R3: Graphical fomulation of SEMs

Yaroslav Mukhin

Feb 21, 2015

Abstract. Structural equations models attempt to capture a causal relationship between modeled variables. Regression models simply summarize statistical properties of random variables. Although it is common to write both types of models as linear equations with additive noise terms, and use orthogonality to identify the decomposition, the interpretation of the coefficients and the set of assumptions (often implicit) behind the two types of models are *completely* different. Here we formulate linear SEMs using a graph of causal/stractural relationships that underpin these models.

1. Angrist-Krueger example

Recall our working example of IV identification:

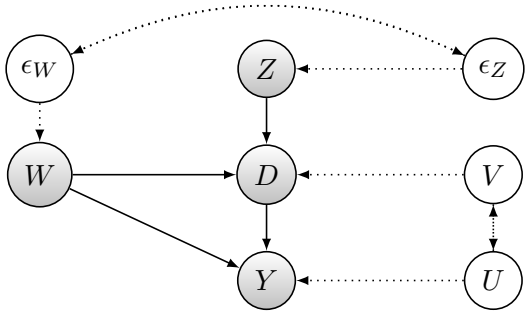
$$Y = \alpha_1 D + \alpha_2 W + U, \quad U \perp W, Z \tag{1}$$

$$D = \beta_1 Z + \beta_2 W + V, \quad V \perp Z, W \tag{2}$$

here Y measures wages, D measures level of education, Z is quarter of birth indicator, W is a vector of observable variables that we wish to include in our analysis (for simplicity I will treat W as a real valued random variable). Our aim is to estimate the effect of education on wages, specifically we would like to use the structural coefficient α_1 to make the following

Statement: A unit increasing in D has the effect of increasing Y by α_1 units.

This statement cannot be made based only on the axioms of probability theory. Question: If one were to define α_1 and the rest of eq. (1) as regression, what prevents us from making the above claim, at least 'on average'? The real problem is not even in how we interpret the *regression* coefficient α_1 [1], but rather that statement does not make sense. A regression equation is purely a summary of the distribution of the data, and increasing D to $D+1$ means *changing* this distribution. In order to proceed, we must postulate a model of how the joint distribution of (Y, D, Z, W) will change. Such mechanism is in the background of every SEM, and we can use a graph to make it explicit:



Each node (or vertex) of the graph represents a random variable, shaded if the variable is observed, clear if the variable is unobserved. Unobserved or exogenous random variables, $U, V, \epsilon_Z, \epsilon_W$, are not explained within the model, but rather summarize everything that is not modeled explicitly. We take their joint distribution P as given by Nature, but also make some assumptions about it. Bidirected edges (and their absence) of the graph, e.g. $U \leftrightarrow V$, capture these assumptions about

[1] $\|\cdot\|_2$-best linear prediction of $Y \ldots$

P. In the A-K example these take the form of orthogonality conditions[2]. Specifically, we allow U and V, to be correlated, same for ϵ_Z and ϵ_W, and assume that $\epsilon_Z, \epsilon_W \perp U, V$. Although these look similar to orthogonality conditions for a regression equation, the latter are obtained by construction, where as the former are indeed assumptions about P.

Observed or endogenous variables, Y, D, Z, W are completely determined within the model. Directed edges (and their absence) of the graph, e.g. $D \to Y$, represent *functional* relationships. For example, in the A-K graph

$$\begin{aligned} Y &= f_Y(D, W, U) & Z &= f_Z(\epsilon_Z) \\ D &= f_D(Z, W, V) & W &= f_W(\epsilon_W) \end{aligned}$$

In particular, if two different realizations of $(U, V, \epsilon_Z, \epsilon_W)$ have the same values of U and ϵ_W, and produce the same value of D according to f_D, then the observed value of Y must be the same. In our working example these functions are assumed to be linear.

In other words, the graph provides a complete analytic description of how endogenous variables are determined. Moreover, if the underlying joint distribution of exogenous variables changes from P to P', the new distribution of endogenous variables must be determined according to the same functional relationships. Under this set of structural assumptions we may make the above Statement. See Pearl [3] for more on causality.

