

# 面板数据

Panel Data

杨点溢

Department of Government  
London School of Economics and Political Science

Spark 社科量化系列课程



# 本节课内容

## 1 回顾

## 2 面板数据 Panel Data

## 3 固定效应回归



# 目前我们学了。。。。

- 使用观察性数据 Observational data 来进行因果推断
  - 需要解决选择偏差 Selection bias
- 在回归分析中我们希望误差项 ( $e_i$ ) 和  $X_i$  不相关。
- 我们研究了在哪些情况下上述假设不成立
  - 也提了一些解决方案
- 目前我们只讨论了**截面** Cross-sectional 数据
  - 每一个单位 unit 只有一次观察 observation
  - 比如：每个州对于芯片法案的投票情况（2022）
  - 比如：学生成绩数据（只考了一次）



# 本节课内容

- 1 回顾
- 2 面板数据 Panel Data
- 3 固定效应回归



# 面板数据的定义 Definition of Panel Data

- 在面板 Panel（或者说纵向 Longitudinal）数据中，我们有若干个单位 units（截面 cross-sections），对于每个单位我们有**好几个**观察量 observations.
  - 例：OECD 国家在 2001-2005 年间**每年**的 GDP 数据
  - 学生**每年**的成绩以及对应的师生比和学校资源信息
- 两个维度：单位 ( $N$ ) 以及每个单位的观察次数 ( $T$ )。
- 叫做 ( $T$ ) 是因为我们总是在不同**时间点**才观察同样的对象单位（但不是必须这样）
  - 在一个研究班级大小对学生的影响研究中，每一个学生在不同大小的班级中被观察 (see Bandiera, Larcinese, and Rasul 2010)
  - 研究对象：某个不愿透露姓名的，拥有 23 个系和 125 个专业的英国顶尖大学。
- 一个**平衡的** **balanced** 面板数据 = 每个单位都有  $T$  个观察，否则就是**不平衡的** **unbalanced** 面板



# 案例：交通死亡和酒精税 /1

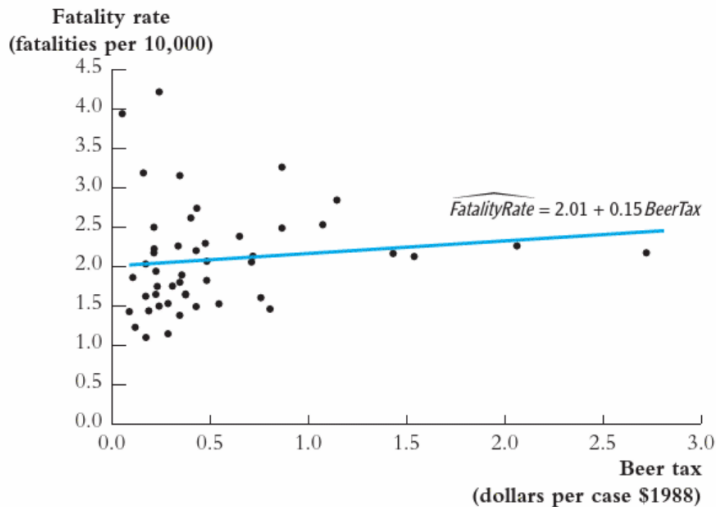
- 对酒收税会减少交通死亡吗？
- 我们有美国大陆 (contiguous) 的 48 个州在 1982 年的数据。
- 设  $FT$ (fataility rate) 为州居民每万人的年度交通死亡，OLS 结果显示：

$$\hat{FT} = \underset{(0.15)}{2.01} + \underset{0.13}{0.15} Tax$$

- 系数为正（尽管结果统计学上不显著）



# 散点回归线图 (Stock and Watson 2014, p.399)



## 案例：交通死亡和酒精税 /2

- 现在采集 1988 年的数据，再跑一次，得到：

$$\hat{FT} = \underset{(0.11)}{1.86} + \underset{(0.13)}{0.44} Tax$$





## 案例：交通死亡和酒精税 /2

- 现在采集 1988 年的数据，再跑一次，得到：

$$\hat{FT} = 1.86 + 0.44 Tax$$

(0.11)      (0.13)

