On the Path Analysis Techniques and Decomposition of Correlation Coefficients

Abstract

Path analysis is an extension of multiple regression technique used to evaluate causal model by examining the relationship among a set of variables. The process of decomposition of path correlation coefficients and estimation of path coefficients was investigated in this work. The outcome of a set of interrelated variables was generated using path diagrams and path coefficients. Decomposition of correlation coefficients into different effects was also carried out through the method of structural equations using path analysis theorem. In order to show the proportion of total variation of the dependent variable, the decision coefficient of a specified pathway was constructed from direct coefficient of determination. The path coefficients were estimated from the specified structural equations by ordinary least squares regression method, and it was deduced that each of the coefficients of determination had large effect on some other variables.

Comment [H1]: One line introduction has to be include

Comment [H2]: Give conclusion

Keywords: Path Analysis, Path Coefficient, Decision Coefficient, Coefficient of Determination, Structural Equations.

1.0 INTRODUCTION

The origin of path analysis technique could be traced to Biologist Sewall Wright who developed it in 1918 as an aid to the development of genetics. This gained popularity in social sciences with further exposition by Duncan (1975) and Land (1969). The technique has been used extensively in different fields such as demography, health, nutrition sciences, education, etc (Sharma and Rutherford, 1990; Michael et al., 2011; Palese et al, 2019; Asance, 2013; Frances, 2004; Shreela et al, 2009; Sandra, 2017).

Comment [H3]: Required recent reference

Path analysis is an interpretational technique used for studying patterns of causation among a set of variables. In other words, path analysis is an analytic tool which is applied to non-experimental data in which correlation are used to infer causation. It is a form or extension of multiple regression statistical analysis for evaluating causal models by examining the relationships between a dependent variable and two or more dependent variable (Chenias, et al., 2001; Barbasa et al.,2017).

The technique of path analysis applies to only sets of relationships among variables that are linear, additive and causal. With the use of the technique, the magnitude and significance of causal relationships among variables can be estimated (Visweswara and Balakrishma, 2007).

The methods of path analysis are significant to the extent that they decompose the correlation coefficient (r) into direct, indirect and total effects; test the relative importance of each causal effect compared to others on the same dependent variable and test the conceptual path model for their adequacy and parsimony (Singh et al., 2000; Kang, 2019; Peter et al., 2012). The methods enable researchers to understand the effect of one variable on the other when the variables are arranged in a causal fashion based on certain assumptions. However, in this study, path equations and path coefficients are estimated on the basis of basic path analysis theorem with correlation coefficient (r_{ij}) and coefficient of determination, (R^2) being decomposed using established illustrations.

Operational Definitions

(a) Exogenous variables: These are variables that cause other variables and whose variability is assumed to be determined by other causes outside the causal model

(b) Endogenous variables: These are variables whose variation is explained by exogenous variables or other variables in the system

(c) Residual variables: These are variables that account for the variance of the endogenous variables not accounted by the prior exogenous variables.

(d) Recursive models: These are models that postulate unidirectional, one way causal flow. This implies that, at a given point in time, a variable is not permitted to be both a cause and an effect of another variable.

(e) Non-recursive models**:** These are models that postulate reciprocal causation between endogenous variables.

(f) Path coefficient: A path coefficient is a standardized regression coefficient. It indicates the direct effect of a variable taken as a cause of a variable taken as an effect and it is denoted by p_{ij}, where i is the effect (dependent or endogenous variable) and j is the cause (independent or exogenous variable).

(g) Net (Direct) effect: The direct effect of factor X_j on weight Y_1 is the amount of variation in weight accounted for by the variation in factor X_j, while the influences of the other factors are removed. It is the product of path coefficients squared and sum of squares of the given variable.

(h) Joint (Indirect) effects: The joint effects of factors X_i and X_j on weight Y is the amount of variation in weight accounted for by the factors X_i and X_j jointly through their mutual correlation apart from their respective net effects.

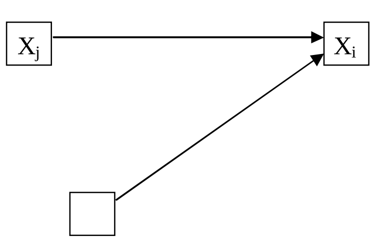
(i) Spurious effects: These are effects that pertain to the effects of common antecedent variables on the correlation between two other variables.

(j) Unanalyzed effects: These effects pertain to components that arise from the correlation between exogenous variables.

