Homework Problems on Structural-Equation Models

Soc. 761

Fall 2014

1. Do the following problems from "Linear structural-equation models" (Ch. 4 in Fox, *Linear Statistical Models and Related Methods*, Wiley, 1984): 4.1 (pp. 230–232) (a)–(g) and (j), parts (i) and (ii); 4.2 (pp. 230–232); 4.5 (p. 234).

2. In addition: Consider the regression equation

$$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

   where the variables $y, x_1,$ and x_2 are expressed as deviations from their expectations (eliminating the regression constant from the model), and the errors ε are normally and independently distributed with constant variance. Suppose that it is not reasonable to assume that the errors ε are independent of the explanatory variables x_1 and x_2, ruling out OLS estimation of the model. Suppose, further, that there are *three* instrumental variables z_1, z_2, and z_3 that are reasonably assumed to be independent of ε but not of the explanatory variables.

   (a) Applying each of the instrumental variables, derive estimating equations for the model.

   (b) What problem for estimation arises from having three rather than two instrumental variables?

   (c) Is this problem insurmountable? How might you deal with it?

3. Do following problems from "Linear structural-equation models" (Ch. 4 in Fox, *Linear Statistical Models and Related Methods*, Wiley, 1984): 4.8 (p. 251) and 4.10 (p. 265), both for models (a)–(g) and (j) (you can simply show your conclusions for these two questions in a table); and either 4.11 (p. 265) or 4.14 (p. 265) (or optionally both).

   For problems 4.11 and 4.14, use the `sem` function in the `sem` package to estimate the models. Covariance matrices for the Rindfuss et al. and Lincoln data sets can be downloaded from the course web page (in the R script files `Rindfuss.R` and `Lincoln.R`); in both cases, the original data sets from which the covariances were computed are not available.

4. Data from a paper by Wheaton, Muthén, Alwin, and Summers ("Assessing reliability and stability in panel models," *Sociological Methodology*, 1977: 84-136) have become a staple for developing examples of latent-variable structural-equation models. The model shown in Figure 1, which uses the Wheaton et al. data, is from the LISREL manual (by Jöreskog and Sörbom).The variables in the model include scales to measure "anomia" and "powerlessness" (both conceived as indicators of a more general, latent variable, "alienation"); "education," measured as years of schooling, and the socio-economic index score ("SEI") of the respondent's occupation (both conceived as indicators of the respondent's socio-economic status, or "SES"). Anomia and powerlessness were measured both in 1967 and 1971; education and SEI were measured only in 1967. Wheaton et al.'s data were a sample comprising 932 respondents drawn from a rural region of Illinois.