- 更高的税还是跟更多事故死亡正相关！



## 案例：交通死亡和酒精税 /2

- 现在采集 1988 年的数据，再跑一次，得到：

$$\hat{FT} = \underset{(0.11)}{1.86} + \underset{(0.13)}{0.44} Tax$$

- 更高的税还是跟更多事故死亡正相关！
- 我们是否应该下结论说对酒精征税**导致**更多的交通事故？



## 案例：交通死亡和酒精税 /2

- 现在采集 1988 年的数据，再跑一次，得到：

$$\hat{FT} = 1.86 + 0.44 Tax$$

(0.11)      (0.13)

- 更高的税还是跟更多事故死亡正相关！
- 我们是否应该下结论说对酒精征税**导致**更多的交通事故？
- 不应该，我们有很严重的**内生性**



## 案例：交通死亡和酒精税 /2

- 现在采集 1988 年的数据，再跑一次，得到：

$$\hat{FT} = 1.86 + 0.44 Tax$$

(0.11)      (0.13)

- 更高的税还是跟更多事故死亡正相关！
- 我们是否应该下结论说对酒精征税**导致**更多的交通事故？
- 不应该，我们有很严重的**内生性**
  - 酗酒文化 OVB



## 案例：交通死亡和酒精税 /2

- 现在采集 1988 年的数据，再跑一次，得到：

$$\hat{FT} = 1.86 + 0.44 Tax$$

(0.11)      (0.13)

- 更高的税还是跟更多事故死亡正相关！
- 我们是否应该下结论说对酒精征税**导致**更多的交通事故？
- 不应该，我们有很严重的**内生性**
  - 酗酒文化 OVB
  - 反向因果



# 交通死亡和酒精税：干扰因素 (confounders)

- 酒精消费大州和有更多酗酒问题的州更有可能加征高酒精税
- 加税有可能是为了财政收入
  - 比如修路没钱
- 越城市化的地区可能越想要高酒精税（扰民），但是也有更多的交通事故（人口密度）
- 总结：有非常多的原因导致内生性



# “事前事后” (Before and After) 比简单截面回归更好

- 在拥有两年的数据之后，我们做的可以比跑两个分别的回归或者是**汇总** **pooled** 回归更好
  - 汇总回归忽略了一个事实：我们对每个单位有不止一个观察



# “事前事后” (Before and After) 比简单截面回归更好

- 在拥有两年的数据之后，我们做的可以比跑两个分别的回归或者是**汇总 pooled**回归更好
  - 汇总回归忽略了一个事实：我们对每个单位有不只一个观察
- 我们分别看两个回归模型，想象有个  $Z_i$  变量代表和交通死亡有关的**固定的州属特征**。
  - 比如说当地对喝酒和开车的文化

$$FT_i^{1982} = \alpha + \beta Tax_i^{1982} + \gamma Z_i + e_i^{1982}$$

$$FT_i^{1988} = \alpha + \beta Tax_i^{1988} + \gamma Z_i + e_i^{1988}$$





# “事前事后” (Before and After) 比简单截面回归更好

- 在拥有两年的数据之后，我们做的可以比跑两个分别的回归或者是**汇总** pooled 回归更好
  - 汇总回归忽略了一个事实：我们对每个单位有不只一个观察
- 我们分别看两个回归模型，想象有个  $Z_i$  变量代表和交通死亡有关的**固定的州属特征**。
  - 比如说当地对喝酒和开车的文化

$$FT_i^{1982} = \alpha + \beta Tax_i^{1982} + \gamma Z_i + e_i^{1982}$$

$$FT_i^{1988} = \alpha + \beta Tax_i^{1988} + \gamma Z_i + e_i^{1988}$$

- 请注意：如果  $Z_i$ **不能**被观察到，那么就被包含在误差项当中了。



# 变化和变化格式 change-on-change specification

- 用 1988 年的等式减去 1982 年的的等式

$$FT_i^{1988} - FT_i^{1982} = \beta(Tax_i^{1988} - Tax_i^{1982}) + (e_i^{1988} - e_i^{1982})$$



# 变化和变化格式 change-on-change specification

- 用 1988 年的等式减去 1982 年的的等式

$$FT_i^{1988} - FT_i^{1982} = \beta(Tax_i^{1988} - Tax_i^{1982}) + (e_i^{1988} - e_i^{1982})$$

- 这样一来我们就从等式中**消掉了**不可被观察的因素  $Z_i$ , 再跑一次回归:

$$\Delta \hat{FT}_i = -0.072 - 1.04 \Delta Tax_i$$

(0.065)      (0.36)



