高级计量经济学

Lecture 10: Models for Panel Data

黄嘉平

工学博士 经济学博士
深圳大学中国经济特区研究中心 讲师

办公室 粤海校区汇文楼1510
E-mail huangjp@szu.edu.cn
Website https://huangjp.com

面板数据

Panel data

- **面板数据（panel data）**，也称为**纵向数据（longitudinal data）**，是指 n 个不同个体在 T 个不同时期上的观测数据。

- 如果数据集包含变量 X 和 Y 的观测值，该数据可以表示为

$$(X_{it}, Y_{it}), \quad i = 1,\ldots,m \text{ and } t = 1,\ldots,T$$

- **平衡面板（balanced panel）**指所有的变量在每个个体和每一时期中都有观测值的面板。否则称为**非平衡面板（unbalanced panel）**。

- m 比较小 T 比较大的面板数据称为**长面板（long panel）**；T 比较小 m 比较大的数据称为**短面板（short panel）**。

美国州级交通事故死亡数据集

The U.S. state traffic fatality data set

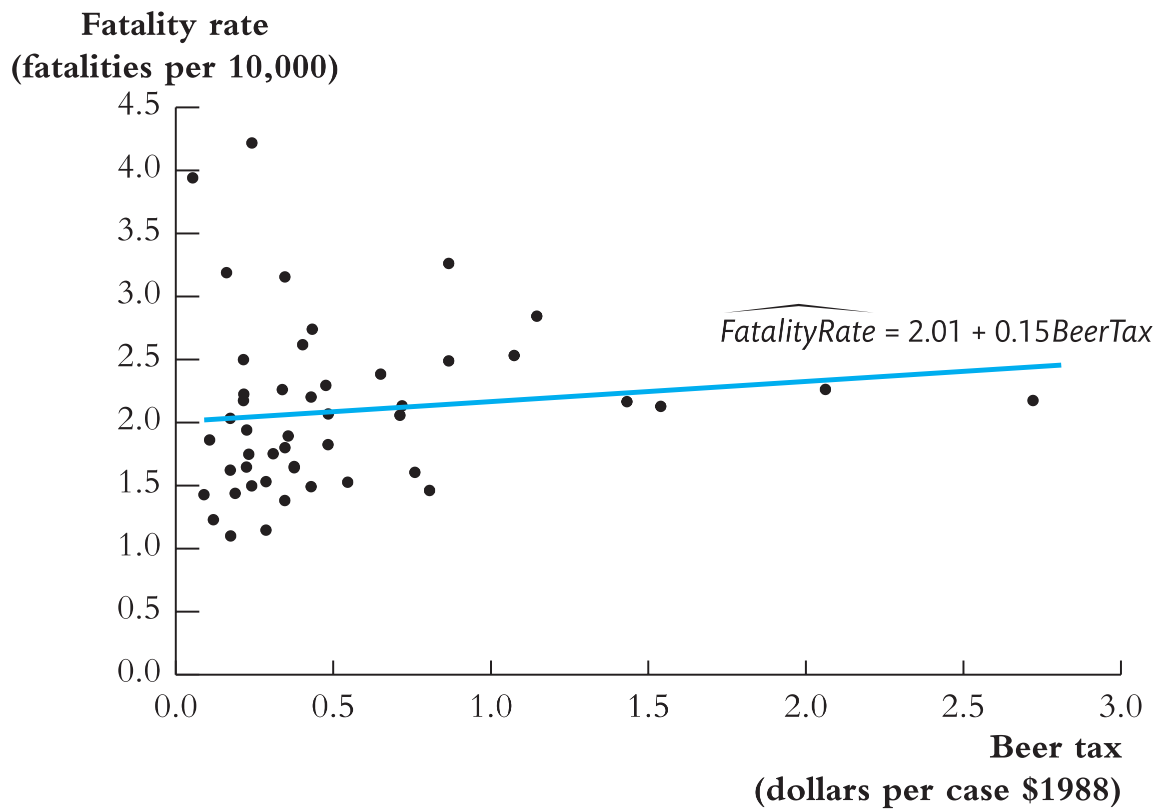
- 数据来源：Ruhm, C. (1996). Alcohol policies and highway vehicle fatalities, Journal of Health Economics, 15:435-454.（可在 Stock & Watson 的教科书中找到，也被收录到 R 的 AER 包中）

