

Please do not distribute without permission.

定量社会科学的因果推断

Causal Inference in Quantitative Social Sciences

江 艇

中国人民大学经济学院

Last updated: June 13, 2020

Lecture 5 线性回归理论回顾（下）

Frisch-Waugh-Lovell 定理

- 考察如下多元回归,

$$y_i = b_1 + b_2x_{i2} + \cdots + b_kx_{ik} + \cdots + b_Kx_{iK} + e$$

- 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,

$$y_i = c_1 + c_2x_{i2} + \cdots + c_{k-1}x_{i,k-1} + c_{k+1}x_{i,k+1} + \cdots + c_Kx_{iK} + \tilde{y}_i$$

$$x_{ik} = d_1 + d_2x_{i2} + \cdots + d_{k-1}x_{i,k-1} + d_{k+1}x_{i,k+1} + \cdots + d_Kx_{iK} + \tilde{x}_{ik}$$

则

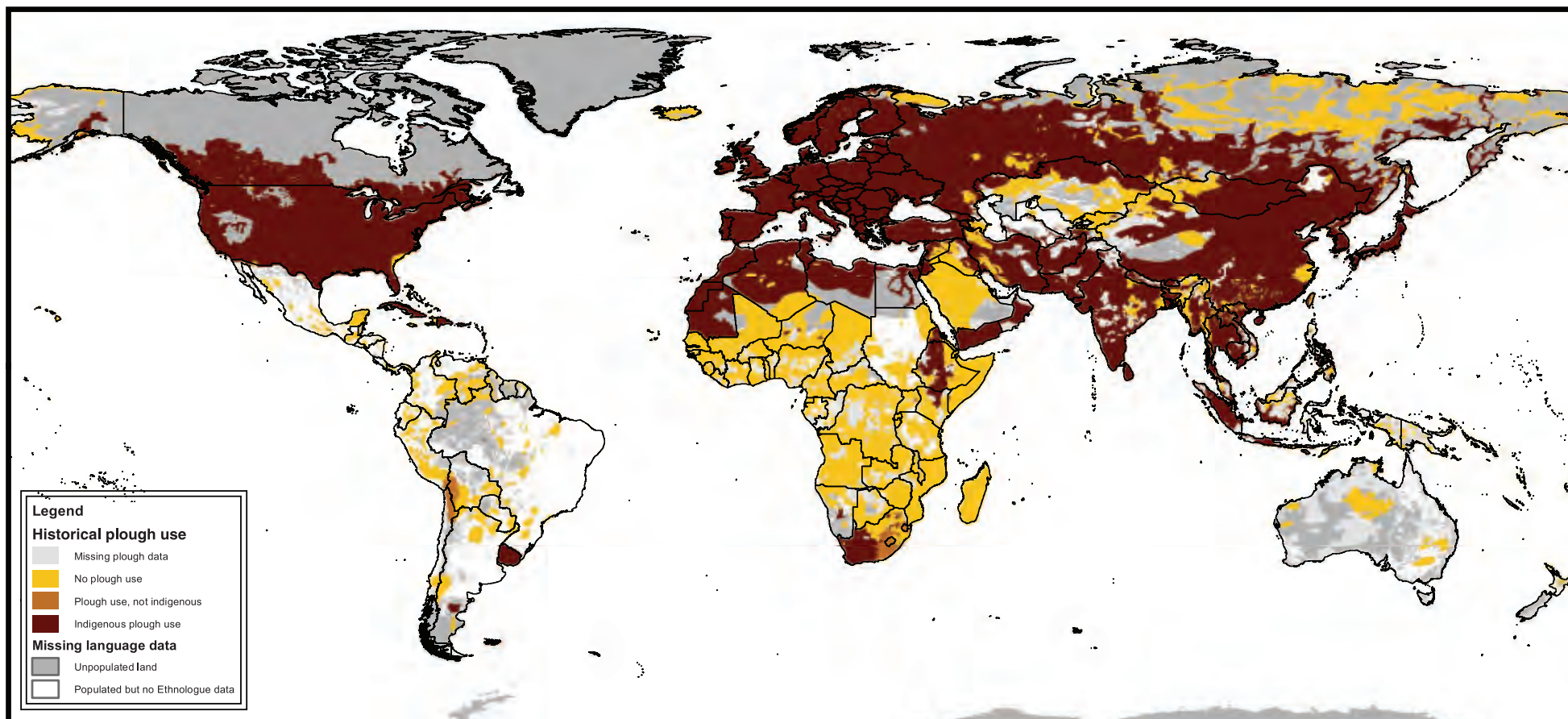
$$b_k = \frac{\widehat{\text{Cov}}(y, \tilde{x}_k)}{\widehat{\text{Var}}(\tilde{x}_k)} = \frac{\widehat{\text{Cov}}(\tilde{y}, \tilde{x}_k)}{\widehat{\text{Var}}(\tilde{x}_k)}$$

