Sources of Bias: Omitted Variables Measurement Error and Fixed Effects

1. **Remember OVB?**
2. **Heterogeneity Bias and Fixed Effects**
3. **Measurement Error as an omitted variable**

1. Remember Omitted Variable "Bias" ?

- Short regression

$y = b_0^s + b_1^s x_1 + e^S$ (SR)

- Long regression

$y = b_0^L + b_1^L x_1 + b_2^L x_2 + e^L$ (LR)

- Claim:

$b_1^s = b_1^L + b_2^L b_{21}$,

b_{21} is slope of a regression of x_2 on x_1

2. Heterogeneity Bias

- Mundlak (1961) was concerned with returns to scale in farming.
  Journal of Farm Economics, Vol. 43, No. 1 (Feb., 1961)
- Estimated a log-linear production function (i.e., Cobb-Douglas)
- If inputs and ability are complements (as implied by Cobb-Douglas), then high ability farmers use more inputs so that OVB is positive on input coefficients.
- So he used the fixed-effects regression to deal with OVB.

Returns to scale estimation

$Y = e^{\alpha} L^{\beta_L} K^{\beta_K} F^{\beta_F} A^{\beta_A} M$ M – Managerial ability

RETURNS TO SCALE $= \beta_L + \beta_K + \beta_K + \beta_A \overset{?}{\gtreqless} 1$

$\ln Y_{it} = \alpha + \beta_L \ln L_{it} + \beta_K \ln K_{it} + \beta_F \ln F_{it} + \beta_A \ln A_{it} + 1 M_i + \varepsilon_{it}$ $i = 1..N$ FARMS

↑ not observed.

Mundlak worried that M complements other inputs ⇒ a) correlation between M and other inputs.
b) coefficient on M is positive (=1)
} O.V.B.

1) $t = 1..T$ more periods

2) Assume $M_{it} = M_i \ \forall t$

"Assume M_i is FIXED OVER TIME"

FIXED EFFECTS

e.g. T=2 (2 P

Fixed effects in 2 periods

- T=2

$$Y_{it} = \beta_1 X_{1it} + \beta_2 X_{2it} + \dots \beta_k X_{kit} + \alpha_i + \varepsilon_{it}$$

$i = 1..N$

$t = 1, 2$

$j = 1..k$

$$Y_{i2} = \beta_1 X_{1i2} + \beta_2 X_{2i2} + \beta_j X_{ji2} \dots \beta_k X_{ki2} + \alpha_i + \varepsilon_{i2}$$

$$-\ (Y_{i1} = \beta_1 X_{1i1} + \beta_2 X_{2i1} + \dots \beta_j X_{ji1} + \dots \beta_k X_{ki1} + \alpha_i + \varepsilon_{i1})$$

$$\Delta Y_i = \beta_1 \Delta X_{1i} + \beta_2 \Delta X_{2i} + \dots \beta_j \Delta X_{ji} + \dots \beta_k \Delta X_{ki} + 0 + \Delta \varepsilon_i$$

ESTIMATING EQN FREE OF O.V. ; $\Delta Y_i = Y_{i2} - Y_{i1}$

Fixed Effects in general

The Fixed Effects Regression Model

The fixed effects regression model is

FIXED EFFECT

$$Y_{it} = \beta_1 X_{1,it} + \cdots + \beta_k X_{k,it} + \alpha_i + u_{it}, \qquad (8.12)$$

where $i = 1, \ldots, n$ and $t = 1, \ldots, T$, where $X_{1,it}$ is the value of the first regressor for entity i in time period t, $X_{2,it}$ is the value of the second regressor, and so forth, and $\alpha_1, \ldots, \alpha_n$ are entity-specific intercepts.

Equivalently, the fixed effects regression model can be written in terms of a common intercept, the X's, and $n - 1$ binary variables representing all but one entity:

$+ \theta_y$ ← could be "year effects"

