Homework #2
Stochastic reading of structural equations

In this homework we will continue to consider the model of Homework #1,

$$
\begin{aligned}
W &= e_0 \\
X &= aW + e_1 \\
Z &= bX + cW + e_2 \\
Y &= dX + tZ + e_3
\end{aligned}
$$

We now assume that the error terms are not observed, but are correlated random variables, sampled from a normal distribution $f(e_0, e_1, e_2, e_3)$ with covariance matrix σ_e (unknown). We will further assume that we are given the expected values and the variance-covariance matrix, σ, of the observed variables X, W, Z and Y. (As in homework #1, the model is assumed to be correct, and the structural coefficient a, b, c, d, t are given to us precisely. So, sit back and relax – we will not consider problems of parameter estimation).

1. Express the covariance matrix, σ, of X, W, Z, Y as a function of the covariance matrix σ_e of e_0, e_1, e_2 and e_3. (This is standard in the literature)

2. Suppose we observe that $X = 1$; what would be our best predictions of W, Z and Y? (This is also standard in the literature, though somewhat more advanced – skip if you have not done this kind of questions before, but identify the structural and covariance elements that are needed for answering this question.)

3. Suppose we intervene and set the value of X to $X = 1$. What would be our best predictions of W, Z and Y? Express your answer in terms of the structural coefficients and the pre-intervention statistic of the observed variables. (This is not standard in the literature, make absolutely sure that you are confident with your answer and your method)

4. (For advanced students) Suppose we observe that $X = 1$, and immediately intervene and change X to $X = 2$. What would be

our best predictions of W, Z and Y (Assume that the error terms do not change between the observation and the intervention).

5. Suppose we intervene to set the value of X to $X = x$, and take a sample of W, Z and Y. Later, we re-set the value of X to $X = x + 1$, and take a another (independent) sample of W, Z and Y. Find the difference between our best predictions of W, Z, and Y in the two samples. (Hint 1: It might be helpful to think of e_0, e_1, e_2, e_3 as unknown but constant characteristics of an individual in the population, and of X, W, Z, and Y as measured behavioral features of such individual. Interventions should then be thought of as a policy changes applied uniformly to the entire population). (Hint 2: For convenient bookkeeping, it might be helpful to use the $do(x)$ notation and/or the models you constructed in Questions 14-15 of Homework #1)

6. Suppose we intervene to set the value of X and Z at $X = x$ and $Z = z$, respectively, and take a sample of Y. Later, we re-set the value of X to $X = x + 1$, while holding Z constant at z, and take a another sample of Y. Find the difference between our best predictions of Y in the two samples.

7. In view of your answer to question 6, formulate a general definition for a structural coefficient, say d, in term of hypothetical experiments.

8. First we intervene to set the value of X at $X = x$, and take a sample of Y. Second, we increase the setting of X to $X = x + 1$, and take a sample of Z. Thirdly, we hold the value of X constant at $X = x$, set Z to the value it had in the second sample, and take another sample of Y. Estimate the difference in Y between the first and third sample.

9. In view of your answers to questions 5, 6 and 8, formulate a general definition for total, direct and indirect effects in SEM models.

10. (Review) Consider the following structural equation:

$$X_0 = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + e, \tag{1}$$

and assume it is part of a larger SEM model M.

(a) Write a verbal interpretation for the empirical meaning of b_1 in Eq. (1). (For this you need to describe either an observation or an experiment whose outcome is predicted to be b1 if the model is correct. Alternatively, you can describe an experiment to test the claim: "$b_1 = 0.5$").

(b) Repeat question 10(a), but replace your verbal description with a mathematical formula

$$b_1 = ...$$

(c) Repeat question 10(a) assuming that Eq. (1) is a regression equation, not necessarily structural. (Recall: in a regression equation, the error term is assumed uncorrelated with any of the regressors (i.e., variables on the right hand side of the equation).

(d) Repeat question 10(c) using a mathematical formula.

$$b_1 = ...$$

(e) Consider the following interpretation (proposed in Hayduk (1986,p245)): "b_1 is the magnitude of the change in X_0 that would be predicted to accompany a unit change in X_1 with X_2 and X_3 left untouched at their original values." Write a mathematical formula that expresses this interpretation,

$$b_1 = ...$$

and construct a structural model that contains Eq. (1), in which this interpretation for b_1 is valid.

(f) Consider the following interpretation of b_1 (modified version of the one proposed in Hayduk (1986,p245)): "b_1 is the magnitude of the change in X_0 that would be predicted to accompany a unit change in X_1, if we find that X_2 and X_3 remain constant, at their original values." Write a mathematical formula that expresses this interpretation,

$$b_1 = ...$$

and construct a structural model that contains Eq. (1), in which this interpretation for b_1 is valid.

(g) Construct a structural model in which X_2 is affected by X_1 and in which the interpretation given in question 10(f) is valid.

(h) Construct structural models in which the interpretations given in questions 10(e) and 10(f) are not valid.

(i) (Advanced) Devise a general criterion for the class of structural models containing Eq. (1), in which the interpretation (for b_1) given in questions 10(f) is valid.