Panel Data Studies: Traffic Fatalities

Dr. Patrick Toche

Textbook:

James H. Stock and Mark W. Watson, *Introduction to Econometrics*, 4th Edition, Pearson.

Other references:

Joshua D. Angrist and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, 1st Edition, Princeton University Press.
Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 7th Edition, Cengage Learning.

The textbook comes with online resources and study guides. Other references will be given from time to time.

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.

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 10, Exercise 1.

This exercise refers to the drunk driving panel data regressions summarized in Table 10.1.

a. New Jersey has a population of 8.1 million people. Suppose New Jersey increased the tax on a case of beer by $\$1$ (in 1988 dollars). Use the results in column (4) to predict the number of lives that would be saved over the next year. Construct a 95% confidence interval for your answer.

b. The drinking age in New Jersey is 21. Suppose New Jersey lowered its drinking age to 18. Use the results in column (4) to predict the change in the number of traffic fatalities in the next year. Construct a 95% confidence interval for your answer.

c. Should time effects be included in the regression? Why or why not?

d. A researcher conjectures that the unemployment rate has a different effect on traffic fatalities in the western states than in the other states. How would you test this hypothesis? (Be specific about the specification of the regression and the statistical test you would use.)

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 10, Exercise 1.

This exercise refers to the drunk driving panel data regressions summarized in Table 10.1.

a. New Jersey has a population of 8.1 million people. Suppose New Jersey increased the tax on a case of beer by $\$1$ (in 1988 dollars). Use the results in column (4) to predict the number of lives that would be saved over the next year. Construct a 95% confidence interval for your answer.

b. The drinking age in New Jersey is 21. Suppose New Jersey lowered its drinking age to 18. Use the results in column (4) to predict the change in the number of traffic fatalities in the next year. Construct a 95% confidence interval for your answer.

c. Should time effects be included in the regression? Why or why not?

d. A researcher conjectures that the unemployment rate has a different effect on traffic fatalities in the western states than in the other states. How would you test this hypothesis? (Be specific about the specification of the regression and the statistical test you would use.)

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 10, Exercise 1.

This exercise refers to the drunk driving panel data regressions summarized in Table 10.1.

a. New Jersey has a population of 8.1 million people. Suppose New Jersey increased the tax on a case of beer by $\$1$ (in 1988 dollars). Use the results in column (4) to predict the number of lives that would be saved over the next year. Construct a 95% confidence interval for your answer.

b. The drinking age in New Jersey is 21. Suppose New Jersey lowered its drinking age to 18. Use the results in column (4) to predict the change in the number of traffic fatalities in the next year. Construct a 95% confidence interval for your answer.

c. Should time effects be included in the regression? Why or why not?

d. A researcher conjectures that the unemployment rate has a different effect on traffic fatalities in the western states than in the other states. How would you test this hypothesis? (Be specific about the specification of the regression and the statistical test you would use.)

Problems and Applications

Stock & Watson, Introduction (4th), Chapter 10, Exercise 1.

This exercise refers to the drunk driving panel data regressions summarized in Table 10.1.

a. New Jersey has a population of 8.1 million people. Suppose New Jersey increased the tax on a case of beer by $\$1$ (in 1988 dollars). Use the results in column (4) to predict the number of lives that would be saved over the next year. Construct a 95% confidence interval for your answer.

b. The drinking age in New Jersey is 21. Suppose New Jersey lowered its drinking age to 18. Use the results in column (4) to predict the change in the number of traffic fatalities in the next year. Construct a 95% confidence interval for your answer.

c. Should time effects be included in the regression? Why or why not?

d. A researcher conjectures that the unemployment rate has a different effect on traffic fatalities in the western states than in the other states. How would you test this hypothesis? (Be specific about the specification of the regression and the statistical test you would use.)

Problems and Applications

a. New Jersey has a population of 8.1 million people. Suppose New Jersey increased the tax on a case of beer by $\$1$ (in 1988 dollars). Use the results in column (4) to predict the number of lives that would be saved over the next year. Construct a 95% confidence interval for your answer.

With a $\$1$ increase in the beer tax, the expected number of lives that would be saved is 0.45 per $10,000$ people. Since New Jersey has a population of 8.1 million, the expected number of lives saved is about

$$0.45 \cdot 8.1 \cdot 10^6/10^4 = 0.45 \cdot 810 \approx 365$$

The coefficient on $BeerTax$ is measured with imprecision, since the standard error is large (0.30 is not much smaller than 0.45), so we expect a relatively wide confidence interval. The 95% confidence interval is:

$$(0.45 \pm 1.96 \cdot 0.30) \cdot 810 \approx [-112, 841]$$

This confidence interval is quite wide and does not exclude the possibility that an increase in the beer tax would increase fatalities!

Problems and Applications

a. New Jersey has a population of 8.1 million people. Suppose New Jersey increased the tax on a case of beer by $\$1$ (in 1988 dollars). Use the results in column (4) to predict the number of lives that would be saved over the next year. Construct a 95% confidence interval for your answer.