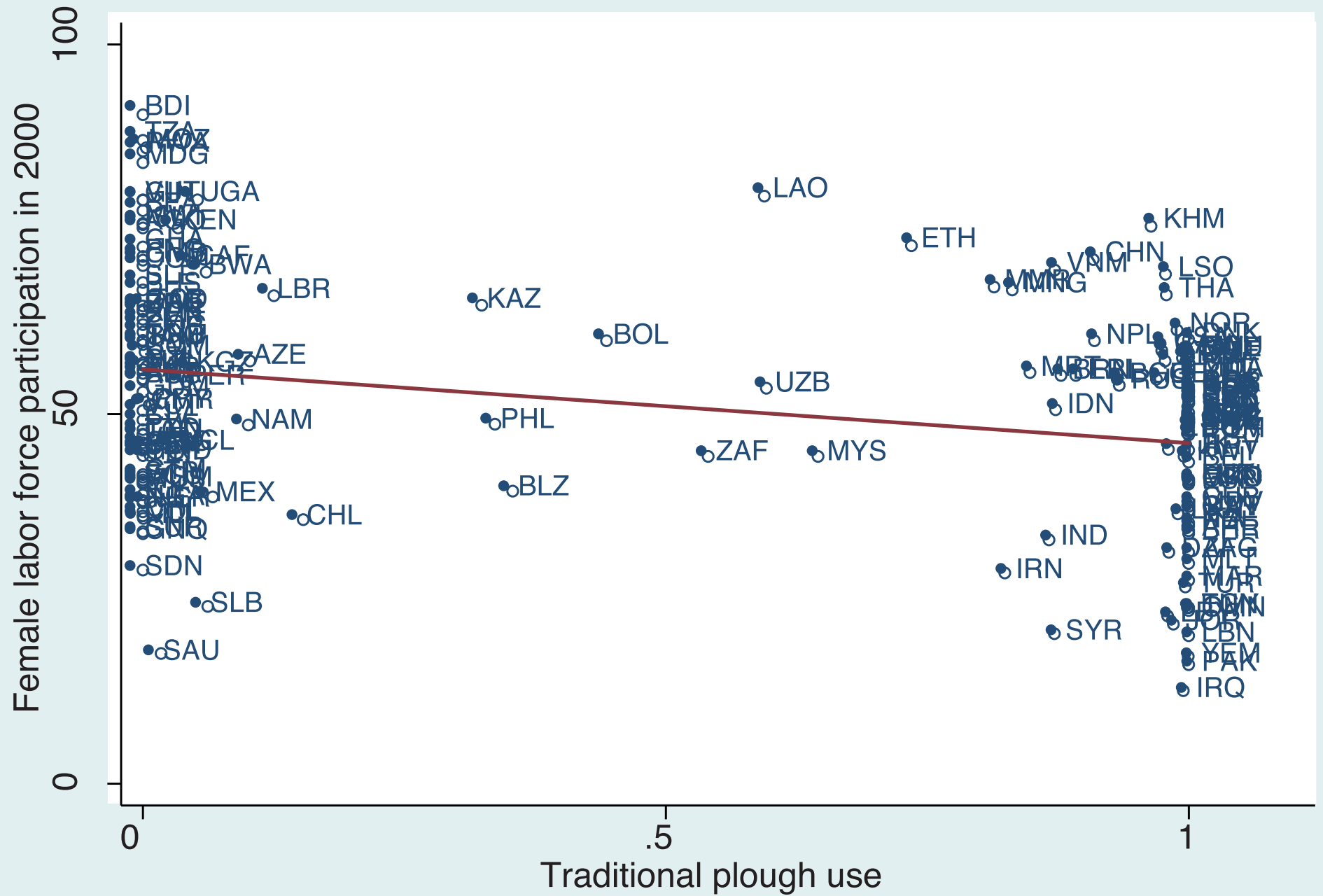
其中后者还能给出正确的标准误。

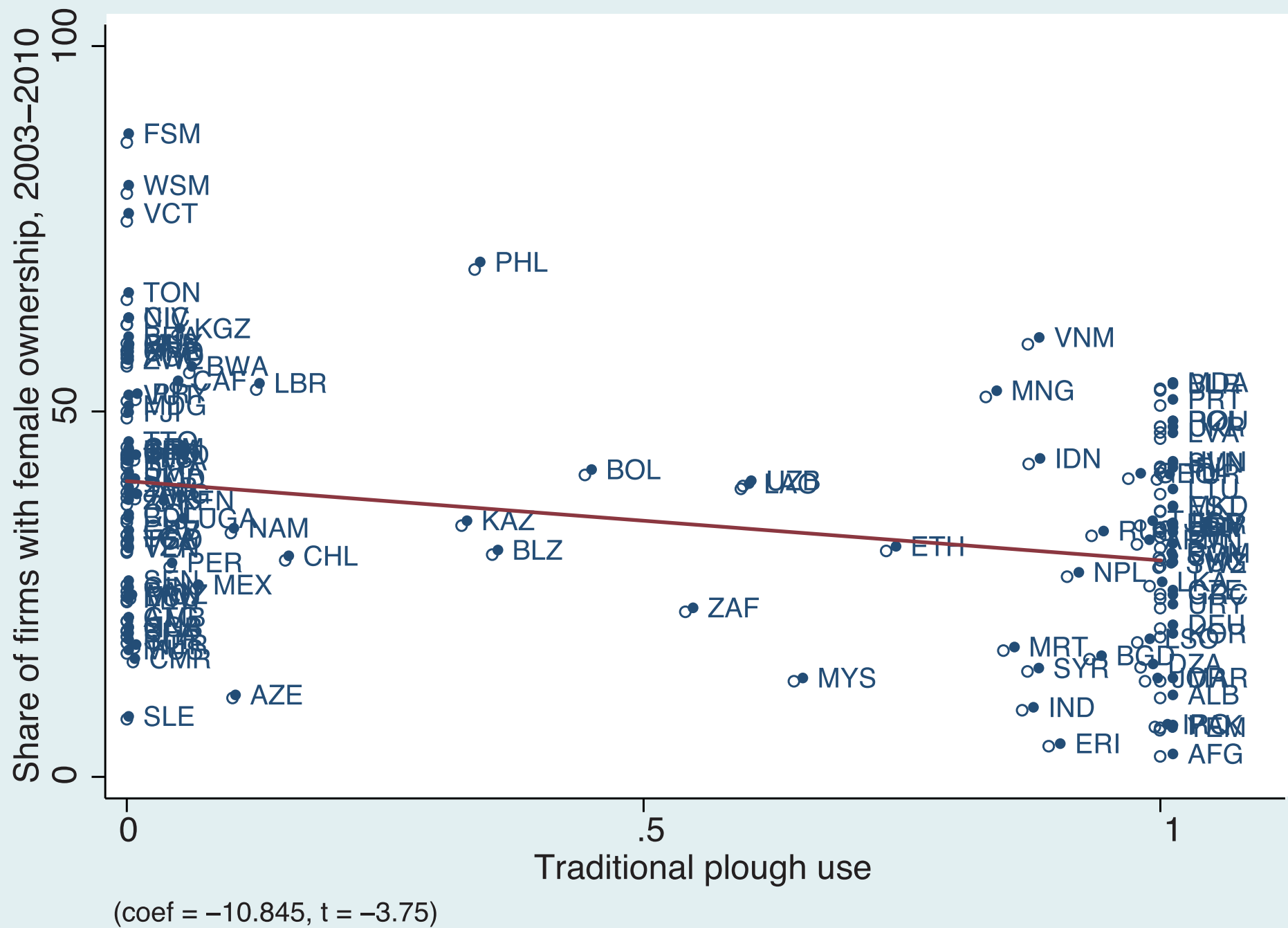
- 应用一：含控制变量的回归中如何直观展示核心解释变量的效应？

示例 17. 女人和犁 (Alesina *et al*, 2013, *QJE*)

- The proportion of respondents in the World Values Survey that “agree” with the statement that “when jobs are scarce, men should have more right to a job than women” varies widely across countries, ranging from 3.6% (in Iceland) to 99.6% (in Egypt).
- Differences in gender roles have their origins in the form of agriculture traditionally practiced in the pre-industrial period (shifting cultivation vs. plough cultivation).







- 应用二：理解虚拟变量控制

- 要想准确估计一个解释变量的系数，这个解释变量必须具有足够的变动性 (variation)。例如，在教育回报率的例子中，

$$\text{wage}_i = b_0 + b_1 \cdot \text{edu}_i + e_i$$

为了估计 b_1 ，要求样本中存在各种不同受教育水平的个体。

- 现在方程右边加入性别控制变量（作用何在？），

$$\text{wage}_i = b_0 + b_1 \cdot \text{edu}_i + c \cdot \text{male}_i + e_i$$

这等价于

$$\widetilde{\text{wage}}_i = b_1 \cdot \widetilde{\text{edu}}_i + e_i$$

其中 $\widetilde{\text{wage}}_i$ 和 $\widetilde{\text{edu}}_i$ 分别是 wage_i 和 edu_i 对 male_i 回归得到的残差。

– 易知

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

因此对于估计 b_1 而言真正有用的变动性不是受教育水平的整体变动性，而是其在性别组内的变动性。我们把控制 male_i 的操作称为控制性别固定效应。

– 类似地，如果在方程右边加入地区控制变量，

$$\text{wage}_i = b_0 + b_1 \cdot \text{edu}_i + \sum_{r=2}^R c^r \cdot \text{region}_i^r + e_i$$

其中

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

等价于

$$\widetilde{\text{wage}}_i = b_1 \widetilde{\text{edu}}_i + e_i$$

其中 $\widetilde{\text{wage}}_i$ 和 $\widetilde{\text{edu}}_i$ 分别是 wage_i 和 edu_i 对所有地区虚拟变量回归得到的残差，也即去地区均值以后工资和受教育水平。控制了地区固定效应以后，真正有用的变动性是地区内部受教育水平的变动性。

重新理解关键假设

- 回到教科书上的结构模型,

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

- 套用关键假设**LS.1**, 要求扰动项与所有 \mathbf{x}_i 不相关,

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

- \mathbf{x}_i 中往往包含控制变量, 而我们在一项研究中不可能去严肃讨论控制变量的外生性, 并且我们通常不关心控制变量的因果效应; 因此在研究中实际采用的假设是条件不相关 (conditional uncorrelatedness) 假设。
- 考虑如下的结构模型,

$$Y = \beta_0 + \beta_1 D + \beta_2 X + \varepsilon$$

其中 D 是核心解释变量, X 是控制变量。

- 若 $\text{Cov}(\varepsilon, D) = \text{Cov}(\varepsilon, X) = 0$, 即 D 和 X 均与扰动项不相关, 则 β_1 和 β_2 的 OLS 估计均反映因果效应。

- 若不给定 X , D 与扰动项相关；一旦给定 X , D 与扰动项条件不相关。则 β_1 的 OLS 估计是一致的, 而 β_2 的 OLS 估计可能不一致。

Assumption LS.2: $\text{Cov}(\varepsilon, D|X) = 0$

- 如何理解条件不相关？

$$Y = \beta_0 + \beta_1 D + \beta_2 X + \varepsilon$$

$$\varepsilon = \delta_0 + \delta_1 D + \delta_2 X + v$$

$$\mathbb{E}(v) = 0, \text{Cov}(D, v) = \text{Cov}(X, v) = 0$$

$$Y = (\beta_0 + \delta_0) + (\beta_1 + \delta_1)D + (\beta_2 + \delta_2)X + v$$

$$b_1 \rightarrow_p \beta_1 + \delta_1$$

条件不相关性即 $\delta_1 = 0$.

