时，应聚类到解释变量所在层级。
  - 聚类层级越高，系数估计标准误一般越大。

– 考虑 ε_{icp}，i 表示企业，c 表示城市，p 表示省份，如果标准误聚类到城市层面，所隐含的对扰动项方差协方差结构的假设是

$$
\begin{pmatrix}
 & \varepsilon_{111} & \varepsilon_{211} & \varepsilon_{321} & \varepsilon_{421} & \varepsilon_{532} & \varepsilon_{632} & \varepsilon_{742} & \varepsilon_{842} \\
\varepsilon_{111} & \times & \times & & & & & & \\
\varepsilon_{211} & \times & \times & & & & & & \\
\varepsilon_{321} & & & \times & \times & & & & \\
\varepsilon_{421} & & & \times & \times & & & & \\
\varepsilon_{532} & & & & & \times & \times & & \\
\varepsilon_{632} & & & & & \times & \times & & \\
\varepsilon_{742} & & & & & & & \times & \times \\
\varepsilon_{842} & & & & & & & \times & \times
\end{pmatrix}
$$

– 如果标准误聚类到省份层面，所隐含的对扰动项方差协方差结构的假设是

$$
\begin{pmatrix}
 & \varepsilon_{111} & \varepsilon_{211} & \varepsilon_{321} & \varepsilon_{421} & \varepsilon_{532} & \varepsilon_{632} & \varepsilon_{742} & \varepsilon_{842} \\
\varepsilon_{111} & \times & \times & \times & \times & & & & \\
\varepsilon_{211} & \times & \times & \times & \times & & & & \\
\varepsilon_{321} & \times & \times & \times & \times & & & & \\
\varepsilon_{421} & \times & \times & \times & \times & & & & \\
\varepsilon_{532} & & & & & \times & \times & \times & \times \\
\varepsilon_{632} & & & & & \times & \times & \times & \times \\
\varepsilon_{742} & & & & & \times & \times & \times & \times \\
\varepsilon_{842} & & & & & \times & \times & \times & \times
\end{pmatrix}
$$

可见，**聚类层级越高，所隐含的假设越弱，标准误估计更稳健**。

– 考虑数据 ε_{icd}，其中 c 表示城市，d 表示行业，此时同一城市内部不同个体的扰动项之间可能相关，同一行业内部不同个体的扰动项之间也可能相关，此时有必要使用**双向聚类 (two-way clustering) 稳健标准误**，其隐含的假设是

$$
\begin{pmatrix}
 & \varepsilon_{111} & \varepsilon_{211} & \varepsilon_{312} & \varepsilon_{412} & \varepsilon_{521} & \varepsilon_{621} & \varepsilon_{722} & \varepsilon_{822} \\
\varepsilon_{111} & \times & \times & \times & \times & \times & \times & & \\
\varepsilon_{211} & \times & \times & \times & \times & \times & \times & & \\
\varepsilon_{312} & \times & \times & \times & \times & & & \times & \times \\
\varepsilon_{412} & \times & \times & \times & \times & & & \times & \times \\
\varepsilon_{521} & \times & \times & & & \times & \times & \times & \times \\
\varepsilon_{621} & \times & \times & & & \times & \times & \times & \times \\
\varepsilon_{722} & & & \times & \times & \times & \times & \times & \times \\
\varepsilon_{822} & & & \times & \times & \times & \times & \times & \times
\end{pmatrix}
$$

要注意其与“聚类到城市 $\times$ 行业层面的稳健标准误”的区别。

2.4 控制变量的作用

- OLS 小样本理论的关键假定**FS.1**要求扰动项均值独立于所有 $\mathbf{x}_i$，

$$E\left(\varepsilon_i|\mathbf{x}_i\right) = E\left(\varepsilon_i\right)$$

- $\mathbf{x}_i$ 中往往包含控制变量，而我们并不关心控制变量的因果效应，因此实际隐含的是条件均值独立假定 (conditional mean independence)。

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

  其中 X_1 是关键解释变量，X_2 是控制变量。

- 若 $E(\varepsilon|X_1, X_2) = E(\varepsilon) = 0$，即 X_1 和 X_2 均外生于 ε，则 β_1 和 β_2 的 OLS 估计均反映因果效应。

- 若 $E(\varepsilon|X_1, X_2) = E(\varepsilon|X_2) \neq E(\varepsilon) = 0$，则称给定 X_2，X_1 条件均值独立于 ε.

  方便起见，假定 $E(\varepsilon|X_2) = \gamma_0 + \gamma_2 X_2$

$$\begin{aligned} Y =& \beta_0 + \beta_1 X_1 + \beta_2 X_2 + E(\varepsilon|X_1, X_2) + \big[\varepsilon - E(\varepsilon|X_1, X_2)\big] \\ =& \beta_0 + \beta_1 X_1 + \beta_2 X_2 + E(\varepsilon|X_2) + \big[\varepsilon - E(\varepsilon|X_1, X_2)\big] \\ =& (\beta_0 + \gamma_0) + \beta_1 X_1 + (\beta_2 + \gamma_2) X_2 + \big[\varepsilon - E(\varepsilon|X_1, X_2)\big] \end{aligned}$$

$$E\big[\varepsilon - E(\varepsilon|X_1, X_2)\big|X_1, X_2\big] = 0$$

  X_1 的因果效应估计是无偏的，但 X_2 的因果效应估计是有偏的。

- 例：X_2 是专业，X_1 是专业内随机分组是否布置作业，Y 是数学考试成绩，专业选择可能和遗漏变量相关（例如高中数学成绩，这就是 $E(\varepsilon|X_2) \neq E(\varepsilon)$ 的原因），因此控制专业后能得到作业的因果效应，但专业本身对 Y 的因果效应估计是有偏的。

