PLSC 503: "Multivariate Analysis for Political Research"

Exercise Six

The topic of the day is models with multiplicative interactions. This exercise has the customary two parts.

Part I

Consider a model like:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + u_i, \tag{1}$$

where both Xs are continuous variables, and (to keep things simple) $N = 400$.

1. *Using simulations*, show that $\min \left[\widehat{\text{s.e.}(\hat{\beta}_1)} \right]$ (that is, the minimum value of the standard error of the "direct effect" coefficient on X_1) occurs when $\bar{X}_2 = 0$.

2. Next, suppose that $\beta_0 = 3$, $\beta_1 = 0.5$, $\beta_2 = 0.5$, and $\beta_3 = -0.5$, with $X_1 \sim U[-1, 1]$, $X_2 \sim U[-1, 1]$, $\text{Cov}(X_1, X_2) = 0$ and $u \sim$ i.i.d. $N(0, 1)$.

   (a) Again using simulations, illustrate and describe the bias in the estimates of β_0, β_1, β_3, and (correspondingly) ψ_1 that results from omitting the "direct effect" of X_2 – that is, from estimating a model of the form:

   $$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_3 X_{1i} X_{2i} + u_i \tag{2}$$

   Briefly describe why you found what you found.

   (b) Repeat the exercise in Part I.2.a, this time with $X_2 \sim U[2, 4]$. What differs? Why?

   (c) Repeat the exercise in Part I.2.a, this time increasing $\text{Cov}(X_1, X_2)$ *gradually*. (Hint: An easy way to do this is to generate:

   $$
   \begin{aligned}
   X_2 &= X_1 + e_i, \text{with} \\
   e_i &\sim \text{Uniform}[-\tau, \tau]
   \end{aligned}
   $$

   for various values of τ). What changes? Why?

Part II

The real-data part of this week's homework is also about interaction terms, and are brought to you by the 2024 American National Election Study. We have around 2,400 over-18 U.S. respondents in our data, who were surveyed immediately before the 2024 U.S. general election. The data we'll use contain the following variables:

- `ID` – A respondent ID variable

- `SexistRemarks` – The respondent's level of agreement with the following sentence: "Many women interpret innocent remarks or acts as being sexist," coded as:

  1 = Disagree strongly
  2 = Disagree somewhat

3 = Neither agree nor disagree

4 = Agree somewhat

5 = Agree strongly

- `GainPower` – The respondent's level of agreement with the following sentence: "Women seek to gain power by getting control over men," coded the same as `SexistRemarks`

- `HarrisTherm` – The respondent's 0-100 *feeling thermometer* rating of then-Vice President Kamala Harris.

- `TrumpTherm` – The respondent's 0-100 *feeling thermometer* rating of then-Former President Donald Trump.

- `RFemale` – a binary variable, coded 1 if the respondent self-identified as female, and 0 if male

- `RAge` – the respondent's age in years, on election day 2024 (note: all ages over 80 are coded as "80")

- `RConservative` – the respondent's self-placement on a seven point ideological scale, coded

1 = Extremely liberal

2 = Liberal

3 = Slightly liberal

4 = Moderate / middle-of-the-road

5 = Slightly conservative

6 = Conservative

7 = Extremely conservative

Using these data, select and answer **two** of the following questions:

1. Does the "gender gap" (the difference in responses between male and female respondents) in respondents' agreement with the `SexistRemarks` item vary with age, and if so, how and by how much?

2. How and to what extent the association between respondents' ideology and their responses on the `GainPower` item different for male respondents than for female respondents?

3. Model and describe empirically the interaction of age and ideology (conservatism) on respondents' "feeling thermometer" ratings of either Vice President Harris or former President Trump.

This homework is due (in the usual way) in electronic (PDF) form, at or before 11:59 p.m. EST on **Wednesday, March 19, 2025**. It is worth 50 possible points.