- 例一： Y 是大学数学考试成绩， D 表示所在专业是否布置作业， X 表示所在专业是否为理工科。假定是否布置作业是在理工科和非理工科内部各专业间随机分配的，理工科专业布置作业的比例较高，非理工科专业布置作业的比例较低。此时如果不控制 X ， D 系数的 OLS 估计可能反映的是理工科专业数学考试成绩较高（因为布置作业的专业中理工科的比例较高）。控制 X 的 OLS 可以得到布置作业的因果效应，但 X 本身对 Y 的因果效应估计是不一致的，因为所在专业是否为理工科并不是随机分配的（可能和扰动项中的数学天赋相关）。

- 例二：在第 1 讲的**示例 6**中， Y 是大学期末考试成绩， D 是是否在大学课堂上使用电脑， X 是高中学习成绩， ε 中包含学习习惯。控制 X 的理由是，一方面，高中成绩反映知识储备，直接影响大学成绩，另一方面，高中成绩反映学习习惯，高中成绩越差，说明学习习惯越差，越倾向于在课堂上使用电脑。如果控制住高中成绩以后，学习习惯不再额外影响是否使用电脑（或者等价地说，高中成绩相同的学生，使用电脑的和不使用电脑的，学习习惯相当），则 OLS 可以得到使用电脑的因果效应的一致估计。

- 例三：在第 1 讲的**示例 2**中，以学区为观测单位， Y 是学生的平均成绩， D 是班级的平均规模， X 是享受午餐补助的学生比例。 ε 中包含学生的经济状况，它既影响 Y （反映课外学习机会的多寡），也影响 D （反映地区的财政实力）。如果控制 X 能够控制住经济状况与 D 的相关性，则可以一致地估计班级规模的因果效应。而 X 的系数很可能是负的，但这并不意味着取消午餐补助能够提高学生平均成绩，而是因为正的 β_2 被负的 δ_2 所抵消。

- 可以从代理变量的角度理解例二和例三：高中成绩作为学习习惯的代理变量，享受午餐补助的学生比例作为地区家庭经济状况的代理变量。

– 假定结构模型为

$$Y = \beta_0 + \beta_1 D + \beta_2 u + v, \text{Cov}(D, v) = \text{Cov}(u, v) = 0$$

其中 u 不可观测或很难度量，考虑使用 X 作为其代理变量 (proxy variable)，代理变量的统计结构为

$$u = \theta_0 + \theta_1 X + \omega, \text{Cov}(X, \omega) = 0$$

$$Y = (\beta_0 + \beta_2 \theta_0) + \beta_1 D + \beta_2 \theta_1 X + (\beta_2 \omega + v)$$

– 可见，合格的代理变量要满足

$$\text{Cov}(D, \omega) = 0$$

u 中剔除了和 X 相关部分之后剩余的部分，与 D 不相关。

– 使用代理变量不但纠正了估计偏误，还提高了估计效率：

$$\text{Var}(\beta_2 \omega + v) < \text{Var}(\beta_2 u + v)$$

- 但控制变量的选择也要慎重：如果用一个无关变量作为代理变量，反而会放大估计偏误。沿用前面的结构模型，

$$Y = \beta_0 + \beta_1 D + (\beta_2 u + v), \text{Cov}(D, v) = \text{Cov}(u, v) = 0$$

$$\hat{\beta}_1^{\text{OLS}} \rightarrow_p \beta_1 + \beta_2 \frac{\text{Cov}(D, u)}{\text{Var}(D)}$$

假定使用了一个无关变量 X 作为代理变量

$$\text{Cov}(X, u) = \text{Cov}(X, v) = 0$$

根据 FWL 定理， Y 对 D 和 X 做回归，相当于 Y 对 \tilde{D} 做回归，

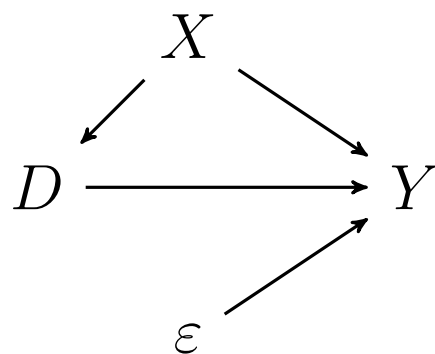
$$D = \pi_0 + \pi_1 X + \tilde{D}$$

$$Y = (\beta_0 + \beta_1 \pi_0) + \beta_1 \tilde{D} + (\beta_1 \pi_1 X + \beta_2 u + v)$$

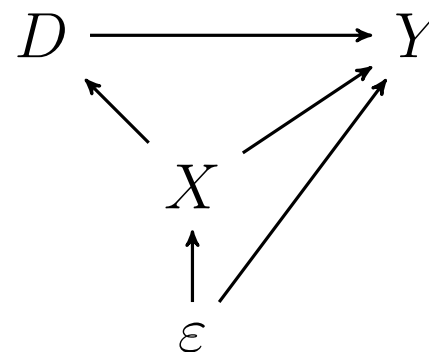
$$\hat{\beta}_1^{\text{OLS, w/ } X} \rightarrow_p \beta_1 + \frac{\text{Cov}(\tilde{D}, \beta_1 \pi_1 X + \beta_2 u + v)}{\text{Var}(\tilde{D})} = \beta_1 + \beta_2 \frac{\text{Cov}(\tilde{D}, u)}{\text{Var}(\tilde{D})}$$

$$\left| \frac{\text{Cov}(D, u)}{\text{Var}(D)} \right| < \left| \frac{\text{Cov}(\tilde{D}, u)}{\text{Var}(\tilde{D})} \right|$$

- 控制变量的双重作用：首先是控制 X 的直接效应 (β_2)，使得对回归系数的估计更准确（回归模型整体拟合程度更高， s^2 更小，从而系数估计的标准误更小）；但更重要的是切断影响 Y 的其它因素（隐藏在扰动项中）与 D 的相关性，研究者希望，这个因素与 X 相关 (δ_2)，并且一旦控制 X 以后，这个因素不再与 D 相关。



$$\delta_2 = 0$$



$$\delta_2 \neq 0$$

- 启示：在一项实证研究中不要过度解读控制变量的系数估计结果。

- 当 X 是连续变量时，线性结构模型依赖于函数形式假设。当关于 $\mathbb{E}(Y^1|X)$ 和 $\mathbb{E}(Y^0|X)$ 的函数形式假设——即关于 $\mathbb{E}(Y|D = 1, X)$ 和 $\mathbb{E}(Y|D = 0, X)$ 的函数形式假设——正确时，可以类似地证明，**假设LS.2和假设ID.2是等价的。**

