

Article

An Entropy-Based Approach to Path Analysis of Structural Generalized Linear Models: A Basic Idea

Nobuoki Eshima ^{1,*}, Minoru Tabata ², Claudio Giovanni Borroni ³ and Yutaka Kano ⁴

¹ Department of Biostatistics, Faculty of Medicine, Oita University, Oita 879-5593, Japan

² Department of Mathematical Sciences, Graduate School of Engineering, Osaka Prefecture University, Osaka 599-8531, Japan; E-Mail: mnrtabata@luck.ocn.ne.jp

³ Department of Statistics and Quantitative Methods, University of Milano Bicocca, via Bicocca degli Arcimboldi 8, Milano 20126, Italy; E-Mail: claudio.borroni@unimib.it

⁴ Division of Mathematical Science, Graduate School of Engineering Science, Osaka University, Osaka 560-8531, Japan; E-Mail: kano@sigmath.es.osaka-u.ac.jp

* Author to whom correspondence should be addressed; E-Mail: eshima@oita-u.ac.jp; Tel.: +81-97-549-4411 (ext. 8372); Fax: +81-97-586-5619.

Academic Editor: Ali E. Abbas

Received: 11 May 2015 / Accepted: 17 July 2015 / Published: 22 July 2015

Abstract: A path analysis method for causal systems based on generalized linear models is proposed by using entropy. A practical example is introduced, and a brief explanation of the entropy coefficient of determination is given. Direct and indirect effects of explanatory variables are discussed as log odds ratios, *i.e.*, relative information, and a method for summarizing the effects is proposed. The example dataset is re-analyzed by using the method.

Keywords: direct effect; entropy coefficient of determination; generalized linear model; indirect effect; log odds ratio

1. Introduction

Path analysis [1] is often applied to causal systems of continuous variables through the linear structural equations model (LISREL) [2,3]. In the LISREL approach, causal relationships among variables are described by a path diagram and translated into linear equations of the variables. Causal effects can then be calculated by regression and correlation coefficients obtained for the linear equations.

In contrast, path analysis of categorical variables is more complex than that of continuous variables because the causal system under consideration cannot be described by linear regression equations. Goodman [4–7] considered path analysis of binary variables by using logit models and discussed the effects of explanatory variables, though without discussing direct and indirect effects. Hagenars [8] discussed path analysis of categorical variables by using a log-linear model, but here as well without discussion of direct and indirect effects. Eshima *et al.* [9] proposed a path analysis method for categorical variables in logit models. Kuha and Goldthorpe [10] gave a two-stage path analysis method for generalized linear models (GLMs) that uses log odds ratios. In their approach, first the total, direct and indirect effects are defined for mean differences of response variables, and then the method is applied to measuring the effects on the basis of log odds ratios. However, additive decomposition of the total effect into the direct and indirect effects only approximately reflects reality, and assessing effects in categorical (polytomous) variable systems become more complicated as the numbers of variable categories are increased [10]. Albert and Nelson [11] proposed a path analysis method to calculate pathway effects for causal systems on the basis of GLMs, but not all pathway effects are identifiable. As in the two-stage cases, when factors, intermediate variables, and response variables are categorical, pathway effects become very complicated because the variable effects are defined for mean differences of response variables.

There are many examples of response variables in practical data that are not normally distributed in various fields of study. There is need for a method of path analysis with responses that are not normally distributed, especially categorical responses, and it is useful to discuss a path analysis approach for causal systems of GLMs [12,13]. When describing causal systems of the variables by GLMs, regression parameters or coefficients are related to log odds ratios [14–16], and so it is natural to consider the effects of factors (explanatory variables) according to odds or log odds ratios. However results become more complicated as the number of categories of the variables increases. In such cases, it is suitable to summarize the effects of factors in GLMs. For this purpose, we use the entropy coefficient of determination (ECD), one of the entropy-based measures of predictive power for GLMs [15,16].

The remainder of this paper is organized as follows: Section 2 presents a practical example of causal systems—British mobility data [10]—and re-analyzes them by a new method of path analysis. Section 3 considers the relation between the log odds ratio and entropy, and ECD is briefly reviewed. Section 4 introduces a path analysis method for causal systems described by GLMs, and in Section 5 a method for testing effects of variables is given. The British mobility data are re-analyzed by the proposed approach in Section 6. Finally, Section 7 provides some discussion and conclusions for the present approach.

2. Practical Example

British mobility data describe the effects of education on social class mobility [10]. There are three variables, which are causally ordered as shown in Figure 1: parents' social class, X ; individual social class, Y ; and education, Z , which intermediates Y between X and Y . The three variables are discrete. Social classes X and Y have three categories, “salaried and employers”, “middle class”, and “working class”; education Z has seven levels. While the effects of X and Z on Y can be discussed through log odds ratios, the results are complicated because the number of variable categories is large. It is important to

summarize causal effects measured with log odds ratios, especially in such practical examples, to assess the intermediate effect of education on social class mobility.

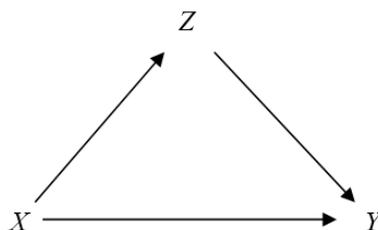


Figure 1. Path diagram of social class mobility.

