HOMEWORK 5

Problem 1. In this problem, you will replicate part of Blau and Duncan's path model in figure 6.1 of Freedman (p. 82). Equation (6.3) explains son's occupation in terms of father's occupation, son's education and son's 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.)

Problem 2. In this problem, you will replicate Gibson's 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 σ^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 understante the SEs, by a bit-but you neeed 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 neeed the off-diagonal element of the convariance matrix. Comment on the result.

Problem 3. Consider the regression equations

$$Y_i = a + bX_i + \delta_i \tag{1a}$$

$$W_i = c + dX_i + eY_i + \varepsilon_i \tag{1b}$$

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

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?