系数稳定性理论

- 当控制了关键控制变量之后，如果核心解释变量的系数估计不再随着更多变量的加入而发生大幅变化，那么就说明潜在的选择性偏误可能很小了。
- 假定结构模型为

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \gamma W + \varepsilon$$

X 为核心解释变量， Z 为不可观测的选择性变量， W 为可观测的选择性变量。如果在已经控制 W 的前提下，回归中忽略了 Z ，则

$$\hat{\beta}_1 \rightarrow_p \beta_1 + \beta_2 \cdot \frac{Cov(\tilde{X}, \tilde{Z})}{Var(\tilde{X})}$$

其中 \tilde{X} 和 \tilde{Z} 是 X 和 Z 对 W 回归得到的残差。

如果 $Cov(\tilde{X}, \tilde{Z}) / Var(\tilde{X})$ 比较小，即使 β_2 不为零，选择性偏误也比较小。

- 考察系数稳定性的原理：控制了关键控制变量之后，新加入的控制变量与核心解释变量的相关性比较小，有理由相信，倘若存在遗漏的选择性变量，其与核心解释变量的相关性也比较小。再看 Dale and Krueger (2002).

$$1.165 \times 0.051 \approx 0.212 - 0.152$$

$$0.066 \times 0.036 \approx 0.034 - 0.031$$

	No selection controls			Selection controls		
	(1)	(2)	(3)	(4)	(5)	(6)
Private school	.212 (.060)	.152 (.057)	.139 (.043)	.034 (.062)	.031 (.062)	.037 (.039)
Own SAT score ÷ 100		.051 (.008)	.024 (.006)		.036 (.006)	.009 (.006)
Log parental income			.181 (.026)			.159 (.025)
Female			-.398 (.012)			-.396 (.014)
Black			-.003 (.031)			-.037 (.035)
Hispanic			.027 (.052)			.001 (.054)
Asian			.189 (.035)			.155 (.037)
Other/missing race			-.166 (.118)			-.189 (.117)
High school top 10%			.067 (.020)			.064 (.020)
High school rank missing			.003 (.025)			-.008 (.023)
Athlete			.107 (.027)			.092 (.024)
Average SAT score of schools applied to ÷ 100				.110 (.024)	.082 (.022)	.077 (.012)
Sent two applications				.071 (.013)	.062 (.011)	.058 (.010)
Sent three applications				.093 (.021)	.079 (.019)	.066 (.017)
Sent four or more applications				.139 (.024)	.127 (.023)	.098 (.020)

	Dependent variable					
	Own SAT score ÷ 100			Log parental income		
	(1)	(2)	(3)	(4)	(5)	(6)
Private school	1.165 (.196)	1.130 (.188)	.066 (.112)	.128 (.035)	.138 (.037)	.028 (.037)
Female		-.367 (.076)			.016 (.013)	
Black		-1.947 (.079)			-.359 (.019)	
Hispanic		-1.185 (.168)			-.259 (.050)	
Asian		-.014 (.116)			-.060 (.031)	
Other/missing race		-.521 (.293)			-.082 (.061)	
High school top 10%		.948 (.107)			-.066 (.011)	
High school rank missing		.556 (.102)			-.030 (.023)	
Athlete		-.318 (.147)			.037 (.016)	
Average SAT score of schools applied to ÷ 100			.777 (.058)			.063 (.014)
Sent two applications			.252 (.077)			.020 (.010)
Sent three applications			.375 (.106)			.042 (.013)
Sent four or more applications			.330 (.093)			.079 (.014)

- **Bellows and Miguel (2008, JPubE)**构造了不可观测变量的选择性偏误强度的测量指标：要把核心解释变量的效应全部归因于不可观测变量的选择性偏误，这一偏误（相对于可观测变量的选择性偏误）必须达到多大：

$$\frac{\hat{\beta}^F}{\hat{\beta}^R - \hat{\beta}^F}$$

其中 $\hat{\beta}^F$ 为包含全部控制变量回归的系数估计， $\hat{\beta}^R$ 为包含部分控制变量回归的系数估计。

分母越小，说明估计值受可观测变量的选择性影响越小，则不可观测变量（相对于可观测变量）的选择性必须更大才能完全解释整个效应；分子越大，说明需要解释的效应越大。

示例 18. 奴隶贸易与人际信任 (Nunn and Wantchekon, 2011, *AER*)

- 非洲历史上的奴隶贸易造成人际间的不信任并延续至今。
- 被解释变量：非洲民意调查中受访者报告的各种信任指标（亲属、邻居、当地政府、族群内、族群间，离散变量按连续变量处理）
- 核心解释变量：受访者所在种族 (ethnicity) 历史上奴隶出口数量
- 计量设定：

$$\text{trust}_{i,e,d,c} = \alpha_c + \beta \text{slave export}_e + \mathbf{X}'_{i,e,d,c} \Gamma_1 + \mathbf{X}'_{e,d,c} \Gamma_2 + \mathbf{X}'_{d,c} \Gamma_3 + \mathbf{X}'_e \Gamma_4 + \varepsilon_{i,e,d,c}$$

其中 i 表示个体， e 表示种族， d 表示地区， c 表示国家；个体特征 ($\mathbf{X}_{i,e,d,c}$) 包括年龄、性别、是否城市、生活条件、受教育水平、宗教、职业；种族-区域特征 ($\mathbf{X}_{e,d,c}$) 为区域内同种族人口；区域特征 ($\mathbf{X}_{d,c}$) 为种族分化程度。

- 识别策略之一：控制种族层面的殖民统治的影响 (\mathbf{X}_e).
 - ▷ 控制决定殖民统治 variation 的变量：初始疾病环境、前殖民时期的发展水平（人口密度、是否有城市、聚落形态、司法层级）
 - ▷ 直接控制反映殖民统治影响的变量：是否有铁路、是否有欧洲探险家经过、欧洲传教士数量

TABLE 2—OLS ESTIMATES OF THE DETERMINANTS OF THE TRUST OF OTHERS

	Trust of relatives (1)	Trust of neighbors (2)	Trust of local council (3)	Intra- group trust (4)	Inter- group trust (5)
$\ln(1 + \text{exports/area})$	−0.133*** (0.037)	−0.159*** (0.034)	−0.111*** (0.021)	−0.144*** (0.032)	−0.097*** (0.028)
Individual controls	Yes	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Number of observations	20,062	20,027	19,733	19,952	19,765
Number of ethnicity clusters	185	185	185	185	185
Number of district clusters	1,257	1,257	1,283	1,257	1,255
R^2	0.13	0.16	0.20	0.14	0.11

Notes: The table reports OLS estimates. The unit of observation is an individual. Standard errors are adjusted for two-way clustering at the ethnicity and district levels. The individual controls are for age, age squared, a gender indicator variable, five living conditions fixed effects, ten education fixed effects, 18 religion fixed effects, 25 occupation fixed effects, and an indicator for whether the respondent lives in an urban location. The district controls include ethnic fractionalization in the district and the share of the district's population that is the same ethnicity as the respondent.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

* Significant at the 10 percent level.