3. Log Odds Ratio and Information

Let \mathbf{X} and Y be a $p \times 1$ explanatory-variable vector and a response variable, respectively, and let $f(y|\mathbf{x})$ be the conditional probability or probability density function of Y given that $\mathbf{X} = \mathbf{x}$. The function $f(y|\mathbf{x})$ is assumed to belong to the following family of exponential distributions:

$$f(y|\mathbf{x}) = \exp\left\{\frac{y\theta - b(\theta)}{a(\varphi)} + c(y, \varphi)\right\}, \tag{1}$$

where θ and φ are parameters, and $a(\varphi)$ (>0), $b(\theta)$, and $c(y, \varphi)$ are known functions. Let $\boldsymbol{\beta}^T = (\beta_1, \beta_2, \dots, \beta_p)^T$. The function θ is a function of $\boldsymbol{\eta} = \boldsymbol{\beta}^T \mathbf{x}$ through a link function $h(u)$.

Remark 1. In general, the systematic component can be extended to be a function of explanatory variable vector \mathbf{x} . Then, the model is referred to as a generalized nonlinear model. For the sake of simplicity, the function is denoted by $\theta = \theta(\mathbf{x})$. The discussion below is applicable to this case.

Let $\boldsymbol{\mu}_x$ and μ_y be the means of \mathbf{X} and Y , respectively. Then, let us consider the following log odds ratio:

$$\begin{aligned} \log \frac{f(y|\mathbf{x})f(\mu_y|\boldsymbol{\mu}_x)}{f(\mu_y|\mathbf{x})f(y|\boldsymbol{\mu}_x)} &= (\log f(y|\mathbf{x}) - \log f(\mu_y|\mathbf{x})) - (f(y|\boldsymbol{\mu}_x) - f(\mu_y|\boldsymbol{\mu}_x)) \\ &= \frac{(y - \mu_y)(\theta(\mathbf{x}) - \theta(\boldsymbol{\mu}_x))}{a(\varphi)}. \end{aligned}$$

The first and second terms of the right hand side of the above equation are the relative information with respect to response variable Y , so the log odds ratio is the change of the relative information in explanatory variable vector \mathbf{X} . In GLMs, taking the expectation of the above log odds ratio with respect to \mathbf{X} and Y , it is reduced to $\frac{\text{cov}(\theta, Y)}{a(\varphi)}$. The quantity $\frac{\text{cov}(\theta, Y)}{a(\varphi)}$ can be expressed as a symmetric type of the Kullback–Leibler (KL) information between a GLM based on (1) and the null model with $\boldsymbol{\beta}^T = \mathbf{0}$ [15]; thus, we denote $\frac{\text{cov}(\theta, Y)}{a(\varphi)}$ by $KL(\mathbf{X}, Y)$ in this paper. Let $f(y)$ be the density or probability function for null model $\boldsymbol{\beta}^T = \mathbf{0}$ and let $g(\mathbf{x})$ be that of \mathbf{X} . Then:

$$KL(\mathbf{X}, Y) = \iint f(y | \mathbf{x})g(\mathbf{x})\log \frac{f(y | \mathbf{x})}{f(y)} \cdot d\mathbf{x}dy + \iint f(y)g(\mathbf{x})\log \frac{f(y)}{f(y | \mathbf{x})} \cdot d\mathbf{x}dy.$$

If the variables in the above are discrete, the related integrals are replaced by the summations. The ECD is then defined as:

$$ECD(\mathbf{X}, Y) = \frac{\text{cov}(Y, \theta)}{\text{cov}(Y, \theta) + a(\varphi)}.$$

Then, ECD can also be expressed as:

$$ECD(\mathbf{X}, Y) = \frac{\frac{\text{cov}(Y, \theta)}{a(\varphi)}}{\frac{\text{cov}(Y, \theta)}{a(\varphi)} + 1} = \frac{KL(\mathbf{X}, Y)}{KL(\mathbf{X}, Y) + 1}.$$

The measure is interpreted as the proportion of the variation in entropy of Y that is explained by \mathbf{X} [15,16]. As shown above, GLMs explain the entropy of response variables, so it is suitable to measure the effects of explanatory variables based on entropy.

Remark 2. Applying ECD to the linear regression model, ECD is the usual coefficient of determination R^2 .

4. Measuring the Total, Direct, and Indirect Effects in Recursive GLM Systems

For simplicity, in the recursive case with X_i ($i = 1, 2, 3$), where X_i precedes X_{i+1} ($i = 1, 2$), we discuss the effects of X_1 and X_2 on X_3 . Let μ_i be the expectation of X_i ($i = 1, 2, 3$). Then, for a GLM with the conditional density or probability function of X_3 when $(X_1, X_2) = (x_1, x_2)$ given by (1), the total effect of $(X_1, X_2) = (x_1, x_2)$ on $X_3 = x_3$ can be defined by using the following log odds ratio:

$$\log \frac{f(x_3 | x_1, x_2)f(\mu_3 | \mu_1, \mu_2)}{f(\mu_3 | x_1, x_2)f(x_3 | \mu_1, \mu_2)} = \frac{(x_3 - \mu_3)(\theta(x_1, x_2) - \theta(\mu_1, \mu_2))}{a(\varphi)}.$$

Taking the expectation of the above effect with respect to X_1, X_2 and X_3 , we have:

$$\frac{\text{cov}(\theta(X_1, X_2), X_3)}{a(\varphi)} = KL((X_1, X_2), X_3).$$

The above KL information is the (summary) total effect of explanatory variables (X_1, X_2) on response variable X_3 . Let $\mu_2(x_1)$ and $\mu_3(x_1)$ be the conditional expectations for X_2 and X_3 , respectively, given that $X_1 = x_1$. The log odds ratio with respect to x_2 and x_3 for a given x_1 is:

$$\log \frac{f(x_3 | x_1, x_2)f(\mu_3(x_1) | x_1, \mu_2(x_1))}{f(\mu_3(x_1) | x_1, x_2)f(x_3 | x_1, \mu_2(x_1))} = \frac{(x_3 - \mu_3(x_1))(\theta(x_1, x_2) - \theta(x_1, \mu_2(x_1)))}{a(\varphi)}.$$

