1

Mathematical Models and SEM

Chapter 7 described approaches to causal modeling and Chapter 8 described approaches to mathematical modeling. It turns out, the framework of structural equation modeling (SEM) that dominates causal modeling data analysis can be viewed as a form of mathematical modeling. The current primer explores the relationships between the two frameworks. We begin by describing how to represent an influence diagram as a set of equations. We then describe parameter estimation using limited information estimation instead of full information estimation as a first step to applying traditional mathematical modeling to structural equation analysis. Next, we translate the traditional linear equations into more general statements of functions that map onto the strategies discussed in Chapter 8. Finally, we describe some non-parametric approaches to estimation that might be used in future statistical work. We assume you have read Chapters 7, 8 and 11 and are familiar with the basics of SEM.

EXPRESSING CAUSAL MODELS MATHEMATICALLY

The example causal model we use is presented in the influence diagram in Figure 1.1.

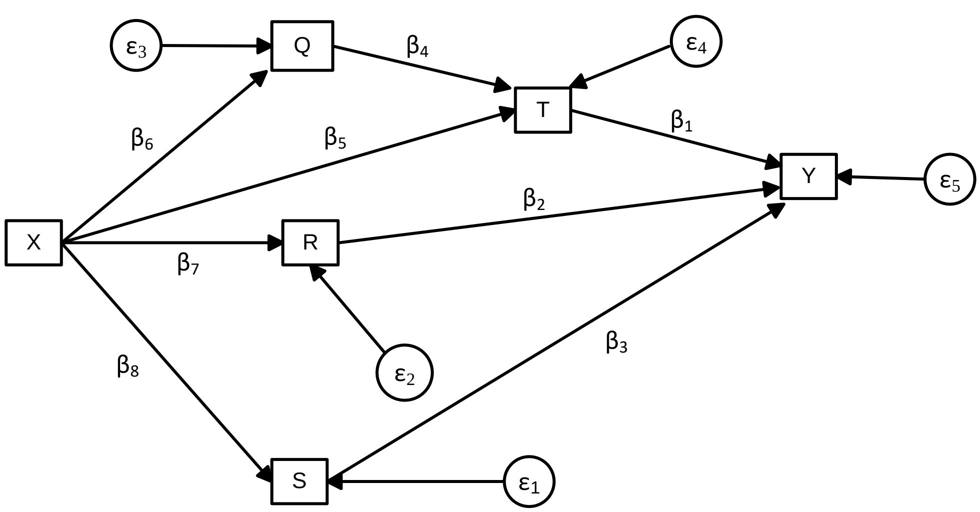

FIGURE 1.1. Causal Model

The model includes disturbance terms, indicating the equations we derive will be stochastic in nature. We use generic labels for the variables for ease of notation. We initially assume that all of the relationships are linear, which is a typical assumption in SEM applications. However, we relax this assumption later. Each endogenous variable is assumed to be a linear function of all variables that have an arrow pointing directly to it. The model, thus, can be expressed as a set of linear equations as follows:

$$Y = \alpha_1 + \beta_1 T + \beta_2 R + \beta_3 S + \varepsilon_5 \quad [1]$$

$$T = \alpha_2 + \beta_4 Q + \beta_5 X + \varepsilon_4 \quad [2]$$

$$Q = \alpha_3 + \beta_6 X + \varepsilon_3 \quad [3]$$

$$R = \alpha_4 + \beta_7 X + \varepsilon_2 \quad [4]$$

$$S = \alpha_5 + \beta_8 X + \varepsilon_1 \quad [5]$$

where α_1 through α_5 are adjustable constants representing intercepts, β_1 through β_8 are adjustable constants representing slopes, and ε_1 through ε_5 are disturbance terms.

The equations yield a model that is over-identified, although constraints must be introduced for estimating the parameters given the presence of disturbance terms (see Bollen, 1989).[1] The adjustable constants represented by the regression or path coefficients in this model reflect the predicted change in the outcome variable given a one unit change in the variable associated with the path coefficient, holding other determinants of the endogenous variable constant. In practice, data on each of the variables would be collected and the model would be fit to the data to determine if the model can account for the observed data. The data are used to estimate the values of the adjustable constants (intercepts, regression/path coefficients, error variances) so as to maximize model fit with the data. If the fit is reasonable, then values of the various adjustable constants are subjected to meaningful interpretation.

FULL INFORMATION ESTIMATION VERSUS LIMITED INFORMATION ESTIMATION

Standard SEM software packages approach model estimation using what is known as a *full information estimation* approach. A set of linear equations is defined for a model (as illustrated above) and the parameter estimates for all the equations are derived in one step taking into account the full system of equations simultaneously. Maximum likelihood criteria typically are used during estimation. An alternative approach is to use what is

[1] Specifically, the path coefficients from a disturbance term to its endogenous variable In Figure 1.1 are fixed at 1.0 so that disturbance variances can be estimated. This is a standard constraint in traditional multiple regression.

known as a *directed regression* or *limited information estimation* method. In this approach, one still works with the equations defined by the model but instead of estimating all parameters simultaneously, the parameters for each equation (or a subset of the equations) are estimated separately using techniques other than those in SEM software (Jaccard et al., 2006). In essence, we break up the model into parts and then estimate each part of the model separately using methods that may be better than applying traditional maximum likelihood methods to every equation. Both full information and limited information approaches have strengths and weaknesses. Although we can't delve into a formal comparison of the approaches here, key differences include:

1. In limited information estimation, the coefficients in an equation within the model are estimated without regard to the parameters in other equations in the model. In full information estimation approaches, the coefficients in all equations in the model are estimated simultaneously.

