PLSC 503: Problem Set 5 Solutions

Thad Dunning

Department of Political Science

Yale University

Spring 2010

1 Theoretical/Conceptual Exercises

Question 1: Interaction terms. Scholars sometimes advise the use of interaction models for testing conditional hypotheses. This exercise gives an example. Assume the following OLS model:

$$Y_i = a + bX_i + cZ_i + dX_iZ_i + \epsilon_i \tag{1}$$

for all $i = 1, ..., n$. Here, X_iZ_i is the product of X_i and Z_i; this is usually called an "interaction term." The usual OLS assumptions apply, e.g., $\epsilon_i \perp\!\!\!\perp (X_i, Z_i, X_iZ_i)$ and the ϵ_i are i.i.d. with $E(\epsilon) = 0$ and $\text{var}(\epsilon_i) = \sigma^2$ for all i.

(a) What is the marginal effect of intervening to change X_i with Z_i held fixed? That is, if

$$E(Y_i|X_i, Z_i) = a + bX_i + cZ_i + dX_iZ_i, \tag{2}$$

what is

$$\frac{\partial E(Y_i|X_i, Z_i)}{\partial X_i}? \tag{3}$$

And what is the marginal effect of intervening to change Z_i with X_i held fixed?

Solution: The marginal effect of intervening to change X_i with Z_i held fixed is

$$\frac{\partial E(Y_i|X_i, Z_i)}{\partial X_i} = b + dZ_i, \tag{4}$$

while the marginal effect of intervening to change Z_i with X_i held fixed is

$$\frac{\partial E(Y_i|X_i, Z_i)}{\partial Z_i} = c + dX_i. \tag{5}$$

(b) Now, suppose you run a regression of Y_i on a constant, X_i, Z_i, and X_iZ_i, obtaining

$$\hat{Y}_i = \hat{a} + \hat{b}X_i + \hat{c}Z_i + \hat{d}X_iZ_i \tag{6}$$

for all i. Here, $\hat{Y}_i$ is the fitted ("predicted") value of Y_i, $\hat{a}$ is the fitted intercept, and $\hat{b}$, $\hat{c}$, and $\hat{d}$ are the fitted slope coefficients. Under the assumptions of the model in (1), $\hat{a}$ estimates a, $\hat{b}$ estimates b, and so on.

What is the estimated marginal effect of intervening to change X_i with Z_i held fixed? How about the estimated effect of intervening to change Z_i with X_i held fixed?

Solution: Just replace the parameters in part (a) with the estimates. Then, the estimated marginal effect of intervening to change X_i with Z_i held fixed is

$$\frac{\partial \hat{Y}_i}{\partial X_i} = \hat{b} + \hat{d}Z_i, \tag{7}$$

while the estimated marginal effect of intervening to change Z_i with X_i held fixed is

$$\frac{\partial \hat{Y}_i}{\partial Z_i} = \hat{c} + \hat{d}X_i. \tag{8}$$

(c) Let $\frac{\partial \hat{Y}_i}{\partial X_i}$ be the estimated marginal effect of intervening to change X_i with Z_i held fixed. Express the variance of $\frac{\partial \hat{Y}}{\partial X_i}$ in terms of variances and covariances of the variables and coefficient estimators in equation (6). (For convenience, treat X_i and Z_i as fixed, rather than as random variables).

Solution: From (7),

$$\frac{\partial \hat{Y}_i}{\partial X_i} = \hat{b} + \hat{d}Z_i. \tag{9}$$

Thus,

$$\begin{aligned}\text{var}(\frac{\partial \hat{Y}_i}{\partial X_i}) &= \text{var}(\hat{b} + \hat{d}Z_i) \\ &= \text{var}(\hat{b}) + Z_i^2\text{var}(\hat{d}) + 2Z_i\text{cov}(\hat{b}, \hat{d}). \end{aligned} \tag{10}$$

where the second line of (10) follows from distributing the variance; here, Z_i factors out, because Z_i is fixed. If Z_i and X_i were random, we would condition on Z_i and X_i when taking the variance in (10).

(d) Suppose that after fitting equation (6), you find that

$$\hat{Y}_i = 1.2 + 2.3X_i + 0.5Z_i - 2.1X_iZ_i. \tag{11}$$

Moreover, the estimated variance-covariance matrix of $\hat{\beta} = (\hat{a}\ \hat{b}\ \hat{c}\ \hat{d})'$ is given by

$$\widehat{\text{cov}}(\hat{\beta}|X, Z, XZ) = \begin{pmatrix} 0.5 & 0.3 & 0.4 & 0.7 \\ 0.3 & 0.9 & 0.5 & -0.3 \\ 0.4 & 0.5 & 0.4 & 0.3 \\ 0.7 & -0.3 & 0.3 & 0.7 \end{pmatrix} \tag{12}$$

(i) Conduct a test of the null hypothesis that $b = 0$.

Solution: The estimated variance of $\hat{b}$ is the (2,2) element of the covariance matrix $\text{cov}(\hat{\beta}|X, Z, XZ)$. This is 0.9, so the estimated standard error is the square root: $\sqrt{0.9} \doteq 0.95$. The t-statistic is the absolute value of the coefficient estimate $\hat{b}$ divided by the estimated standard error:

$$t = |\frac{2.3}{0.95}| = 2.42 \tag{13}$$

So the t-test suggests that $\hat{b}$ is highly significant: we can reject the null hypothesis that $b = 0$.

(ii) Conduct a test of the null hypothesis that $d = 0$.