# 变化和变化格式 change-on-change specification

- 用 1988 年的等式减去 1982 年的的等式

$$FT_i^{1988} - FT_i^{1982} = \beta(Tax_i^{1988} - Tax_i^{1982}) + (e_i^{1988} - e_i^{1982})$$

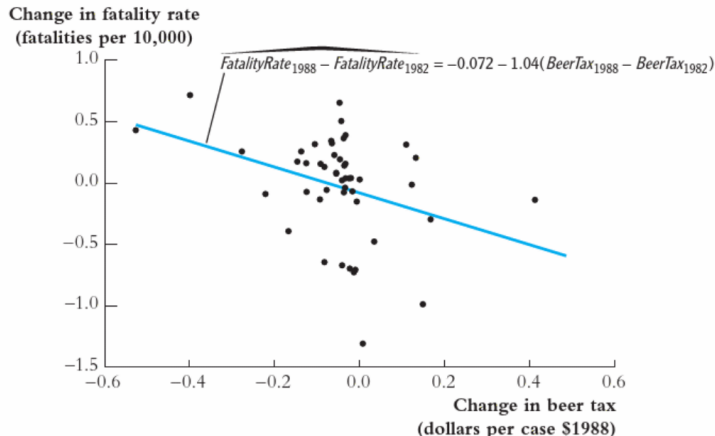
- 这样一来我们就从等式中**消掉了**不可被观察的因素  $Z_i$ , 再跑一次回归:

$$\Delta \hat{FT}_i = \underset{(0.065)}{-0.072} - \underset{(0.36)}{1.04} \Delta Tax_i$$

- 现在 OLS 告诉我们酒精税和交通事故死亡呈显著负相关



# Change-on-change 散点回归线图 (Stock and Watson 2014, p.402)



# 其他例子：经济发展和民主 (Acemoglu et al. 2009)

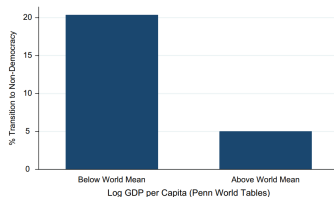


Fig. 3. Transition to non-democracy and income level, 1960–1995. Sample includes all countries in five year intervals between 1955 and 1990 which are democratic according to the Przeworski Measure of Democracy for which at least two observations are available. Observations are grouped depending on whether log income per capita is above or below the average log income per capita in the world for the observation year. Each column measures the fraction of observations within each group which experience a transition to non-democracy five years later. See text for data definitions and sources.

1046

D. Acemoglu et al. / Journal of Monetary Economics 56 (2009) 1043–1058

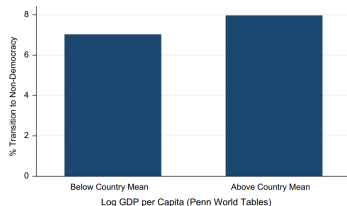


Fig. 4. Transition to non-democracy and income level, 1960–1995. Sample is the same as in Fig. 3. Log income per capita for every observation is demeaned from the average log income per capita in the world for the observation year. Observations are grouped depending on whether demeaned log income per capita is above or below the average demeaned log income per capita in the country between 1955 and 1990. Each column measures the fraction of observations within each group which experience a transition to non-democracy five years later. See text for data definitions and sources.



# 本节课内容

- 1 回顾
- 2 面板数据 Panel Data
- 3 固定效应回归



# 回到酒精税和交通死亡

- 我们可以从 changes-on-changes 回归模型结果中进行因果推断吗？
- 记住：我们消掉了所有固定（或者改变缓慢）的州固定因素的影响。但是，其他因素仍然有可能在 1982-1988 年当中改变，和税收的改变相关
- 比如：一个州在加税的同时决定在媒体上宣传酒精的坏处 (downward or upward bias in  $\hat{\beta}$ ?)
- 有可能因为经济衰退而减税（经济衰退也导致交通事故变少） (downward or upward bias in  $\hat{\beta}$ ?)
- 如果我们不控制这些因素的话内生性仍然存在。





# 长格式 long form

$T > 2$  的面板数据往往以长格式显现

YEAR	STATE	TAX
1982	Alabama	1.5
1988	Alabama	1.8
1982	California	2
1988	California	1.9
1982	Texas	0.3
1988	Texas	0.3
1982	Vermont	2.3
1988	Vermont	3.7
etc	etc	etc

我们有  $N \times T$  个观察 (observations).