The total effect of $X_2 = x_2$ on $X_3 = x_3$ when $X_1 = x_1$ is defined by the above log odds ratio because the effect expresses the total effect of $X_2 = x_2$ on $X_3 = x_3$ when the effect of the preceding variable $X_1 = x_1$ is excluded. From this, the total effect of $X_1 = x_1$ on $X_3 = x_3$ is defined by

$$\begin{aligned} & (\text{the total effect of } x_1 \text{ and } x_2 \text{ on } x_3) - (\text{the total effect of } x_2 \text{ on } x_3 \text{ when } X_1 = x_1) \\ &= \log \frac{f(x_3 | x_1, x_2) f(\mu_3 | \mu_1, \mu_2)}{f(\mu_3 | x_1, x_2) f(x_3 | \mu_1, \mu_2)} - \log \frac{f(x_3 | x_1, x_2) f(\mu_3(x_1) | x_1, \mu_2(x_1))}{f(\mu_3(x_1) | x_1, x_2) f(x_3 | x_1, \mu_2(x_1))} \\ &= \frac{(x_3 - \mu_3)(\theta(x_1, x_2) - \theta(\mu_1, \mu_2))}{a(\varphi)} - \frac{(x_3 - \mu_3(x_1))(\theta(x_1, x_2) - \theta(x_1, \mu_2(x_1)))}{a(\varphi)}. \end{aligned}$$

By taking the expectation of the above information with respect to X_1, X_2 and X_3 , the (summary) total effect of X_1 on X_3 is given by

$$\frac{\text{cov}(\theta(X_1, X_2), X_3)}{a(\varphi)} - \frac{\text{cov}(\theta(X_1, X_2), X_3 | X_1)}{a(\varphi)} = KL((X_1, X_2), X_3) - KL(X_2, X_3 | X_1),$$

where $KL(X_2, X_3 | X_1)$ is given by

$$\begin{aligned} KL(X_2, X_3 | X_1) &= \iiint f(x_3 | x_1, x_2) g(x_1, x_2) \log \frac{f(x_3 | x_1, x_2)}{f(x_3 | x_1)} dx_1 dx_2 dx_3 \\ &+ \iiint f(x_3 | x_1) g(x_1, x_2) \log \frac{f(x_3 | x_1)}{f(x_3 | x_1, x_2)} dx_1 dx_2 dx_3, \end{aligned} \tag{2}$$

The second term implies the effect of X_2 by itself, that is, the effect of X_2 on X_3 when the effect of X_1 is excluded, and is defined as the (summary) total effect of X_2 on X_3 . The direct effect of $X_1 = x_1$ on $X_3 = x_3$ can be understood according to the following odds ratio:

$$\log \frac{f(x_3 | x_1, x_2) f(\mu_3(x_2) | \mu_1(x_2), x_2)}{f(\mu_3(x_2) | x_1, x_2) f(x_3 | \mu_1(x_2), x_2)} = \frac{(x_3 - \mu_3(x_2))(\theta(x_1, x_2) - \theta(\mu_1(x_2), x_2))}{a(\varphi)}.$$

The above effect is derived by excluding the effect of $X_2 = x_2$, so it is defined as the direct effect of $X_1 = x_1$ on $X_3 = x_3$. Taking the expectation of the above effect, we have the (summary) direct effect of X_1 on X_3 , expressed as follows:

$$\frac{\text{cov}(\theta(X_1, X_2), X_3 | X_2)}{a(\varphi)} = KL(X_1, X_3 | X_2),$$

where $KL(X_1, X_3 | X_2)$ is defined as in (2). The above quantity is the amount of entropy of X_3 explained by X_1 alone, that is, excluding the effect of X_2 . By subtracting the direct effect of $X_1 = x_1$ on $X_3 = x_3$ from the total effect, we have the indirect effect of $X_1 = x_1$ on $X_3 = x_3$:

$$\begin{aligned} & \log \frac{f(x_3 | x_1, x_2) f(\mu_3 | \mu_1, \mu_2)}{f(\mu_3 | x_1, x_2) f(x_3 | \mu_1, \mu_2)} - \log \frac{f(x_3 | x_1, x_2) f(\mu_3(x_1) | x_1, \mu_2(x_1))}{f(\mu_3(x_1) | x_1, x_2) f(x_3 | x_1, \mu_2(x_1))} \\ & \quad - \log \frac{f(x_3 | x_1, x_2) f(\mu_3(x_2) | \mu_1(x_2), x_2)}{f(\mu_3(x_2) | x_1, x_2) f(x_3 | \mu_1(x_2), x_2)} \\ &= \frac{(x_3 - \mu_3)(\theta(x_1, x_2) - \theta(\mu_1, \mu_2))}{a(\varphi)} - \frac{(x_3 - \mu_3(x_1))(\theta(x_1, x_2) - \theta(x_1, \mu_2(x_1)))}{a(\varphi)} \\ & \quad - \frac{(x_3 - \mu_3(x_2))(\theta(x_1, x_2) - \theta(\mu_1(x_2), x_2))}{a(\varphi)}. \end{aligned} \tag{3}$$

Taking the expectation of the above effect, the (summary) indirect effect is given by

$$KL((X_1, X_2), X_3) - KL(X_2, X_3 | X_1) - KL(X_1, X_3 | X_2). \tag{4}$$

As in the previous section, to standardize the above effects by ECD, we define the standardized total, direct, and indirect effects of X_1 and X_2 on X_3 as follows:

