

## HOMEWORK 5

### MATH 463 - SPRING 2017

ALEX THIES

INTRODUCTION. I collaborated with Sienna Allen, Joel Bazzle, Torin Brown, Andy Heeszal, Ashley Ordway, and Seth Temple on this assignment.

#### 1. ASSIGNMENT

**Problem 1.** In this problem, you will replicate part of Blau and Duncans path model in figure 6.1 of Freedman (p. 82). Equation (6.3) explains sons occupation in terms of fathers occupation, sons education and sons first job. Variables are standardized. Correlations are given in table 6.1

- (a) Estimate the path coefficients in (6.3) and the standard deviation of the error term. How do your results compare with those in figure 6.1?
- (b) Compute SEs for the estimated path coefficients (Assume there are 20,000 subjects.)

*Solution.*

- (a) We proceed by following the process outlined in the text for determining the path coefficients of (6.1), and the standard deviation of the error term. In order to take advantage of the fact that the data is standardized<sup>1</sup> we translate the given model  $\mathbf{Y} = e\mathbf{U} + f\mathbf{X} + g\mathbf{W} + \eta$  into the language of linear algebra as follows,

$$\mathbf{Y} = \mathbf{M} \begin{pmatrix} e \\ f \\ g \end{pmatrix} + \eta,$$

where  $\mathbf{M} = (\mathbf{U} \quad \mathbf{X} \quad \mathbf{W})$ . Computing estimates for the coefficients boils down to solving the following equation

$$\begin{pmatrix} \hat{e} \\ \hat{f} \\ \hat{g} \end{pmatrix} = (\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{Y}.$$

Let's compute the two pieces of the above equation,  $\mathbf{M}'\mathbf{M}$ , and  $\mathbf{M}'\mathbf{Y}$ .<sup>2</sup> Note that  $\rho(i, j)$  corresponds to the  $(i, j)^{th}$  entry of the correlation matrix on page

---

<sup>1</sup>Thus, we can use the correlation matrix rather than the dataset itself

<sup>2</sup>We compute  $\mathbf{M}'\mathbf{M}$  by hand and let R do the heavy-lifting of inverting it.

83 of the text.

$$\begin{aligned}
\mathbf{M}'\mathbf{M} &= \begin{pmatrix} \mathbf{U} \\ \mathbf{X} \\ \mathbf{W} \end{pmatrix} (\mathbf{U} \quad \mathbf{X} \quad \mathbf{W}), \\
&= \begin{pmatrix} \langle \mathbf{U}, \mathbf{U} \rangle & \langle \mathbf{U}, \mathbf{X} \rangle & \langle \mathbf{U}, \mathbf{W} \rangle \\ \langle \mathbf{X}, \mathbf{U} \rangle & \langle \mathbf{X}, \mathbf{X} \rangle & \langle \mathbf{X}, \mathbf{W} \rangle \\ \langle \mathbf{W}, \mathbf{U} \rangle & \langle \mathbf{W}, \mathbf{X} \rangle & \langle \mathbf{W}, \mathbf{W} \rangle \end{pmatrix}, \\
&= \begin{pmatrix} \sum_i U_i^2 & \sum_i U_i X_i & \sum_i U_i W_i \\ \sum_i X_i U_i & \sum_i X_i^2 & \sum_i X_i W_i \\ \sum_i W_i U_i & \sum_i W_i X_i & \sum_i W_i^2 \end{pmatrix}, \\
&= n \begin{pmatrix} 1 & \rho(\mathbf{U}, \mathbf{X}) & \rho(\mathbf{U}, \mathbf{W}) \\ \rho(\mathbf{X}, \mathbf{U}) & 1 & \rho(\mathbf{X}, \mathbf{W}) \\ \rho(\mathbf{W}, \mathbf{U}) & \rho(\mathbf{W}, \mathbf{X}) & 1 \end{pmatrix}, \\
&= n \begin{pmatrix} 1 & 0.438 & 0.538 \\ 0.438 & 1 & 0.417 \\ 0.538 & 0.417 & 1 \end{pmatrix}. \\
(\mathbf{M}'\mathbf{M})^{-1} &= \begin{pmatrix} 7.63 \times 10^{-5} & -1.97 \times 10^{-5} & -3.28 \times 10^{-5} \\ -1.97 \times 10^{-5} & 6.56 \times 10^{-5} & -1.68 \times 10^{-5} \\ -3.28 \times 10^{-5} & -1.68 \times 10^{-5} & 7.46 \times 10^{-5} \end{pmatrix}.
\end{aligned}$$