TABLE 3—OLS ESTIMATES OF THE DETERMINANTS OF THE TRUST OF OTHERS,
WITH ADDITIONAL CONTROLS

	Trust of relatives (1)	Trust of neighbors (2)	Trust of local council (3)	Intragroup trust (4)	Intergroup trust (5)
$\ln(1 + \text{exports/area})$	−0.178*** (0.032)	−0.202*** (0.031)	−0.129*** (0.022)	−0.188*** (0.033)	−0.115*** (0.030)
Colonial population density	Yes	Yes	Yes	Yes	Yes
Ethnicity-level colonial controls	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes
District controls	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes
Number of observations	16,709	16,679	15,905	16,636	16,473
Number of ethnicity clusters	147	147	146	147	147
Number of district clusters	1,187	1,187	1,194	1,186	1,184
R^2	0.13	0.16	0.21	0.16	0.12

Notes: The table reports OLS estimates. The unit of observation is an individual. Standard errors are adjusted for two-way clustering at the ethnicity and district levels. The individual controls are for age, age squared, a gender indicator variable, five living conditions fixed effects, ten education fixed effects, 18 religion fixed effects, 25 occupation fixed effects, and an indicator for whether the respondent lives in an urban location. The district controls include ethnic fractionalization in the district and the share of the district's population that is the same ethnicity as the respondent. Ethnicity-level colonial controls include the prevalence of malaria, a 1400 urbanization indicator variable, eight fixed effects for the sophistication of precolonial settlement, the number of jurisdictional political hierarchies beyond the local community in the precolonial period, an indicator for integration with the colonial rail network, an indicator for contact with precolonial European explorers, and the number of missions per square kilometer during colonial rule. Colonial population density is the natural log of an ethnicity's population density during the colonial period.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

TABLE 4—USING SELECTION ON OBSERVABLES TO ASSESS THE BIAS FROM UNOBSERVABLES

		Trust of relatives	Trust of neighbors	Trust of local council	Intragroup trust	Intergroup trust
Controls in the restricted set	Controls in the full set	(1)	(2)	(3)	(4)	(5)
None	Full set of controls from equation (1)	4.31	4.23	3.03	4.13	3.32
None	Full set of controls from equation (1), ethnicity-level colonial controls, and colonial population density	11.54	6.98	2.65	9.22	3.80
Age, age squared, gender	Full set of controls from equation (1)	4.17	3.99	2.89	3.91	3.12
Age, age squared, gender	Full set of controls from equation (1), ethnicity-level colonial controls, and colonial population density	10.93	6.52	2.57	8.44	3.59

Notes: Each cell of the table reports ratios based on the coefficient for $\ln(1 + \text{exports/area})$ from two individual-level regressions. In one, the covariates include the “restricted set” of control variables. Call this coefficient β^R . In the other, the covariates include the “full set” of controls. Call this coefficient β^F . In both regressions, the sample sizes are the same, and country fixed effects are included. The reported ratio is calculated as: $\beta^F/(\beta^R - \beta^F)$. See Table 3 for the description of the full set of controls from equation (1), the ethnicity-level colonial controls, and colonial population density.

- Altonji et al (2005, *JPE*)

$$\begin{aligned} Y &= \tau D + \mathbf{W}'\mathbf{\Gamma} \\ &= \tau D + \mathbf{X}'\mathbf{\Gamma}_X + \xi \\ &= \tau D + \mathbf{X}'\boldsymbol{\gamma} + \epsilon \end{aligned}$$

\mathbf{X} 为可观测的选择性变量； ξ 为不可观测的选择性变量； $\boldsymbol{\gamma}$ 同时捕捉了 \mathbf{X} 对 y 的直接效应 $\mathbf{\Gamma}_X$ 以及 \mathbf{X} 和 ξ 的相关性，因此 $Cov(\mathbf{X}, \epsilon) = 0$ 。
(标准化后的) 基于不可观测变量的选择性

$$SU \equiv \frac{\mathbb{E}(\epsilon|D=1) - \mathbb{E}(\epsilon|D=0)}{\text{Var}(\epsilon)}$$

(标准化后的) 基于可观测变量的选择性

$$SO \equiv \frac{\mathbb{E}(\mathbf{X}'\boldsymbol{\gamma}|D=1) - \mathbb{E}(\mathbf{X}'\boldsymbol{\gamma}|D=0)}{\text{Var}(\mathbf{X}'\boldsymbol{\gamma})}$$

第二类识别假设相当于 $SU = 0$ ；如果 \mathbf{X} 是从影响 Y 的全部变量集合 \mathbf{W} 中随机选取，则 $SU = SO$ 。

我们可以计算在 $SU = SO$ 假设下 OLS 的偏误；等价地，也可以计算要想使得 D 对 Y 的效应 τ 完全消失 SU 至少是 SO 的多少倍。

先用 D 对 \mathbf{X} 回归：

$$D = \mathbf{X}'\boldsymbol{\beta} + \tilde{D}$$
$$Y = \tau\tilde{D} + \mathbf{X}'(\boldsymbol{\gamma} + \tau\boldsymbol{\beta}) + \epsilon$$

因为 \tilde{D} 与 \mathbf{X} 不相关，

$$\begin{aligned}\hat{\tau} &\rightarrow_p \tau + \frac{\text{Cov}(\tilde{D}, \epsilon)}{\text{Var}(\tilde{D})} = \tau + \frac{\text{Cov}(D, \epsilon)}{\text{Var}(\tilde{D})} \\ &= \tau + \frac{\text{Var}(D)}{\text{Var}(\tilde{D})} \cdot \frac{\text{Cov}(D, \epsilon)}{\text{Var}(D)} \\ &= \tau + \frac{\text{Var}(D)}{\text{Var}(\tilde{D})} [\mathbb{E}(\epsilon|D=1) - \mathbb{E}(\epsilon|D=0)]\end{aligned}$$

在原假说 $\tau = 0$ 下，我们可以一致地估计 $\boldsymbol{\gamma}$ 。

若 $SU = SO$,

$$\begin{aligned}\text{Bias} &= \frac{\text{Var}(D)}{\text{Var}(\tilde{D})} [\mathbb{E}(\epsilon|D = 1) - \mathbb{E}(\epsilon|D = 0)] \\ &= \frac{\text{Var}(D)}{\text{Var}(\tilde{D})} \frac{\text{Var}(\epsilon)}{\text{Var}(\mathbf{X}'\boldsymbol{\gamma})} [\mathbb{E}(\mathbf{X}'\boldsymbol{\gamma}|D = 1) - \mathbb{E}(\mathbf{X}'\boldsymbol{\gamma}|D = 0)] \\ &= \frac{\text{Var}(D)}{\text{Var}(\tilde{D})} \frac{\text{Var}(\epsilon)}{\text{Var}(\mathbf{X}'\boldsymbol{\gamma})} \frac{\text{Cov}(\mathbf{X}'\boldsymbol{\gamma}, D)}{\text{Var}(D)} \\ &= \frac{\text{Var}(\epsilon)}{\text{Var}(\tilde{D})} \left(\frac{\text{Cov}(\mathbf{X}'\boldsymbol{\gamma}, D)}{\text{Var}(\mathbf{X}'\boldsymbol{\gamma})} \right)\end{aligned}$$

$$\text{Selection ratio} = \frac{\hat{\tau}}{\text{Bias}}$$

示例 19. 革命的威胁与民主化 (Aidt and Franck, 2015, *ECMA*)

- 被解释变量：1831 年英国下议会选举中 244 个选区支持改革的辉格党席次
- 核心解释变量：1830-1831 年各选区附近斯温暴动数量

LOCAL SWING RIOTS AND THE OUTCOME OF THE 1831 ELECTION. BASELINE RESULTS^a

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A</i>						
	<i>Whig Share 1831 (%)</i>					
	Least Squares					
Riots within 10 km	0.57 (0.32)* [0.25]**	0.37 (0.22)* [0.19]*	0.44 (0.18)** [0.18]**	0.47 (0.18)** [0.18]**	0.47 (0.18)** [0.19]**	0.44 (0.18)** [0.18]**
Whig share 1826		0.87 (0.19)***	0.32 (0.19)	0.35 (0.20)*	0.38 (0.20)*	0.38 (0.071)***
(Whig share 1826) ²		−0.0045 (0.0019)**	0.00055 (0.0020)	0.00035 (0.0020)	−6.8e−06 (0.0020)	
Reform support 1830		12.0 (5.60)**	12.1 (4.97)**	11.2 (5.09)**	12.1 (5.14)**	12.6 (4.77)**
County constituency			33.0 (5.14)***	37.2 (6.50)***	35.2 (7.04)***	31.6 (4.68)***
University constituency			−60.8 (9.39)***	−58.1 (10.7)***	−58.1 (8.60)***	−61.8 (10.50)***
Narrow franchise			−3.35 (5.62)	−2.85 (5.39)	−3.62 (5.26)	
Patronage index			−17.0 (3.42)***	−13.5 (3.94)***	−12.2 (3.86)***	−15.3 (3.52)***
Emp. fract. index				7.52 (30.9)	7.83 (29.49)	
Agriculture (emp. share)				−28.4 (27.5)	−27.2 (27.0)	
Trade (emp. share)				11.4 (30.9)	14.0 (31.1)	
Professionals (emp. share)				−143 (120)	−119 (120)	
Population					0.00028 (0.009)	
Population density					0.15 (2.68)	
Thriving economy					−10.1 (5.91)*	
Declining economy					−10.6 (5.86)*	−10.3 (5.72)*
Selection ratio	N.A.	0.67	2.26	2.54	2.56	2.59
Adjusted R ²	0.021	0.27	0.44	0.44	0.45	0.45
Obs. (constituencies)	244	244	244	244	244	244