(k) Flowgraph analysis: This is the presentation of a system of structural equations in an iconic, causal diagram that guides the mathematical analysis of the variables relations.

Path Diagrams and Path Coefficients

Path diagrams display the outcome of a set of linearly inter-related variables and the assumed causal relationships among them. The set of variables that are involved in a path diagram are shown :



P_{ie}

R_{ie}

Figure 1: Exogenous, Endogenous and Residual Variables in Standardized Form.

The variables X_i, X_j, ..., R_d, R_e are the variable values in their standardized form. The values p_{ij} is the path coefficients and it is not a symmetric relation between variables like a coefficient of correlation. The limits of path coefficients exceed +1 or -1 in absolute value. The residual path coefficient decides the strength of that particular structural equation in a model, and varies between 0 and +1. If the coefficient is closer to zero, the variables in the equation are sufficient enough in explaining the entire variation in the endogenous variable; the reverse is the case if the coefficient value is far away from zero (Wright, 1960; Diao, 2016; Lombardi, 2015).

Comment [H4]: What it mean?

METHODOLOGY

Basic Theorem of Path Analysis

The basic theorem of path analysis expresses the correlation among the variables in a given model in terms of the path coefficients of the model. The basic theorem of path analysis is written as

$$r_{ij} = \sum k\rho_{ik} \,.\, r_{jk} \,, \qquad (1)$$

where i denotes the endogenous variable and j denotes the exogenous variable in the system and k includes all variables from which paths lead directly to variable X_i (Duncan, 1975; Farias, et al., 2016)

Path Analysis Assumptions

(i) There exists a causal framework interlinking individual predictor variable with the response variables;
(ii) All variables are linearly related. If any variable is non linear, such is made linear by taking the log value of the variable;
(iii) Each equation in the path analysis model is additive;
(iv) The residual variables (error terms) are uncorrelated with each other and also with all priori exogenous variables in the model. This allows equations in the model to be solved independently using ordinary least method;
(v) The endogenous variables are measured on an interval scale;
(vi) The observed variables are assumed to be measured without error.

Estimation of Path Coefficients

Estimating path coefficient is a function of the number of variables that are direct causal antecedents of an endogenous variable. If only one exogenous variable variable, X, has a direct causal impact on the endogenous variable, Y, then the corresponding path coefficient ρ_{XY} is estimated to be the correlation between X and Y ($\rho_{XY} = r_{XY}$). Figure 2 shows this relationship.

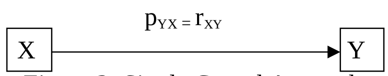


Figure 2: Single Causal Antecedent

Suppose Y has two causal antecedents X_a and X_b, then the path coefficients for X_a and X_b are beta weights in the regression of Y on X_a and X_b which are absolutely uncorrelated. This relationship is shown by figure 3.

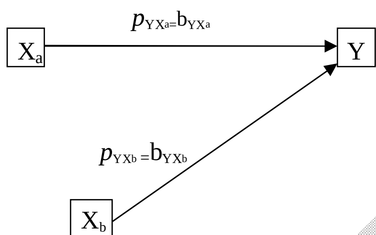


Figure 3: Two Causal Antecedents

The path coefficients are computed by considering a recursive path analytic model involving four variables in figure 4 by expressing all variables in standard score form.

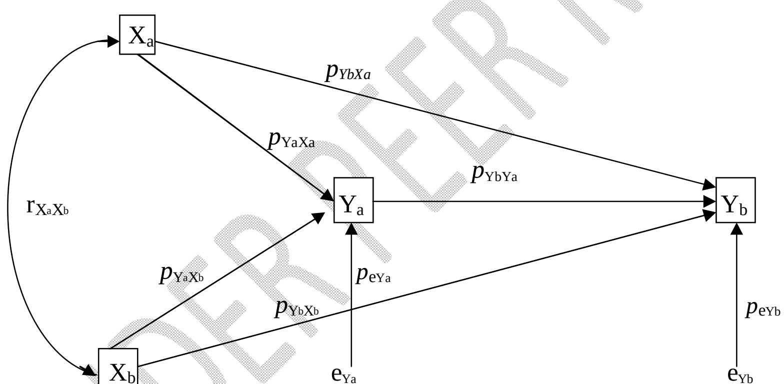


Figure 4: Path analytic model with two exogenous and two endogenous variables.