The total effect of X_1 and X_2 on X_3 is:

$$\begin{aligned} e_T((X_1, X_2) \rightarrow X_3) &= \frac{\text{cov}(\theta(X_1, X_2), X_3)}{\text{cov}(\theta(X_1, X_2), X_3) + a(\phi)} \\ &= \frac{KL((X_1, X_2), X_3)}{KL((X_1, X_2), X_3) + 1}. \end{aligned}$$

The total effect of X_1 on X_3 :

$$\begin{aligned} e_T(X_1 \rightarrow X_3) &= \frac{\text{cov}(\theta(X_1, X_2), X_3) - \text{cov}(\theta(X_1, X_2), X_3 | X_1)}{\text{cov}(\theta(X_1, X_2), X_3) + a(\phi)} \\ &= \frac{KL((X_1, X_2), X_3) - KL(X_2, X_3 | X_1)}{KL((X_1, X_2), X_3) + 1}. \end{aligned}$$

The direct effect of X_1 on X_3 :

$$\begin{aligned} e_D(X_1 \rightarrow X_3) &= \frac{\text{cov}(\theta(X_1, X_2), X_3 | X_2)}{\text{cov}(\theta(X_1, X_2), X_3) + a(\phi)} \\ &= \frac{KL(X_1, X_3 | X_2)}{KL((X_1, X_2), X_3) + 1}. \end{aligned}$$

The indirect effect of X_1 on X_3 :

$$e_I(X_1 \rightarrow X_3) = e_T(X_1 \rightarrow X_3) - e_D(X_1 \rightarrow X_3).$$

The total (direct) effect of X_2 on X_3 :

$$\begin{aligned} e_T(X_2 \rightarrow X_3) = e_D(X_2 \rightarrow X_3) &= \frac{\text{cov}(\theta(X_1, X_2), X_3 | X_1)}{\text{cov}(\theta(X_1, X_2), X_3) + a(\phi)} \\ &= \frac{KL(X_2, X_3 | X_1)}{KL((X_1, X_2), X_3) + 1}. \end{aligned}$$

In this case:

$$e_T((X_1, X_2) \rightarrow X_3) = e_T(X_1 \rightarrow X_3) + e_T(X_2 \rightarrow X_3).$$

A general approach based on the above discussion is given below. Let X_i ($i = 1, 2, \dots, K$) be variables such that the parents of X_k are $\mathbf{X}_{\text{pa}(k)} = (X_1, X_2, \dots, X_{k-1})$ ($k = 2, 3, \dots, K$), that is, X_i precedes X_{i+1} ($i = 1, 2, \dots, K - 1$). Let $f(x_K | x_1, x_2, \dots, x_{K-1})$ be the conditional density or probability of X_K given $\mathbf{X}_{\text{pa}(K)} = (X_1, X_2, \dots, X_{K-1})$ such that:

$$f(x_K | x_1, x_2, \dots, x_{K-1}) = \exp\left\{\frac{x_K \theta - b(\theta)}{a(\varphi)} + c(x_K, \varphi)\right\}.$$

Explaining response variable X_K in a GLM framework by explanatory variables $\mathbf{X}_{pa(K)}$, the effects of the explanatory variables on the response variable can be treated in terms of entropy as discussed above. From this the standardized (summary) total effect of X_1 on X_K is defined by:

$$e_T(X_1 \rightarrow X_K) = \frac{KL(\mathbf{X}_{pa(K)}, X_K) - KL(\mathbf{X}_{pa(K)}, X_K | X_1)}{KL(\mathbf{X}_{pa(K)}, X_K) + 1}.$$

Second, the total effect of X_2 is defined as:

$$e_T(X_2 \rightarrow X_K) = \frac{KL(\mathbf{X}_{pa(K)}, X_K | X_1) - KL(\mathbf{X}_{pa(K)}, X_K | X_1, X_2)}{KL(\mathbf{X}_{pa(K)}, X_K) + 1}.$$

Then, we can find the total effects of X_i by induction, which yields:

$$e_T(X_i \rightarrow X_K) = \frac{KL(\mathbf{X}_{pa(K)}, X_K | X_{pa(i)}) - KL(\mathbf{X}_{pa(K)}, X_K | X_{pa(i+1)})}{KL(\mathbf{X}_{pa(K)}, X_K) + 1}, \quad (i = 1, 2, \dots, K - 1),$$

where $KL(X_{pa(K)}, X_K | X_{pa(i)})$ and $KL(X_{pa(K)}, X_K | X_{pa(i+1)})$ can be defined as in (2). In the above formulae, we have:

$$e_T(X_i \rightarrow X_K) \geq 0 \quad (i = 1, 2, \dots, K - 1). \tag{5}$$

and:

$$e_T(\mathbf{X}_{pa(K)} \rightarrow X_K) = \sum_{i=1}^{K-1} e_T(X_i \rightarrow X_K). \tag{6}$$

