CHAPTER 10
Regression with Panel Data

Kazu Matsuda

IBEC PHBU 430

Econometrics

Definition of Panel Data

- **Panel data** = consist of observations on the same *n* entities at two or more time periods *T*. If the data set contains observations on the variables *X* and *Y*, then the data are denoted:

- where the first subscript, *i*, refers to the entity being observed, and the second subscript, *t*, refers to the date at which it is observed.

- Those data are for **?** entities (states), where each entity is observed in **?** time periods, for a total of **?** observations.

Example: Traffic Fatalities and Alcohol Taxes

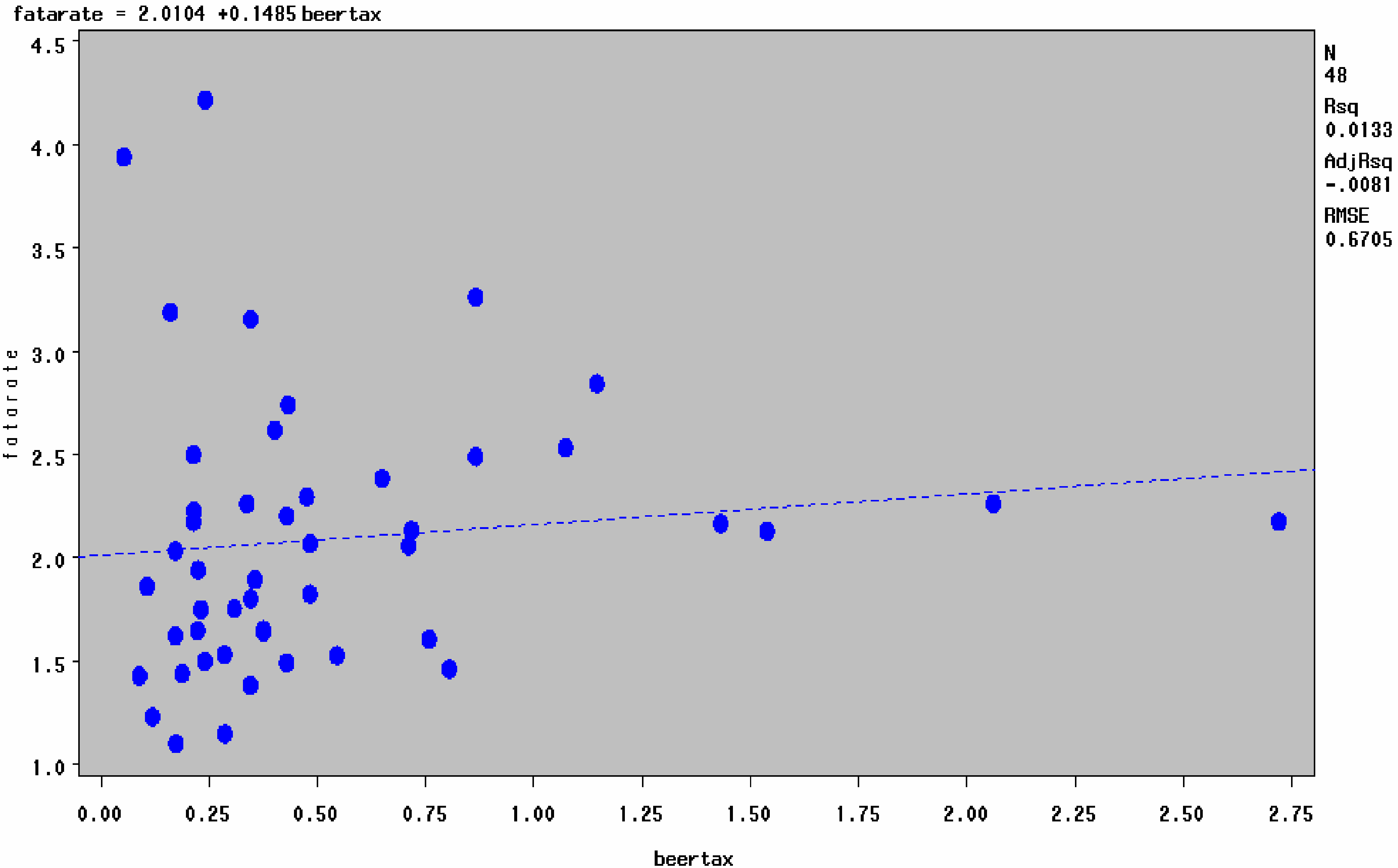
- There are approximately 40,000 highway traffic fatalities each year in the U.S.
- Approximately one-third of fatal crashes involve a driver who was drinking, and this fraction rises during peak drinking periods.
- One study estimates that as many as **?** of drivers on the road between 1 A.M. and 3 A.M. have been drinking, and that a driver who is legally drunk is at least **?** times as likely to cause a fatal crash as a driver who has not been drinking.