Expressing all the variables in standard score form, the following equations corresponding to figure 4 are generated as follows:

$$x_a = e_{x_a} \tag{2}$$

$$x_b = e_{x_b} \tag{3}$$

$$y_a = \rho Y_a X_a x_a + \rho Y_a X_b x_b + eY_a \tag{4}$$

$$y_b = \rho Y_b X_a x_a + \rho Y_b X_b x_b + \rho Y_b Y_a y_a + eY_a \tag{5}$$

where the e's are also expressed in standard score form. The variables X_a and X_b are represented by the residuals, eX_a and eX_b respectively because they are exogenous.

To now estimate the path coefficient, we start with $\rho Y_a X_a$ noting the relationship:

$rY_aX_a = \frac{1}{N}\sum x_a y_a$ (6)

Substituting equation (4) for y_a results to

$$rX_aY_a = \frac{1}{N}\sum x_a(pY_a X_a x_a + \rho Y_b X_b x_b + eY_a)$$
$$= \rho Y_a X_a \frac{\sum x_a x_a}{N} + \rho Y_a X_b \frac{\sum x_a x_b}{N} + \frac{\sum x_a eY_a}{N}$$
$$= \rho Y_a X_a + \rho Y_a X_b r X_a X_b \quad (7)$$

Equation (7) holds since $\frac{\sum x_a x_a}{N} = \frac{\sum x_a^2}{N} = 1$, $\frac{\sum x_a x_b}{N} = rX_aX_b$ and $E(x_a eY_a) = 0$. Thus, the term $\rho Y_a X_b r X_a X_b$ indicates that Y_a is affected by X_a and X_b which are both correlated.
Equation (7) also involves two unknowns, = $\rho Y_a X_a$ and $\rho Y_a X_b$ and therefore, intractable to solve. However, a solution is proffered by constructing another equation similar to the same unknowns.

Let $rX_bY_a = \frac{1}{N}\sum x_b y_a$ (8)

Substituting equation (4) for y_a results to

$$rX_bY_a = \frac{1}{N}\sum x_b(\rho Y_a X_a x_a + \rho Y_a X_b x_b + eY_a)$$
$$= \rho Y_a X_a \frac{\sum x_b x_a}{N} + \rho Y_a X_b \frac{\sum x_b x_b}{N} + \frac{\sum x_b eY_a}{N}$$
$$= \rho Y_a X_a\, r X_a X_b + \rho Y_a X_b \quad (9)$$

Since $\frac{\sum x_b x_a}{N} = rX_aX_b$; $\frac{\sum x_b x_b}{N} = \frac{\sum x_b^2}{N} = 1$ and $E(x_b eY_a) = 0$, we have two equations involving path coefficients that lead to Y_a:

$$rX_aY_a = \rho Y_a X_a + \rho Y_a X_b r X_a X_b \quad (10)$$
$$rX_bY_a = \rho Y_a X_a r X_a X_b + \rho Y_a X_b \quad (11)$$

Since path coefficients are equivalent to regression weights, we appropriately rewrite equations (10) and (11) as (12) and (13) respectively:

$$rX_aY_a = bY_a X_a + bY_a X_b r X_a X_b \quad (12)$$
$$rX_bY_a = bY_a X_a r X_a X_b + bY_a X_b \quad (13)$$

Following this procedure for y_1 yields these equations:

$$rX_aY_b = \rho Y_b X_a + \rho Y_b Y_a \rho Y_a X_a + \rho Y_b X_b r X_a X_b + \rho Y_b Y_a \rho Y_a X_b r X_a X_b \quad (14)$$
$$rX_bY_b = \rho Y_b X_b + \rho Y_b Y_a \rho Y_a X_b + \rho Y_b X_a r X_a X_b + \rho Y_b Y_a \rho Y_a X_a r X_a X_b \quad (15)$$
$$rY_aY_b = \rho Y_b X_a + \rho Y_a Y_a \rho Y_a X_a + \rho Y_b X_a \rho Y_a X_b r X_a X_b + \rho Y_b X_b \rho Y_a X_b + \rho Y_b X_b \rho Y_a X_a r X_a X_b \quad (16)$$

The path coefficient estimates are obtained by regressing Y_b on Y_a, X_a and X_b.
The causal system from figure 4 requires two regression analyses for all path coefficients to be estimated. The paths from X_a and X_b to Y_a ($\rho Y_a X_a$ and $\rho Y_a X_b$) are obtained by regressing Y_a on X_a and X_b. Similarly, the paths from X_a and X_b and Y_a to Y_b ($\rho Y_b X_a$ and $\rho Y_b X_b$ *and* $\rho Y_b Y_a$) are obtained by regressing Y_b on X_a , X_b and Y_a jointly.