Remark 3. The total effect of $X_i = x_i$ on $X_K = x_K$ is given by:

$$\begin{aligned} & \log \frac{f(x_K | \mathbf{x}_{pa(i)}, x_i, \dots, x_{K-1}) f(\mu_K(\mathbf{x}_{pa(i)}) | \mathbf{x}_{pa(i)}, \mu_i(\mathbf{x}_{pa(i)}), \dots, \mu_{K-1}(\mathbf{x}_{pa(i)}))}{f(\mu_K(\mathbf{x}_{pa(i)}) | \mathbf{x}_{pa(i)}, x_i, \dots, x_{K-1}) f(x_K | \mathbf{x}_{pa(i)}, \mu_i(\mathbf{x}_{pa(i)}), \dots, \mu_{K-1}(\mathbf{x}_{pa(i)}))} \\ & - \log \frac{f(x_K | \mathbf{x}_{pa(i+1)}, x_{i+1}, \dots, x_{K-1}) f(\mu_K(\mathbf{x}_{pa(i+1)}) | \mathbf{x}_{pa(i+1)}, \mu_{i+1}(\mathbf{x}_{pa(i+1)}), \dots, \mu_{K-1}(\mathbf{x}_{pa(i+1)}))}{f(\mu_K(\mathbf{x}_{pa(i+1)}) | \mathbf{x}_{pa(i+1)}, x_{i+1}, \dots, x_{K-1}) f(x_K | \mathbf{x}_{pa(i+1)}, \mu_{i+1}(\mathbf{x}_{pa(i+1)}), \dots, \mu_{K-1}(\mathbf{x}_{pa(i+1)}))} \\ & = \frac{(x_K - \mu_K(\mathbf{x}_{pa(i)})) (\theta(\mathbf{x}_{pa(i)}, x_i, \dots, x_{K-1}) - \theta(\mathbf{x}_{pa(i)}, \mu_i(\mathbf{x}_{pa(i)}), \dots, \mu_{K-1}(\mathbf{x}_{pa(i)})))}{a(\varphi)} \\ & \quad - \frac{(x_K - \mu_K(\mathbf{x}_{pa(i+1)})) (\theta(\mathbf{x}_{pa(i+1)}, x_{i+1}, \dots, x_{K-1}) - \theta(\mathbf{x}_{pa(i+1)}, \mu_{i+1}(\mathbf{x}_{pa(i+1)}), \dots, \mu_{K-1}(\mathbf{x}_{pa(i+1)})))}{a(\varphi)}, \end{aligned}$$

where $\mu_k(\mathbf{x}_{pa(i)})$ and $\mu_k(\mathbf{x}_{pa(i+1)})$ be the conditional expectations of X_k given $\mathbf{X}_{pa(i)} = \mathbf{x}_{pa(i)}$ and $\mathbf{X}_{pa(i+1)} = \mathbf{x}_{pa(i+1)}$, respectively.

Let $\mathbf{X}_{pa(K)}^i = (X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_{K-1})$ be parent variables of X_K excluding X_i . The direct effect of X_i on X_K is defined by:

$$e_D(X_i \rightarrow X_K) = \frac{KL(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(K)}^i)}{KL(\mathbf{X}_{pa(K)}, X_K) + 1} \quad (i = 1, 2, \dots, K - 1).$$

From this, we have the indirect effect of X_i :

$$\begin{aligned}
 e_i(X_i \rightarrow X_K) &= e_T(X_i \rightarrow X_K) - e_D(X_i \rightarrow X_K) \\
 &= \frac{KL(\mathbf{X}_{pa(K)}, X_K | X_{pa(i)}) - KL(\mathbf{X}_{pa(K)}, X_K | X_{pa(i+1)}) - KL(\mathbf{X}_{pa(K)}, X_K | X_{pa(K)}^{/i})}{KL(\mathbf{X}_{pa(K)}, X_K) + 1} \tag{7} \\
 &\hspace{15em} (i = 1, 2, \dots, K - 2)
 \end{aligned}$$

Remark 4. The direct effect of $X_i = x_i$ on $X_K = x_K$ is given by

$$\begin{aligned}
 &\log \frac{f(x_K | \mathbf{x}_{pa(i)}, x_i, x_{i+1}, \dots, x_{K-1}) f(\mu_K(\mathbf{x}_{pa(K)}^{/i}) | \mathbf{x}_{pa(i)}, \mu_i(\mathbf{x}_{pa(K)}^{/i}), x_{i+1}, \dots, x_{K-1})}{f(\mu_K(\mathbf{x}_{pa(K)}^{/i}) | \mathbf{x}_{pa(i)}, x_i, x_{i+1}, \dots, x_{K-1}) f(x_K | \mathbf{x}_{pa(i)}, \mu_i(\mathbf{x}_{pa(K)}^{/i}), x_{i+1}, \dots, x_{K-1})} \\
 &= \frac{(x_K - \mu_K(\mathbf{x}_{pa(K)}^{/i})) (\theta(\mathbf{x}_{pa(i)}, x_i, x_{i+1}, \dots, x_{K-1}) - \theta(\mathbf{x}_{pa(i)}, \mu_i(\mathbf{x}_{pa(K)}^{/i}), x_{i+1}, \dots, x_{K-1}))}{a(\varphi)},
 \end{aligned}$$

where $\mu_i(\mathbf{x}_{pa(K)}^{/i})$ is the conditional expectation of X_i given $\mathbf{X}_{pa(K)}^{/i} = \mathbf{x}_{pa(K)}^{/i}$.