Example: Traffic Fatalities and Alcohol Taxes

- The measure of traffic deaths: **fatality rate** = the number of annual traffic deaths per 10,000 people in the population in the state.

- The measure of alcohol taxes: **“real” tax on a case of beer** = the beer tax put into 1988 dollars by adjusting for inflation.

1982 only



fatarate = 2.0104 +0.1485 beertax
N
48
Rsq
0.0133
AdjRsq
-.0081
RMSE
0.6705
fatarate
beertax
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75

1982 only

$$\widehat{FataRate} = 2.01 + 0.15 BeerTax$$
$$(0.15) \quad (0.13)$$

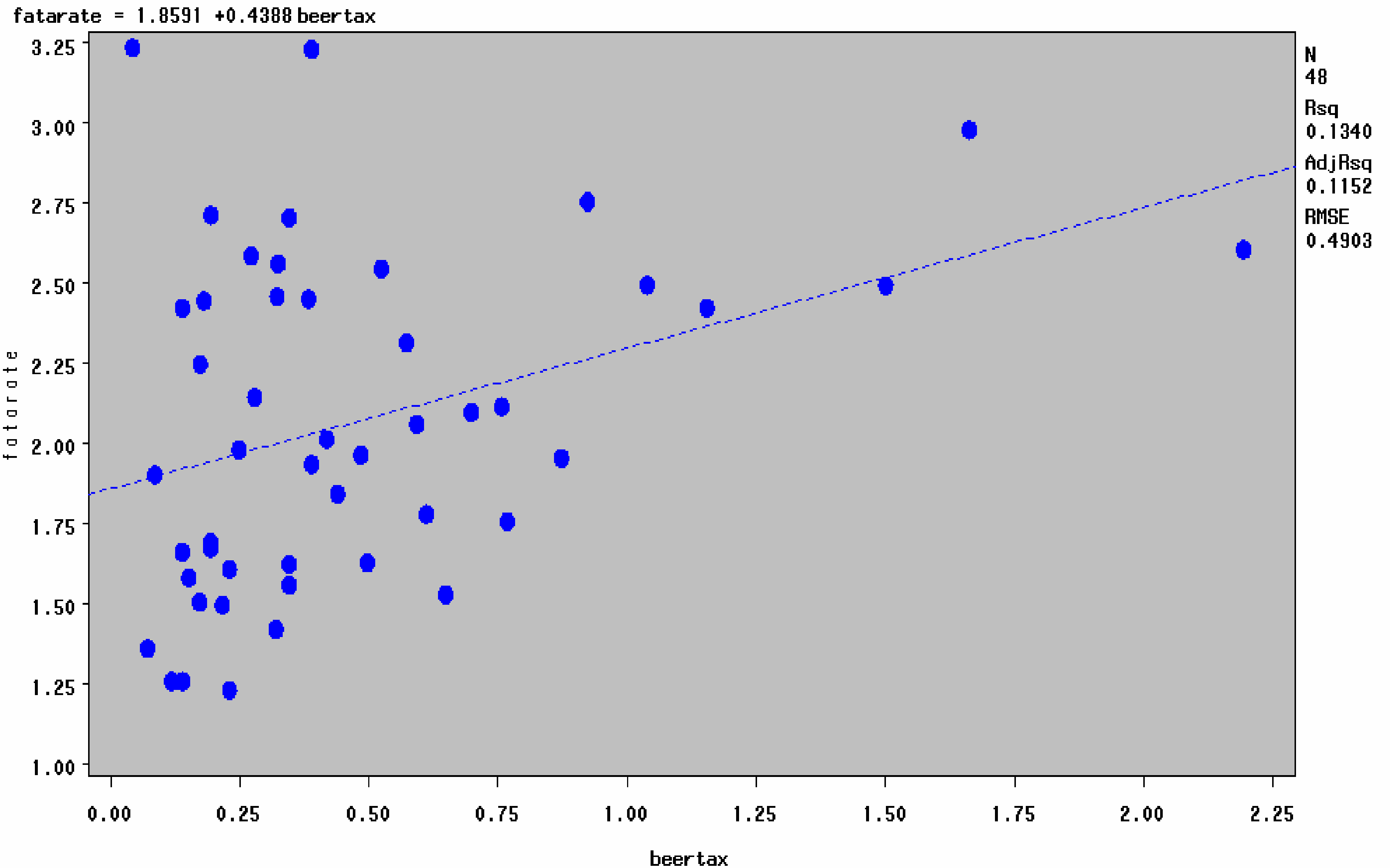
$$F = 0.62$$

$$R^2 = 0.01$$

$$\text{Standard Error of Regression} = 0.67$$

$$\overline{FataRate} = 2.09$$

1988 only


fatarate = 1.8591 +0.4388 beertax
N
48
Rsq
0.1340
AdjRsq
0.1152
RMSE
0.4903
fatarate
beertax
3.25
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25

1988 only

$$\widehat{FataRate} = 1.86 + 0.44 BeerTax$$