- 包含美国 48 个本土州（不包含阿拉斯加和夏威夷）1982 至 1988 年间的交通事故及相关的平衡面板数据。

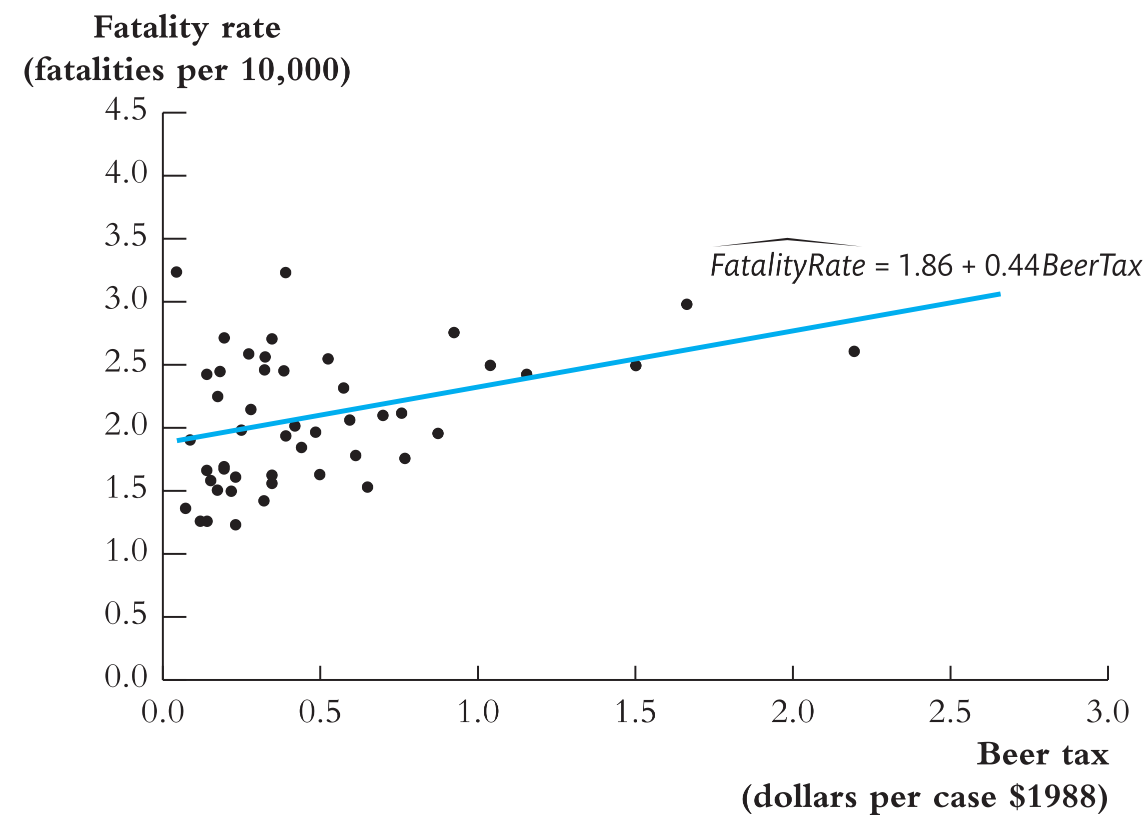
- 主要变量：
    - 交通事故死亡率 *FatalityRate*：每州每年每一万人中死于交通事故的人数
    - 啤酒税 *BeerTax*：单位为美元/箱（1988年美元价值），可作为酒精税的代理变量
    - 其他变量：法定饮酒年龄、酒驾处罚、人均收入、失业率等。

交通事故死亡率与酒精税

Stock & Watson, *Introduction to Econometrics*, 3rd.