Residual Path Terms Estimation

In a recursive system, the path coefficient from the residual, e, to an endogenous variable, $Y_i = \sqrt{1 - R^2_{Y_{i\,.jkl\ldots p}}}$, where $R^2_{Y_{i\,.jkl\ldots p}}$ is the squared multiple correlation of the endogenous variable Y_i with all those variables, both exogenous and endogenous that affect it. Thus, according to the causal system in figure 4, the residual path terms are estimated as:

$$\rho e_{Y_a} = \sqrt{1 - R^2_{Y_a.X_aX_b}} \qquad (17)$$

$$\rho e_{Y_b} = \sqrt{1 - R^2_{Y_b.Y_aX_aX_b}} \qquad (18)$$

Decomposition of Correlation Coefficient

The method of structural equations decomposes the correlation coefficient, r_{ij} by first of all deriving the path estimation equations. The correlation between two variables can be decomposed into a direct and indirect effects or joint effects shared with other variables in the system.

Wright(1960), suggested a method equivalent to structural equations procedure for decomposition of correlation coefficient into direct and several indirect effects between the endogenous and exogenous variables using a path diagram.

Alwin and Hauser (1975) systematically used a set of reduced forms of equations by defining the total effect as a combination of direct and indirect effects. The method does not consider non causal effect, but can be obtained by deviation of total effect from the total association.
An application of flowgraph method to both recursive and non-recursive was suggested by Chen(1983). The flowgraph analysis presented the properties of a complicated system as a whole in a readily visualised form and it facilitates the manipulation and solution of the system of equation.
The correlation coefficient between two variables $i = a$ and j is considered and expressed by the theorem of path analysis as

$$r_{ij} = r_{ik} = \sum_k \rho_{ik} . r_{jk} \qquad (19)$$

The general form of the path model is

$$X_a = \rho_{ab}X_b + \rho_{ac}X_c + \cdots + \rho_{an}X_n + \rho_{az}X_Z, \qquad (20)$$

where X_a is the endogenous variable, $X_b, \ldots X_n$ are the exogenous variable and X_Z is the residual variable.

Equation (19) is rewritten as

$$r_{ij} = r_{aj} = \rho_{aj} = \sum_{k=b,k\neq j} \rho_{ak}.r_{jk}, \text{ since } n = a \quad (21)$$

The first term in equation (21) is the direct effect of exogenous variable X_j on X_a. The total effect of X_j on X_a is obtained as

$$r_{aj} = \rho_{aj} = \sum_{k=b,k\neq j} \rho_{ak}.r_{jk} \quad (22)$$

Thus, the correlation coefficient r_{1j} between endogenous variable X_a and the exogenous variable X_j is decomposed into direct effect and indirect effect.

Using figure 4, the correlation coefficient r_{ij} is decomposed into four components: Direct Effect (DE); Indirect Effect (IE); Spurious Effects (SE) due to common causes and Unanalyzed Effect (UE) due to correlated exogenous variable as follows:

$$rX_aY_a = \rho VX_a + \rho Y_aX_brX_aX_b$$
$$= DE + UE \quad (23)$$

$$rX_bY_a = \rho Y_aX_b + \rho Y_aX_arX_aX_b$$
$$= \text{DE} + \text{UE} \quad (24)$$

$$rX_aY_b = \rho Y_bX_a + \rho Y_bY_a\rho Y_aX_a + \rho Y_bX_brX_aX_b + \rho Y_bY_a\rho Y_aX_brX_aX_b$$
$$= \text{DE} + \text{IE} + \text{UE} \quad (25)$$

$$rX_bY_b = \rho Y_bX_b + \rho Y_bY_a\rho Y_aX_b + \rho Y_bX_arX_aX_b + \rho Y_bY_a\rho Y_aX_arX_aX_b$$
$$= \text{DE} + \text{IE} + \text{UE} \quad (26)$$

$$rY_aY_b = \rho Y_bX_a + \rho Y_bY_a\rho Y_aX_a + \rho Y_bX_a\rho Y_aX_brX_aX_b + \rho Y_bX_b\rho Y_aX_b + \rho Y_bX_b\rho Y_aX_arX_aX_2$$
$$= \text{DE} + \text{SE} \quad (27)$$

Component Analysis of Coefficient of Determination (R^2)

The coefficient of determination, R^2 $(0 \leq R^2 \leq 1)$ measures the proportion of variation in the dependent variable explained by all the independent variables involved in the regression equation. It shows the strength of the regression equation.