For canonical links:

$$\theta = \sum_{i=1}^{K-1} \beta_i x_i,$$

we have:

$$\begin{aligned}
 \text{cov}(\theta, X_K | \mathbf{X}_{pa(i)}) &= \text{cov}\left(\sum_{j=1}^{K-1} \beta_j X_j, X_K | \mathbf{X}_{pa(i)}\right) \\
 &= \sum_{j=i}^{K-1} \beta_j \text{cov}(X_j, X_{K-1} | \mathbf{X}_{pa(i)})
 \end{aligned}$$

and:

$$\text{cov}(\theta, X_K | \mathbf{X}_{pa(i+1)}) = \sum_{j=i+1}^{K-1} \beta_j \text{cov}(X_j, X_{K-1} | \mathbf{X}_{pa(i+1)})$$

From (5) we have:

$$\begin{aligned}
 e_T(X_i \rightarrow X_K) &= \frac{\text{cov}(\theta, X_K | \mathbf{X}_{pa(i)}) - \text{cov}(\theta, X_K | \mathbf{X}_{pa(i+1)})}{a(\varphi)(KL(\mathbf{X}_{pa(K)}, X_K) + 1)} \\
 &= \frac{\beta_i \text{cov}(X_i, X_K | \mathbf{X}_{pa(i)}) + \sum_{j=i+1}^{K-1} \beta_j (\text{cov}(X_j, X_K | \mathbf{X}_{pa(i)}) - \text{cov}(X_j, X_K | \mathbf{X}_{pa(i+1)}))}{a(\varphi)(KL(\mathbf{X}_{pa(K)}, X_K) + 1)} \tag{8} \\
 &\hspace{15em} (i = 1, 2, \dots, K - 2).
 \end{aligned}$$

The direct effect of X_i on X_K is given by:

$$e_D(X_i \rightarrow X_K) = \frac{\beta_i \text{cov}(X_i, X_K | \mathbf{X}_{pa(K)}^{/i})}{a(\varphi)(KL(\mathbf{X}_{pa(K)}, X_K) + 1)}, \tag{9}$$

and the indirect effect is calculated by (8) minus (9).

The present approach is different from the usual approach for linear equation models and from the approach in [10], because it is based on the log odds ratio and entropy by using all the variables concerned.

Remark 5. The total effects of variables by Kuha and Goldthorpe [10] are defined with the marginal distributions of response variables and explanatory variables. Meanwhile the present approach defines the total effects of explanatory variables based on a recursive structure of all the variables concerned and we have (6).

Remark 6. Indirect effects are defined by the total effects minus the direct effects as (3), (4) and (7); however the interpretation can be done in terms of entropy. On the other hand, direct and indirect effects are defined in an approach by [10], though the sum of the effects does not equal to the total effect.

Remark 7. Assessing the model identification and testing the goodness-of-fit of the model are based on the discussion of GLMs.

5. Statistical Test for Effects

Let $\widehat{KL}(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i)})$, $\widehat{KL}(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i+1)})$, and $\widehat{KL}(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(K)}^i)$ be the ML estimators of $KL(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i)})$, $KL(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i+1)})$, and $KL(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(K)}^i)$, respectively.

A similar result presented in Eshima & Tabata [16] can be used to show that:

$$n\widehat{KL}(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i)}) - n\widehat{KL}(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i+1)}) \quad (10)$$

is asymptotically distributed according to a chi-squared distribution with the degrees of freedom equal to the number of parameters in the conditional independent model with $\mathbf{X}_{pa(i)}$ minus that with $\mathbf{X}_{pa(i+1)}$.

By using statistic (10), the total effects can be tested. Similarly, the statistic:

$$n\widehat{KL}(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(K)}^i) = n\widehat{KL}(X_i, X_K | \mathbf{X}_{pa(K)}^i)$$

is asymptotically distributed according to a chi-squared distribution with degrees of freedom equal to the number of regression coefficients (parameters) related to variable X_i .

The following statistic is asymptotically distributed according to a non-central chi-squared distribution with degree of non-centrality:

$$\lambda = nKL(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i)}) - nKL(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i+1)})$$

and an appropriate degrees of freedom ν , found as the number of parameters in the conditional independent model with $\mathbf{X}_{pa(i)}$ minus that with $\mathbf{X}_{pa(i+1)}$. Let:

$$\chi_T^2 = n\widehat{KL}(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i)}) - n\widehat{KL}(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i+1)})$$

and let $c = 1 + \frac{\lambda}{\nu + \lambda}$ and $\nu' = \nu + \frac{\lambda^2}{\nu + 2\lambda}$. The statistic $\frac{\chi_T^2}{c}$ is asymptotically distributed according to

the chi-squared distribution with ν' degrees of freedom. As ν' becomes large, the chi-squared distribution tends to a normal distribution with mean ν' and variance $2\nu'$. From this, for sufficiently large sample sizes n , the statistic:

$$\frac{\chi_T^2}{n} = \widehat{KL}(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i)}) - \widehat{KL}(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i+1)})$$

is asymptotically normally distributed with mean $\frac{c\nu'}{n}$ and variance $\frac{2c^2\nu'}{n^2}$ [17]. For sufficiently large n , we have that:

$$\begin{aligned} \frac{c\nu'}{n} &\approx KL(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i)}) - KL(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i+1)}) \\ \frac{2c^2\nu'}{n} &\approx \frac{4}{n} (KL(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i)}) - KL(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i+1)})) \approx \frac{4}{n^2} \chi_T^2. \end{aligned}$$

From this, the asymptotic standard error (ASE) of $\frac{\chi_T^2}{n}$ is $\frac{2\sqrt{\chi_T^2}}{n}$. Similarly, the asymptotic standard error of:

$$\frac{\chi_D^2}{n} = \widehat{KL}(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(K)}^v)$$

is $\frac{2\sqrt{\chi_D^2}}{n}$. Moreover:

$$\frac{\chi_T^2}{n} - \frac{\chi_D^2}{n} = \widehat{KL}(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i)}) - \widehat{KL}(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i+1)}) - \widehat{KL}(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(K)}^v)$$

is asymptotically equal to a normal distribution with mean:

$$KL(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i)}) - KL(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(i+1)}) - KL(\mathbf{X}_{pa(K)}, X_K | \mathbf{X}_{pa(K)}^v)$$

and variance $\frac{4}{n^2}(\chi_T^2 - \chi_D^2)$. By using the above results, ASEs of the estimates of the summary total and direct effects can be calculated.