- 由此可以看出控制变量的双重作用：首先是控制 X_2 的直接效应 (β_2)，使得对回归模型的估计更准确；但更重要的是控制与 X_2 相关且可能与 X_1 相关的因素 (γ_2)，研究者希望，一旦控制 X_2 以后，ε 中的剩余部分不再和 X_1 相关。
- 类似地，OLS 大样本理论的关键假定**LS.1**要求所有 $\mathbf{x}_i$ 均与扰动项不相关，

$$Cov(\mathbf{x}_i, \varepsilon_i) = E(\mathbf{x}_i \varepsilon_i) = 0$$

但实际隐含的往往是条件不相关性 (conditional uncorrelatedness).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

$$\varepsilon = \delta_0 + \delta_1 X_1 + \delta_2 X_2 + v$$

$$\mathsf{E}(v) = 0,\ Cov(X_1, v) = Cov(X_2, v) = 0$$

$$y = (\beta_0 + \delta_0) + (\beta_1 + \delta_1) X_1 + (\beta_2 + \delta_2) X_2 + v$$

$$\hat{\beta}_1^{OLS} \to_p \beta_1 + \delta_1$$

- 条件不相关性意味着 $Cov(\varepsilon, X_1 | X_2) = 0$，即 $\delta_1 = 0$.

- 控制变量的重要性举例：辛普森悖论 (Simpson's Paradox)
  - 如果不控制造成选择性（影响分配机制）的协变量，对因果效应的估计甚至会出现方向的背离。

	Drug ($D=1$)	No drug ($D=0$)
Men ($X=0$)	$81/87$ (93%)	$234/270$ (87%)
Women ($X=1$)	$192/263$ (73%)	$55/80$ (69%)
Combined data	$273/350$ (78%)	$289/350$ (83%)

$$E(Y|X=1,D=1)-E(Y|X=1,D=0)=73\%-69\%=4\%$$
$$E(Y|X=0,D=1)-E(Y|X=0,D=0)=93\%-87\%=6\%$$
$$E(Y|D=1)-E(Y|D=0)=78\%-83\%=-5\%$$

  - 如果知道患者性别，应该开药；如果不知道患者性别，则不应该开药？！这个结论显然是荒唐的——如果该药对男性和女性都有效，则应该对任何人都有效。

$$
\begin{aligned}
E(Y|D=1) &= E(Y|X=1,D=1)\cdot P(X=1|D=1) \\
&\quad + E(Y|X=0,D=1)\cdot P(X=0|D=1) \\
&= 73\% \times 75\% + 93\% \times 25\% = 78\% \\
E(Y|D=0) &= E(Y|X=1,D=0)\cdot P(X=1|D=0) \\
&\quad + E(Y|X=0,D=0)\cdot P(X=0|D=0) \\
&= 69\% \times 23\% + 87\% \times 77\% = 83\%
\end{aligned}
$$

- 直观解释：无论是否使用药物，女性的治愈率都较低；女性使用药物的几率较高（等价地，药物使用者中女性的几率较高）。因此，药物在总体上似乎无效的原因在于：当我们随机挑选一位药物使用者时，该对象为女性的几率较高，因此平均而言治愈率较低。换言之，负向的女性效应抵消了正向的用药效应。

$$P(D=1|X=1) = P(X=1|D=1) \cdot \frac{P(D=1)}{P(X=1)}$$
$$P(D=1|X=0) = P(X=0|D=1) \cdot \frac{P(D=1)}{P(X=0)}$$

若 $P(X=1) = P(X=0)$，则

$$\underbrace{P(D=1|X=1) > P(D=1|X=0)}_{\text{女性使用药物的几率较高}}$$
$$\Leftrightarrow \underbrace{P(X=1|D=1) > P(X=0|D=1)}_{\text{药物使用者中女性的几率较高}}$$

– 从线性回归的角度来思考这一问题。

$$E(Y|X, D = d) = \alpha_d + \beta_d X$$

– 可以分组回归或使用交互项模型,

$$E(Y|D, X) = \alpha_0 + (\alpha_1 - \alpha_0)D + \beta_0 X + (\beta_1 - \beta_0)D \cdot X$$

$$\begin{cases} E(Y|X = 1, D = 1) = \alpha_1 + \beta_1 & = 73\% \\ E(Y|X = 0, D = 1) = \alpha_1 & = 93\% \\ E(Y|X = 1, D = 0) = \alpha_0 + \beta_0 & = 69\% \\ E(Y|X = 0, D = 0) = \alpha_0 & = 87\% \end{cases}$$

– 条件边际效应

$$\begin{aligned}\frac{\partial E(Y|D,X)}{\partial D} &= E(Y|X,D=1) - E(Y|X,D=0) \\ &= (\alpha_1 - \alpha_0) + (\beta_1 - \beta_0) \cdot X\end{aligned}$$

– 平均边际效应

$$\begin{aligned}E_X\left[\frac{\partial E(Y|D,X)}{\partial D}\right] &= E_X\left[E(Y|X,D=1) - E(Y|X,D=0)\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[(\alpha_1 - \alpha_0) + (\beta_1 - \beta_0)X_i\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\left[\alpha_1 + \beta_1 X_i\right] - \frac{1}{n}\sum_{i=1}^{n}\left[\alpha_0 + \beta_0 X_i\right]\end{aligned}$$

– 辛普森悖论相当于犯了什么错误？

$$
\begin{aligned}
& E(Y|D=1) - E(Y|D=0) \\
&= E\big[E(Y|X,D=1)|D=1\big] - E\big[E(Y|X,D=0)|D=0\big] \\
&= \frac{1}{n_T}\sum_{t\in T}\big[\alpha_1 + \beta_1 X_t\big] - \frac{1}{n_C}\sum_{c\in C}\big[\alpha_0 + \beta_0 X_c\big] \\
&\neq \frac{1}{n}\sum_{i=1}^{n}\big[\alpha_1 + \beta_1 X_i\big] - \frac{1}{n}\sum_{i=1}^{n}\big[\alpha_0 + \beta_0 X_i\big]
\end{aligned}
$$

2.5 面板数据

- 面板数据可以用来处理一类特殊的假定**LS.1**被违背的情形——不随时间变化的不可观测因素。

$$y_{it} = \beta x_{it} + u_i + \varepsilon_{it},\ i = 1\ldots, n,\ t = 1\ldots, T$$