Solution: The estimated variance of $\hat{d}$ is the (4,4) element of $\widehat{\text{cov}}(\hat{\beta}|X, Z, XZ)$. This is 0.7, so the estimated standard error is the square root: $\sqrt{0.7} \doteq 0.84$. The t-statistic is the absolute value of $\hat{d}$ divided by the estimated standard error:

$$t = |\frac{-2.1}{0.84}| = 2.5. \tag{14}$$

So the t-test suggests that $\hat{d}$ is highly significant: we can reject the null hypothesis that $d = 0$.

(iii) Conduct a test of the null hypothesis that the marginal effect of intervening to change X_i, with Z_i held fixed at $Z_i = 1$, is zero.

Solution: The estimated marginal effect of intervening to change X_i is $\hat{b} + \hat{d}Z_i$ (see b above). So when $Z_i = 1$, the estimated marginal effect is $\hat{b} + \hat{d} = 2.3 - 2.1 = 0.2$. From (c), the variance of the estimated marginal effect is $\text{var}(\hat{b}) + Z_i^2\text{var}(\hat{d}) + 2Z_i\text{cov}(\hat{b}, \hat{d})$. Using the (2,2), (4,4), and (4,2)—or equivalently, the (2,4)—elements of $\widehat{\text{cov}}(\hat{\beta}|X, Z, XZ)$, we have

$$\begin{aligned}
\text{var}(\hat{b}) + Z_i^2\text{var}(\hat{d}) + 2Z_i\text{cov}(\hat{b}, \hat{d}) &= 0.9 + Z_i^2 0.7 - 2Z_i 0.3 \\
&= 0.9 + 0.7 - 2(0.3) \\
&= 1.0
\end{aligned} \tag{15}$$

where the second line follows from $Z_i = 1$. Thus, the estimated standard error is $\sqrt{1.0} = 1.0$. Finally, the t-statistic is the absolute value of the estimated marginal effect over the estimated standard error:

$$t = |\frac{0.2}{1.0}| = 0.2. \tag{16}$$

Thus, the estimated marginal effect is statistically insignificant: we cannot reject the null hypothesis that the marginal effect on Y_i of intervening to change X_i with Z_i held fixed at $Z_i = 1$ is zero.

(iv) Conduct a test of the null hypothesis that the marginal effect of intervening to change X_i, with Z_i held fixed at $Z_i = 10$, is zero.

Solution: As before, the estimated marginal effect of intervening to change X_i is $\hat{b} + \hat{d}Z_i$, so when $Z_i = 10$, the estimated marginal effect is $\hat{b} + 10\hat{d} = 2.3 - 10(2.1) = -18.7$. Again, the variance of the estimated marginal effect is $\text{var}(\hat{b}) + Z_i^2\text{var}(\hat{d}) + 2Z_i\text{cov}(\hat{b}, \hat{d})$. Here, with $Z_i = 10$, we have

$$\begin{aligned}
\text{var}(\hat{b}) + Z_i^2\text{var}(\hat{d}) + 2Z_i\text{cov}(\hat{b}, \hat{d}) &= 0.9 + Z_i^2 0.7 - 2Z_i 0.3 \\
&= 0.9 + (10)^2 0.7 - 2(10)(0.3) \\
&= 0.9 + 70 - 6 \\
&= 64.9 \qquad (17)
\end{aligned}$$

Thus, the standard error is $\sqrt{64.9} \doteq 8.06$. Finally, the t-statistic is the absolute value of the estimated marginal effect over the estimated standard error:

$$t = |\frac{-18.7}{8.06}| = 2.32 \qquad (18)$$

Thus, the estimated marginal effect is highly significant: we can reject the null hypothesis that the marginal effect on Y_i of intervening to change X_i, with Z_i held fixed at $Z_i = 10$, is zero.

(v) Comment on your results. Do they suggest that—given the model—the effect of X_i is conditional on Z_i?

Solution: The estimated coefficients $\hat{b}$ and $\hat{d}$ are both highly significant. However, this does not imply that the marginal effect of intervening to change X_i is different from zero. Indeed, this depends on the level of Z_i. When Z_i is negative (or small and positive), the overall marginal effect may be positive (or zero). When Z_i is large and positive, however, the estimated marginal effect is significantly negative. Thus, if the model is correct, the results do suggest that the effect of X_i is conditional on Z_i.

Question 2: Work exercise 21 in the discussion questions in Freedman (2009), Chapter 4 and then look at the data in the "answers to exercises" question on p. 252.

First, answer these questions. Which subjects are the "Always-Treats" in the control group? Which subjects are the "Never-Treats" in the invitation (assigned-to-treatment) group?

Now, estimate the effect of treatment on Compliers. (Show your work). Compare your results to the intention-to-treat analysis. If intention-to-treat analysis shows zero effect of treatment assignment, can the estimated effect of treatment on Compliers be non-zero? Why or why not?

Solution: The "Always-Treats" in the control group are the 1,122 men in the control group who were screened. The "Never-Treats" in the assigned-to-treatment group are the 23,785 men in the invitation group who were not screened.

The estimated effect of treatment on Compliers is

$$\frac{49 - 49}{0.236 - 0.073}. \tag{19}$$

The numerator of (19) is the death rate in assigned-to-treatment group, minus the death rate in the assigned-to-control group. This is the estimate of the intention-to-treat parameter. The denominator is the fraction of the assigned-to-treatment group that was treated ($\frac{7348}{31133} \doteq 0.236$) minus the fraction of the assigned-to-control group that was treated ($\frac{1122}{15353} \doteq 0.073$).

If the numerator of (19) is zero, then so is the whole expression: intention-to-treat analysis shows zero effect of treatment assignment, then estimated effect of treatment on Compliers will be zero as well.

2 Computer exercises

Code for the computer exercises is posted to the Classes v2 server (with the filename "ps 5.do").