6. Path Analysis of the British Morbidity Data

The British mobility data described in Section 2 were analyzed in detail by using log odds ratios [10]. Here, the proposed path analysis method is applied to summarize the effects of parental class X and education Z on destination class Y , measured by log odds ratios as in the previous section, and to give a simple interpretation from the summary effects of X and Z on Y . The three variables are random, and the GLM system can be composed of logit models. In this example, the employed logistic model can be expressed as follows. Let X be a categorical factor; Z a score that take levels $\{1,2,3\}$ and $\{1,2,\dots,7\}$, respectively, and let Y be a categorical response variable with levels $\{1,2,3\}$. Let:

$$X_i = \begin{cases} 1 & (X = i) \\ 0 & (X \neq i) \end{cases} \text{ and } Y_j = \begin{cases} 1 & (Y = j) \\ 0 & (Y \neq j) \end{cases}$$

Then, dummy variable vectors $\mathbf{X} = (X_1, X_2, X_3)^T$ and $\mathbf{Y} = (Y_1, Y_2, Y_3)^T$ are identified with categorical variables X and response Y , respectively. From this, the systematic component of the above model can be expressed as follows:

$$\boldsymbol{\theta} = \boldsymbol{\alpha} + \mathbf{B}_{(1)}\mathbf{X} + \boldsymbol{\beta}_{(2)}\mathbf{Z},$$

where:

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}, \mathbf{B}_{(1)} = \begin{pmatrix} \beta_{(1)11} & \beta_{(1)12} & \beta_{(1)13} \\ \beta_{(1)21} & \beta_{(1)22} & \beta_{(1)23} \\ \beta_{(1)31} & \beta_{(1)32} & \beta_{(1)33} \end{pmatrix}, \text{ and } \boldsymbol{\beta}_{(2)} = \begin{pmatrix} \beta_{(2)1} \\ \beta_{(2)2} \\ \beta_{(2)3} \end{pmatrix}$$

Then, the logit model is described as:

$$\Pr(\mathbf{Y} = \mathbf{y} \mid \mathbf{x}, \mathbf{z}) = \frac{\exp(\mathbf{y}^T \boldsymbol{\theta})}{\sum_{\mathbf{u}} \exp(\mathbf{u}^T \boldsymbol{\theta})} = \frac{\exp(\mathbf{y}^T \boldsymbol{\alpha} + \mathbf{y}^T \mathbf{B}_{(1)} \mathbf{x} + \mathbf{y}^T \boldsymbol{\beta}_{(2)} \mathbf{z})}{\sum_{\mathbf{u}} \exp(\mathbf{u}^T \boldsymbol{\alpha} + \mathbf{u}^T \mathbf{B}_{(1)} \mathbf{x} + \mathbf{u}^T \boldsymbol{\beta}_{(2)} \mathbf{z})}$$

where $\sum_{\mathbf{u}}$ implies the summation over all \mathbf{u} . Then, from Table 4 in [10], the estimated regression parameters for men are calculated as follows:

$$\hat{\boldsymbol{\alpha}} = \begin{pmatrix} 0.60 \\ 1.20 \\ 1.79 \end{pmatrix}, \hat{\mathbf{B}}_{(1)} = \begin{pmatrix} 0.35 & -0.10 & -0.25 \\ -0.05 & 0.16 & -0.11 \\ -0.30 & -0.06 & 0.36 \end{pmatrix}, \text{ and } \hat{\boldsymbol{\beta}}_{(2)} = \begin{pmatrix} 0.32 \\ -0.07 \\ -0.25 \end{pmatrix}$$

Similarly, we have the estimated parameters for women as follows:

$$\hat{\boldsymbol{\alpha}} = \begin{pmatrix} 1.04 \\ 2.20 \\ 3.01 \end{pmatrix}, \hat{\mathbf{B}}_{(1)} = \begin{pmatrix} 0.18 & 0.03 & -0.21 \\ 0.04 & 0.00 & -0.03 \\ -0.22 & -0.03 & 0.24 \end{pmatrix}, \text{ and } \hat{\boldsymbol{\beta}}_{(2)} = \begin{pmatrix} 0.41 \\ -0.03 \\ -0.38 \end{pmatrix}$$

From Tables 1 and 5 in [10], the joint distributions of parental class X and education Z for men and women are calculated, respectively, in Table 1.

Table 1. The estimated joint distributions of parental class X and education level Z .

Sex	Parental Class X	Education Level Z						
		1	2	3	4	5	6	7
Men	S	0.038	0.015	0.009	0.053	0.050	0.047	0.082
	I	0.069	0.017	0.015	0.051	0.032	0.037	0.025
	W	0.189	0.018	0.037	0.088	0.055	0.051	0.028
Women	S	0.040	0.014	0.020	0.079	0.034	0.048	0.048
	I	0.072	0.013	0.023	0.075	0.021	0.034	0.021
	W	0.216	0.018	0.046	0.110	0.023	0.032	0.009

On the basis of the estimated parameters shown above and the estimated joint distribution of X and Z in Table 1, the joint distributions of $X, Y,$ and Z by sex can be estimated. The effects of X and Z on Y for men are shown in Tables 2–4, for example, the effects of $X = S$ and $Z = 5$ on $Y = S$ illustrated in Table 2 are as follows:

- the total effect of $X = S$ and $Z = 5$ on $Y = S$ is calculated as follows: 0.51;
- the total effect of $Z = 5$ is 0.04;
- the total effect of $X = S$ is 0.47 when $Z = 5$;
- the direct effect of $X = S$ is 0.16 when $Z = 5$;
- the indirect effect of $X = S$ is 0.31 when $Z = 5$.