# 总是想一想误差项里有什么！

- 使用面板数据我们可以估计以下形式模型：

$$Y_{it} = \alpha + \beta X_{it} + \gamma Z_{it} + e_{it} \quad (1)$$

- 着重看误差项

$$e_{it} = \eta_i + \eta_t + \eta_{it} \quad (2)$$

- 如果有多个单位  $i$  的话我们可以把  $\eta_i$  放到回归模型当中，这样一来我们可以把所有固定单位因素从误差项中消去。



# 固定效应模型 /1

- **固定效应模型**(fixed effects specification) 可以通过设  $\alpha_i = \alpha + \eta_i$  来实现：

$$Y_{it} = \alpha_i + \beta X_{it} + \gamma Z_{it} + u_{it} \quad (3)$$

- 我们估计了  $N$  个不同的截距 intercepts, 每个单位都估计了一个, 囊括了所有影响  $Y$  的固定单位特征。
- 很多这些因素是不可被观察的：所以固定效应模型是解决 OVB 的一个有力工具



# 固定效应模型 /1

- **固定效应模型**(fixed effects specification) 可以通过设  $\alpha_i = \alpha + \eta_i$  来实现：

$$Y_{it} = \alpha_i + \beta X_{it} + \gamma Z_{it} + u_{it} \quad (3)$$

- 我们估计了  $N$  个不同的截距 intercepts, 每个单位都估计了一个, 囊括了所有影响  $Y$  的固定单位特征。
- 很多这些因素是不可被观察的：所以固定效应模型是解决 OVB 的一个有力工具
- 或者我们也可以把固定效应模型写成

$$Y_{it} = \alpha + \beta X_{it} + \gamma Z_{it} + \sum_{i=2}^N \delta_i D_i + u_{it} \quad (4)$$

- $D_i$  是个虚拟变量 (所以从 2 开始避免陷阱), 当观察值来自单位  $i$  时取值为 1, 否则为 0.



## 固定效应模型 /2

- 估计的  $\hat{\beta}$  和  $\hat{\gamma}$  都是一样的，但是对虚拟变量的解读不同
- 在公式 (4) 中， $\alpha$  就是被遗漏的单位的截距， $\delta_i$  则是第  $i$  个单位的截距和  $\alpha$  之差。所以第  $i$  个单位的截距就是  $\alpha + \delta_i$ 。
- 两种形式都有很多个系数 coefficients.
- 在  $N = 48$  时不是个大问题，但是  $N$  在某些情况下可以远大于这个数  $\Rightarrow$  会有问题
- 想象我们用县取代州，或者用一个 4000 人样本的重复问卷数据。



# The within transformation /1

- 就像我们在交通事故的那个案例做的那，我们可以移除固定的单位特征效应 (fixed unit-specific effects) 和推导 derive 出一个计算上 computationally 更简单的公式
- 考虑固定效应公式

$$Y_{it} = \alpha_i + \beta X_{it} + e_{it} \quad (5)$$

然后定义

$$Y_i^* = \alpha_i + \beta X_i^* + e_i^* \quad (6)$$

where

$$Y_i^* = \frac{1}{T} \sum_{t=1}^T Y_{it}, \quad X_i^* = \frac{1}{T} \sum_{t=1}^T X_{it}, \quad e_i^* = \frac{1}{T} \sum_{t=1}^T e_{it} \quad (7)$$



# The within transformation /2

- 当我们做 (5) 减去 (6), 可得

$$Y_{it} - Y_i^* = \alpha_i - \alpha_i^* + \beta(X_{it} - X_i^*) + (e_{it} - e_i^*)$$

这可以被写成

$$\tilde{Y}_{it} = \beta\tilde{X}_{it} + \tilde{e}_{it}$$

- 这叫做“within transformation”: we exploit deviations from (over time) averages for each unit, and therefore use the within unit variation.
- 用这个方法估计的  $\hat{\beta}$  和使用 (3) 和 (4) 得到的结果一样。



## 重新考虑公式 (2)

$$e_{it} = \eta_i + \eta_t + \eta_{it}$$

- 目前来讲我们把  $\eta_t$  还是留在了误差项。我们通过单位固定效应解决了不同单位差异带来的 OVB.
- 但是，也有根据时间变化的遗漏变量
- 拿交通事故举例说明的话，联邦法律的变化，以及车辆安全性的提升。





## 时间固定效应 Time fixed effects /2

- 我们可以把  $\eta_t$  包含在我们的等式中，从而消除掉对所有单位有相同影响。
- 这种包含时间固定效应的形式如下：

$$Y_{it} = \beta X_{it} + \gamma Z_{it} + \lambda_t + e_{it}$$

where  $\lambda_t = \alpha + \eta_t$ .