2. Maximum likelihood methods in SEM are based on asymptotic theory. As such, they require a sample size large enough to produce sampling distributions that approximate the theoretical sampling distributions assumed by asymptotic theory as N approaches infinity. The limited information approach does not require asymptotic theory and can be appropriate with small sample sizes (though statistical power with small N is an issue in both approaches).

3. Traditional SEM assumes multivariate normality among the variables (except in the case of fixed, categorical exogenous variables). Limited information estimation can relax these assumptions. Indeed, one can apply robust methods of regression to parameter estimation that circumvent the need for standard assumptions of normality and variance homogeneity and can protect against outliers in a theoretically informed way as well.

4. Traditional SEM can adjust for measurement error by using multiple indicators and latent variables. Limited information estimation relies on cruder approaches to accommodate measurement error.

5. Limited information estimation can take advantage of robust regression methods derived from the literature on robust statistics and methods appropriate for limited dependent variables. SEM is evolving similar approaches, but it is not nearly as far along as methods in robust statistics.

6. For correctly specified models, full information estimators are generally more efficient

than limited information estimators (in a technical sense of the term). Thus, traditional SEM can yield more efficient estimates than limited information estimation as long as the model is correctly specified.

7. Specification error in one part of the model in traditional SEM can reverberate through the model and affect other parameter estimates. In limited information estimation, the consequences of specification error in one part of the model are limited to that part of the model. Thus, specification error is more compartmentalized in limited information estimation.

8. For over-identified models, SEM methods provide indices of model fit separate from significance tests of the path coefficients (e.g., CFI RMSEA). In limited information estimation, model fit is restricted to tests involving the predicted statistical significance or non-significance of path/regression coefficients; there are no global fit indices.

9. In SEM, complex correlated error structures can be readily accommodated. In limited information estimation, strategies for dealing with correlated error are more restricted and somewhat more difficult to implement.

10. As discussed in the main text, directed regression can “mix” analytic strategies, using traditional multiple regression in those parts of the model where doing so is appropriate, multinomial logistic regression for those parts of the model where doing so is appropriate, and so on. Traditional SEM applies the same estimation algorithm to all parts of the model.

MATH MODELING AND SEM WITH LIMITED INFORMATION ESTIMATION

We can apply the math modeling approaches discussed in Chapter 8 to SEM by using limited information estimation strategies. We do not assume linearity. Using stochastic models (which include the disturbances within the functions, but which we omit here for the sake of pedagogy), we rewrite equations 1 through 5 as

$$Y = f\,(T, R, S) \qquad [6]$$

$$T = f\,(Q, X) \qquad [7]$$

$$Q = f\,(X) \qquad [8]$$

$$R = f\,(X) \qquad [9]$$

$$S = f\,(X) \qquad [10]$$

The task of the math modeler is to creatively apply the strategies in Chapter 8 to isolate viable functions for each equation, be they linear or non-linear in character. As well, the modeler may need to include functions for correlated errors across equations, which can occur in multi-equation models. Note that the function for any given equation can include interaction (moderation) dynamics between variables specified by the function, as appropriate.

NON-PARAMETRIC SEM

Judea Pearl (2012) has been an advocate of what he calls non-parametric SEM. By non-parametric, he essentially refers to the use of equations 6 to 10. He recognizes that there may be scenarios where the types of functions discussed in Chapter 8 may be applicable, but he prefers to use frameworks grounded in probability theory coupled with specialized notation that traditional probability theory lacks. For a wealth of information about his approach, see his web page at http://bayes.cs.ucla.edu/jp_home.html. Another potential useful semi-nonparametric approach is the use of multi-predictor running interval smoothers as well as the general additive model on a per function basis (see Wilcox, 2017).

CONCLUDING COMMENTS

Structural equation modeling can be pursued from the framework of math modeling. To be sure, there are challenges in doing so which we do not consider here (e.g., modeling latent variables), but it certainly has promise. We encourage those of you with such interests to pursue this line of work, especially in the spirit of the non-parametric modeling framework advocated by Pearl (2012).

REFERENCES

Bollen, K. (1989). *Structural equations with latent variables*. New York: Wiley.

Jaccard, J., Guilamo-Ramos, V. Johansson, M. & Bouris, A. (2006). Multiple regression analyses in clinical child and adolescent psychology. *Journal of Clinical Child and Adolescent Psychology*, 35, 456-479.

Pearl, J. (2012). The causal foundations of structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling*. (pp. 68-91). New York: Guilford Press.

Wilcox, R. (2017). *Introduction to robust estimation and hypothesis testing*. San Diego: Academic Press (Fourth edition).