其中 n 较大，T 较小，适用 $n \to \infty$ 的大样本理论。（n 和 T 都趋于无穷且 n/T 为非零常数的情形下，面板数据的计量理论仍是未决的课题。）

- 可以把面板数据看作样本容量为 $n \times T$ 的混合横截面，此时 OLS 估计量的一致性要求

$$\mathsf{E}(x_{it}\varepsilon_{it}) = 0 \text{ 且 } \mathsf{E}(x_{it}u_i) = 0$$

但 x_{it} 和 u_i 的潜在相关性正是我们关心且致力于解决的问题，因此我们通常不使用混合横截面回归。

- 固定效应估计量：组内去平均变换

$$\bar{y}_i = \beta \bar{x}_i + u_i + \bar{\varepsilon}_i$$

其中 $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$, $\bar{\varepsilon}_i = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}$

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

然后采用 OLS 估计。

- 固定效应模型的关键假设：

$$\mathsf{E}\left(x_{is}\varepsilon_{it}\right) = 0, \ \ s, t = 1, \ldots, T$$

这一假设允许 x_{it} 和 u_i 之间具有任意形式的相关性, 但比 $\mathsf{E}(x_{it}\varepsilon_{it}) = 0$ 限制性更强。

- 如果 $\mathsf{E}(x_{it}u_i) = 0$ 成立，那么采用随机效应估计量可以改进估计效率 (efficiency)。但渐进效率的改进是有限的, 特别是在有限样本下。更重要的是，不能把对解决内生性的指望寄托在 Hausman 检验的 power 上。所以，**不要使用随机效应模型，无需做 Hausman 检验。**

- 固定效应模型的另一种估计方法是把固定效应看作参数而非随机变量, 从而进行虚拟变量回归 (least squares dummy variable regression).

$$y_{it} = \beta x_{it} + \sum_{j=1}^{n} \delta_j D_{it}^j + \varepsilon_{it},\ D_{it}^j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

- 应用 Frisch-Waugh-Lovell 定理，很容易证明 $\hat{\beta}_{\mathsf{LSDV}} = \hat{\beta}_{\mathsf{FE}}$.
- STATA 不直接报告 u_i (即 $\hat{\delta}_j$) 的“估计值”，通过 `predict` 命令可以得到。从 LSDV 回归关于虚拟变量的一阶条件也可以看到

$$\sum_{t=1}^{T} (y_{it} - \beta x_{it} - \delta_i) = 0$$

$$\hat{u}_i = \bar{y}_i - \hat{\beta}_{\mathsf{FE}} \bar{x}$$

而 STATA 报告的常数项估计则是

$$\hat{\beta}_0 = \frac{1}{n} \sum_{i=1}^{n} \hat{u}_i$$

- 固定效应的“估计值”是不一致的（因为 T 是固定的），但 $\hat{\beta}_0$ 却是一致的。因此关于特定的某个固定效应，我们没法讨论更多，我们并不知道它的含义，只能笼统称之为“不可观测的异质性 (unobserved heterogeneity)”，但我们可以分析这种异质性的分布性质，比如中位数、分位数等。

  示例 2. Bertrand and Schoar (2003, *QJE*)

  - 在公司层面的面板数据模型中加入了经理人的虚拟变量，然后将这种固定效应称之为“经理人风格”，他们发现经理人固定效应提高了模型的解释力，并且固定效应是联合显著的。

$$y_{it} = \alpha_t + \gamma_i + \beta x_{it} + \lambda_{CEO} + \lambda_{CFO} + \lambda_{Others} + \varepsilon_{it}$$

  - 样本涵盖 1992-1999 年约 600 家美国上市公司 500 余位经理人。
  - 固定效应的识别依赖于一个经理人任职于多个公司，但经理人的更换不是随机的，因此固定效应反映的是相关关系而非因果关系。

Executive Transitions between Positions and Industries

to:	CEO	CFO	Other
from:			
CEO	117	4	52
	63%	75%	69%
CFO	7	58	30
	71%	71%	57%
Other	106	0	145
	60%		42%

Panel A: Investment policy

	F-tests on fixed effects for				
	CEOs	*CFOs*	*Other executives*	*N*	*Adjusted* R^2
Investment				6631	.91
Investment	16.74 (, .0001, 198)			6631	.94
Investment	19.39 (, .0001, 192)	53.48 (, .0001, 55)	8.45 (, .0001, 200)	6631	.96
Inv to Q sensitivity				6631	.95
Inv to Q sensitivity	17.87 (, .0001, 223)			6631	.97
Inv to Q sensitivity	5.33 (, .0001, 221)	9.40 (, .0001, 58)	20.29 (, .0001, 208)	6631	.98
Inv to CF sensitivity				6631	.97
Inv to CF sensitivity	2.00 (, .0001, 205)			6631	.98
Inv to CF sensitivity	0.94 (.7276, 194)	1.29 (.0760, 55)	1.28 (.0058, 199)	6631	.98
N of acquisitions				6593	.25
N of acquisitions	2.01 (, .0001, 204)			6593	.28
N of acquisitions	1.68 (, .0001, 199)	1.74 (.0006, 55)	4.08 (, .0001, 203)	6593	.36

Panel B: Financial policy

	F-tests on fixed effects for							
	CEOs		*CFOs*		*Other executives*		*N*	*Adjusted* R^2
Leverage							6563	.39
Leverage	0.99	(.5294, 203)					6563	.39
Leverage	0.86	(.9190, 199)	1.43	(.0225, 54)	1.21	(.0230, 203)	6563	.41
Interest coverage							6278	.31
Interest coverage	0.56	(.99, 193)					6278	.31
Interest coverage	0.35	(.99, 192)	13.85	(, .0001, 50)	2.61	(, .0001, 192)	6278	.41
Cash holdings							6592	.77
Cash holdings	2.52	(, .0001, 204)					6592	.78
Cash holdings	2.48	(, .0001, 201)	3.68	(, .0001, 54)	2.53	(, .0001, 202)	6592	.80
Dividends/earnings							6580	.65
Dividends/earnings	5.78	(, .0001, 203)					6580	.71
Dividends/earnings	4.95	(, .0001, 199)	1.07	(.3368, 54)	1.74	(, .0001, 203)	6580	.72

