Economics 20: Econometrics

CONCEPTUAL EXERCISE #3

Part I. Multiple Choice Problems. For the following questions, circle all true responses.

1 and 2. Suppose that Vermont has passed a law requiring employers to provide 6 months of paid maternity leave. You are concerned that women's wages will drop in order to pay for this new benefit. You find a data set that samples men and women in Vermont and in New Hampshire and has information on wages. You pool 2 cross-sections, one from the year before the law took effect and one from the year after and find that the mean wage for various groups is as follows:

	New Hampshire		Vermont	
	Before	After	Before	After
Women	$9	$12	$8	$10
Men	$12	$14	$10	$12

1. Suppose you estimate the following model using only data from Vermont:

$wage = \beta_0 + \beta_1 after + \beta_2 women + \beta_3 after*women + u$

where *after* and *women* are dummy variables for the second period and being a woman respectively. Your estimate of β_3 will be:
a) 2
b) 0
c) -1
d) impossible to say

2. (Harder – but I bet you can figure it out!) Suppose instead you estimate the following model on all of the data:

$wage = \beta_0 + \beta_1 after + \beta_2 women + \beta_3 Vermont + \beta_4 after*women + \beta_5 after*Vermont + \beta_6 Vermont*women + \beta_7 after*women*Vermont + u$,

where *after* and *women* are as before and *Vermont* is a dummy variable for Vermont. Your estimate of β_7 will be:
a) 2
b) 0
c) -1
d) impossible to say

Part II. Interpreting Research Tables. There is a large amount of research on employer discrimination against minorities and women, much less on discrimination by other physical characteristics, such as weight. Christopher Carpenter investigated the impact of a 1993 court case *Cook v. Rhode Island*, which reinstated a nurse (Ms. Cook) who had been fired by her employer because of being obese. After this court decision, it essentially became illegal for employers to discriminate against obese people. Did this change raise employment of the obese? To find out, Kitt Carpenter obtained data on a large number of US workers both before and after the *Cook* decision.[1]

Doctors classify peoples' weights into five categories based on "body mass index" or BMI (a function of weight adjusted for height): **underweight, normal weight, overweight, obese, morbidly obese**. On the next page is a table of regression results from that paper, which uses repeated cross-section data on people's BMI and employment status. Each column gives an estimate of a different regression specification, with standard errors beneath each coefficient. The dependent variable is employ which is a dummy variable equal to 1 if you are employed, zero otherwise. The X's are:

> **Post** - a dummy variable equal to 1 for observations after the *Cook* decision (i.e., after discrimination against the obese became illegal)
> **Underweight, overweight, obese, morbid** - dummy variables for weight categories
> **Various interactions** - morbid*post, etc.
> **Other controls -** The "Yes"s at the bottom of the table indicate other groups of controls are included without the being shown in the table. Column (2) of the table adds linear controls for age and education, column (3) adds one race ("white") dummy, one marital status dummy, and column (4) adds 50 dummies for state.

a. Based on inspection of this table, what is the excluded weight category among the dummy variables included in the regressions? Interpret the coefficient on "underweight" in column (2), keeping in mind that there is also an "underweight*post" interaction included in the regression. (For this purpose outcome can be thought of as the "share employed.")

b. In column (1) the control for age was omitted from the regression, and as a result the coefficient on "underweight" was larger in magnitude (more negative). Age has a positive relationship with employment status. What is the sign of its relationship with "underweight?" (Hint: think about the omitted variables bias formula).

---

1 Carpenter, Christopher S. "The Effects of Employment Protection for Obese People." *Industrial Relations* 45(3): July 2006, pp. 393-415.

Table 1

Regression Table from Carpenter (2006)

Dependendent Variable: Employ = 1 if employed, 0 Otherwise

	(1)	(2)	(3)	(4)
Post	0.008	0.009	0.014	0.013
	(0.006)	(0.005)	(0.005)	(0.005)
Underweight	-0.098	-0.042	-0.036	-0.037
	(0.013)	(0.013)	(0.013)	(0.012)
Overweight	0.05	0.027	0.018	0.018
	(0.006)	(0.005)	(0.005)	(0.005)
Obese	0.024	0.002	-0.007	-0.008
	(0.009)	(0.008)	(0.008)	(0.007)
Morbid	-0.07	-0.093	-0.093	-0.093
	(0.030)	(0.031)	(0.030)	(0.028)
Underweight*Post	-0.002	-0.007	-0.008	-0.008
	(0.017)	(0.017)	(0.016)	(0.017)
Overweight*Post	-0.001	-0.001	-0.001	-0.001
	(0.008)	(0.007)	(0.007)	(0.006)
Obese*Post	0.0025	0.022	0.021	0.021
	(0.010)	(0.010)	(0.010)	(0.009)
Morbid*Post	0.021	0.022	0.023	0.024
	(0.038)	(0.037)	(0.037)	(0.037)
Controls For				
Age?		Yes	Yes	Yes
Education?		Yes	Yes	Yes
Marital Status?			Yes	Yes
Race?			Yes	Yes
50 State Dummies?				Yes
R^2	0.0065	0.0486	0.0579	0.0617
N	292,469	292,469	292,469	292,469

c. Using the column of your choice what is the "difference-in-differences" estimate of the court decision on employment rates for the obese and morbidly obese?

i. What is the key assumption underlying causal inference in a difference-in-difference estimator? Try to put that in words for these estimates.
ii. If you can answer i, what kind of evidence might you look at to validate that assumption?

d. Test whether the 50 “state dummies,” added in column (4), are jointly significant. There are a total of 64 independent variables included in column (4), including the 50 state dummies.

e. Embedded in these regressions is an example of a “placebo test.” What is it, and does the analysis “pass” the test?