- **Oster (2017, JBES)**

- 直接观察核心解释变量的系数估计稳定性是不够的，因为即使可观测变量完全不能反映关于不可观测变量的信息，但是由于可观测变量对 Y 的解释力比较弱，在加入可观测变量前后，也能观察到核心解释变量的系数估计很稳定。
- 因此还需要观察可观测变量的加入对模型拟合程度的影响。Altonji *et al* (2005) 的做法实际上体现了这一思想，但仅当核心解释变量系数为零的原假设成立时才可行。
- Oster 证明，在一定的限制性假设下，核心解释变量系数的一致估计量表达式为

$$\hat{\beta}^* = \hat{\beta}^F - \delta \left(\hat{\beta}^R - \hat{\beta}^F \right) \frac{R_{max}^2 - R_F^2}{R_F^2 - R_R^2}$$

其中 R_F^2 为包含全部控制变量回归的拟合优度， R_R^2 为包含部分控制变量回归的拟合优度， R_{max}^2 为包含全部控制变量及遗漏变量回归的拟合优度， δ 为（相对于可观测变量而言）不可观测变量的选择性。

- Altonji *et al* (2005) 隐含假定 $R_{max}^2 = 1$, 而 Nunn and Wantchekon (2011) 相当于假定 $R_{max}^2 - R_F^2 = R_F^2 - R_R^2$, 从而计算 δ .
- 更合理的假设是 $\delta = 1$ (即可观测变量至少和不可观测变量同等重要), 从而根据 R_{max}^2 判断 $\hat{\beta}^*$ 的大小。

		b_F	b_R	R2_F	R2_R	R2_max=2*R2_F-R2_R	R2_max=1	max{R2_max}
none	+baseline	-0.133	-0.164	0.133	0.115	-0.102	1.358	0.210
none	++colonial	-0.178	-0.194	0.130	0.106	-0.163	0.368	0.414
age+gender	+baseline	-0.133	-0.165	0.133	0.117	-0.101	1.625	0.199
age+gender	++colonial	-0.178	-0.195	0.130	0.108	-0.162	0.457	0.375
		b_F	b_R	R2_F	R2_R	R2_max=2*R2_F-R2_R	R2_max=1	max{R2_max}
none	+reform	0.374	0.569	0.284	0.025	0.179	-0.165	0.780
none	++base	0.439	0.569	0.461	0.025	0.310	0.279	1.000
none	++demographic	0.468	0.569	0.471	0.025	0.367	0.348	1.000
none	++structure	0.467	0.569	0.483	0.025	0.366	0.353	1.000

- Oster 给出了放松限制性假设以后的遗漏变量偏误形式, 可以通过 STATA 包 `psacalc` 实现 (仅适用于 `reg`、`xtreg` 和 `areg` 估计方法)。

交互项模型

- 两个离散变量的交互。

- 其本质是组间均值差异之差异。

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

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

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

- 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；

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

这两种解释在数学上是等价的，采用哪种解释取决于 D_1 和 D_2 哪个是核心解释变量。

- 注意到, $\mathbb{E}(Y|D_1, D_2)$ 最多只能取四个值, 这也是一个饱和模型的例子。系数估计量可以很直观得到:

$$\hat{\beta}_0 = \bar{y}_{FL}$$

$$\hat{\beta}_1 = \bar{y}_{ML} - \bar{y}_{FL}$$

$$\hat{\beta}_2 = \bar{y}_{FH} - \bar{y}_{FL}$$

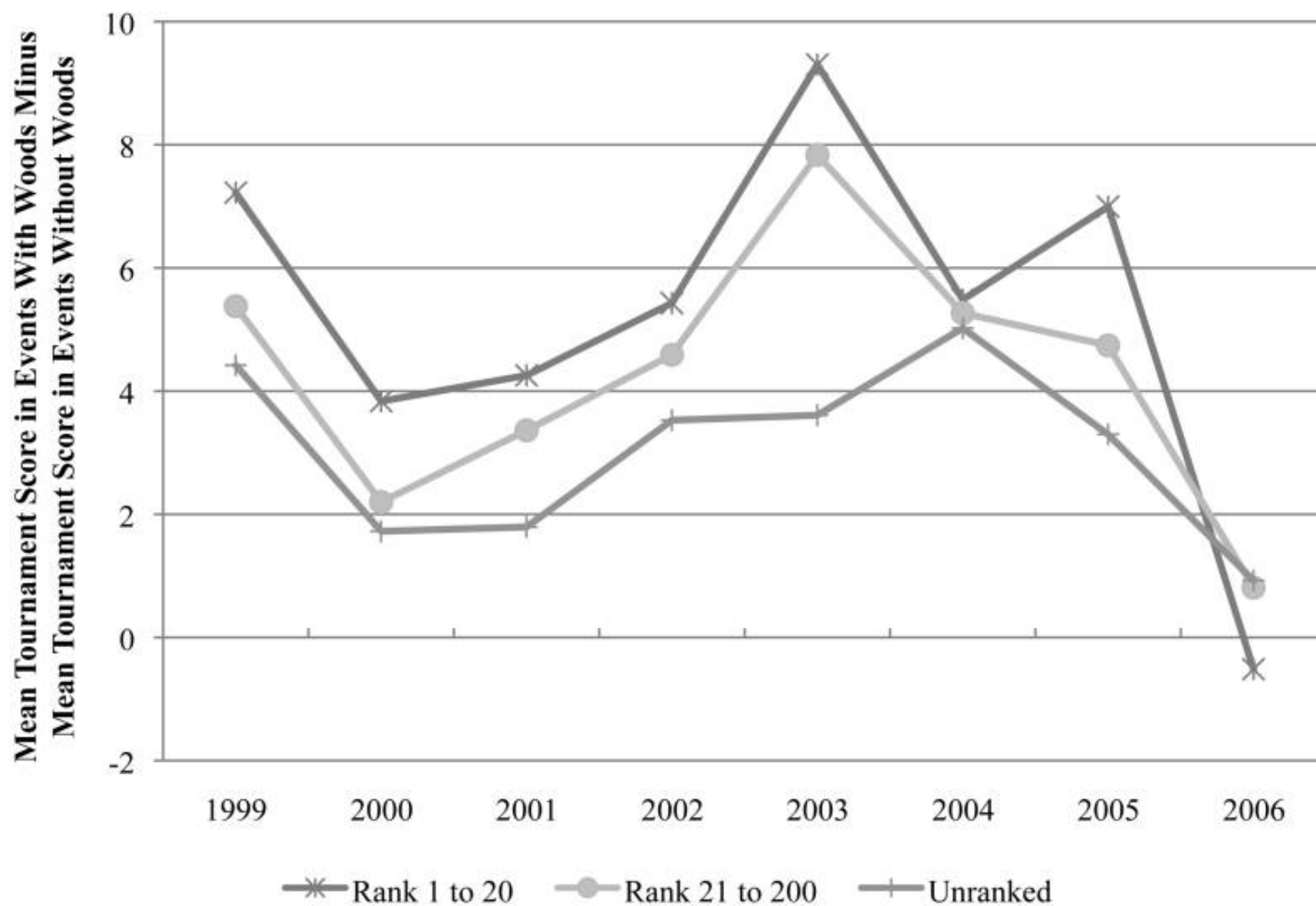
$$\hat{\beta}_3 = (\bar{y}_{MH} - \bar{y}_{ML}) - (\bar{y}_{FH} - \bar{y}_{FL})$$