R^2 is subdivided into direct coefficient of determination, $\sum_{j=1} R_j^2$ and indirect coefficient of determination, $\sum_{\substack{j=1 \\ j<t}}^{p} R_{jt}$ so that

$$R^2 = \sum_{j=1} R_j^2 + \sum_{\substack{j=1 \\ j<t}}^{p} R_{jt} \text{ (see figure 5)}$$
$$= \sum_{j=1}^{p} (b_j^+)^2 + \sum_{\substack{j=1 \\ j<t}}^{p} 2b_jr_jb_t^+ \quad (28)$$

where SS*residual* specifies the discrepancy between the data and the model estimation, SS *total* indicates the total variation of data set.

For this work, coefficient of determination is split into its components, which are net effects and joint effects. For example, we consider the structural equation

$$X_f = \rho_{fa}X_a + \rho_{fb}X_b + \rho_{fc}X_c + \rho X_d + \rho_{fe}X_e + \rho_{fw}R_w \quad (29)$$

where $X_a, X_b, X_c \ldots R_w$ are the variable values in their standard form. The symbolic values $\rho_{fa}, \rho_{fb}, \rho_{fc}, \ldots$, are the path coefficients.

Thus, the coefficient of determination $R^2_{f.abcde}$ of the structural equation X_f is estimated using the relationship:

$$R^2_{f.abcde} = \rho^2_{fa} + \rho^2_{fb} + \rho^2_{fc} + \rho^2_{fd} + \rho^2_{fe} + 2\rho_{fa}\rho_{fb}r_{ab} + 2\rho_{fa}\rho_{fc}r_{ac} + 2\rho_{fa}\rho_{fd}r_{ad} + 2\rho_{fa}\rho_{fe}r_{ae} + 2\rho_{fb}\rho_{fc}r_{bc} + 2\rho_{f2}\rho_{fd}r_{bd} + 2\rho_{fb}\rho_{fe}r_{be} + 2\rho_{fc}\rho_{f4}r_{cd} + 2\rho_{fc}\rho_{fe}r_{ce} + 2\rho_{fd}\rho_{fe}r_{de} \quad (30)$$

In equation (29), the first five terms on the RHS are the path coefficients squared with each of them denoting a net effect in the components of the coefficient of determination, $R^2_{f.abcde}$. There are five net effect terms in $R^2_{f.abcde}$ since there are five independent variables, X_a, X_b, X_c, X_d, X_e in the structural equation, X_f. The net effect is the direct effect of the variable $X_j, j = a, b, c, \ldots e$, on the dependent variable X_f. It measures the amount of variation in X_f accounted for by the variation in the variable, X_j, while the influences of other variables are removed.

Each of the other terms in the RHS of equation (29) denotes the effect in the components of the coefficient of determination.

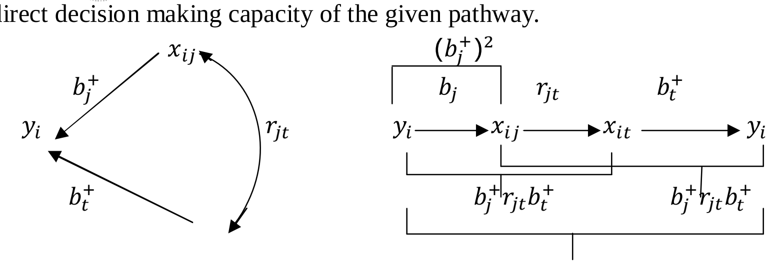
The joint effects are the joint influence of two variables on the dependent variable X_f through their mutual correlation. For example, the joint effect of variable $X_a \; and \; X_c$ on the dependent variable X_f is $2\rho_{fa}\rho_{fc}r_{ac}$. This is the amount of variation in the variable X_f accounted for by the variables X_a and X_c jointly through their mutual correlation apart from their individual net effects.