$$\begin{aligned}
\mathbf{M}'\mathbf{Y} &= \begin{pmatrix} \mathbf{U}'\mathbf{Y} \\ \mathbf{X}'\mathbf{Y} \\ \mathbf{W}'\mathbf{Y} \end{pmatrix}, \\
&= \begin{pmatrix} \sum_i^n U_i Y_i \\ \sum_i^n X_i Y_i \\ \sum_i^n W_i Y_i \end{pmatrix}, \\
&= n \begin{pmatrix} \rho(\mathbf{U}, \mathbf{Y}) \\ \rho(\mathbf{X}, \mathbf{Y}) \\ \rho(\mathbf{W}, \mathbf{Y}) \end{pmatrix}, \\
&= n \begin{pmatrix} 0.596 \\ 0.405 \\ 0.541 \end{pmatrix}.
\end{aligned}$$

Hence, we compute the vector of estimated coefficients  $(\hat{e} \quad \hat{f} \quad \hat{g})'$  by taking the product of the two pieces which we have just constructed,

$$\begin{aligned}
\begin{pmatrix} \hat{e} \\ \hat{f} \\ \hat{g} \end{pmatrix} &= (\mathbf{M}'\mathbf{M})^{-1} \mathbf{M}'\mathbf{Y}, \\
&= \begin{pmatrix} 0.395 \\ 0.115 \\ 0.281 \end{pmatrix}.
\end{aligned}$$

A quick glance at the path model on page 82 of the text shows us that our estimates agree with those computed by the authors of the study.

To compute the standard deviation of  $\eta$  we again take advantage of the standardization of the data, in this case the fact that  $n^{-1} \sum_i \mathbf{Y}_i^2 = 1$ , allowing us to simply solve for  $\hat{\sigma}^2$ . Recall that  $\sum \mathbf{U}_i^2 = \sum \mathbf{X}_i^2 = \sum \mathbf{W}_i^2 = 1$ , that  $n^{-1} 2\hat{e} \sum \mathbf{U}_i \hat{\eta}_i = n^{-1} 2\hat{f} \sum \mathbf{X}_i \hat{\eta}_i = n^{-1} 2\hat{g} \sum \mathbf{W}_i \hat{\eta}_i = 0$ , and that  $n^{-1} \sum \mathbf{U}_i \mathbf{X}_i = \rho(\mathbf{U}, \mathbf{X})$ ,  $n^{-1} \sum \mathbf{U}_i \mathbf{W}_i = \rho(\mathbf{U}, \mathbf{W})$  and  $n^{-1} \sum \mathbf{X}_i \mathbf{W}_i = \rho(\mathbf{X}, \mathbf{W})$ .

$$\begin{aligned}
1 &= \frac{\sum \mathbf{Y}_i^2}{n}, \\
&= \frac{\sum (e\mathbf{U}_i + f\mathbf{X}_i + g\mathbf{W}_i + \eta)^2}{n}, \\
&= n^{-1} \left( \hat{e}^2 \sum \mathbf{U}_i^2 + \hat{f}^2 \sum \mathbf{X}_i^2 \right. \\
&\quad + \hat{g}^2 \sum \mathbf{W}_i^2 + 2\hat{e}\hat{f} \sum \mathbf{U}_i \mathbf{X}_i \\
&\quad + 2\hat{f}\hat{g} \sum \mathbf{X}_i \mathbf{W}_i + 2\hat{e}\hat{g} \sum \mathbf{U}_i \mathbf{W}_i \\
&\quad \left. + 2\hat{e} \sum \mathbf{U}_i \hat{\eta}_i + 2\hat{f} \sum \mathbf{X}_i \hat{\eta}_i + 2\hat{g} \sum \mathbf{W}_i \hat{\eta}_i \right), \\
&= \hat{e}^2 + \hat{f}^2 + \hat{g}^2 + 2\hat{e}\hat{f}\rho(\mathbf{U}, \mathbf{X}) \\
&\quad + 2\hat{f}\hat{g}\rho(\mathbf{X}, \mathbf{W}) + 2\hat{e}\hat{g}\rho(\mathbf{U}, \mathbf{W}) + \hat{\sigma}^2.
\end{aligned}$$

We solve for  $\hat{\sigma}^2$ .