## 时间固定效应 Time fixed effects /2

- 我们可以把  $\eta_t$  包含在我们的等式中，从而消除掉对所有单位有相同影响。
- 这种包含时间固定效应的形式如下：

$$Y_{it} = \beta X_{it} + \gamma Z_{it} + \lambda_t + e_{it}$$

where  $\lambda_t = \alpha + \eta_t$ .

- 每个时间端都有一个截距  $\lambda_t$ （时间固定效应）。或者你也可以写成：

$$Y_{it} = \alpha + \beta X_{it} + \gamma Z_{it} + \sum_{t=2}^T \delta_t D_t + e_{it}$$

此时  $\delta_t$  代表着与基准  $\alpha$  的差距；时间段  $t$  对应的截距为  $\alpha + \delta_t$ .



# 双固定效应 Two-way fixed effects

在很多情况中，**同时**包含时间**和**单位固定效应是可取的

$$Y_{it} = \alpha_i + \beta X_{it} + \gamma Z_{it} + \lambda_t + e_{it}$$

- 可以写成有一个共同的截距， $(N - 1)$  个单位固定效应， $(T - 1)$  个时间固定效应的形式，也可以把共同截距放到  $\alpha_i$ **或者** $\lambda_t$  中。
- 可以使用 OLS 计算，即把所有固定效应加为虚拟变量，或者使用 within transformation.



# 重新思考交通事故 /1

- 我们现在拥有 1982 到 1988 年每年 48 个州的数据



# 重新思考交通事故 /1

- 我们现在拥有 1982 到 1988 年每年 48 个州的数据
- 一个有州固定效应的回归模型给出以下估计:

$$\hat{FT}_{it} = \hat{\alpha}_i - \underset{(0.20)}{0.66} Tax_{it}$$



# 重新思考交通事故 /1

- 我们现在拥有 1982 到 1988 年每年 48 个州的数据
- 一个有州固定效应的回归模型给出以下估计:

$$\hat{FT}_{it} = \hat{\alpha}_i - \underset{(0.20)}{0.66} Tax_{it}$$

- 这个关系式负的
  - 加入州固定效应相较于普通截面 cross-sectional 方法有很大的区别
  - 将  $T$  从 2 升为 7 也导致了很多变化



- 现在同时加入州和时间固定效应：

$$\hat{FT}_{it} = \hat{\alpha}_i + \hat{\lambda}_t - \underset{(0.25)}{0.64} Tax_{it}$$

- 在这个案例中时间固定效应没有造成较大的区别。可能时间跨度 time span 太短导致没有较重大的时间固定差别。



# 控制变量 Control Variables

- 面板 Panel 数据允许我们控制时间固定以及单位固定的因素的影响
  - 非常有用——这些因素经常是不可被观察的





# 控制变量 Control Variables

- 面板 Panel 数据允许我们控制时间固定以及单位固定的因素的影响
  - 非常有用——这些因素经常是不可被观察的
- 但是面板数据还是需要我们控制在单位和时间跨度上都有变化的其他因素
- 比如说：



# 控制变量 Control Variables

- 面板 Panel 数据允许我们控制时间固定以及单位固定的因素的影响
  - 非常有用——这些因素经常是不可被观察的
- 但是面板数据还是需要我们控制在单位和时间跨度上都有变化的其他因素
- 比如说：
  - 各州每年的道路保养预算
  - 各州每年的公共交通支出
  - 各州每年的私家车数量
  - 各州每年的失业率
- 上述既不时间固定也不单位固定的遗漏变量应该加到回归模型中（如果可以的话）。



# 固定单位特征 Fixed unit characteristics 不能加!

- 在固定效应 FE 模型里可以加入如下控制变量：
  - 对于一些（不见得是全部）截面 cross-section 单位，会随着时间变化而变化的变量
  - 对于一些（不见得是全部）时间段，会随着单位变化而变化的变量