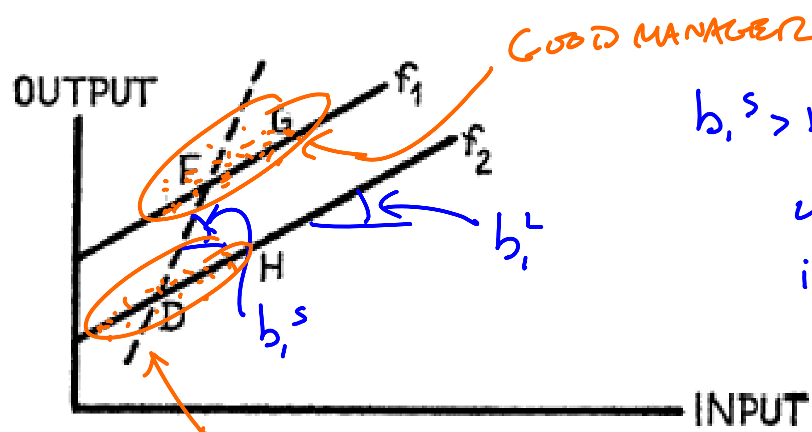
$$D2_i = \begin{cases} 1 & i = 2 \\ 0 & i \neq 2 \end{cases}$$

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \cdots + \beta_k X_{k,it} + \gamma_2 D2_i + \gamma_3 D3_i + \cdots + \gamma_n Dn_i + u_{it}, \qquad (8.13)$$

where $D2_i = 1$ if $i = 2$ and zero otherwise, and so forth.

Heterogeneity bias graphically

PRODUCTION FUNCTION FREE OF MANAGEMENT BIAS



NOT SO GOOD MANAGER FARMS

$Y \uparrow$

FIG. 1

$X_1 \rightarrow$

$b_1^s > b_1^L$ because of O.V.B. where omitted variable is M - managerial ability.

The variables included in the regression are:

Y = value of product
X_1 = number of labor days
X_2 = variable expense
X_3 = value of livestock at the beginning of the year
X_4 = value of livestock and poultry barns measured in I£. The value was derived by first measuring the capacity of the barns disregarding age differences. The capacity values were then multiplied by the market price.
X_i = amount of land on irrigated area basis (1 irrigated dunam = 4 dunams of dry land; 1 dunam = $\frac{1}{4}$ acre).

Heterogeneity bias – Managerial ability

TABLE 1

Item	X_1	X_2	X_3	X_4	X_5	Output
1. Geometric means	539	9900	1792	6621	255	18973
2. Estimated elasticities						Σb
a. Data pooled	.130	.692	.0043**	.103	.037	.967
b. Allowing for a year effect	.153	.679	.0042**	.101	.032*	.969
c. Allowing for a firm effect	.083*	.635	.0021**	.156	.002**	.878
d. Allowing for both year & firm effects	.115	.582	.005**	.100*	−.007**	.795
3. Absolute bias						
(a) Assumption of no year effect	.047	.057	.0022	−.053	.035	
(b) Allowing for year effect	.038	.097	−.0008	.001	.039	

C.S. Regression

Cannot reject Ho of C.R.S.

Decreasing R.T.S.
Sum of coefficients
$\Sigma b < 1$
Reject Ho of C.R.S.

Evidence of heterogeneity/managerial/ability bias. Or not. What about measurement error?

"Constant Returns to Scale" = ($\Sigma b = 1$)