$$(0.11)\quad(0.13)$$

$$F = 7.12$$

$$R^2 = 0.134$$

$$\text{Standard Error of Regression} = 0.49$$

$$\overline{FataRate_{1988}} = 2.07$$

- Should we conclude that an increase in the tax on beer leads to more traffic deaths?

- These regressions could have substantial ?

  .

Many factors affect the fatality rate

- The quality of the automobiles driven in the state.
- Whether the state highways are in good repair.
- The density of cars on the road.
- Whether it is socially acceptable to drink and drive.

- Some of these variables might be very hard or even impossible to measure.

- If these factors remain constant over time in a given state, then another route is available. Because we have panel data, we can in effect hold these factors constant, even though we cannot measure them.

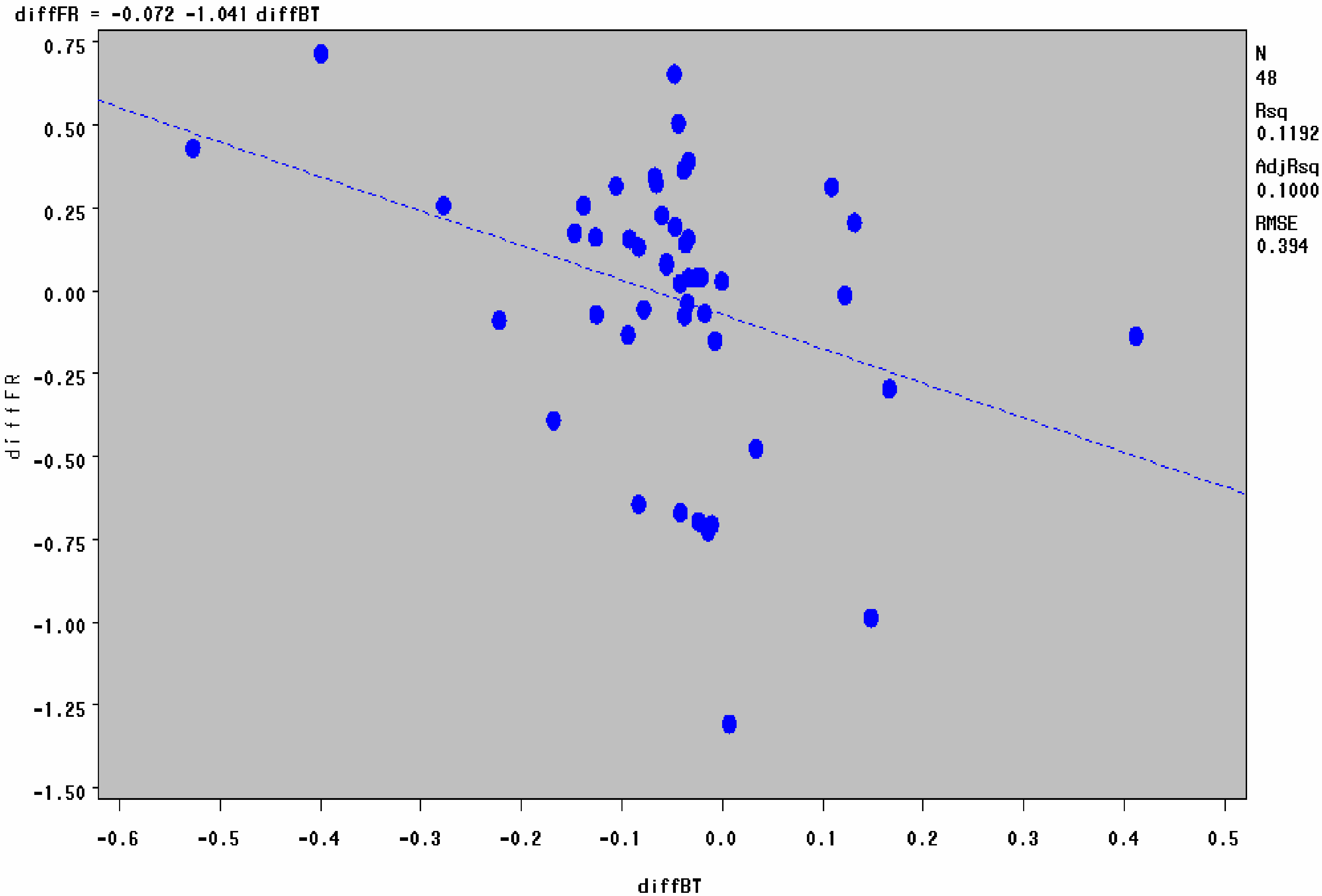
10.2 Panel Data with Two Time Periods: "Before and After" Comparisons