- 交互项模型不止一种写法, 可以根据需要 (希望得到组间均值差异的绝对值还是相对值) 选择哪些主效应项和交互效应项进入回归。

$$Y = \gamma_0 + \gamma_1 D_1 + \gamma_2 (D_1 \times D_2) + \gamma_3 ((1 - D_1) \times D_2) + \varepsilon$$

$$E(Y|D_1, D_2) = \begin{cases} \gamma_0 + \gamma_1 + \gamma_2 & \text{if } D_1 = 1 \text{ \& } D_2 = 1 \\ \gamma_0 + \gamma_1 & \text{if } D_1 = 1 \text{ \& } D_2 = 0 \\ \gamma_0 + \gamma_3 & \text{if } D_1 = 0 \text{ \& } D_2 = 1 \\ \gamma_0 & \text{if } D_1 = 0 \text{ \& } D_2 = 0 \end{cases}$$

— 回忆一下**示例 4**超级明星效应的例子,



– 可以采用如下的模型来实现,

$$\begin{aligned} \text{strokes}_{ij} = & \alpha_0 + \alpha_1 \text{star}_j + \alpha_2 \text{HRanked}_i + \alpha_3 \text{LRanked}_i \\ & + \alpha_4 \text{star}_j \times \text{HRanked}_i + \alpha_5 \text{star}_j \times \text{LRanked}_i + \varepsilon_{ij} \end{aligned}$$

$$\mathbb{E}(\text{strokes}|\cdot) = \begin{cases} \alpha_0 + \alpha_1 + \alpha_2 + \alpha_4 & \text{HRanked w/ star} \\ \alpha_0 + \alpha_2 & \text{HRanked w/o star} \\ \alpha_0 + \alpha_1 + \alpha_3 + \alpha_5 & \text{LRanked w/ star} \\ \alpha_0 + \alpha_3 & \text{LRanked w/o star} \\ \alpha_0 + \alpha_1 & \text{URanked w/ star} \\ \alpha_0 & \text{URanked w/o star} \end{cases}$$

α_1 反映了 URanked 选手的超级明星效应；

α_4 反映了超级明星效应在 HRanked 和 URanked 选手之间的差异；

α_5 反映了超级明星效应在 LRanked 和 URanked 选手之间的差异。

– 但作者实际使用的模型是下面这种,

$$\text{strokes}_{ij} = \beta_0 + \beta_1 \text{HRanked}_i + \beta_2 \text{LRanked}_i + \beta_3 \text{star}_j \times \text{HRanked}_i \\ + \beta_4 \text{star}_j \times \text{LRanked}_i + \beta_5 \text{star}_j \times \text{URanked}_i + \varepsilon_{ij}$$

$$\mathbb{E}(\text{strokes}|\cdot) = \begin{cases} \beta_0 + \beta_1 + \beta_3 & \text{HRanked w/ star} \\ \beta_0 + \beta_1 & \text{HRanked w/o star} \\ \beta_0 + \beta_2 + \beta_4 & \text{LRanked w/ star} \\ \beta_0 + \beta_2 & \text{LRanked w/o star} \\ \beta_0 + \beta_5 & \text{URanked w/ star} \\ \beta_0 & \text{URanked w/o star} \end{cases}$$

β_3 反映了 HRanked 选手的超级明星效应；

β_4 反映了 LRanked 选手的超级明星效应；

β_5 反映了 URanked 选手的超级明星效应。

	TOURNAMENT TYPES			
	First Round Regulars and Majors (1)	First Round Regulars (2)	Tournament Regulars and Majors (3)	Tournament Regulars (4)
Superstar effect for players:				
Ranked 1–20	.596** (.281)	.535* (.302)	1.358** (.726)	.996 (.786)
Ranked 21–200	.161 (.113)	.141 (.117)	.804*** (.318)	.672** (.328)
Unranked	.202 (.126)	.212 (.131)	.596 (.396)	.311 (.400)
Observations	34,986	29,167	18,805	15,651
Adjusted R^2	.29	.21	.48	.38

– 也可以进行分组回归。

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

示例 20. 官员更替与企业慈善捐赠（戴亦一等，2014，经济研究）

	假设 1(总样本)		假设 2(分组样本)			
	Lndonate (1)	Donatedum (2)	国有企业 Lndonate(3)	民营企业 Lndonate(4)	国有企业 Donatedum(5)	民营企业 Donatedum(6)
Gt	0.4796 *** (3.31)	0.0899 *** (2.85)	0.3010 (1.58)	0.6076 *** (2.97)	0.0445 (0.94)	0.1340 *** (2.58)
Lnasset _{t-1}	1.2946 *** (7.32)	0.2087 *** (5.01)	0.4244 * (1.77)	2.7254 *** (11.89)	0.0035 (0.07)	0.5520 *** (9.03)
Lncash _{t-1}	0.3100 ** (2.41)	0.0582 ** (2.32)	0.6206 *** (3.46)	-0.1755 (-0.78)	0.1219 *** (2.98)	-0.0411 (-0.76)
ROA _{t-1}	0.0622 *** (5.27)	0.0057 ** (2.29)	0.1016 *** (5.01)	0.0672 *** (4.81)	0.0148 *** (2.97)	0.0097 *** (2.87)
Lev _{t-1}	0.0278 *** (5.66)	0.0060 *** (4.17)	0.0353 *** (3.85)	0.0170 ** (2.21)	0.0075 *** (3.56)	0.0041 ** (2.04)
Big	-0.0309 *** (-5.54)	-0.0065 *** (-4.01)	-0.0398 *** (-3.98)	-0.0032 (-0.36)	-0.0086 *** (-3.47)	-0.0009 (-0.38)
Herfindal	1.1661 (0.63)	0.4962 (1.02)	2.9676 (1.14)	0.1438 (0.05)	0.9791 (1.39)	0.3651 (0.59)
Consumer	0.3436 (1.33)	0.0798 (1.34)	0.6694 ** (2.01)	-0.5596 (-1.55)	0.1701 ** (2.08)	-0.1318 * (-1.67)
Year&Provin Dummy	控制	控制	控制	控制	控制	控制
R ² /Pseudo R ²	0.1127	0.0552	0.0795	0.1887	0.0340	0.1139
N	7517	7517	3975	3542	3975	3542

• 离散变量和连续变量的交互

- 我们已经知道，当核心解释变量 D 为 0-1 变量时，核心解释变量和控制变量的交互是为了灵活设定 Y 关于控制变量的条件期望函数在 $D = 0$ 和 $D = 1$ 组间的不同形式（也就是允许 D 对 Y 的因果效应随着 X 而变化）。这里讨论核心解释变量为连续变量的情形。
- 也等价于分组回归。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。

$$Y = \beta_0 + \beta_1 X + \beta_2 D + \beta_3 (X \times D) + \varepsilon$$

其中 X 为核心解释变量， D 为分组变量。

$$Y = \begin{cases} \beta_0 + \beta_1 X + \varepsilon & \text{if } D = 0 \\ (\beta_0 + \beta_2) + (\beta_1 + \beta_3)X + \varepsilon & \text{if } D = 1 \end{cases}$$

