Financial Econometrics
Shanghai Jiao Tong University
Summer School – 2025
Empirical Exercise 5A
With Solutions

Fatih Kansoy

Reference: Stock, J.H. and Watson, M.W., Introduction to Econometrics (4th Edition)

Data: `Birthweight_Smoking.csv`

Empirical Exercise 7.1. Use the `Birthweight_Smoking` data set to answer the following questions. To begin, run three regressions:

(1) *Birthweight* on *Smoker*

(2) *Birthweight* on *Smoker*, *Alcohol*, and *Nprevist*

(3) *Birthweight* on *Smoker*, *Alcohol*, *Nprevist*, and *Unmarried*

(a) What is the value of the estimated effect of smoking on birth weight in each of the regressions?

(b) Construct a 95% confidence interval for the effect of smoking on birth weight, using each of the regressions.

(c) Does the coefficient on *Smoker* in regression (1) suffer from omitted variable bias? Explain.

(d) Does the coefficient on *Smoker* in regression (2) suffer from omitted variable bias? Explain.

(e) Consider the coefficient on *Unmarried* in regression (3).

(i) Construct a 95% confidence interval for the coefficient.

(ii) Is the coefficient statistically significant? Explain.

(iii) Is the magnitude of the coefficient large? Explain.

(iv) A family advocacy group notes that the large coefficient suggests that public policies that encourage marriage will lead, on average, to healthier babies. Do you agree? (*Hint:* Review the discussion of control variables in Section 6.8.)

(f) Consider the various other control variables in the data set. Which do you think should be included in the regression? Using a table like Table 7.1, examine the robustness of the confidence interval you constructed in (b). What is a reasonable 95% confidence interval for the effect of smoking on birth weight?

Solution:

Key Concepts

95% Confidence Interval:

$$\text{CI} = \hat{\beta} \pm 1.96 \times SE(\hat{\beta})$$

Omitted Variable Bias (OVB): Bias occurs when both conditions hold:

1. The omitted variable is correlated with an included regressor
2. The omitted variable is a determinant of Y

Control Variables: Variables included to reduce OVB, not necessarily for causal interpretation. Their coefficients may reflect selection effects, not causal impacts.

Data Overview

Variable	Description	Mean	SD	Min	Max
birthweight	Birth weight (grams)	3,382.9	592.2	425	5,755
smoker	Mother smoked (1=Yes)	0.194	0.395	0	1
alcohol	Mother drank (1=Yes)	0.019	0.138	0	1
nprevist	Prenatal visits	10.99	3.67	0	35
unmarried	Unmarried (1=Yes)	0.227	0.419	0	1
age	Mother's age	26.89	5.36	14	44
educ	Education (years)	12.91	2.17	0	17

$n = 3{,}000$. Source: 1989 Pennsylvania Natality-Mortality files.

Three Base Regressions

	Reg (1)	**Reg (2)**	**Reg (3)**
Smoker	-253.23^{***}	-217.58^{***}	-175.38^{***}
	(26.95)	(26.68)	(27.10)
Alcohol		-30.49	-21.08
		(76.23)	(75.61)
Nprevist		34.07^{***}	29.60^{***}
		(2.85)	(2.90)
Unmarried			-187.13^{***}
			(26.01)
Constant	3432.06^{***}	3051.25^{***}	3134.40^{***}
	(11.87)	(34.02)	(35.66)
R^2	0.029	0.073	0.089
$\bar{R}^2$	0.028	0.072	0.087

Standard errors in parentheses. $^{***}p < 0.01$

(a) Estimated Effect of Smoking

Answer:

- Regression (1): $\hat{\beta}_{\text{Smoker}} = -253.23$ grams
- Regression (2): $\hat{\beta}_{\text{Smoker}} = -217.58$ grams
- Regression (3): $\hat{\beta}_{\text{Smoker}} = -175.38$ grams

In all three regressions, smoking is associated with **lower birthweight**. The magnitude decreases as more controls are added, suggesting the simple regression may overstate the effect due to OVB.

(b) 95% Confidence Intervals for Smoking Effect

Using the formula $\hat{\beta} \pm 1.96 \times SE(\hat{\beta})$:

Regression	$\hat{\beta}$	SE	**95% CI Lower**	**95% CI Upper**
(1) Smoker only	-253.23	26.95	-306.1	-200.4
(2) + Alcohol, Nprevist	-217.58	26.68	-269.9	-165.3
(3) + Unmarried	-175.38	27.10	-228.5	-122.3

Answer: All three confidence intervals exclude zero, confirming the smoking effect is statistically significant at the 5% level in each specification. The intervals overlap substantially, suggesting the estimated effect is relatively stable.

(c) & (d) Omitted Variable Bias Analysis

Correlations with Smoker:

	Smoker	Alcohol	Nprevist	Unmarried
Smoker	1.000			
Alcohol	**0.121**	1.000		
Nprevist	−0.109	−0.043	1.000	
Unmarried	**0.238**	0.051	−0.233	1.000

(c) Does Regression (1) Suffer from OVB?