(a) 1982 data



(b) 1988 data

$$\widehat{FatalityRate} = \underset{(0.15)}{2.01} + \underset{(0.13)}{0.15}\, BeerTax \quad (1982 \text{ data}).$$

$$\widehat{FatalityRate} = \underset{(0.11)}{1.86} + \underset{(0.13)}{0.44}\, BeerTax \quad (1988 \text{ data}).$$

线性模型的回归结果是反常识的：啤酒税越高死亡率也越高

⇒ 存在遗漏变量偏差?：道路设施状态、社会对饮酒的容忍度等

很难度量，却不随时间而变化 → 利用面板数据

面板数据模型

面板数据的线性回归模型

面板数据的线性回归模型可以写成

$$y_{it} = \boldsymbol{X}_{it}\boldsymbol{\beta} + u_{it}, \quad i = 1,\ldots,m, \quad t = 1,\ldots,T$$

此处，$\boldsymbol{X}_{it}$ 是 $1 \times k$ 的行向量，总样本量为 $n = mT$。

- 如果误差项满足 $E[u_{it} \mid \boldsymbol{X}_{it}] = 0$，则可以用 OLS 估计，否则 OLS 估计量不满足一致性。
- 在面板数据中，误差项一般无法满足 i.i.d. 条件。例如全国范围内的政策可以使 u_{it} 和 u_{jt} 间产生相关性，而个体层面的特征可以使 u_{is} 和 u_{it} 间产生相关性。

  此时 $\boldsymbol{\beta}$ 的 OLS 估计量不满足有效性（Gauss-Markov 定理）。而通常用来估计 $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ 的协方差矩阵的估计量 $s^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$ 也不满足一致性（但是存在稳健估计量）。

误差成分模型

Error component models

我们可以考虑将误差项 u_{it} 分解为几个独立的成分，即

$$u_{it} = e_t + v_i + \varepsilon_{it}$$

这里 e_t 作用于所有发生在时间 t 的观测值；v_i 作用于所有个体 i；ε_{it} 仅作用于 it 观测值。

e_t 和 v_i 可以假设为固定变量或随机变量，分别对应**固定效应模型**（**fixed effect model**）和**随机效应模型**（**random effect model**）。

- FE 模型：可以将 e_t 和 v_i 当作参数进行 OLS 估计
- RE 模型：无法用 OLS 估计 e_t 和 v_i，但可以用 feasible GLS

固定效应模型

Fixed effects model

这里将模型简化为时间效应 $e_t = 0$，并假设 $\boldsymbol{X}$ 是外生的。假设 v_i 独立于 ε_{it}，但可以和 $\boldsymbol{X}$ 相关。

将数据按（个体，时间）的顺序进行排列，并定义虚拟变量 $d_1, \ldots, d_m$，其矩阵计为 $\boldsymbol{D}$。

i	t	X_{it}	d_1	d_2	$\cdots$	d_m
1	1	$\cdots$	1	0	$\cdots$	0
1	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
1	T	$\cdots$	1	0	$\cdots$	$\vdots$
2	1	$\cdots$	0	1	$\cdots$	$\vdots$
2	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
2	T	$\cdots$	$\vdots$	1	$\cdots$	$\vdots$
$\cdots$	$\cdots$	$\cdots$	$\vdots$	0	$\cdots$	0
m	1	$\cdots$	$\vdots$	$\vdots$	$\cdots$	1
m	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$
m	T	$\cdots$	0	0	$\cdots$	1

$= D$

则固定效应模型可以表达为

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{D\eta} + \boldsymbol{\varepsilon}, \quad E[\boldsymbol{\varepsilon\varepsilon}^\top] = \sigma_\varepsilon^2 \boldsymbol{I}_n$$

其中 $\boldsymbol{\eta} = (v_1, v_2, \ldots, v_m)^\top$ 为个体固定效应（individual fixed effects）。注意 $\boldsymbol{X}$ 中不应包含常数项，否则会出现共线性问题。