- Let Z_i be a variable that determines the fatality rate in the i^{th} state, but does not change over time (so the t subscript is omitted). Ohio specific factor.
- For example, Z_i might be the local cultural attitude toward drinking and driving, which changes slowly and thus could be considered to be constant between 1982 and 1988.

$$\left(FatalityRate_{i1988} - FatalityRate_{i1982}\right)$$

$$= \beta_1\left(BeerTax_{i1988} - BeerTax_{i1982}\right) + \left(\varepsilon_{i1988} - \varepsilon_{i1982}\right)$$

- Specifying the regression in changes eliminates the effect of the unobserved variables Z_i that are constant over time. In other words, analyzing changes in Y and X has the effect of controlling for variables that are constant over time, thereby eliminating this source of ? bias.



diffFR = -0.072 -1.041 diffBT
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00
-1.25
-1.50
diffFR
-0.6
-0.5
-0.4
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
0.4
0.5
diffBT
N
48
Rsq
0.1192
AdjRsq
0.1000
RMSE
0.394

Estimation result

$$\left(FatalityRate_{i1988} - FatalityRate_{i1982}\right)$$

$$= -0.072 - 1.04\left(BeerTax_{i1988} - BeerTax_{i1982}\right)$$

$$(0.065)\ (0.36)$$

$$F = 6.22$$

$$R^2 = 0.12$$

$$\text{Standard Error of Regression} = 0.394$$

$$\overline{\Delta FataRate} = -0.0195$$

- An increase in the real beer tax by $1 per case reduces the traffic fatality rate by **?** deaths per 10,000 people.

- By examining changes in the fatality rate over time, the regression controls for fixed factors such as cultural attitudes toward drinking and driving.

10.3 Fixed Effects Regression

Purpose: Fixed effects regression is a method for controlling for **?** in panel data when the omitted variables vary across entities (states, individuals, groups) but do not change over time.

Model

$$FatalityRate_{it} = \beta_0 + \beta_1 BeerTax_{it} + \beta_2 Z_i + \varepsilon_{it}$$

- Z_i is an unobserved variable that varies from one state to the next but does not change over time (for example, cultural attitudes toward drinking & driving).

$$FatalityRate_{it} = \beta_0 + \beta_1 BeerTax_{it} + \beta_2 Z_i + \varepsilon_{it}$$

- Because Z_i varies from one state to the next but is constant over time, the model can be interpreted as having *n* intercepts, one for each state.
- Rewrite the model as:

- α_i =
- The slope coefficient of the population regression line, β_1, is the same for all states, but the intercept of the population regression line varies from one state to the next.

$$FatalityRate_{it} = \beta_1 BeerTax_{it} + \alpha_i + \varepsilon_{it}$$

- The state-specific intercepts α_i*'s* can be expressed using binary variables to denote the individual states.
- We cannot include all *n* binary variables plus a common intercept (perfect multicollinearity called dummy variable trap). So we omit the dummy variable for the first group.