Answer: Yes, likely.
Check OVB conditions for Nprevist:

1. Corr(Smoker, Nprevist) $= -0.109 \neq 0$ ✓
2. Nprevist affects Birthweight: $\hat{\beta}_{\text{Nprevist}} = 34.07 > 0$ in Reg (2) ✓

Direction of bias:

$$\text{Bias} \approx \underbrace{\hat{\beta}_{\text{Nprevist}}}_{>0} \times \underbrace{\frac{\text{Cov}(\text{Smoker}, \text{Nprevist})}{\text{Var}(\text{Smoker})}}_{<0} < 0$$

The bias is **negative**, meaning the simple regression **overstates** the negative effect of smoking (makes it more negative than the true effect). This is consistent with $\hat{\beta}$ changing from -253 to -218 when controls are added.

(d) Does Regression (2) Suffer from OVB?

Answer: Yes, likely.
Check OVB conditions for Unmarried:

1. Corr(Smoker, Unmarried) $= 0.238 > 0$ ✓
2. Unmarried affects Birthweight: $\hat{\beta}_{\text{Unmarried}} = -187.13 < 0$ in Reg (3) ✓

Direction of bias:

$$\text{Bias} \approx \underbrace{\hat{\beta}_{\text{Unmarried}}}_{<0} \times \underbrace{\frac{\text{Cov}(\text{Smoker}, \text{Unmarried})}{\text{Var}(\text{Smoker})}}_{>0} < 0$$

Again, the bias is **negative**—Reg (2) still overstates the negative smoking effect. Adding Unmarried reduces $|\hat{\beta}|$ from 218 to 175 grams.

(e) Analysis of the Unmarried Coefficient

From Regression (3): $\hat{\beta}_{\text{Unmarried}} = -187.13$, $SE = 26.01$

(e)(i) 95% Confidence Interval

$$\text{CI} = -187.13 \pm 1.96 \times 26.01 = (-238.1, -136.2)$$

(e)(ii) Statistical Significance

$$t = \frac{-187.13}{26.01} = -7.20$$

Since $|t| = 7.20 > 1.96$, the coefficient is **statistically significant** at the 5% level (and at 1%).

(e)(iii) Magnitude

Answer: Yes, the magnitude is large. Being unmarried is associated with **187 grams lower birthweight**—comparable in magnitude to the smoking effect (−175 grams). Given average birthweight of 3,383 grams, this represents about a 5.5% reduction.

(e)(iv) Does This Imply Marriage Policies Would Improve Birth Outcomes?

Answer: No, this conclusion is not warranted.
The coefficient on Unmarried should be interpreted as a **control variable**, not a causal effect. The coefficient likely reflects **selection effects** and **omitted variables**:

- Unmarried mothers may differ systematically from married mothers in unobserved ways (income, stress, social support, health behaviors)
- Unmarried is correlated with other factors that affect birthweight
- The coefficient captures the *association* between marital status and birthweight, conditional on other controls—not the causal effect of marriage itself

A policy that "encourages marriage" would not necessarily change the underlying factors that drive the association. This is a classic example of confusing correlation with causation.
Section 6.8 insight: Control variables are included to reduce OVB in the coefficient of interest (Smoker), not because their own coefficients have causal interpretations.

(f) Robustness Analysis (Table 7.1 Style)

Regression	$\hat{\beta}_{\text{Smoker}}$	**SE**	**95% CI**	**Controls**
(1)	−253.23	26.95	(−306.1, −200.4)	None
(2)	−217.58	26.68	(−269.9, −165.3)	Alc, Nprevist
(3)	−175.38	27.10	(−228.5, −122.3)	+ Unmarried
(4)	−176.96	27.47	(−230.8, −123.1)	+ Age, Educ
(5)	−177.91	27.53	(−231.9, −124.0)	+ Young, Educ
(6)	−178.27	27.43	(−232.0, −124.5)	+ Tripre dummies

Answer: The smoking coefficient is **robust** to the inclusion of additional controls.

Key observations:

- The coefficient stabilizes around -175 to -178 grams once Unmarried is included
- Adding Age, Education, Young indicator, and Tripre dummies has minimal additional effect
- All 95% CIs overlap substantially and exclude zero

Reasonable 95% CI for smoking effect: Based on the robustness analysis, a reasonable confidence interval is approximately $(-230, -125)$ **grams**. The effect of smoking appears to reduce birthweight by roughly **150–230 grams**.

Summary

1. **Smoking Effect:** Consistently negative and significant across all specifications. Best estimate: -175 to -180 grams reduction in birthweight.
2. **OVB Pattern:** Simple regression overstates the smoking effect. Adding controls reduces the coefficient from -253 to -175, consistent with negative OVB.
3. **Unmarried:** Large, significant coefficient (-187 grams) but should not be interpreted causally. It serves as a control variable.
4. **Robustness:** The smoking coefficient is stable once key controls (Unmarried, prenatal care) are included. Additional demographic controls have minimal impact.