$$\begin{aligned}
\hat{\sigma}^2 &= 1 - \hat{e}^2 - \hat{f}^2 - \hat{g}^2 - 2\hat{e}\hat{f}\rho(\mathbf{U}, \mathbf{X}) \\
&\quad - 2\hat{f}\hat{g}\rho(\mathbf{X}, \mathbf{W}) - 2\hat{e}\hat{g}\rho(\mathbf{U}, \mathbf{W}), \\
&= 1 - 0.394^2 - 0.115^2 - 0.281^2 \\
&\quad - 2(0.394)(0.115)(0.438) \\
&\quad - 2(0.115)(0.281)(0.417) \\
&\quad - 2(0.394)(0.281)(0.538) \\
&= 0.753.
\end{aligned}$$

- (b) Determining the standard errors for  $\hat{e}$ ,  $\hat{f}$ ,  $\hat{g}$  can be done by the following computation which follows from Theorem 4.3 in *Freedman*,  $\mathbf{SE} = \sqrt{\hat{\sigma}^2 n (\mathbf{M}' \mathbf{M})^{-1}}$ , where the standard errors for the estimated coefficients are found along the diagonal. Thus, we have the following standard errors,

$$\begin{aligned}
SE_{\hat{e}} &= 6.57 \times 10^{-3}, \\
SE_{\hat{f}} &= 6.10 \times 10^{-3}, \\
SE_{\hat{g}} &= 6.50 \times 10^{-3}.
\end{aligned}$$

□

**Problem 2.** In this problem, you will replicate Gibsons path diagram, which explains repression in terms of mass and elite tolerance (section 6.3 of *Freedman*). The correlation between mass and elite tolerance scores is 0.52; between mass tolerance scores and repression score, -0.26; between elite tolerance scores and repression scores, -0.42. (Tolerance scores were averaged within state.)

- (a) Compute the path coefficients in figure 6.2, using the method of section 6.1.

- (b) Estimate  $\sigma^2$ . Gibson had repression scores for all the states. He had mass tolerance scores for 26 states and elite tolerance scores for 26 states this will underestimate the SEs, by a bit-but you need to decide if  $p$  is 2 or 3.
- (c) Compute SEs for the estimates.
- (d) Compute the SE for the difference of the two path coefficients. You will need the off-diagonal element of the covariance matrix. Comment on the result.

*Solution.*

- (a) We will attack this problem in much the same way as Problem 1, in this case we solve  $\begin{pmatrix} \hat{a} & \hat{b} \end{pmatrix}' = (\mathbf{N}'\mathbf{N})^{-1}\mathbf{N}'\mathbf{R}$ , where  $\mathbf{N} = \begin{pmatrix} \mathbf{M} & \mathbf{E} \end{pmatrix}$  and  $\mathbf{R} = a\mathbf{M} + b\mathbf{E}$ . We compute the following,

$$\begin{aligned}
 \mathbf{N}'\mathbf{N} &= \begin{pmatrix} \mathbf{M} \\ \mathbf{E} \end{pmatrix} \begin{pmatrix} \mathbf{M} & \mathbf{E} \end{pmatrix}, \\
 &= \begin{pmatrix} \langle \mathbf{M}, \mathbf{M} \rangle & \langle \mathbf{M}, \mathbf{E} \rangle \\ \langle \mathbf{E}, \mathbf{M} \rangle & \langle \mathbf{E}, \mathbf{E} \rangle \end{pmatrix}, \\
 &= n \begin{pmatrix} 1 & \rho(\mathbf{M}, \mathbf{E}) \\ \rho(\mathbf{E}, \mathbf{M}) & 1 \end{pmatrix}, \\
 &= n \begin{pmatrix} 1 & 0.52 \\ 0.52 & 1 \end{pmatrix}. \\
 (\mathbf{N}'\mathbf{N})^{-1} &= \begin{pmatrix} 0.053 & -0.027 \\ -0.027 & 0.053 \end{pmatrix}.
 \end{aligned}$$

$$\begin{aligned}
 \mathbf{N}'\mathbf{R} &= \begin{pmatrix} \mathbf{M}'\mathbf{R} \\ \mathbf{E}'\mathbf{R} \end{pmatrix}, \\
 &= n \begin{pmatrix} \rho(\mathbf{M}, \mathbf{R}) \\ \rho(\mathbf{E}, \mathbf{R}) \end{pmatrix}, \\
 &= n \begin{pmatrix} -0.26 \\ 0.42 \end{pmatrix}.
 \end{aligned}$$

Now we compute the estimated coefficient vector,

$$\begin{aligned}
 \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} &= (\mathbf{N}'\mathbf{N})^{-1}\mathbf{N}'\mathbf{R}, \\
 &= \begin{pmatrix} -0.057 \\ -0.39 \end{pmatrix}.
 \end{aligned}$$