Rewrite

Estimation result

$$FatalityRate =$$

$$3.48 - 0.66 BeerTax - 0.57 D_{AZ} + 0.52 D_{GA} - 2.19 D_{NY} + 0.56 D_{SC} + ...$$

$$(0.31) \quad (0.2) \qquad\qquad (0.27) \qquad (0.18) \qquad (0.3) \qquad (0.11)$$

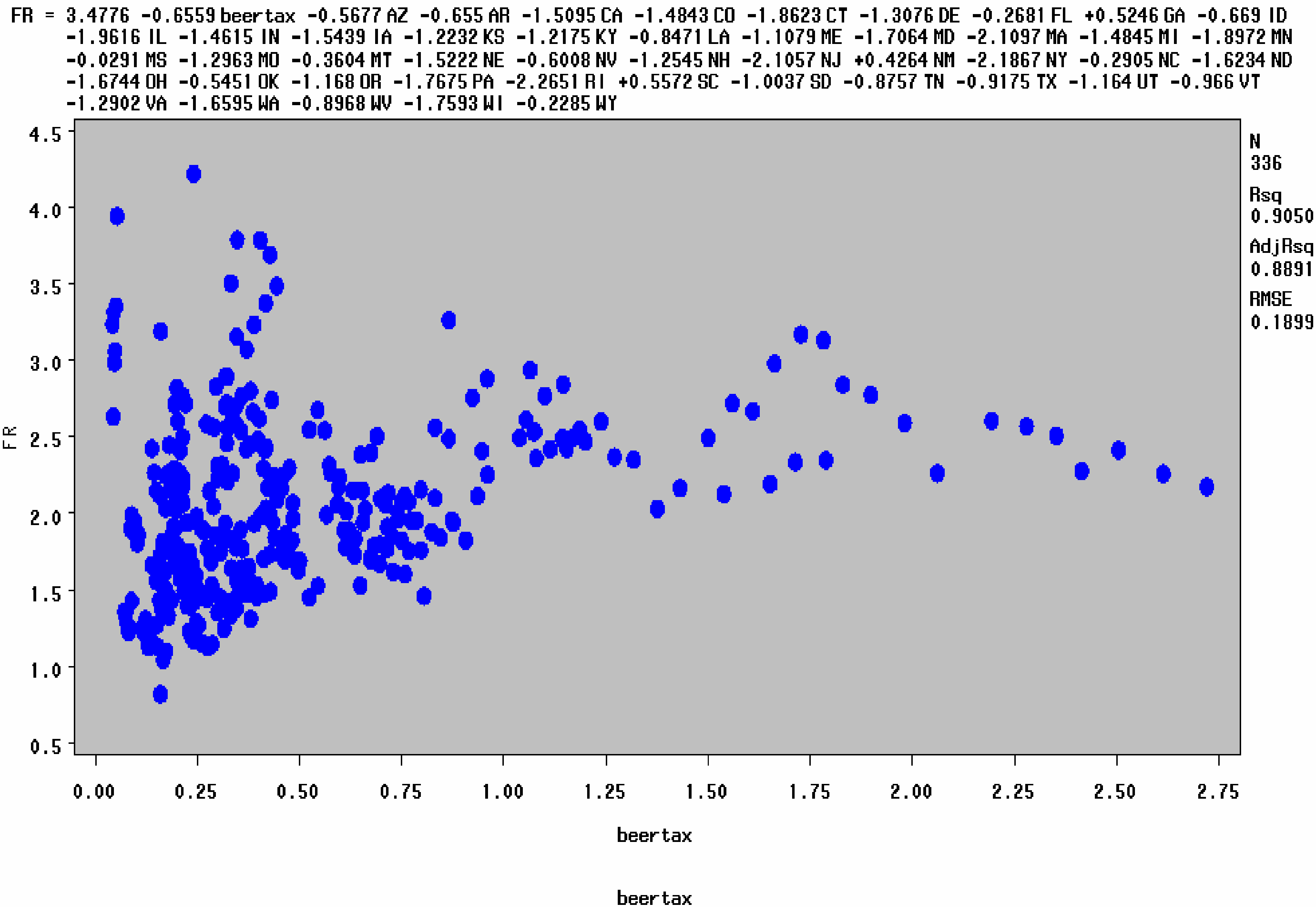
$$F = 57$$

$$R^2 = 0.905$$

$$\text{Standard Error of Regression} = 0.19$$

$$\overline{FataRate} = 2.04$$

- An increase in the real beer tax by $1 per case reduces the traffic fatality rate by **?** deaths per 10,000 people.