Decision Coefficient

The decision coefficient, R_j of a specified pathway is constructed to show the proportion of total variation of dependent variable, y_k, determined by a specified pathway, $x.x_{kj}$ $(j = 1,2, \ldots m)$. The decision coefficient consists of two terms [see figures 5 (i) and 5 (ii)]:

$$R_j = (b_j^+)^2 + 2\sum_{\substack{t=1 \\ j\neq t}} b_j^+ r_{jt}\, b_j^+, \quad (32)$$

where the first term is the direct determination factor that demonstrates the direct decision making capacity of the given pathway, the square root of the first term is the direct effect in the path analysis model and the second term is the indirect determination factor that shows the indirect decision making capacity of the given pathway.



x_{it}

(i) (ii) $2b_j^+ r_{jt} b_t^+ \ , t = 1,2,3, \dots, m \ and \ j \neq t$

Figure 5: Decision Coefficient of Specified Pathways

Derivation of Path Estimation Equations from Structural Equations

Since structural equations are different from path estimation equations, basic theorem of path analysis is applied to some recursive system of regression equations leading to formation of set of path estimation equations. The following regression equations are considered to form a set of path estimation equations in equations (36) to (47):

$X_d = \rho_{da}X_a + \rho_{db}X_b + \rho_{dc}X_c + \rho_{dk}X_k$ (33)

$X_e = \rho X_a + \rho X_b + \rho_{ec}X_c + \rho X_e + \rho_{el}X_l$ (34)

$X_f = \rho_{fa}X_a + \rho_{fb}X_b + \rho_{fc}X_c + \rho_{fd}X_d + \rho_{fe}X_f + \rho_{fm}R_m$ (35)

$r_{da} = \rho_{da} + \rho_{db}.r_{ab} + \rho_{dc}.r_{ac}$ (36)

$r_{db} = \rho_{da}.r_{ba} + \rho_{db} + \rho_{dc}.r_{bc}$ (37)

$r_{dc} = \rho_{da}.r_{ca} + \rho_{db}.r_{cb} + \rho_{dc}$ (38)

$r_{ea} = \rho_{ea} + \rho_{eb}.r_{ab} + \rho_{ec}.r_{ac} + \rho_{ed}.r_{ad}$ (39)

$r_{eb} = \rho_{ea}.r_{ba} + \rho_{eb} + \rho_{ec}.r_{bc} + \rho_{ed}.r_{bd}$ (40)

$r_{ec} = \rho_{ea}.r_{ca} + \rho_{eb}.r_{cb} + \rho_{ec} + \rho_{ed}.r_{cd}$ (41)

$r_{ed} = \rho_{ea}.r_{da} + \rho_{eb}.r_{db} + \rho_{ec}.r_{dc} + \rho_{ed}$ (42)

$r_{fa} = \rho_{fa} + \rho_{fb}.r_{ab} + \rho_{fc}.r_{ac} + \rho_{fd}.r_{ad} + \rho_{fe} + r_{ae}$ (43)

$r_{fb} = \rho_{fa}.r_{ba} + \rho_{fb} + \rho_{fc}.r_{bc} + \rho_{fd}.r_{bd} + \rho_{fe}.r_{be}$ (44)

$r_{fc} = \rho_{fa}.r_{ca} + \rho_{fb}.r_{cb} + \rho_{fc} + \rho_{fd}.r_{cd} + \rho_{fe}.r_{ce}$ (45)

$r_{fd} = p_{fa}.r_{da} + p_{fb}.r_{db} + p_{fc}.r_{dc} + p_{fd} + p_{fe}.r_{ce}$ (46)

$r_{fe} = \rho_{fa}.r_{ea} + \rho_{fb}.r_{eb} + \rho_{fc}.r_{dc} + \rho_{fd}.r_{ed} + \rho_{fe}$ (47)

Equation (43) is written as

$$r_{fa} = \rho_{fa} + \rho_{fd}.\rho_{da} + \rho_{fe}.\rho_{ea} + \rho_{fe}.\rho_{ed}.\rho_{da} + r_{ab}(\rho_{fb} + \rho_{fd}.\rho_{db} + \rho_{fe}.\rho_{eb} + \rho_{fe}.\rho_{ed}.\rho_{db}) + r_{ac}(\rho_{fc} + \rho_{fd}.\rho_{dc} + \rho_{fe}.\rho_{ec} + \rho_{fe}.\rho_{ed}.\rho_{dc}) \quad (48)$$

The path estimation equations (43) - (47) are re-written in matrix form as

$A = PC,$ (49)

where $A = (r_{fa}, r_{fb}, r_{fc}, r_{fd}, r_{fe})$

$P = (\rho_{fa}, \rho_{fb}, \rho_{fc}, \rho_{fd}, \rho_{fe})$

$$C = \begin{bmatrix} 1 & r_{ab} & r_{ac} & r_{ad} & r_{ae} \\ r_{ba} & 1 & r_{bb} & r_{bc} & r_{bd} \\ r_{ca} & r_{cb} & 1 & r_{cd} & r_{ce} \\ r_{da} & r_{db} & r_{dc} & 1 & r_{de} \\ r_{ea} & r_{eb} & r_{ec} & r_{ed} & 1 \end{bmatrix}$$

The path coefficients ρ_{fa} and ρ_{fe} are estimated as

$P = C^{-1}A$ (50)

The residual path coefficient $\rho_{fm} = \sqrt{1 - R^2_{f.abcde}}$ (51)

The general form of p_{fm} is generated as

$$p^2_{fm} = r_{ff} - \sum_{j=a} p^2_{fj} + \sum_{j=a} \sum_{k=j+a} p_{fj} p_{fk} r_{jk}, \qquad (52)$$

where f denotes the endogenous variable, m the residual variable and j the exogenous variables.