Panel A: Organizational strategy

	F-tests on fixed effects for				
	CEOs	*CFOs*	*Other executives*	*N*	*Adjusted* R^2
N of diversifying acquis.				6593	.22
N of diversifying acquis.	2.06 (, .0001, 204)			6593	.25
N of diversifying acquis.	1.23 (.0163, 202)	1.74 (.0007, 53)	3.97 (, .0001, 202)	6593	.33
R&D				4283	.78
R&D	1.86 (, .0001, 145)			4283	.79
R&D	2.27 (, .0001, 143)	3.60 (, .0001, 45)	4.46 (, .0001, 143)	4283	.83
Advertising				2584	.79
Advertising	2.88 (, .0001, 95)			2584	.81
Advertising	4.03 (, .0001, 95)	0.84 (.6665, 21)	6.10 (, .0001, 80)	2584	.84
SG&A				2397	.46
SG&A	33.55 (, .0001, 123)			2397	.83
SG&A	13.80 (, .0001, 118)	0.82 (.7934, 42)	0.77 (.9777, 146)	2397	.83

Panel B: Performance

	F-tests on fixed effects for				
	CEOs	*CFOs*	*Other executives*	*N*	*Adjusted* R^2
Return on assets				6593	.72
Return on assets	2.04 (, .0001, 217)			6593	.74
Return on assets	2.46 (, .0001, 201)	3.39 (, .0001, 54)	4.46 (, .0001, 202)	6593	.77
Operating return on assets				5135	.34
Operating return on assets	2.61 (, .0001, 217)			5135	.39
Operating return on assets	1.60 (, .0001, 216)	0.66 (.9788, 58)	1.01 (.4536, 217)	5135	.39

	Median	Standard deviation	25th percentile	75th percentile
Investment	0.00	2.80	20.09	0.11
Inv to Q sensitivity	20.02	0.66	20.16	0.12
Inv to CF sensitivity	0.04	1.01	20.17	0.28
N of acquisitions	20.04	1.50	20.54	0.41
Leverage	0.01	0.22	20.05	0.09
Interest coverage	0.00	860.0	256.0	51.7
Cash holdings	0.00	0.06	20.03	0.02
Dividends/earnings	20.01	0.59	20.13	0.11
N of diversifying acquis.	20.04	1.05	20.28	0.21
R&D	0.00	0.04	20.10	0.02
SG&A	0.00	0.66	20.09	0.09
Advertising	0.00	0.04	20.01	0.01
Return on assets	0.00	0.07	20.03	0.03
Operating return on assets	0.00	0.08	20.02	0.03

	Manager-firm matched sample	
	Mean	St. dev.
Total sales	5606.5	11545.6
Investment	0.39	2.94
Average Tobin's Q	2.40	3.85
Cash flow	0.44	1.91
N of acquisitions	0.77	1.48
Leverage	0.35	0.39
Interest coverage	35.0	875.1
Cash holdings	0.11	0.16
Dividends/earnings	0.11	0.79
N of diversifying acquisitions	0.32	1.09
R&D	0.05	0.07
Advertising	0.05	0.06
SG&A	0.26	0.98
Return on assets	0.16	0.11
Operating return on assets	0.09	0.12
Sample size	6766	

	Investment	Inv to Q	Inv to CF	Cash holdings	Leverage	R&D	Return on assets
Investment							0.00
							(0.00)
Inv to Q sensitivity	**6.8**						**0.03**
	(0.92)						**(0.01)**
Inv to CF sensitivity	0.02	**−0.23**					20.01
	(0.6)	**(0.11)**					(0.01)
Cash holdings	21.10	20.79	20.46				**−0.12**
	(1.62)	(1.71)	(1.72)				**(0.05)**
Leverage	20.39	20.28	20.63	**−0.40**			20.02
	(0.55)	(0.59)	(0.60)	**(0.17)**			(0.02)
R&D	**0.07**	**0.08**	**−0.03**	**−0.23**	**−0.02**		0.11
	(0.00)	**(0.02)**	**(0.01)**	**(0.04)**	**(0.01)**		(0.11)
Advertising	0.01	**0.02**	20.01	20.01	0.00	0.25	**0.31**
	(0.01)	**(0.01)**	(0.01)	(0.04)	(0.01)	(0.15)	(0.15)
N of acquisitions	**−0.27**	0.08	**0.23**	**0.01**	**0.02**	**−0.01**	**−0.01**
	(0.11)	(0.10)	**(0.10)**	**(0.00)**	**(0.01)**	**(0.00)**	**(0.00)**
N of divers. acquis.	**−0.30**	20.14	0.14	0.01	0.01	**−0.01**	**−0.01**
	(0.13)	(0.15)	(0.14)	(0.01)	(0.02)	**(0.00)**	**(0.00)**
SG&A	**−0.22**	**−0.30**	**0.10**	0.54	0.06	**−4.32**	**−3.36**
	(0.01)	**(0.04)**	**(0.03)**	(0.56)	(0.21)	**(0.90)**	**(0.62)**

- Wooldridge (2010) 指出，当固定效应的数目随着样本容量的增大而增大时（意味着样本容量的增大并不带来更多的信息），其联合显著性的 F 检验需要依赖于很强的假设，不一定可靠。
- Fee *et al* (2013, *RFS*) 发现，打乱经理人的跳槽数据或者随机生成数据，F 检验也很容易显著。
- 这一文献激发了很多后续的研究，比如 Yao and Zhang (2015, *JEG*) 和 Xi et al (*forthcoming, JCE*)；更多的研究则转而讨论可观测特征的因果效应。