FR = 3.4776 -0.6559 beertax -0.5677 AZ -0.655 AR -1.5095 CA -1.4843 CO -1.8623 CT -1.3076 DE -0.2681 FL +0.5246 GA -0.669 ID
-1.9616 IL -1.4615 IN -1.5439 IA -1.2232 KS -1.2175 KY -0.8471 LA -1.1079 ME -1.7064 MD -2.1097 MA -1.4845 MI -1.8972 MN
-0.0291 MS -1.2963 MO -0.3604 MT -1.5222 NE -0.6008 NV -1.2545 NH -2.1057 NJ +0.4264 NM -2.1867 NY -0.2905 NC -1.6234 ND
-1.6744 OH -0.5451 OK -1.168 OR -1.7675 PA -2.2651 RI +0.5572 SC -1.0037 SD -0.8757 TN -0.9175 TX -1.164 UT -0.966 VT
-1.2902 VA -1.6595 WA -0.8968 WV -1.7593 WI -0.2285 WY
N
336
Rsq
0.9050
AdjRsq
0.8891
RMSE
0.1899
FR
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
beertax
beertax

$$FatalityRate_{it} = \beta_1 BeerTax_{it} + \alpha_i + \varepsilon_{it}$$

- What are other sources of omitted variable bias?
- Over this period cars were getting ? and occupants were increasingly wearing ? .
- If the real tax on beer rose on average during the mid-1980s, then it could be picking up the effect of overall automobile safety improvements.
- If safety improvements evolved over time but were the same for all states, then we can eliminate their influence by including time fixed effects.

10.4 Regression with Time Fixed Effects

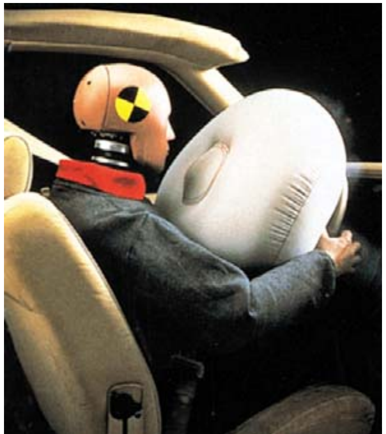
- Time fixed effects = control for variables that are constant across entities but evolve over time.

- It is plausible to think of automobile safety as an omitted variable that changes over time but has the same value for all states.

- Entity & Time fixed effect regression model

$$FatalityRate_{it} = \beta_0 + \beta_1 BeerTax_{it} + \beta_2 Z_i + \beta_3 S_t + \varepsilon_{it}$$






- ✓ Where α_i is the **?** and λ_t is the **?** .
- ✓ This model can equivalently be represented using $n-1$ **?** variables and $T-1$ **?** variables, along with an intercept:

$$FatalityRate_{it} = \beta_0 + \beta_1 BeerTax_{it}$$
$$+\gamma_{AZ} D_{AZ} \ldots + \gamma_{WY} D_{WY} + \delta_{83} D_{83} \ldots + \delta_{88} D_{88} + \varepsilon_{it}$$

Estimation result

$$FatalityRate = 3.51 - \underset{(0.25)}{0.64} BeerTax$$

$$-0.55 D_{AZ} + \ldots - 0.20 D_{WY} - 0.08 D_{1983} \ldots\ldots - 0.05 D_{1988}$$

$$AdjR^2 = 0.89 \quad F = 51.93 \quad SER = 0.188$$

• This specification includes the beer tax, 47 state dummy variables, 6 year dummy variables, and an intercept, so that this regression actually has right hand variables.

• An increase in the real beer tax by $1 per case reduces the traffic fatality rate by **?** deaths per 10,000 people.