Traffic Fatality Data

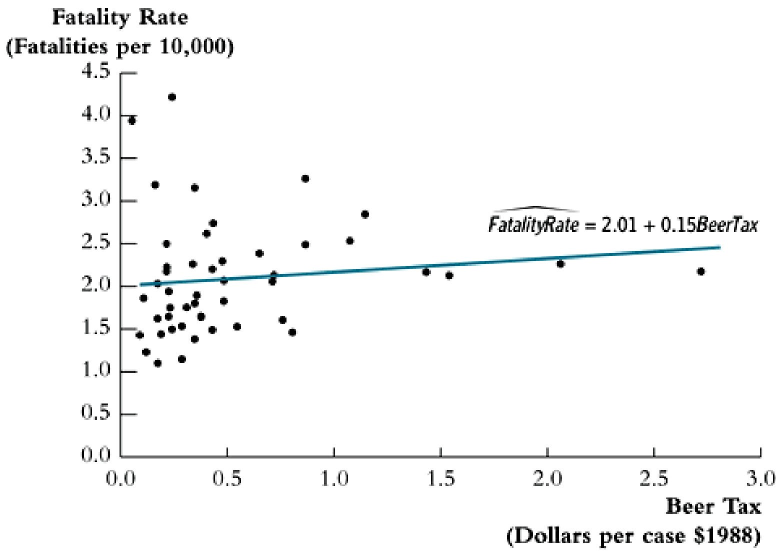
The State Traffic Fatality Data Set

The data are for the "lower 48" U.S. states (excluding Alaska and Hawaii), annually for 1982 through 1988. The traffic fatality rate is the number of traffic deaths in a given state in a given year, per 10,000 people living in that state in that year. Traffic fatality data were obtained from the U.S. Department of Transportation Fatal Accident Reporting System. The beer tax is the tax on a case of beer, which is a measure of state alcohol taxes more generally. The drinking age variables in Table 8.1 are binary variables indicating whether the legal drinking age is 18, 19, or 20. The two binary punishment variables in Table 8.1 describe the state's minimum sentencing requirements for an initial drunk driving conviction: "Mandatory jail?" equals one if the state requires jail time and equals zero otherwise, and "Mandatory community service?" equals one if the state requires community service and equals zero otherwise. Data on the total vehicle miles traveled annually by state were obtained from the Department of Transportation. Personal income was obtained from the U.S. Bureau of Economic Analysis, and the unemployment rate was obtained from the U.S. Bureau of Labor Statistics.

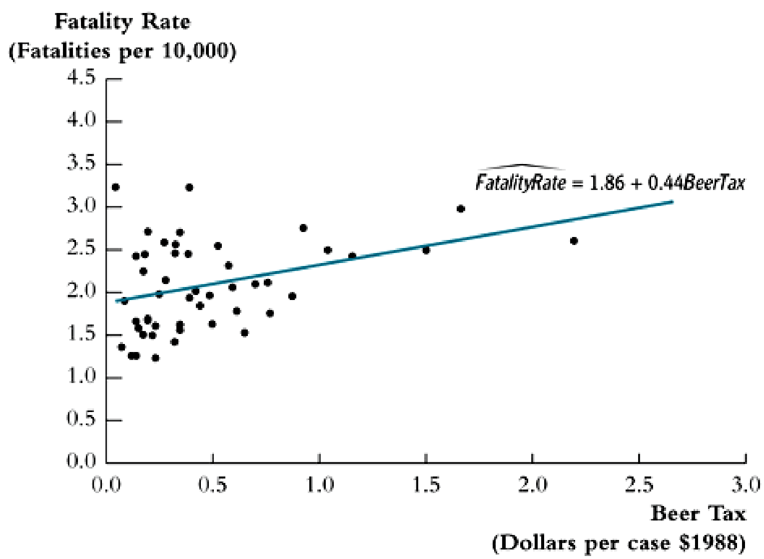
These data were graciously provided to us by Professor Christopher J. Ruhm of the Department of Economics at the University of North Carolina.

FIGURE 8.1 The Traffic Fatality Rate and the Tax on Beer

Panel a is a scatterplot of traffic fatality rates and the real tax on a case of beer (in 1988 dollars) for 48 states in 1982. Panel b shows the data for 1988. Both plots show a positive relationship between the fatality rate and the real beer tax.



(a) 1982 data



(b) 1988 data

The next two slides each present one half of Figure 8.1.

FIGURE 8.2 **Changes in Fatality Rates and Beer Taxes, 1982–1988**

This is a scatterplot of the *change* in the traffic fatality rate and the *change* in real beer taxes between 1982 and 1988 for 48 states. There is a negative relationship between changes in the fatality rate and changes in the beer tax.