CEO style effects for job movers

	Asset growth	Capital expend.	Leverage	ROA
Panel A: F-tests on manager fixed effects				
F-statistic: Actual data (N=2,990)	1.27	1.58	2.67	3.70
(p-value)	(0.027)	(0.000)	(0.000)	(0.000)
F-statistic: Scrambled data (N=2,990)	1.28	1.56	2.42	3.89
(p-value)	(0.022)	(0.000)	(0.000)	(0.000)
F-statistic: Generated data (N=2,400)	1.17	1.51	2.31	2.57
(p-value)	(0.057)	(0.000)	(0.000)	(0.000)

- 实践中更经常采用的模型是双向乃至多向模型。

$$y_{it} = \beta x_{it} + u_i + \eta_t + \varepsilon_{it}$$

- 对于“三维”面板，理论上最细致的控制固定效应的方式是

$$y_{ict} = \beta x_{ict} + u_{ic} + \eta_{it} + \theta_{ct} + \varepsilon_{ict}$$

  固定效应数目增长很快，对计算性能要求很高，必要时有所取舍。
- 当数据层级较多时，有更丰富的控制固定效应的方式。例如，因城市而异的时间趋势 `i.city#c.year`；因城市而异的时间固定效应 `i.city#i.year`

			个体固定效应						城市固定效应				时间固定效应			时间趋势	因城市而异的时间趋势				因城市而异的时间固定效应									
firm	city	year	D1	D2	D3	D4	D5	D6	C1	C2	C3	C4	T1	T2	T3	T	C1T	C2T	C3T	C4T	C1T1	C1T2	C1T3	C2T1	C2T2	C2T3	C3T1	C3T2	C3T3	C4T1	C4T2	C4T3
1	1	1	1						1				1			1	1				1											
1	1	2	1						1					1		2	2					1										
1	1	3	1						1						1	3	3						1									
2	1	1		1					1				1			1	1				1											
2	1	2		1					1					1		2	2					1										
2	1	3		1					1						1	3	3						1									
3	2	1			1					1			1			1		1						1								
3	2	2			1					1				1		2		2							1							
3	2	3			1					1					1	3		3								1						
4	3	1				1					1		1			1			1									1				
4	3	2				1					1			1		2			2										1			
4	3	3				1					1				1	3			3											1		
5	3	1					1				1		1			1			1								1					
5	3	2					1				1			1		2			2									1				
5	3	3					1				1				1	3			3										1			
6	4	1						1				1	1			1				1										1		
6	4	2						1				1		1		2				2											1	
6	4	3						1				1			1	3				3												1

- 对于面板数据 ε_{ict}，没有理由认为同一个体不同时期的扰动项不相关，因此系数估计标准误至少聚类到个体层面。

$$
\begin{array}{c|cccccccc}
 & \varepsilon_{111} & \varepsilon_{112} & \varepsilon_{211} & \varepsilon_{212} & \varepsilon_{321} & \varepsilon_{322} & \varepsilon_{421} & \varepsilon_{422} \\
\hline
\varepsilon_{111} & \times & \times & & & & & & \\
\varepsilon_{112} & \times & \times & & & & & & \\
\varepsilon_{211} & & & \times & \times & & & & \\
\varepsilon_{212} & & & \times & \times & & & & \\
\varepsilon_{321} & & & & & \times & \times & & \\
\varepsilon_{322} & & & & & \times & \times & & \\
\varepsilon_{421} & & & & & & & \times & \times \\
\varepsilon_{422} & & & & & & & \times & \times
\end{array}
$$

如果有必要，可以聚类到城市层面，

	ε_{111}	ε_{112}	ε_{211}	ε_{212}	ε_{321}	ε_{322}	ε_{421}	ε_{422}
ε_{111}	$\times$	$\times$	$\times$	$\times$				
ε_{112}	$\times$	$\times$	$\times$	$\times$				
ε_{211}	$\times$	$\times$	$\times$	$\times$				
ε_{212}	$\times$	$\times$	$\times$	$\times$				
ε_{321}					$\times$	$\times$	$\times$	$\times$
ε_{322}					$\times$	$\times$	$\times$	$\times$
ε_{421}					$\times$	$\times$	$\times$	$\times$
ε_{422}					$\times$	$\times$	$\times$	$\times$

示例 3. Paravisini *et al* (2015, *RES*)

- 企业-产品-目的地-时间层面的出口量对企业-时间层面的银行信贷数量的回归：

$$\ln\left(X_{ipdt}\right) = \beta \cdot \ln\left(C_{it}\right) + \delta_{ipd} + \alpha_{pdt} + \varepsilon_{ipdt}$$

- δ_{ipd} 表示（例如）i 企业对 d 国 p 产品市场的了解程度；α_{pdt} 表示（例如）d 国对 p 产品的需求。
- C_{it} 可能是内生的（均衡结果），工具变量为

$$F_i \cdot Post_t \triangleq \left(\sum_b \omega_{ib} FD_b\right) \cdot \mathbf{1}(\text{After July 2008})$$

其中 ω_{ib} 为 i 企业在 b 银行负债占其总负债比例；FD_b 为 b 银行负债中外国负债的比例（事前）。

$$\ln\left(C_{it}\right) = \theta F_i \cdot Post_t + \mu_{ipd} + \gamma_{pdt} + \epsilon_{it}$$

- t 可以是月份层面的，但作者将其合并成前后两期。

$$\ln\left(C_{iPost}\right)-\ln\left(C_{iPre}\right)=\theta F_i+\gamma'_{pd}+\epsilon'_i$$

$$\ln\left(X_{ipdPost}\right)-\ln\left(X_{ipdPre}\right)=\beta\cdot\left[\ln\left(C_{iPost}\right)-\ln\left(C_{iPre}\right)\right]+\alpha'_{pd}+\varepsilon'_{ipd}$$

- 标准误聚类到产品-目的地层面。

- 报告 R^2：可以报告 `xtreg` 的 within R^2, 也可以报告 `reghdfe` 的 R^2, 但不建议报告 `xtreg` 的 overall R^2.