- 分组回归与交互项模型能够得到相同的系数估计，但标准误估计不同。交互项模型适用于对因果效应的组间差异进行检验。

示例 21. 民营资本的宗族烙印（潘越等，2019，经济研究）

综上，宗族文化聚合的宗族关系网络是开展经济活动的组织基础，而宗族文化丰富的内涵则可以促成族内成员以及族间成员之间的相互信任，促进市场中的信息交互，并形成对企业家的道德约束和道德激励，最终帮助企业克服融资难题。并且，在宗族文化越为浓厚的地区，个体间的信任度更高，信息交流更多，道德约束和道德激励的作用更加明显，从而能够更好地帮助缓解企业面临的融资约束。据此，本文提出以下假设：

假设：宗族文化有助于缓解企业融资约束，即地区内宗族文化越浓厚，企业的融资约束越小。

（三）企业家在本地与异地经营的差异

中国人历来擅长于构建关系，尤其是在宗族文化的影响下，同姓个体间很容易超越血缘关系，通过难以考证的共同祖先（认本家）、结拜兄弟等方式虚构亲缘关系，将没有直接血缘联系的个体纳入宗族网络。然而，和出生并成长在当地的个体相比，外来的同姓个体虽然可以凭借对于宗族文化的认同融入当地的宗族网络，但由于缺乏可以考证的血缘关系，本地个体对其形成的情感认同会相对较弱，对他的信任程度也更低；与此同时，来自异地的个体也更少受到当地宗族文化的道德约束，这些都将影响其获取市场资源。为此，本文通过公开数据库和网络渠道，搜集整理了企业实际控制人的籍贯地信息，^②设置 *Local* 变量，如果企业的实际控制人是本地人，则 *Local* 设为 1，否则为 0；同时引入 *Clan* 与 *Local* 的交乘项 $Clan \times Local$ 。表 6 列(2)的结果显示，交乘项显著为负，说明相比来自异地的企业家，土生土长的企业家更容易从宗族文化中受益。

• 两个连续变量的交互

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2) + \varepsilon$$

$$\frac{\partial \mathbb{E}(Y|X_1, X_2)}{\partial X_1} = \beta_1 + \beta_3 X_2$$
$$\frac{\partial \mathbb{E}(Y|X_1, X_2)}{\partial X_2} = \beta_2 + \beta_3 X_1$$

- 有时候 X_1 和 X_2 都是核心解释变量，此时该模型适用于考察两者对被解释变量作用的互补性 ($\beta_3 > 0$) 或替代性 ($\beta_3 < 0$)，但更多时候只有其中一个是核心解释变量（例如 X_1 ），此时该模型仍然是为了考察因果关系的作用机制。
- 当 X_1 为核心解释变量时， X_2 也被称作调节变量。有时候，不加入 X_2 （及其与 X_1 的交互项）时， X_1 的系数估计是显著的，当加入 X_2 以及交互项后， X_1 的系数估计变得不显著，这是很正常的。因为此时 X_1 的系数含义已经变化了。

示例 22. 官员能力和关系对晋升的影响 (Jia *et al*, 2015, *JEEA*)

TABLE 2. Complementarity between connections and growth.

(Dependent variable: Promoted)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Connection	0.049** (0.023)		0.048** (0.023)	0.032 (0.023)	0.011 (0.024)	0.029 (0.024)	0.043 (0.027)
Growth		0.702 (0.465)	0.690 (0.447)	0.356 (0.427)	3.730 (5.865)	0.210 (0.471)	0.317 (0.430)
Connection * Growth				2.195** (0.889) [0.056]*	2.025*** (0.733) [0.038]**	1.983** (0.769) [0.058]*	2.222** (1.004) [0.098]*
Age					−0.009*** (0.003)		
Age * Growth					−0.061 (0.106)		
Served in center						0.026 (0.025)	
Served in center * Growth						0.607 (0.670)	
PSC work province							−0.039 (0.034)
PSC work province * Growth							−0.400 (1.237)
Fixed Effects	Y	Y	Y	Y	Y	Y	Y
No. of provinces	31	31	31	31	31	31	31
No. of observations	966	966	966	966	966	966	966

示例 23. 金融发展与经济增长 (Rajan and Zingales, 1998, *AER*)

$$\text{Growth}_{jk} = \beta \text{ExtDep}_j \times \text{FinaDev}_k + \gamma \text{FracManu}_{jk} \\ + \text{Industry}_j + \text{Country}_k + \varepsilon_{jk}$$

Growth _{jk}	1980-1990 年 k 国 j 行业增加值年均实际增长率
FracManu _{jk}	1980 年 j 行业增加值在 k 国制造业总增加值占比
ExtDep _j	j 行业外部融资依存度 (资本性支出中外部融资占比)
FinaDev _k	k 国金融发展水平 (1990 年各国会计标准)

	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

TABLE 4—INDUSTRY GROWTH AND VARIOUS MEASURES OF DEVELOPMENT

Variable	Financial development measured as					
	Total capitalization	Bank debt	Accounting standards	Accounting standards in 1983	Accounting standards and capitalization	Instrumental variables
Industry's share of total value added in manufacturing in 1980	-0.912 (0.246)	-0.899 (0.245)	-0.643 (0.204)	-0.587 (0.223)	-0.443 (0.135)	-0.648 (0.203)
Interaction (external dependence \times total capitalization)	0.069 (0.023)	—	—	—	0.012 (0.014)	—
Interaction (external dependence \times domestic credit to private sector)	—	0.118 (0.037)	—	—	—	—
Interaction (external dependence \times accounting standards)	—	—	0.155 (0.034)	—	0.133 (0.034)	0.165 (0.044)
Interaction (external dependence \times accounting standards 1983)	—	—	—	0.099 (0.036)	—	—
R^2	0.290	0.290	0.346	0.239	0.419	0.346
Number of observations	1217	1217	1067	855	1042	1067
Differential in real growth rate	1.3	1.1	0.9	0.4	1.3	1.0

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.

• 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 在某些子总体 (subpopulation) 中存在，在另一些子总体中不存在，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。

- 此前我们处理内生性/选择性的思路一直聚焦在寻找合适的 conditioning variables 和 conditioning strategy，即找到导致内生性/选择性的原因，然后正式地刻画它、测量它、控制它。交互项模型则提供了另一种处理内生性/选择性的思路，即尝试挖掘被解释变量和核心解释变量之间更丰富的相关性，如果这种相关性是其它因果故事所不能解释的，那么即使此时内生性/选择性或许仍然存在，但至少证明我们所感兴趣的因果关系是存在的，否则这种更丰富的相关性不会出现。这种思路体现在第 1 讲的流程图中就是寻找因果模型的可验证含义 (testable implications)。
- 在超级明星效应的例子中，无法完全控制“赛事难度”这一导致核心解释变量“伍兹是否参赛”存在内生性/选择性的因素（尽管已经最大限度地控制了赛事级别、场地质量、奖金总额等），作者转而去挖掘更丰富的相关性：伍兹是否参赛导致的比赛杆数差异对于高水平选手而言是否更大？这个事实可以被超级明星效应所解释，但不能被伍兹参加的都是高难度赛事所解释，因果论证的目的就达到了。

- 同样地，在金融促进增长的例子中，无法完全控制反向因果或第三方混淆因素（尽管已经控制了不变的行业特征和国家特征，但仍然可能存在同时随行业和国家而变的选择性变量），作者转而去挖掘更丰富的相关性：金融和增长的正相关性在外部融资依存度更高的行业是否更强？这个事实可以被金融通过缓解企业外部融资约束从而促进增长的故事所解释，但不能被其它故事合理解释，因果论证的目的就达到了。
- 这两个例子启发我们，交互项模型是一种很重要的因果论证手段，但在使用这种手段之前，首先要发展出一个说得通的理论：如果超级明星效应这个故事成立，那么就应该看到不同人的效应大小不同（因为激励强度不同）；如果金融促进增长这个故事成立，那么就应该看到不同行业的效应大小不同（因为对金融的需求不同），如此之类。然后再构建相应的交互项模型去验证这个理论。这就是因果推断理论先行 (theory driven) 的含义。