[21] Wold, S. (2001). Personal memories of the early PLS development, *Chemometrics and Intelligent Laboratory Systems*, **58**, 83–84.

PAUL D. SAMPSON AND FRED L. BOOKSTEIN

Path Analysis and Path Diagrams

History

Path analysis was first proposed by Sewall Wright [12] as a method for separating sources of variance in skeletal dimensions of rabbits (see [10, 11], for historical accounts). He extended this method of *path coefficients* to be a general method for calculating the association between variables in order to separate sources of environmental and genetic variance [14]. Wright was also the first recognize the principle on which path analysis is based when he wrote, 'The correlation between two variables can be shown to equal the sum of the products of the chains of path coefficients along all of the paths by which they are connected' [13, p. 115].

Path analysis was largely ignored in the behavioral and social sciences until Duncan [3] and later Goldberger [4] recognized and brought this work to the attention the fields of econometrics and psychometrics. This led directly to the development of the first modern **structural equation modeling** (SEM) software tools ([5] (*see* **Structural Equation Modeling: Software**)).

Path diagrams were also first introduced by Wright [13], who used them in the same way as they are used today; albeit using drawings of guinea pigs rather than circles and squares to represent variables (see [6], for an SEM analysis of Wright's data). Duncan argued that path diagrams should be '. . .isomorphic with the algebraic and statistical properties of the postulated system of variables. . .' ([3], p. 3). Modern systems for path diagrams allow a one-to-one translation between a path diagram and computer scripts that can be used fit the implied structural model to data [1, 2, 9].

Path Diagrams

Path diagrams express the regression equations relating a system of variables using basic elements such as squares, circles, and single- or double-headed arrows (see Figure 1). When these elements are correctly combined in a path diagram the algebraic relationship between the variables is completely specified and the predicted covariance matrix between the measured variables can be unambiguously calculated [8]. The (Reticular Action Model) RAM method is discussed below as a recommended practice for constructing path diagrams.

Figure 2 is a path diagram expressing a bivariate regression equation with one outcome variable, y, such that

$$y = b_0 + b_1 x + b_2 z + e \tag{1}$$

where e is a residual term with a mean of zero. There are four single-headed arrows, b_0, b_1, b_2, and b_3 pointing into y and likewise four terms are added together on the right hand side of 1. In a general linear model, one would use a column of ones to allow the estimation of the intercept b_0. Here a constant variable is denoted by a triangle that maps onto that column of ones. For simplicity of presentation, the two predictor variables x and z in this example have means of zero. If x and z had

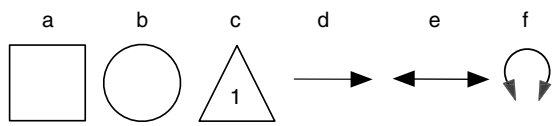

Figure 1 Graphical elements composing path diagrams. a. Manifest (measured) variable. b. Latent (unmeasured) variable. c. Constant with a value of 1. d. Regression coefficient. e. covariance between variables. f. Variance (total or residual) of a variable

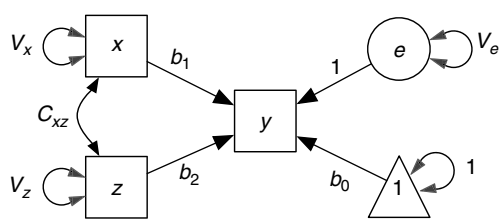

Figure 2 Bivariate regression expressed as a path diagram (*see* **Multivariate Multiple Regression**)

nonzero means, then single-headed arrows would be drawn from the triangle to x and from the triangle to z.

There are double-headed variance arrows for each of the variables x, z, and e on the right hand side of the equation, representing the variance of the predictor variables, and residual variance respectively. The constant has, by convention, a nonzero variance term fixed at the value 1.0. While this double-headed arrow may seem counterintuitive since it is not formally a variance term, it is required in order to provide consistency to the path tracing rules described below. In addition, there is a double-headed arrow between x and z that specifies the potential covariance, C_{xz} between the predictor variables.