$$\begin{aligned}
\text{within } R^2 &= \rho^2(y_{it} - \bar{y}_i, \mathbf{x}_{it}'\hat{\boldsymbol{\beta}} - \bar{\mathbf{x}}_i'\hat{\boldsymbol{\beta}}) \\
\text{between } R^2 &= \rho^2(\bar{y}_i, \bar{\mathbf{x}}_i'\hat{\boldsymbol{\beta}}) \\
\text{overall } R^2 &= \rho^2(y_{it}, \mathbf{x}_{it}'\hat{\boldsymbol{\beta}})
\end{aligned}$$

- 固定效应模型的参数识别依赖于同一个体随时间的变化，需要关注这种变化的来源；换句话说，固定效应模型得不到显著的结果，可能并不是因为 x 不影响 y，而是 x 的逐期变动中的信息含量较少；极端情况下，当 x 不随时间变化时，无法识别其对 y 的影响（与固定效应完全共线性），可以将 x 与时间虚拟变量做交互，但只是转移了焦点，并没有解决问题。

本文使用省级(包括直辖市)党代会数据和地级市本级政府财政数据考察中国地方政府是否存在政治预算周期,设定了如下的估计模型:

$$g_{it} = \alpha_0 + \alpha_1 \mathrm{CCP4}_{it} + \alpha_2 \mathrm{CCP0}_{it} + \alpha_3 \mathrm{CCP1}_{it} + \alpha_4 \mathrm{CCP2}_{it} + X'\beta + \lambda_i + \mu_t + \varepsilon_{it}$$

其中,角标 i 表示地级市,角标 t 表示时间,被解释变量 g_{it} 表示市本级政府一般预算支出的增长率,计算公式为:当年一般预算支出增长率=ln(当年一般预算支出/上年一般预算支出)。在后文中,由于研究需要,我们还会将被解释变量换为一般预算收入增长率等增长率变量,具体的被解释变量将在回归部分予以说明。$\mathrm{CCP4}_{it}$ 表示党代会召开前一年(本届党代会召开前一年即上一届党代会开完的第四年),$\mathrm{CCP0}_{it}$ 表示党代会召开当年,$\mathrm{CCP1}_{it}$ 和 $\mathrm{CCP2}_{it}$ 分别表示党代会召开后第一年和党代会召开后第二年。回归方程中没有加入的党代会召开后第三年($\mathrm{CCP3}_{it}$),以其为比较的基准。X 为表示其他控制变量的矩阵,在地级

本文将采用1999年至2011年281个城市的面板数据，考察省级党代会的召开对土地出让面积的影响，基本的计量模型设定如下：

$$\ln land_{it} = \alpha_0 + \alpha_1 PPCpre2_{it} + \alpha_2 PPCpre1_{it} + \alpha_3 PPC_{it} + \alpha_4 PPCpost1_{it} + X_{it}\beta + \mu_i + \alpha_5 trend_t + \varepsilon_{it} \quad (9)$$

其中，$\ln land_{it}$为地级市 i 在 t 年国有土地出让总面积对数值。$PPCpre2_{it}$、$PPCpre1_{it}$、PPC_{it}和 $PPCpost1_{it}$是四个虚拟变量，用于刻画省党代会（Provincial Party Congress）的影响。如果省党代会在当年召开，$PPC_{it}=1$，否则取0；省党代会在未来两年召开，$PPCpre2_{it}=1$，否则取0；省党代会在未来一年召开，$PPCpre1_{it}=1$，否则取0；省党代会在去年召开，$PPCpost1_{it}=1$，否则取0。X_{it}是一组可能影响土地出让

① 为了考察省党代会的影响，年份效应没有在回归模型中控制。原因在于如果在回归模型中同时放入省党代会虚拟变量和年份虚拟变量，我们无法区分省党代会的影响。因此，模型控制时间趋势，而没有像通常情况下控制年份虚拟变量。

- 不能通过将解释变量替换成其滞后项来缓解内生性问题，也不能使用解释变量的滞后项作为其工具变量。
- 对于非线性模型，消除固定效应的方法失效，此时要得到解释变量系数的一致估计需要限制性很强的假设，因此对于结构复杂的面板数据，研究者更多地采用线性概率模型。

2.6 交互项模型

两个离散变量的交互

- 本质：Difference in differences in means.

$$y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 (D_1 \times D_2) + \varepsilon$$

- 例如：y 表示工资，$D_1 = 1$ 表示男性，$D_2 = 1$ 表示上大学，

$$E(y) = \begin{cases} \beta_0 + \beta_1 + \beta_2 + \beta_3 & \text{if } D_1 = 1 \ \& \ D_2 = 1 \\ \beta_0 + \beta_1 & \text{if } D_1 = 1 \ \& \ D_2 = 0 \\ \beta_0 + \beta_2 & \text{if } D_1 = 0 \ \& \ D_2 = 1 \\ \beta_0 & \text{if } D_1 = 0 \ \& \ D_2 = 0 \end{cases}$$

易知 β_0 表示没上大学女性的平均工资；β_1 表示没上大学人群的工资性别溢价；β_2 表示女性的上大学回报；

$$\begin{aligned}
\beta_3 = & \underbrace{[E(y|D_1=1,D_2=1) - E(y|D_1=1,D_2=0)]}_{\text{男性的上大学回报}} \\
& - \underbrace{[E(y|D_1=0,D_2=1) - E(y|D_1=0,D_2=0)]}_{\text{女性的上大学回报}} \\
= & \text{ 上大学回报的性别差异} \\
= & \underbrace{[E(y|D_1=1,D_2=1) - E(y|D_1=0,D_2=1)]}_{\text{上大学人群的工资性别歧视}} \\
& - \underbrace{[E(y|D_1=1,D_2=0) - E(y|D_1=0,D_2=0)]}_{\text{没上大学人群的工资性别歧视}} \\
= & \text{ 工资性别歧视在不同受教育程度人群中的差异}
\end{aligned}$$

- 交互项模型不止一种写法, 可以根据需要（希望得到边际效应的绝对值还是相对值）选择哪些 main terms 和 interaction terms 进入回归。
- 常见的双重差分模型是这类模型的特例,

$$y_{it} = \beta_0 + \beta_1 \cdot \texttt{treat}_i + \beta_2 \cdot \texttt{post}_t + \beta_3 \cdot \texttt{treat}_i \times \texttt{post}_t + \varepsilon_{it}$$