此模型满足矩条件 $E\left[\boldsymbol{X}_{it}^\top(y_{it} - \boldsymbol{X}_{it}\boldsymbol{\beta} - v_i)\right] = 0$ 和 $E\left[y_{it} - \boldsymbol{X}_{it}\boldsymbol{\beta} - v_i\right] = 0$，此时的 $\hat{\boldsymbol{\beta}}$ （即 OLS 估计量）称为固定效应估计量（fixed effects estimator），也称为最小二乘虚拟变量估计量（least squares dummy variables, **LSDV**, estimator）。

Kronecker 积

Kronecker product

在很多文献中会看到 $\otimes$ 记号。在计量经济学中，这个记号通常代表矩阵的 Kronecker 积（也称为矩阵直积 matrix direct product）。

注意：集合的直积是笛卡尔积 Cartesian product

对于任意的 $m \times n$ 矩阵 $\boldsymbol{A}$ 和 $p \times q$ 矩阵 $\boldsymbol{B}$，$\boldsymbol{A} \otimes \boldsymbol{B}$ 是 $(mp) \times (nq)$ 矩阵，定义为

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{pmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{pmatrix}$$

很显然，前面的矩阵 $\boldsymbol{D}$ 可以写成 $\boldsymbol{D} = \boldsymbol{I}_m \otimes \boldsymbol{\iota}_T$。 $\boldsymbol{\iota}_T$ 代表维度为 T，所有要素都为 1 的向量

Kronecker 积的一些重要性质： D&M Section 12.2: pp.504-505

- $(\boldsymbol{A} \otimes \boldsymbol{B})^\top = \boldsymbol{A}^\top \otimes \boldsymbol{B}^\top$
- $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC}) \otimes (\boldsymbol{BD})$
- $(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$

中心化与组内估计量

Demeaning and within-group estimator

求 LSDV 估计量需要用到的虚拟变量较多，当 m 较大时计算难度大。

定义 $\boldsymbol{M_D} = \boldsymbol{I} - \boldsymbol{D}(\boldsymbol{D}^\top\boldsymbol{D})^{-1}\boldsymbol{D}^\top$，即以 $\boldsymbol{D}$ 为解释变量的残差生成矩阵。根据 FWL 定理，

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{D\eta} + \boldsymbol{\varepsilon} \ \text{ 和 } \ \boldsymbol{M_D y} = \boldsymbol{M_D X b} + \boldsymbol{\epsilon}$$

中的估计量 $\hat{\boldsymbol{\beta}}$ 和 $\hat{\boldsymbol{b}}$ 相同。由右侧模型可得

$$\hat{\boldsymbol{\beta}}_{\mathrm{FE}} = (\boldsymbol{X}^\top\boldsymbol{M_D}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{M_D}\boldsymbol{y}$$

如何理解 $\boldsymbol{M_D y}$? 该向量的第 i 项是回归模型 $y_{it} = v_i + u_i$ 的 OLS 残差，即 $y_{it} - \frac{1}{T}\sum_{t=1}^{T} y_{it} = y_{it} - \overline{y}_i$。也就是说 $\boldsymbol{M_D y}$ 从 y_{it} 中减去了第 i 组的组内均值。$\boldsymbol{M_D X}$ 也可以给出同样的解释。因此，$\hat{\boldsymbol{\beta}}_{\mathrm{FE}}$ 也称为 within-group estimator。

固定效应模型的优缺点

- within-group 估计量很容易计算：只需计算每一组的均值，然后从观测值中减去即可。
- 固定效应的估计量 $\hat{v}_i = \bar{y}_i - \overline{\boldsymbol{X}}_i \hat{\boldsymbol{\beta}}_{\mathrm{FE}}$。
- 由于加入了固定效应，所有不随时间变化的变量的系数都无法估计（例如性别、出生地、教育程度等），因为加入这些变量会引起共线性问题。
- 如果 $\boldsymbol{D}$ 能很好地解释 $\boldsymbol{X}$，则 $\boldsymbol{M_D X}$ 的变动范围很小，此时 $\hat{\boldsymbol{\beta}}_{\mathrm{FE}}$ 的估计精确度会大幅减小。