Starting from the graph, we may reconstruct the system of linear structural equations:

$$\begin{aligned} Y &= \alpha_1 D + \alpha_2 W + U & Z &= \epsilon_Z \\ D &= \beta_1 Z + \beta_2 W + V & W &= \epsilon_W \end{aligned}$$

Noting the fact that linearity is crucial in this step, we can solve the above system for the vector of endogenous variables:

$$\begin{aligned} Y &= \underbrace{\alpha_1\beta_1}_{=:\gamma_1}\epsilon_Z + (\alpha_1\beta_2 + \alpha_2)\epsilon_W + \underbrace{U + \alpha_1 V}_{=:\varepsilon_1} & Z &= e_Z \\ D &= \beta_1\epsilon_Z + \beta_2\epsilon_W + V & W &= e_W \end{aligned}$$

and observe that the orthogonality assumptions on the distribution P of exogenous variables imply that γ_1 is the regression coefficient in projection of Y onto $\text{span}(Z, W)$, and β_1 is the regression coefficient of orthogonal projection of D onto $\text{span}(Z, W)$. This implies that we can identify

$$\alpha_1 = \frac{\gamma_1}{\beta_1}$$

provided that $\beta_1 \neq 0$.

2. Linear SEMs

Here we sketch a general graph-theoretic formulation of linear SEMs with orthogonality restrictions. We follow Drton, Foygel, and Sullivant [1] and Fox, Kaufl, and Drton [2]. Let $\{Y_j \ ; \ j = 1, \ldots, m\}$ denote all the observed variables we wish to model. Each of these variables is identified with a vertex of a mixed graph $G = (V, D, B)$, so $V = \{1, .., m\}$. There are two sets of edges

[2] A stronger alternative is independence. Orthogonality combines well with linear models like the A-K example, whereas independence becomes necessary in non-linear models.

$B, D \subset V \times V$. The edges in D are directed and represent structural/causal dependences among the nodes. There are no self-loops in D and we will assume that there are no directed cycles. No explicit reference is made to the unobserved exogenous variables (error terms), but their presence in the model is encoded through the set of biderected edges B.

Let $\mathbb{R}^D$ denote the set of matrices $\Lambda = (\lambda_{ij}) \in \mathbb{R}^{m \times m}$ with entries $\lambda_{ij} = 0$ whenever $i \to j$ in not in D. These are all the possible linear structural relationships among the endogenous variables. Let $PD(B)$ denote the set of all positive definite $m \times m$ matrices $\Omega = (\omega_{ij})$ with $\omega_{ij} = 0$ whenever $i \neq j$ and $i \leftrightarrow j$ in not in B. These are all the possible correlation structures among the error terms.

Definition. The linear structural equation model $\mathcal{M}(G)$ induced by an acyclic mixed graph $G(V, D, B)$ is the family of all probability distributions L on $\mathbb{R}^m$ of the form

$$L = P \circ (I - \Lambda) * \delta_{(I-\Lambda)^{-1} l_0}$$

where P is the joint distribution of the error terms $(\epsilon_1, \ldots, \epsilon_m)$, assumed to be centered and have covariance structure Ω. In particular, the implied covariance matrix of the endogenous variables is

$$\Sigma = (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1}.$$

The graphical model $\mathcal{M}(G)$ corresponds to the system of linear equations

$$Y_j = l_{0j} + \sum_i \lambda_{ij} Y_i + \epsilon_i, \quad j = 1, \ldots, m,$$

which can be solved for $Y = (I - \Lambda)^{-1}(l_0 + \epsilon)$ using our assumption that the graph is acyclic.

The following graph induces our working example:

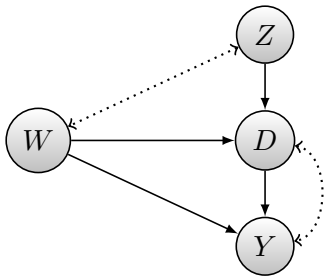


Figure 1: Graph that induces the A-K linear SEM.

References

[1] M. Drton, R. Foygel, and S. Sullivant. "Global identifiability of linear structural equation models". *The Annals of Statistics* 39.2 (2011), pp. 865–886.

[2] C. Fox, A. Kaufl, and M. Drton. "On the causal interpretation of acyclic mixed graphs under multivariate normality". *arXiv:1308.3728* (2013).

[3] J. Pearl. "Causal inference in statistics: An overview". *Statistics Surveys* 3 (2009), pp. 96–146.

14.382 Econometrics
Spring 2017