  我们通常会控制个体固定效应和时间固定效应, 此时两个 main terms 会被吸收,

$$y_{it} = \gamma \cdot \texttt{treat}_i \times \texttt{post}_t + \mu_i + \eta_t + \varepsilon_{it}$$

离散变量和连续变量的交互

$$y = \beta_0 + \beta_1 x + \beta_2 D + \beta_3 (x \times D) + \varepsilon$$

- 系数估计结果等价于分组回归，但标准误不同。在包含控制变量的模型中，与所有解释变量的完全交互才能得到等价的系数估计结果。
- 尽管分组回归更直观，但交互项模型可以对斜率系数的差异进行统计检验。
  - 不能仅比较两组系数估计值的大小就得出结论，例如，两个估计值一个显著，另一个不显著，其差异可能是不显著的。
  - 如果差异不显著，诚实报告结果："在其中一组发现了显著的效应，但在另一组没发现"，"差异接近显著（p 值为 0.11）"或"差异在 12% 水平下显著"。

- 如何报告结果？以 Bai and Jia (2016, *ECMA*) 为例。

$$R_{pt} = \beta \ln Q_p \times Post_t + \lambda_p + \gamma_t + \varepsilon_{pt}$$

$$\hat{\beta} = 0.112$$

as the effect of quotas per capita. On average, a one standard deviation increase in the logged quota (0.57 after controlling for logged population size) implies about a six percentage point higher probability of revolution participation, which is about 40% of the mean probability (16 percentage points). Col-

为了看得更清楚，稍微改动一下回归方程：

$$R_{pt} = \beta \ln Q_p \times Post_t + \lambda \ln Q_p + \gamma Post_t + \varepsilon_{pt}$$

$$\frac{\partial R_{pt}}{\partial \ln Q_p} = \beta Post_t + \lambda$$

解释一：$\ln Q_p$ 增加一个标准差 σ，革命概率增加 $\beta\sigma$.

解释二：$\ln Q_p$ 增加一个标准差 σ，取消科举后革命概率增加 $\beta\sigma$.

正确解释：

$$\frac{\partial E(R_{pt}|Post_t = 1)}{\partial \ln Q_p} - \frac{\partial E(R_{pt}|Post_t = 0)}{\partial \ln Q_p} = \beta$$

$\ln Q_p$ 增加一个标准差 σ，由此带来的革命概率增加，在取消科举后比取消科举前要高 $\beta\sigma$.

$$\frac{\partial \left[E(R_{pt}|Post_t = 1) - E(R_{pt}|Post_t = 0)\right]}{\partial \ln Q_p} = \beta$$

取消科举带来的革命概率增加，当 $\ln Q_p$ 增加一个标准差 σ 时，会提高 $\beta\sigma$.

- **交互项的系数永远要在双重差分的意义上进行解释。**

- 适用于考察因果关系的组间异质性，有时候异质性本身是重要的，有时候是为了从中发现因果关系作用的机制。

  Insignificane may indicate not the end of the world, but the light of hope: Countervailing mechanisms or group-specific heterogeneities are waiting for your call.

- 估计 Flexible nonlinear relationship，近似非参数回归的效果。

两个连续变量的交互

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 \times x_2) + \varepsilon$$

- 适用于考察两个解释变量对被解释变量作用的互补性 ($\beta_3 > 0$) 或替代性 ($\beta_3 < 0$)，例如，官员能力和关系对晋升的影响 (Jia *et al*, 2015, *JEEA*)。但更多时候仍然是为了考察因果关系的作用机制。
- 如何报告结果？

  示例 4. 金融发展与经济增长 (Rajan and Zingales, 1998, *AER*)

$$\begin{aligned}\texttt{Growth}_{jk} =& \beta\ \texttt{ExtDep}_j \times \texttt{FinaDev}_k + \gamma\ \texttt{FracManu}_{jk} \\ &+ \texttt{Industry}_j + \texttt{Country}_k + \varepsilon_{jk}\end{aligned}$$

  - $\texttt{Growth}_{jk}$: 1980-1990 年 k 国 j 行业增加值年均实际增长率。
  - $\texttt{FracManu}_{jk}$: 1980 年 j 行业增加值在 k 国制造业总增加值占比。
  - $\texttt{ExtDep}_j$: j 行业外部融资依存度，以美国上市公司的实际外部融资依存度作为代理指标。（因此估计样本中不包含美国。）
  - $\texttt{FinaDev}_k$: 金融发展指标——1990 年各国会计标准（测度各国企业年报信息透明度）。

	Countries below the median in accounting standards	Countries above the median in accounting standards
Least financially dependent industries		
Tobacco	0.53	−0.60
Pottery	0.25	−0.30
Leather	0.77	−0.77
Most financially dependent industries		
Drug	−1.11	1.30
Plastics	−0.21	0.21
Computers	−2.00	1.80

– Differential in real growth rate.

$$\texttt{ExtDep}\ 75\text{-}25\ 分位差 \times \texttt{FinaDev}\ 75\text{-}25\ 分位差 \times \hat{\beta}$$

The interaction term is akin to a second derivative. One way to get a sense of its magnitude is as follows. The industry at the 75th percentile of dependence (high dependence) is Machinery. The industry at the 25th percentile (low dependence) is Beverages. The country at the 75th percentile of development as measured by capitalization is Italy, while the country at the 25th percentile is the Philippines. We set the industry's initial share of manufacturing at its overall mean. The coefficient estimate then predicts that Machinery should grow 1.3 percent faster than Beverages annually, and in real terms, in Italy as compared to the Philippines. For comparison, the real annual growth rate is, on average, 3.4 percent per year. So a differential of 1.3 percent is a large number.