随机效应模型

Random effects model

现在我们假设 v_i 独立于 ε_{it} 和 $\boldsymbol{X}$。根据 $\boldsymbol{X}$ 的外生性，$E[u_{it} \mid \boldsymbol{X}] = E[v_i + \varepsilon_{it} \mid \boldsymbol{X}] = 0$ 成立。在这个条件下，回归模型 $y_{it} = \boldsymbol{X}_{it}\boldsymbol{\beta} + u_{it}$ 的 OLS 估计量满足非偏性。

如果假设 $v_i \sim \mathrm{IID}(0, \sigma_v^2)$，则有

$$
\begin{aligned}
\mathrm{Var}[u_{it}] &= \sigma_v^2 + \sigma_\varepsilon^2 \\
\mathrm{Cov}[u_{it}, u_{is}] &= \sigma_v^2 \\
\mathrm{Cov}[u_{it}, u_{js}] &= 0 \ \ \text{for } i \neq j
\end{aligned}
$$

此时误差项 u_{it} 的协方差矩阵 $\boldsymbol{\Omega}$ 可以写成

$$
\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Sigma} & \boldsymbol{O} & \cdots & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{\Sigma} & \cdots & \boldsymbol{O} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{O} & \boldsymbol{O} & \cdots & \boldsymbol{\Sigma} \end{pmatrix}, \qquad \boldsymbol{\Sigma} \equiv \sigma_\varepsilon^2 \boldsymbol{I}_T + \sigma_v^2 \boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top
$$

$\boldsymbol{\Omega} \neq \sigma^2 \boldsymbol{I}$，因此 OLS 估计量不满足有效性。GLS 估计量是更好的选择。

随机效应模型的 GLS 估计

如何估计 σ_ε^2 和 σ_v^2?

σ_ε^2 可以通过 $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{D\eta} + \boldsymbol{\varepsilon}$ 模型的 OLS 残差进行估计，其非偏且一致估计量是 $\hat{\sigma}_\varepsilon^2 = \mathrm{SSR}\big/(n-k-m)$。

当 m 随着 n 增加而增加时，σ_v^2 可以通过下面的模型进行估计

$$\boldsymbol{P_D y} = \boldsymbol{P_D X b} + \boldsymbol{\epsilon} \quad \Rightarrow \quad \hat{\boldsymbol{\beta}}_{\mathrm{BG}} = (\boldsymbol{X}^\top \boldsymbol{P_D X})^{-1} \boldsymbol{X}^\top \boldsymbol{P_D y}$$

估计量 $\hat{\boldsymbol{\beta}}_{\mathrm{BG}}$ 称为 between-group 估计量（组间估计量），因为 $\boldsymbol{P_D y}$ 将 y_{it} 都替换为 $\bar{y}_i$，使得每一组（即每个个体 i ）内部的变化都消失了，只剩下每组均值之间的变化。

可知 $\boldsymbol{P_D y} = \boldsymbol{P_D X b} + \boldsymbol{\epsilon}$ 中的误差项的方差是 $\mathrm{Var}(\epsilon_i) = \sigma_v^2 + \sigma_\varepsilon^2\big/T$，因此 $\hat{\sigma}_v^2 = \mathrm{SSR}\big/(m-k) - \hat{\sigma}_\varepsilon^2\big/T$ 是 σ_v^2 的一致估计量。