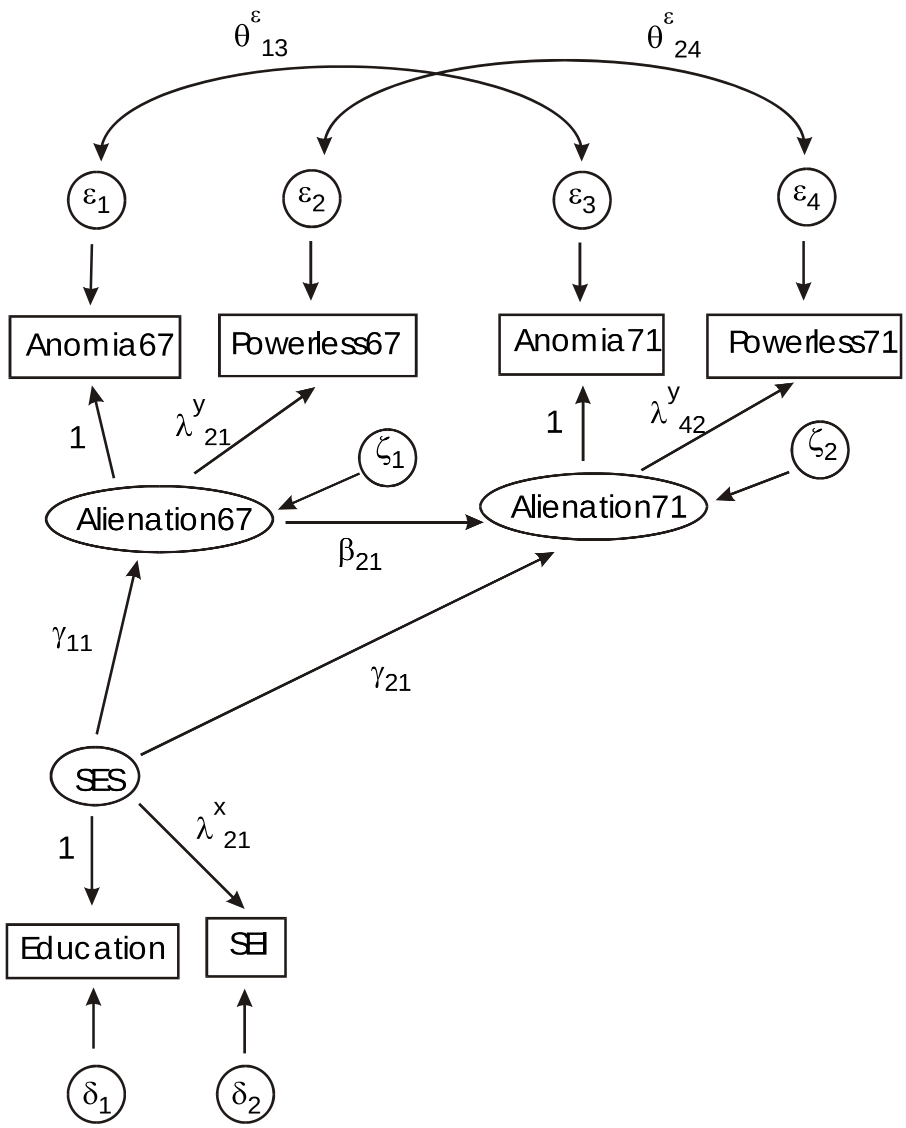

Figure 1: A latent-variable SEM for the Wheaton et. al. alienation data.

The variance-covariance matrix of the observed variables is as follows (and is also in the file `Wheaton.R` on the course web site):

	Anomia67	*Powerless67*	*Anomia71*	*Powerless71*	*Education*	*SEI*
Anomia67	11.834					
Powerless67	6.947	9.364				
Anomia71	6.819	5.091	12.532			
Powerless71	4.783	5.028	7.495	9.986		
Education	−3.839	−3.889	−3.841	−3.625	9.610	
SEI	−21.899	−18.831	−21.748	−18.775	35.522	450.288

(a) Write out the equations for the structural submodel and the measurement submodel of the SEM in Figure 1.

(b) Comment briefly on the specification of the model.

(c) Figuring out whether this model is identified is not simple, but the model is identified. It is simple, however, to check that the number of unknown parameters in the model is not greater than the number of variances and covariances among the observed variables. Perform this check, and, given that the model is identified, determine whether it is just-identified or overidentified.

Hints: In counting the parameters make sure that you include the variances of the measurement errors, the variances of the structural disturbances, and the variance of the latent exogenous variable; you should find that there are *four* fewer parameters than observable variances and covariances.

(d) Use the `sem` function to fit the model to the data and briefly comment on the results.

Hints: Be careful to include among the parameters to be estimated: the structural parameters; the variances of the measurement errors (these will by default be added automatically); those measurement-error covariances that are part of the model; the variances of the structural disturbances (also included by default); and the variance of the latent exogenous variable. Remember that in the RAM formulation of the model, measurement-error covariances are associated with observable variables; for example, the measurement-error variance $\theta^{\varepsilon}_{13}$ of ε_1 and ε_3 is specified as `Anomia67 <-> Anomia71` using `specifyModel()`, or as `C(Anomia67, Anomia71)` using the more convenient `specifyEquations()`. Similarly, the variance ψ_{11} of the disturbance ζ_1 is specified as `Alienation67 <-> Alienation67` or `V(Alienation67)`, though it is easier just to let `specifyModel()` or `specifyEquations()` add this variance to the model automatically.

(e) *[optional] Fit a version of the model in which the measurement-submodel parameters for Alienation are the same in 1967 and 1971. That is, impose the equality constraints $\lambda^{y}_{11} = \lambda^{y}_{32}$, $\theta^{\varepsilon}_{11} = \theta^{\varepsilon}_{33}$, and $\theta^{\varepsilon}_{22} = \theta^{\varepsilon}_{44}$. Check the equality constraints by performing a likelihood-ratio test comparing this version of the model with the previous version. What conclusions do you draw?