Multivariate models expressed as a system of linear equations can also be represented as path diagrams. As an example, a factor model (*see* **Factor Analysis: Confirmatory**) with two correlated factors is shown in Figure 3. Each variable with one or more single-headed arrows pointing to it defines one equation in the system of simultaneous linear equations. For instance, one of the six simultaneous equations implied by Figure 3 is

$$y = b_1 F_1 + u_y, \tag{2}$$

where y is an observed score, b_1 is a regression coefficient, F_1 is an unobserved common factor score, and u_y is an unobserved unique factor. In order to identify the scale for the factors F_1 and F_2, one path

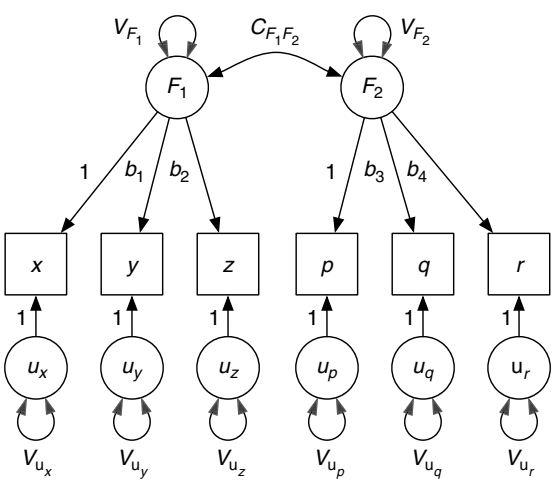

Figure 3 Simple structure confirmatory factor model expressed as a path diagram

leading from each is fixed to a numeric value of 1. All of the covariances between the variables in path diagrams such as those in Figures 1 and 3 can be calculated using rules of path analysis.

Path Analysis

The predicted covariance (or correlation) between any two variables v_1 and v_2 in a path model is the sum of all *bridges* between the two variables that satisfy the form shown in Figure 4 [7]. Each bridge contains one and only one double-headed arrow. From each end of the double-headed arrow leads a sequence of zero or more single-headed arrows pointing toward the variable of interest at each end of the bridge. All of the regression coefficients from the single-headed arrows in the bridge as well as the one variance or covariance from the double-headed arrow are multiplied together to form the component of covariance associated with that bridge. The sum of these components of covariance from all bridges between any two selected variables v_1 and v_2 equals the total covariance between v_1 and v_2.

If a variable v is at both ends of the bridge, then each bridge beginning and ending at v calculates a component of the variance v. The sum of all bridges between a selected variable v and itself calculates the total variance of v implied by the model.

As an example, consider the covariance between x and y predicted by the bivariate regression model shown in Figure 2. There are two bridges between x and y. First, if the double-headed arrow is the variance V_x, then there is a length zero sequence of single-headed arrows from one end of V_x and pointing to x and a length one sequence single-headed arrows leading from the other end of V_x and pointing to y. This bridge is illustrated in Figure 5-a and leads to a covariance component of $V_x b_1$, the product of the

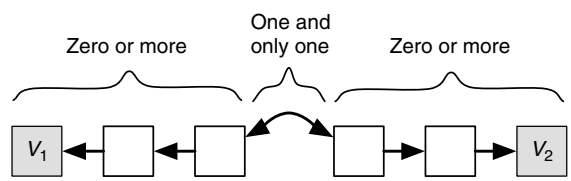

Figure 4 Schematic of the rule for forming a bridge between two variables: exactly one double-headed arrow with zero or more single-headed arrows pointing away from each end towards the selected variable(s)