# 固定单位特征 Fixed unit characteristics 不能加!

- 在固定效应 FE 模型里可以加入如下控制变量：
  - 对于一些（不见得是全部）截面 cross-section 单位，会随着时间变化而变化的变量
  - 对于一些（不见得是全部）时间段，会随着单位变化而变化的变量
- **不可以加入时间/单位固定特征**——会固定效应有完美多重共线性 perfect multicollinearity.
  - 不可以 在 面板问卷 survey 数据含有个人固定效应的模型中加入性别作为控制变量
    - 除非有变性人？
  - 不可以 在 有时间固定效应的模型中加入联邦失业率作为控制变量



# 固定单位特征 Fixed unit characteristics 不能加!

- 在固定效应 FE 模型里可以加入如下控制变量：
  - 对于一些（不见得是全部）截面 cross-section 单位，会随着时间变化而变化的变量
  - 对于一些（不见得是全部）时间段，会随着单位变化而变化的变量
- **不可以加入时间/单位固定特征**——会固定效应有完美多重共线性 perfect multicollinearity.
  - 不可以 在 面板问卷 survey 数据含有个人固定效应的模型中加入性别作为控制变量
    - 除非有变性人？
  - 不可以 在 有时间固定效应的模型中加入联邦失业率作为控制变量
- 可以作为**分类**指标来加入
  - 男女分别回归
  - 交互项 interaction



## 标准误：自回归 SE: autocorrelation

- 固定效应模型中 SE 有效的假设为：

$$\text{cov}(e_{it}, e_{is} | X_{i1}, \dots, X_{iT}, \alpha_i) = 0 \text{ for } t \neq s$$

- 以回归量和固定效应为条件的误差项随时间不相关 the error terms, conditional on the regressors and the fixed effect, are uncorrelated over time.
- 现实往往不符合  $\Rightarrow$  自回归 (autocorrelation or serial correlation)
- 与 OVB 不同：不与模型内的自变量相关。在交通死亡的例子中，天气可能是个自回归的原因，但是它不见得导致 OVB（如果我们使用月度数据而不是年度数据的话）



# 聚类标准误 Clustered standard errors

- 如果误差是序列相关 (serially correlated) 的, 那么稳健标准误也是不靠谱的 (无效的)
- 解决办法: 使用**聚类标准误**。
  - 聚类标准误允许聚类内的误差项随意的相关, 但是要求聚类和聚类之间没有相关联
  - 如果有相关联  $\Rightarrow$  考虑 Driscoll-Kraay(DK) 标准误 (需要大  $T$ )
- 注意: 标准误依赖的**渐进性质**(asymptotic properties) 需要聚类的数量比较大
- 较小的话  $\rightarrow$  **野聚类**wild cluster
- 当聚类的数量占人口数量的比重打的时候 (且已知比例) + Binary treatment: CCV 或者 TSCB (Abadie et al. 2022)



- 当加入控制变量和普通的稳健 (robust) 标准误时:

$$\hat{F}T_{it} = \hat{\alpha}_i - \underset{(0.22)}{0.45} Tax_{it} + \hat{\lambda}_t + \hat{\gamma} Controls$$

- 现在使用按州聚类的标准误:

$$\hat{F}T_{it} = \hat{\alpha}_i - \underset{(0.32)}{0.45} Tax_{it} + \hat{\lambda}_t + \hat{\gamma} Controls$$





**TABLE 10.1** Regression Analysis of the Effect of Drunk Driving Laws on Traffic Deaths**Dependent variable: Traffic fatality rate (deaths per 10,000).**

Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Beer tax	0.36** (0.05)	-0.66* (0.29)	-0.64+ (0.36)	-0.45 (0.30)	-0.69* (0.35)	-0.46 (0.31)	-0.93** (0.34)
Drinking age 18				0.028 (0.070)	-0.010 (0.083)		0.037 (0.102)
Drinking age 19				-0.018 (0.050)	-0.076 (0.068)		-0.065 (0.099)
Drinking age 20				0.032 (0.051)	-0.100+ (0.056)		-0.113 (0.125)
Drinking age						-0.002 (0.021)	
Mandatory jail or community service?				0.038 (0.103)	0.085 (0.112)	0.039 (0.103)	0.089 (0.164)
Average vehicle miles per driver				0.008 (0.007)	0.017 (0.011)	0.009 (0.007)	0.124 (0.049)
Unemployment rate				-0.063** (0.013)		-0.063** (0.013)	-0.091** (0.021)
Real income per capita (logarithm)				1.82** (0.64)		1.79** (0.64)	1.00 (0.68)
Years	1982–88	1982–88	1982–88	1982–88	1982–88	1982–88	1982 & 1988 only
State effects?	no	yes	yes	yes	yes	yes	yes
Time effects?	no	no	yes	yes	yes	yes	yes
Clustered standard errors?	no	yes	yes	yes	yes	yes	yes