Illustrative Example

The illustrative example is anchored on real life data adapted from six demographic and economic variables for fifteen states in India (Pathak and Murthy, 2007). The variables were defined as follows:

X_a = Per capital net state domestic product at factor cost
X_b = Percentage of urban population to total population
X_c = Female literacy rate for ages 7 and above
X_d = Infant mortality rate
X_e = Effective couple protection rate
X_f = Crude birth rate per 1000 population

The first three variables are exogenous variables while the last three variables are endogenous variables.

Table 1: Data on Demographic and Socioeconomic Variables in India (1989).

States	X_a	X_b	X_c	X_d	X_e	X_f
S_1	1692	25.8	30.9	83	39.0	25.9
S_2	1558	10.7	43.7	99	26.2	29.4
S_3	1071	13.0	21.1	97	22.9	34.3
S_4	2506	33.4	45.5	90	53.2	28.7
S_5	3086	23.9	36.7	90	56.4	35.2
S_6	2041	30.3	41.0	74	42.3	28.0
S_7	1447	24.1	83.6	28	46.4	20.3
S_8	680	22.3	25.6	121	36.2	35.5
S_9	2960	37.6	47.7	68	54.7	28.5
S_{10}	1455	12.9	31.6	122	37.5	30.5
S_{11}	3552	29.1	46.7	62	68.2	28.3
S_{12}	1620	22.3	18.8	103	27.8	34.2
S_{13}	2030	33,8	48.7	74	52.6	23.1
S_{14}	1547	19.3	23.4	124	28.8	37.0
S_{15}	1930	27.1	43.9	69	31.3	27.2

Table 2: Correlation Coefficients Matrix among the Six Socioeconomic Variables in India (1989)

Variables	X_a	X_b	X_c	X_d	X_e	X_f
X_a	1	0.603	0.286	-0.421	0.821	-0.168
X_b		1	0.378	-0.535	0.703	0.434
X_c			1	-0.847	0.502	0.823

X_d				1	-0.5109	0.810
X_e					1	-0.380
X_f						1

Using the data under consideration, the path coefficients are estimated by the regression method of ordinary least squares. The generated fitted form of the path model in equations 33-35 is given as:

$$X_d = 0.86X_a - 0.20X_b - 0.75X_c \quad (53)$$

$$X_e = -0.63X_a + 0.31X_b + 0.49X_c + 0.34X_d \quad (54)$$

$$X_f = -0.28X_a - 0.17X_b - 0.45X_c - 0.43X_d - 0.043X_e \quad (55)$$

The estimated R square values are given as:

$$R^2_{f.abcde} = 0.76;\ R^2_{e.abcd} = 0.81;\ R^2_{d.abc} = 0.78;\ \rho_{fm} = 0.24$$

RESULTS and DISCUSSION

From equation (48), the correlation between the variables decomposed into direct, indirect or joint effects shared with other variables were observed. Hence, the first term in the right hand side of the equation is the direct effect of X_a on X_f, the second term is the indirect effect of X_a on X_f through X_d, the third term is the indirect effect of X_a on X_a through X_e, the fourth term is the indirect effect of X_a on X_f through X_d and X_e, the fifth term is the joint effect on X_f which X_a shared with X_b and the sixth term is the joint effect on X_f which X_a shared with X_c.

Each of the coefficients of determination in respect of the structural equations, X_d ,X_e X_f has large effect on the other variables. Clearly, 76% of the variance in variable X_f is predicted by other 5 variables while 24% of the variance in the variable is unexplained by the model.

Also, 81% of the variance in variable X_e is predicted by other 4 variables while 19% of the variance in the variable is unexplained by the model. Similarly, 78% of the variance in variable X_d is predicted by other 3 variables while 22% of the variance in the variable is unexplained by the model. This shows that the larger the value of R^2, the better the model under consideration.