With a $\$1$ increase in the beer tax, the expected number of lives that would be saved is 0.45 per $10,000$ people. Since New Jersey has a population of 8.1 million, the expected number of lives saved is about

$$0.45 \cdot 8.1 \cdot 10^6/10^4 = 0.45 \cdot 810 \approx 365$$

The coefficient on $BeerTax$ is measured with imprecision, since the standard error is large (0.30 is not much smaller than 0.45), so we expect a relatively wide confidence interval. The 95% confidence interval is:

$$(0.45 \pm 1.96 \cdot 0.30) \cdot 810 \approx [-112, 841]$$

This confidence interval is quite wide and does not exclude the possibility that an increase in the beer tax would increase fatalities!

Problems and Applications

b. The drinking age in New Jersey is 21. Suppose New Jersey lowered its drinking age to 18. Use the results in column (4) to predict the change in the number of traffic fatalities in the next year. Construct a 95% confidence interval for your answer.

If New Jersey lowered the drinking age from 21 to 18, the expected fatality rate would increase by 0.03 deaths per $10,000$.

The standard error is large, so we expect a relatively wide confidence interval. The 95% confidence interval for the change in death rate is:

$$(0.03 \pm 1.96 \cdot 0.07) \cdot 810 \approx [-87, 136]$$

Problems and Applications

b. The drinking age in New Jersey is 21. Suppose New Jersey lowered its drinking age to 18. Use the results in column (4) to predict the change in the number of traffic fatalities in the next year. Construct a 95% confidence interval for your answer.

If New Jersey lowered the drinking age from 21 to 18, the expected fatality rate would increase by 0.03 deaths per $10,000$.

The standard error is large, so we expect a relatively wide confidence interval. The 95% confidence interval for the change in death rate is:

$$(0.03 \pm 1.96 \cdot 0.07) \cdot 810 \approx [-87, 136]$$

Problems and Applications

c. Should time effects be included in the regression? Why or why not?

The F-statistic associated with time fixed effects is 10.12, with associated p-value smaller than 0.001, suggesting that the time effects should be included in the regression.

Problems and Applications

c. Should time effects be included in the regression? Why or why not?

   The F-statistic associated with time fixed effects is 10.12, with associated p-value smaller than 0.001, suggesting that the time effects should be included in the regression.

Problems and Applications

d. A researcher conjectures that the unemployment rate has a different effect on traffic fatalities in the western states than in the other states. How would you test this hypothesis?

The hypothesis singles out the western states, so define a binary variable $West = 1$ for the western states and $West = 0$ for all other states. The hypothesis singles out the unemployment rate in the western states, so we include the interaction term $West \times Unemp$, where $Unemp$ is the unemployment rate variable. Our baseline regression is in column (4), so we include the control variables listed in that column, which we denote Z.

$$\begin{aligned} FatalityRate = \beta_0 + \beta_1\, BeerTax + \beta_Z\, Z \\ + \beta_W\, West + \beta_U\, Unemp + \beta_{WU}\, West \times Unemp + u \end{aligned}$$

The coefficient β_W captures the fixed effect associated with the western states, that is the effect that is orthogonal to the effect of the unemployment rate $Unemp$. The coefficient β_U captures the average effect in the unemployment rate in the other states, while $\beta_U + \beta_{WU}$ captures the effect of the unemployment rate in the western states. The difference in the effect of the unemployment rate in the western and eastern states is therefore β_{WU}.

d. A researcher conjectures that the unemployment rate has a different effect on traffic fatalities in the western states than in the other states. How would you test this hypothesis?

The hypothesis singles out the western states, so define a binary variable $West = 1$ for the western states and $West = 0$ for all other states. The hypothesis singles out the unemployment rate in the western states, so we include the interaction term $West \times Unemp$, where $Unemp$ is the unemployment rate variable. Our baseline regression is in column (4), so we include the control variables listed in that column, which we denote Z.

$$FatalityRate = \beta_0 + \beta_1\, BeerTax + \beta_Z\, Z \\ + \beta_W\, West + \beta_U\, Unemp + \beta_{WU}\, West \times Unemp + u$$

The coefficient β_W captures the fixed effect associated with the western states, that is the effect that is orthogonal to the effect of the unemployment rate $Unemp$. The coefficient β_U captures the average effect in the unemployment rate in the other states, while $\beta_U + \beta_{WU}$ captures the effect of the unemployment rate in the western states. The difference in the effect of the unemployment rate in the western and eastern states is therefore β_{WU}.

Problems and Applications

d. How would you test this hypothesis?

The conjecture can be tested with a test of the significance of the coefficient β_{WU}.

After estimating the regression, the t-statistic would be computed from the estimated coefficients and standard errors.

$$H_0 : \beta_{WU} = 0$$
$$H_1 : \beta_{WU} \neq 0$$
$$t = \frac{\hat{\beta}_{WU}}{\text{SE}(\hat{\beta}_{WU})}$$
$$t > t_{\text{crit}} \implies \text{Reject } H_0$$

where the critical value t_{crit} for significance level 0.05 is $t_{\text{crit}} = 1.96$.

d. How would you test this hypothesis?

The conjecture can be tested with a test of the significance of the coefficient β_{WU}.

After estimating the regression, the t-statistic would be computed from the estimated coefficients and standard errors.

$$H_0 : \beta_{WU} = 0$$
$$H_1 : \beta_{WU} \neq 0$$
$$t = \frac{\hat{\beta}_{WU}}{\text{SE}(\hat{\beta}_{WU})}$$
$$t > t_{\text{crit}} \implies \text{Reject } H_0$$

where the critical value t_{crit} for significance level 0.05 is $t_{\text{crit}} = 1.96$.