**F-Statistics and p-Values Testing Exclusion of Groups of Variables**

Time effects = 0			4.22 (0.002)	10.12 ( $< 0.001$ )	3.48 (0.006)	10.28 ( $< 0.001$ )	37.49 ( $< 0.001$ )
Drinking age coefficients = 0				0.35 (0.786)	1.41 (0.253)		0.42 (0.738)
Unemployment rate, income per capita = 0				29.62 ( $< 0.001$ )		31.96 ( $< 0.001$ )	25.20 ( $< 0.001$ )
$\bar{R}^2$	0.091	0.889	0.891	0.926	0.893	0.926	0.899

These regressions were estimated using panel data for 48 U.S. states. Regressions (1) through (6) use data for all years 1982 to 1988, and regression (7) uses data from 1982 and 1988 only. The data set is described in Appendix 10.1. Standard errors are given in parentheses under the coefficients, and  $p$ -values are given in parentheses under the  $F$ -statistics. The individual coefficient is statistically significant at the \*10%, \*5%, or \*\*1% significance level.

# 我们复现一下吧！

	(1)	(2)	(3)	(4)	(5)	(6)
Beer tax	0.365*** (0.053)	-0.656* (0.292)	-0.640+ (0.357)	-0.445 (0.297)	-0.690+ (0.352)	-0.456 (0.307)
Drinking age 18				0.028 (0.070)	-0.010 (0.083)	
Drinking age 19				-0.018 (0.050)	-0.076 (0.068)	
Drinking age 20				0.032 (0.051)	-0.100+ (0.056)	
Drinking age						-0.002 (0.022)
Mandatory jail or community service?				0.038 (0.103)	0.085 (0.112)	0.039 (0.103)
Average vehicle (thousand) miles per driver				0.008 (0.007)	0.017 (0.011)	0.009 (0.007)
Unemployment rate				-0.063*** (0.013)		-0.063*** (0.013)
Real income per capita (logarithm)				1.816** (0.636)		1.786** (0.643)
FE: State	No	Yes	Yes	Yes	Yes	Yes
FE: Year	No	No	No	No	No	No
Clustered SE	No	Yes	Yes	Yes	Yes	Yes
N	336	336	336	335	335	335
Time effects=0			4.22 (0.002)	10.12 (<0.001)	3.48 (0.006)	10.28 (<0.001)
Drinking age coefficients=0				0.35 (0.786)	1.41 (0.253)	
u/e and income(pc)=0				29.62 (<0.001)		31.96 (<0.001)
R <sup>2</sup>	0.091	0.889	0.891	0.926	0.893	0.926

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001



# References I



Abadie, Alberto et al. (Oct. 2022). “When Should You Adjust Standard Errors for Clustering?” In: *The Quarterly Journal of Economics* 138.1, pp. 1–35. DOI: 10.1093/qje/qjac038.



Acemoglu, Daron et al. (Nov. 2009). “Reevaluating the modernization hypothesis”. en. In: *Journal of Monetary Economics* 56.8, pp. 1043–1058. ISSN: 0304-3932. DOI: 10.1016/j.jmoneco.2009.10.002. URL: <https://www.sciencedirect.com/science/article/pii/S0304393209001299> (visited on 08/01/2023).



Bandiera, Oriana, Valentino Larcinese, and Imran Rasul (Nov. 2010). “Heterogeneous Class Size Effects: New Evidence from a Panel of University Students”. In: *The Economic Journal* 120.549, pp. 1365–1398. DOI: 10.1111/j.1468-0297.2010.02364.x.



# References II



Stock, James and Mark Watson (2014). *Econometrics, Update PDF Ebook, Global Edtion*. Harlow, United Kingdom, UNITED KINGDOM: Pearson Education, Limited. ISBN: 9781292071367. URL: <http://ebookcentral.proquest.com/lib/londonschoolecons/detail.action?docID=5174962> (visited on 07/19/2023).