- 不论何种交互项模型，都要尽可能控制固定效应，而非仅仅 main terms（如果 x 在两个维度上都有 variation，则还可以同时控制 main terms）。
- 不论何种交互项模型，都必须包含所有的 main terms 和 interaction terms（或 interactive fixed effects），否则可能得到虚假的交互效应。

如何使用交互项模型论证因果关系？

- 我们发现了 x_1 与 y 的相关性，并且想要主张 x_1 是 y 的原因，可以通过检验 x_1 影响 y 的某个具体机制来对从 x_1 到 y 的因果关系进行论证，论证的逻辑如下：

1. 提出一个 x_1 影响 y 的理论 T。根据这个理论，x_1 通过某个机制 M 影响 y，并且可以识别出 M 在某些 subpopulation 中存在，在另一些 subpopulation 中不存在，令 $x_2 = 1$ 表示存在 M，$x_2 = 0$ 表示不存在 M。
2. 在 $x_2 = 1$ 组，x_1 与 y 的相关性继续存在，而在 $x_2 = 0$ 组，x_1 与 y 的相关性不复存在。
3. 可能导致 x_1 与 y 的相关性的竞争性解释包括 y 影响 x_1 的反向因果理论 R，或者有混淆因素同时影响 x_1 和 y 的遗漏变量理论 C。如果无法想象理论 R 或理论 C 发挥作用的机制在 $x_2 = 1$ 和 $x_2 = 0$ 组存在显著差异，则很可能理论 R 或理论 C 不成立。否则，我们应该在 $x_2 = 0$ 组也观察到 x_1 与 y 的相关性。

4. 有时候，两个组中 x_1 与 y 的相关性都存在，但在 $x_2 = 1$ 组这种相关性更强，表现在对 y 的回归中，x_1 的系数估计在 $x_2 = 1$ 组更大，且组间差异在统计上显著。这时我们至少可以说，x_1 与 y 的相关性不全是理论 R 或理论 C 所带来的，否则这种相关性应该在 $x_2 = 1$ 和 $x_2 = 0$ 组无差异。
5. 当 x_2 是连续变量时，可以做类似的理解，相当于机制 M 在 x_2 较大时存在（或更明显），在 x_2 较小时不存在（或更不明显）。

- 在 Rajan and Zingales (1998) 中，金融发展水平 (x_1) 与经济增长 (y) 强相关，文章想说金融发展是经济增长的原因，并检验了金融发展通过缓解企业的外部融资约束 (M) 从而促进了企业成长这一理论 (T)。文章将行业分成两组，一组是外部融资依存度 (x_2) 较高的行业（存在 M），另一组是外部融资依存度较低的行业（不存在 M），发现在对行业增长的回归中，金融发展水平与外部融资依存度的交互项显著，表明金融发展水平与行业增长之间的相关性在外部融资依存度不同的组间存在显著差异。金融与增长之间的相关性可能是因为增长影响金融，高增长引发了融资需求从而导致金融市场发展（理论 R），也可能是因为某个混淆因素（例如节俭传统）同时影响金融发展和经济增长（理论 C），那么除非理论 R 和理论 C 在外部融资依存度不同的组间发挥作用的程度不同，否则就证明了理论 T。

- 好的 x_2 本身应该比较稳定，或者其变动是外生的，尤其是 x_2 不受 x_1 或 y 影响。
- 坏的 x_2 相当于在 DID 研究中，处理组和控制组的构成一直在变化 (compositional change)，并且导致这种变化的因素和 y 相关（隐藏在扰动项之中），那么处理组和控制组的平行趋势假定就不再成立。把 Rajan and Zingales (1998) 看作一个 DID，低外部融资依存组在不同金融发展水平国家的增长差异为 α，高外部融资依存组在不同金融发展水平国家的增长差异为 $\alpha + \beta$，但如果外部融资依存度本身在变化，那么高外部融资依存组的反事实趋势不一定是 α（可能是因为高增长潜力的行业同时是高外部融资依存的，用本国本行业实际的外部融资依存度——而不是美国该行业的外部融资依存度——作为 x_2 就可能存在这一问题）。
- 从另一个角度看，如果 x_1 和 x_2 相关，$x_1 \cdot x_2$ 代表的可能是 x_2^2。

如何在交互项模型中讨论遗漏变量偏误和反向因果？仍以 Rajan and Zingales (1998) 为例。

- 遗漏变量偏误

> Do external dependence or financial development proxy for something else? In principle, there is a long list of sources of comparative advantage that may dictate the presence, absence, or growth of industries in a country. Our results, though, cannot be explained unless the dependence of industries on this source of comparative advantage is strongly correlated with their dependence on external funding *and* financial development is a good proxy for the source of comparative advantage. We rule out two such possibilities below.

如果交互项反映的是别的因素，那么只有当该因素与外部融资依存度的相关性很强，同时与金融发展的相关性也很强时，才会对研究设计造成威胁。

– 外部融资依存度高的行业可能人力资本需求也高，而人力资本发展水平有可能跟金融发展水平相关。交互项效应反映的可能是对人力资本需求越高的行业在人力资本发展水平越高的国家增长越快。

  做法：在回归中控制平均受教育年限与外部融资依存度的交互项。

– 外部融资依存度低的行业可能是成熟行业，金融发展水平低的国家可能是发展中国家，交互项效应反映的可能是越成熟的行业在发展中国家增长越快（技术转移越充分）。

  做法：在回归中控制人均收入水平与外部融资依存度的交互项。

- 反向因果
  - 可能有别的因素（例如自然资源禀赋）导致某些行业相对快速增长，而这些行业恰好又是外部融资依存度高的，因此对金融发展产生需求，如果这种因素持续存在，则会导致交互项效应显著。

    做法：将样本限制在增加值在该国占比高于中位数的行业（占比较高，则有理由认为该国拥有适合该行业发展的条件，在这一子样本中，潜在有利因素的 variation 不复存在）。

  - 美国历史表明，金融有可能确实是应某些行业的需要而发展的，继而带动了另一些更年轻行业的发展。

    做法：将样本限制在增加值在该国占比低于中位数的行业（对这些年轻行业而言，金融发展水平是前定的，更不容易存在反向因果）。