- (b) We compute the standard deviation of the error term in the same manner as in Problem 1 with the additional step of multiplying by  $n/n - p$  due to the small sample size. As mentioned in *Freedman*,  $p = 3$  in this case because we are estimating things about the intercept ‘behind the scenes’.

$$\begin{aligned}
 \hat{\sigma}^2 &= \left( 1 - \hat{a}^2 - \hat{b}^2 - 2\hat{a}\hat{b}\rho(\mathbf{E}, \mathbf{M}) \right) (n/n - p), \\
 &= \left( 1 - (-0.057)^2 - (-0.39)^2 - 2(0.057)(0.39)(0.52) \right) (26/23), \\
 &= 0.928.
 \end{aligned}$$

- (c) As might be expected, we proceed as we did for the corresponding task in Problem 1.

$$SE_{\hat{a}} = 0.221,$$

$$SE_{\hat{b}} = 0.221.$$

- (d) To compute the standard error of the difference we compute the following

$$\begin{aligned} SE_{\hat{b}-\hat{a}} &= \sqrt{\text{Var}(\hat{b}) + \text{Var}(\hat{a}) + \text{Cov}(\hat{a}, \hat{b})}, \\ &= \sqrt{0.0489 + 0.0489 + 2(-0.0254)}, \\ &= 0.2168. \end{aligned}$$

□

**Problem 3.** Consider the regression equations

$$(1) \quad Y_i = a + bX_i + \delta_i,$$

$$(2) \quad W_i = c + dX_i + eY_i + \varepsilon_i$$

If  $\varepsilon_i$  and  $\delta_i$  are correlated, are the OLS estimators of  $(c, d, e)$  unbiased? If yes, show it, otherwise, provide a counterexample.

*Solution.* We proceed via a counterexample. Suppose  $a = b = 0$ , and  $\delta_i$  and  $\varepsilon_i$  are correlated such that  $\delta_i = \gamma_i$ , and  $\varepsilon_i = \gamma_i + \psi_i$ . Additionally, suppose  $\gamma_i$ ,  $\psi_i$  are I.I.D. and let  $\mathbf{A} = (\mathbf{1} \quad \mathbf{X} \quad \mathbf{Y})$  so that  $\hat{\beta} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{W}$ . We compute the following,

$$\begin{aligned} \mathbb{E}[\hat{B}|A] &= \mathbb{E}[(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{W}] \\ &= (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbb{E}[\mathbf{W}] \\ &= (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbb{E}[\mathbf{A}\beta + \varepsilon_i] \\ &= (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbb{E}[\mathbf{A}\beta] + (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbb{E}[\varepsilon_i] \\ &= \beta + (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbb{E}[\varepsilon_i] \\ &= \beta + (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbb{E}[\gamma_i + \psi_i] \end{aligned}$$

Because  $\gamma_i$  and  $\psi_i$  are I.I.D. but not necessarily normal,

$$\mathbb{E}[\hat{B}|A] = \beta + (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbb{E}[\gamma_i + \psi_i] \neq \beta.$$

Thus, the OLS estimators are biased.

□

**Problem 4.** In the same set-up as problem 3, suppose that  $(a, b, c, d, e) = (1, 2, 1, 3, 2)$ , and that  $(X_i, \delta_i, \varepsilon_i)$  is multivariate Normal with mean vector  $(0, 0, 0)$  and covariance matrix

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{bmatrix}$$

Simulate 1000 times  $(X_i, Y_i, W_i, \delta_i, \varepsilon_i)$  according to the equations (1) in Problem 3. Note: The R package `mvtnorm` is useful for generating from the multivariate Normal distribution. Alternatively find  $\Sigma^{1/2}$  (e.g., using the spectral decomposition) and write

$$(X_i, \delta_i, \varepsilon_i)' = \Sigma^{1/2}(Z_{i,1}, Z_{i,2}, Z_{i,3})',$$

where  $(Z_{i,1}, Z_{i,2}, Z_{i,3})'$  are i.i.d.  $N(0, 1)$  random variables. Estimate the bias of the OLS estimators using the simulated data. How accurate is your estimate?

*Solution.* We run the requested simulation in R and determine that the bias for  $\hat{c}$ ,  $\hat{d}$ , and  $\hat{e}$  are  $-0.503$ ,  $-1.007$ , and  $0.503$ , respectively. To determine how accurate these estimates of the biases are we provide the following 95% confidence intervals,

$$CI_{\hat{c}} = (-0.507, -0.500),$$

$$CI_{\hat{d}} = (-1.013, -1.000),$$

$$CI_{\hat{e}} = (0.500, 0.506).$$

□

*E-mail address:* `athies@uoregon.edu`