需要注意的是，这样获得的 $\hat{\sigma}_v^2$ 有可能为负。如果出现了负的估计值，说明真实的模型可能不是误差成分模型。

基于以上方差的估计量，我们可以得到 $\hat{\boldsymbol{\Omega}}$，并对随机效应模型进行 GLS 估计。

随机效应模型的 GLS 估计

由 $\boldsymbol{\Sigma} \equiv \sigma_\varepsilon^2 \boldsymbol{I}_T + \sigma_v^2 \boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top$ 可以推出

$$\boldsymbol{\Sigma}^{-\frac{1}{2}} = \frac{1}{\sigma_\varepsilon}\left(\boldsymbol{I}_T - \frac{\lambda}{T}\boldsymbol{\iota}_T \boldsymbol{\iota}_T^\top\right), \quad \lambda = 1 - \left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_v^2}\right)^{\frac{1}{2}}$$

由此可知，$\boldsymbol{\beta}$ 的 GLS 估计量可以通过对

$$(\boldsymbol{I} - \lambda \boldsymbol{P_D})\boldsymbol{y} = (\boldsymbol{I} - \lambda \boldsymbol{P_D})\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{w}$$

进行 OLS 回归获得。用 $\hat{\sigma}_v^2$ 和 $\hat{\sigma}_\varepsilon^2$ 替代理论值，即可获得 feasible GLS 估计量 $\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$。

由此可知 $\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ 是 $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$ 的 OLS 估计量 $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ 与 between-group 估计量 $\hat{\boldsymbol{\beta}}_{\mathrm{BG}}$ 的线性结合。

- 当 $\lambda = 0$ 时，$\hat{\boldsymbol{\beta}}_{\mathrm{GLS}} = \hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$，此时 $\sigma_v^2 = 0$。
- 当 $\lambda = 1$ 时，$\hat{\boldsymbol{\beta}}_{\mathrm{GLS}} = \hat{\boldsymbol{\beta}}_{\mathrm{FE}}$，此时 $\sigma_\varepsilon^2 = 0$ 或 $T \to \infty$。
- 当 $0 < \lambda < 1$ 时，$\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ 比 $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ 和 $\hat{\boldsymbol{\beta}}_{\mathrm{FE}}$ 更有效。

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) [0.26, 0.46]	−0.66 (0.29) [−1.23, −0.09]	−0.64 (0.36) [−1.35, 0.07]	−0.45 (0.30) [−1.04, 0.14]	−0.69 (0.35) [−1.38, 0.00]	−0.46 (0.31) [−1.07, 0.15]	−0.93 (0.34) [−1.60, −0.26]
Drinking age 18		0.10		0.03 (0.07) [−0.11, 0.17]	−0.01 (0.08) [−0.17, 0.15]		0.04 (0.10) [−0.16, 0.24]
Drinking age 19				−0.02 (0.05) [−0.12, 0.08]	−0.08 (0.07) [−0.21, 0.06]		−0.07 (0.10) [−0.26, 0.13]
Drinking age 20				0.03 (0.05) [−0.07, 0.13]	−0.10 (0.06) [−0.21, 0.01]		−0.11 (0.13) [−0.36, 0.14]
Drinking age						0.00 (0.02) [−0.05, 0.04]	
Mandatory jail or community service?				0.04 (0.10) [−0.17, 0.25]	0.09 (0.11) [−0.14, 0.31]	0.04 (0.10) [−0.17, 0.25]	0.09 (0.16) [−0.24, 0.42]
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, 95% confidence intervals are given in square brackets under the coefficients, and *p*-values are given in parentheses under the *F*-statistics.

补充

- 只考虑 v_i 或 e_t 的模型称为 one-way error component model，同时考虑 v_i 和 e_t 的称为 two-way error component model。
- 也可以加入个体和时间之外的组别效应，例如地区。但需要注意的是，如果一个个体只在同一个地区内有观测值，则固定效应模型无法加入地区效应（共线性问题）。
- 固定效应还是随机效应？当 T 很大时，二者基本没有区别；当 T 很小时，可以考虑 Hausman 检验等方法。
- 参考书：
  - Hsiao, C. (萧政) (2014). *Analysis of Panel Data*, 3rd Edition. Cambridge University Press.
  - Baltagi, B. H. (2013). *Econometric Analysis of Panel Data*, 5th Edition. Wiley.