CONCLUSION

In this work, the principles of path analysis techniques and decomposition of correlation coefficients have been presented. Path estimation equations from specified structural equations were also generated with real life applications. It is established that in path analysis method, causal linkages and directions are determined by some predefined factors. We also therefore assert that path analysis reduces inter-correlation matrix for a set of variables anchored on the framework of predicted causal associations.

REFERENCES

Comment [H5]: Much recent references to be included.

- Duncan, O.D. (1975). Introduction to Structural Equation Models. Academic Press, New York.

- Land, K.C. (1969). Principles of Path Analysis: In Borgatta, E.F, Bohmstdt, G.W. (Eds). Sociological Methodology, San Francisco Jossey-Bass, pp 3-37.
- Wright, S. (1960). Path Coefficient and Path Regressions: Alternative or Complementary Concepts. Biometrics, 16, pp 189- 202.
- Alwin, D. & Hauser, R.M. (1975). The Decomposition of Effects in Path Analysis: American Sociological Review, 40, pp37-47.
- Chen, H.T. (1983). Flowgraph Analysis for Effect Decomposition: Use in Recursive and Non-recursive Models. Sociological Methods and Research, 12(1), pp 3-29.
- Pathak, K.B. & Murthy, P.K. (2007). Path Analysis and Demographic Examples: A manual of Statistical Methods for Use in Health, Nutrition and Anthropology, 2^{nd} Ed., Jaypee Brothers Medical Publishers Ltd, New Delhi.
- Micheal, G. Gorm, B.J. & Thorkild, I.A (2011). Dynamic Path Analysis in Life Course Epidemiology. American Journal of Epidemiology , 173(10), pp 1131-1139.
- Palese, A., Luca, G. & Valentina, B. (2019). A path Analysis on the Direct and Indirect Effects of the Unit Environment on Eating Dependence among Congnitively impaired Nursing Home Residents. BMC Health Services Research, 19(775), pp 1-14.
- Asance, T. (2013). A Path Analysis of Relationship between Factors with Achievement Motivation of Students of Private Universities in Bangkok, Thailand Procedia, Social and Behavioural Sciences, 881, pp 229-238.
- Frances, K.S., Hasani, C.C. & Amany, N. (2004). Path Analysis: An Introduction and Analysis of a Decade of Research. Journal of Educational Research, 98(1), pp 5-13.
- Shreela, V.S, Deanna, M.H.& Steven, H.K. (2009). A Path Analysis to identify the Psyco-Social Factors Influencing Physical Activity and Bone Health in Middle Schools Girls. Journal of Physical Health, 6(5), pp 606-616.
- Sandra, G.L., Leila, K. & Timothy, B. (2017). A Path Analysis Study of Factors Influencing Hospital Perception of Quality of Care Factors Associated with Patient Satisfaction and Patient Experience.
- Singh, K. & Ozturk, M. (2000). Effect of Part- time on High School Mathematics and Science Course. Journal of Educational Research, 94, pp 67 -75.
- Chenias, M.M., Hu, L. T. & Garaa, B.F. (2001). Academic Self Efficacy and First Year College Students' Performance and Adjustment. Journal of Educational Psycology, 93(1), pp 55-64.
- Kang, D.H. (2019). A Path Analysis of the Elderly Deprivation Experience on the Thinking of Suicide. Journal of Social Science, 58, pp 197-245.

- Diao, Z.J. & Chen, B. (2016). Correlation and Path Coefficient Analysis between Thermal ExtractionYield and Coal Properties. Energy Sources Part A Recovery. Util Environment Eff. 38(22), pp 3412 -3416.
- Barbasa, R.P., Alcantara, N.F. & Gravina, L.M. (2017). Early Selection of Sugarcane using Path Analysis Genetics and Molecular Research, 6(1), pp 1-8.
- Peter, F .A., Alcantara, N.F. & Barros, W.S. (2012). Phenotypic Correlations and Genotypic and Phenotypic Path Analysis of Cane and Sugar Yield in Sugarcane. Indian Journal Genet, Plant Breed. 46, pp 550-570.
- Lombardi, G.M., Nunes, J.A. & Parrella, R.A. (2015). Path Analysis of Agro- Industrial Traits in Sweet Sorgum. Gene. Mol. Research, 14, pp 16392-16402.
- Farias, F.J.C, Carvalho, L.P. & Silva, F.J.L. (2016). Correlations and Path Analysis among Agronomic and Technological Traits of uplant Cotton. Genet, Mol. Research, 15, pp 1-7.