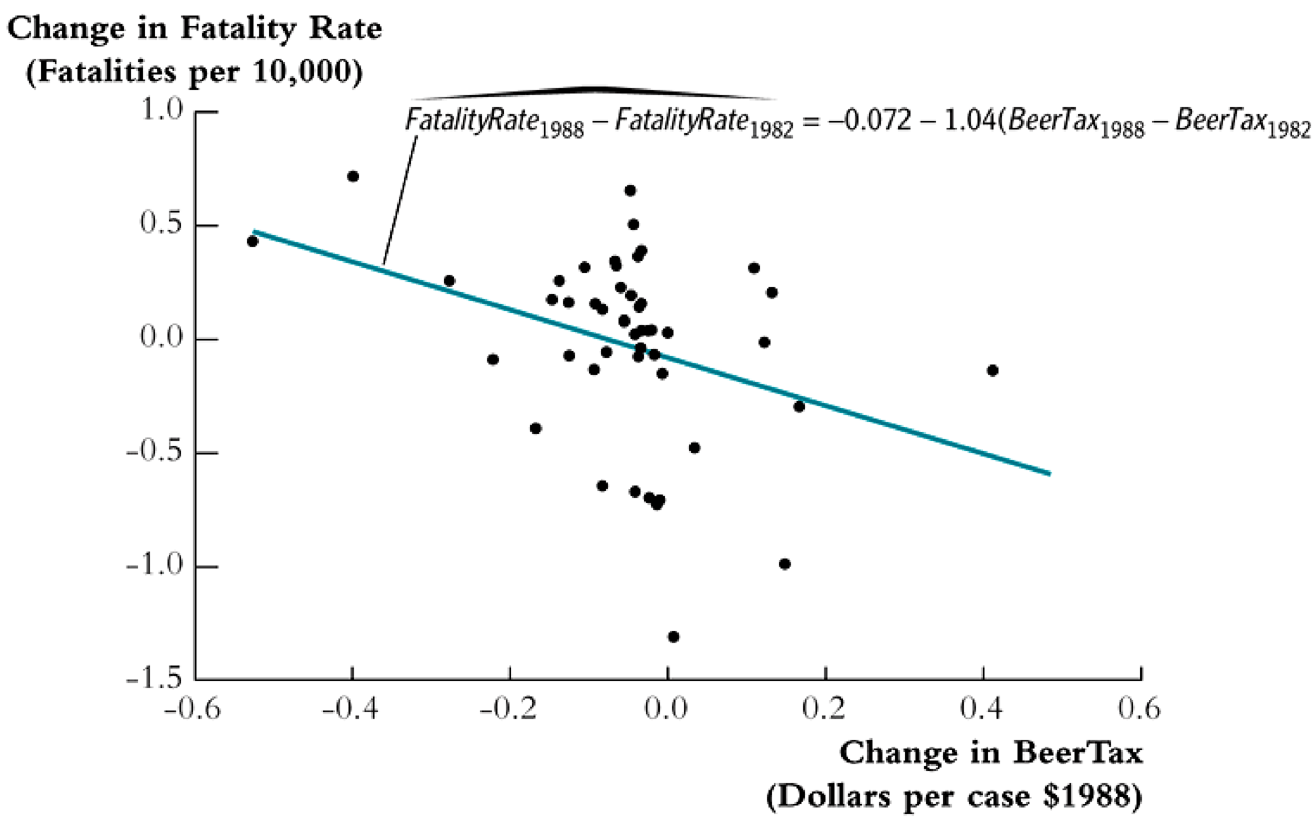


TABLE 8.1 Regression Analysis of the Effect of Drunk Driving Laws on Traffic Deaths

Dependent Variable: Traffic Fatality Rate (Deaths Per 10,000).

Regressor	(1)	(2)	(3)	(4)	(5)	(6)
Beer tax	0.36** (0.05)	−0.66** (0.20)	−0.64* (0.25)	−0.45* (0.22)	−0.70** (0.25)	−0.46* (0.22)
Drinking age 18				0.028 (0.066)	−0.011 (0.064)	
Drinking age 19				−0.019 (0.040)	−0.078 (0.049)	
Drinking age 20				0.031 (0.046)	−0.102* (0.046)	
Drinking age						−0.002 (0.017)
Mandatory jail?				0.013 (0.032)	−0.026 (0.065)	
Mandatory community service?				0.033 (0.115)	0.147 (0.137)	
Mandatory jail or community service?						0.031 (0.076)
Average vehicle miles per driver				0.008 (0.008)	0.017 (0.010)	0.009 (0.008)
Unemployment rate				−0.063** (0.012)		−0.063** (0.012)
Real income per capita (logarithm)				1.81** (0.47)		1.79** (0.45)
State effects?	no	yes	yes	yes	yes	yes
Time effects?	no	no	yes	yes	yes	yes
***F*-statistics and *p*-values Testing Exclusion of Groups of Variables:**						
Time effects = 0			2.47 (0.024)	11.44 (<0.001)	2.28 (0.037)	11.59 (<0.001)
Drinking age coefficients = 0				0.48 (0.696)	2.09 (0.102)	
Jail, community service coefficients = 0				0.17 (0.845)	0.59 (0.557)	
Unemployment rate, income per capita = 0				38.29 (<0.001)		40.12 (<0.001)
$\bar{R}^2$	0.090	0.889	0.891	0.926	0.893	0.926

These regressions were estimated using panel data for 48 U.S. states from 1982 to 1988 (336 observations total), described in Appendix 8.1. 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 *5% level or **1% significance level.

The next two slides each present one half of Table 8.1.

3. Measurement error bias

•

Measurement error bias as OVB

•