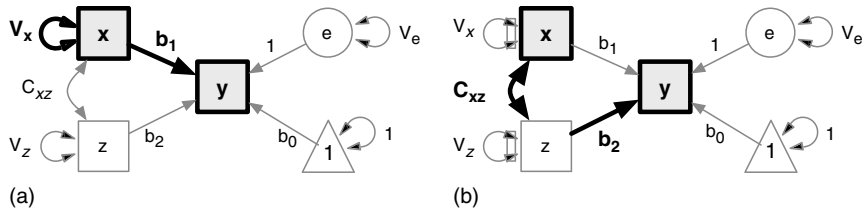

Figure 5 Two bridges between the variables x and y

coefficients used to form the bridge. Second, if the double-headed arrow is the covariance C_{xz} between x and z, then there is a length zero sequence of single-headed arrows from one end of C_{xz} leading to x and a length one sequence of single-headed arrows leading from the other end of C_{xz} to y. This bridge is illustrated in Figure 5-b and results in a product $C_{xz}b_2$. Thus the total covariance C_{xy} is the sum of the direct effect $V_x b_1$ and the background effect $C_{xz}b_2$

$$C_{xy} = b_1 V_x + b_2 C_{xz} \tag{3}$$

Path analysis is especially useful in gaining a deeper understanding of the covariance relations implied by a specified structural model. For instance, when a theory specifies mediating variables, one may work out all of the background covariances implied by the theory. Concepts such as suppression effects or the relationship between measurement interval and cross-lag effects can be clearly explained using a path analytic approach but may seem more difficult to understand from merely studying the equivalent algebra.

References

[1] Arbuckle, J.L. (1997). *Amos User's Guide*, Version 3.6 SPSS, Chicago.

[2] Boker, S.M., McArdle, J.J. & Neale, M.C. (2002). An algorithm for the hierarchical organization of path diagrams and calculation of components of covariance between variables, *Structural Equation Modeling* **9**(2), 174–194.

[3] Duncan, O.D. (1966). Path analysis: sociological examples, *The American Journal of Sociology* **72**(1), 1–16.

[4] Goldberger, A.S. (1971). Econometrics and psychometrics: a survey of communalities, *Econometrica* **36**(6), 841–868.

[5] Jöreskog, K.G. (1973). A general method for estimating a linear structural equation system, in *Structural Equation Models in the Social Sciences*, A.S. Goldberger & O.D. Duncan, eds, Seminar, New York, pp. 85–112.

[6] McArdle, J.J. & Aber, M.S. (1990). Patterns of change within latent variable structural equation modeling, in *New Statistical Methods in Developmental Research*, A. von Eye ed., Academic Press, New York, pp. 151–224.

[7] McArdle, J.J. & Boker, S.M. (1990). *Rampath*, Lawrence Erlbaum, Hillsdale.

[8] McArdle, J.J. & McDonald, R.P. (1984). Some algebraic properties of the reticular action model for moment structures, *The British Journal of Mathematical and Statistical Psychology* **87**, 234–251.

[9] Neale, M.C., Boker, S.M., Xie, G. & Maes, H.H. (1999). *Mx: Statistical Modeling*, Box 126 MCV, 23298: 5th Edition, Department of Psychiatry, Richmond.

[10] Wolfle, L.M. (1999). Sewall Wright on the method of path coefficients: an annotated bibliography, *Structural Equation Modeling* **6**(3), 280–291.

[11] Wolfle, L.M. (2003). The introduction of path analysis to the social sciences, and some emergent themes: an annotated bibliography, *Structural Equation Modeling* **10**(1), 1–34.

[12] Wright, S. (1918). On the nature of size factors, *The Annals of Mathematical Statistics* **3**, 367–374.

[13] Wright, S. (1920). The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs, *Proceedings of the National Academy of Sciences* **6**, 320–332.

[14] Wright, S. (1934). The method of path coefficients, *The Annals of Mathematical Statistics* **5**, 161–215.

(*See also* **Linear Statistical Models for Causation: A Critical Review**; **Structural Equation Modeling: Checking Substantive Plausibility**; **Structural Equation Modeling: Nontraditional Alternatives**)

STEVEN M. BOKER AND JOHN J. MCARDLE

Path-length Trees *see* Additive Tree