Similarly, the effects of X and Z on Y for women can be calculated. The results are omitted to avoid redundancy of the discussion.

Table 2. The effects of X and Y on $Y = S$.

X	Z	Effect on $Y = S$				
		(X,Z) (total)	Z (total)	X (total)	X (direct)	X (indirect)
S	1	-0.79	-0.89	0.10	0.22	-0.12
	2	-0.49	-0.65	0.19	0.23	-0.04
	3	-0.14	-0.42	0.29	0.24	0.04
	4	0.19	-0.19	0.38	0.17	0.21
	5	0.51	0.04	0.47	0.16	0.31
	6	0.84	0.28	0.56	0.16	0.40
	7	1.16	0.51	0.65	0.11	0.55
I	1	-0.93	-0.89	-0.04	-0.04	-0.01
	2	-0.93	-0.57	-0.04	-0.05	0.01
	3	-0.28	-0.25	-0.03	-0.05	0.02
	4	0.04	0.07	-0.03	-0.04	0.02
	5	0.37	0.39	-0.02	-0.05	0.03
	6	0.69	0.71	-0.01	-0.05	0.03
	7	1.02	1.02	-0.01	-0.05	0.04
W	1	-0.99	-0.84	-0.15	0.17	-0.32
	2	-0.67	-0.45	-0.21	0.01	-0.23
	3	-0.34	-0.07	-0.23	0.03	-0.30
	4	-0.02	0.32	-0.34	0.06	-0.39
	5	0.31	0.71	-0.40	0.03	-0.43
	6	0.63	1.09	-0.46	0.03	-0.49
	7	0.96	1.48	-0.52	0.01	-0.53

Table 3. The effects of X and Y on $Y = I$.

X	Z	Effect on $Y = I$				
		(X,Z) (Total)	Z (Total)	X (Total)	X (Direct)	X (Indirect)
S	1	0.07	0.60	-0.52	-0.07	-0.45
	2	0.01	0.44	-0.43	-0.09	-0.34
	3	-0.06	0.28	-0.34	-0.09	-0.25
	4	-0.12	0.13	-0.25	-0.10	-0.15
	5	-0.19	-0.03	-0.16	-0.11	-0.04
	6	-0.25	-0.19	-0.06	-0.12	0.06
	7	-0.32	-0.34	0.03	-0.12	0.15
I	1	0.15	0.24	-0.09	0.06	-0.16
	2	0.08	0.17	-0.09	0.10	-0.18
	3	0.02	0.10	-0.08	0.10	-0.18
	4	-0.05	0.03	-0.07	0.07	-0.15
	5	-0.11	-0.05	-0.07	0.09	-0.15
	6	-0.18	-0.12	-0.06	0.08	-0.14
	7	-0.24	-0.19	-0.05	0.09	-0.15

Table 3. Cont.

X	Z	Effect on $Y = I$				
		(X,Z) (Total)	Z (Total)	X (Total)	X (Direct)	X (Indirect)
W	1	0.32	0.01	0.31	0.07	0.24
	2	0.25	0.00	0.25	0.00	0.25
	3	0.19	0.00	0.19	0.01	0.18
	4	0.12	0.00	0.13	0.01	0.12
	5	0.06	-0.01	0.07	0.00	0.06
	6	-0.01	-0.01	0.00	0.00	0.00
	7	-0.07	-0.01	-0.06	0.00	-0.05

Table 4. The effects of X and Y on $Y = W$.

X	Z	Effect on $Y = W$				
		(X,Z) (Total)	Z (Total)	X (Total)	X (Direct)	X (Indirect)
S	1	0.67	1.28	-0.62	-0.05	-0.57
	2	0.42	0.95	-0.52	-0.09	-0.44
	3	0.18	0.61	-0.43	-0.09	-0.34
	4	-0.07	0.27	-0.34	-0.07	-0.27
	5	-0.31	-0.06	-0.25	-0.09	-0.16
	6	-0.56	-0.40	-0.16	-0.10	-0.06
	7	-0.80	-0.74	-0.06	-0.08	0.01
I	1	0.74	0.65	0.09	-0.02	0.11
	2	0.49	0.39	0.10	-0.04	0.14
	3	0.25	0.14	0.11	-0.05	0.15
	4	0.00	-0.11	0.11	-0.03	0.14
	5	-0.24	-0.36	0.12	-0.04	0.16
	6	-0.49	-0.61	0.12	-0.04	0.16
	7	-0.73	-0.86	0.13	-0.04	0.17
W	1	0.64	0.40	0.24	-0.10	0.34
	2	0.39	0.21	0.18	-0.01	0.19
	3	0.15	0.03	0.12	-0.03	0.14
	4	-0.10	-0.15	0.06	-0.07	0.12
	5	-0.34	-0.34	-0.01	-0.05	0.04
	6	-0.59	-0.52	-0.07	-0.05	-0.02
	7	-0.83	-0.70	-0.13	-0.03	-0.10

The standardized summary effects are shown in Table 5. For men, the total effect of X and Z on Y is 0.276, and so 27.6% of the variation of Y's entropy is explained by X and Z. The indirect effect of X is about twice the direct effect, and the total (direct) effect of Z on Y is about 1.5-fold that of X. Therefore, the effect of education Z on the destination class Y is large. For women, the total effect of X and Z on Y is 0.289, meaning that 28.9% of the variation of Y's entropy is explained by X and Z. The indirect effect of X on Y is about 6-fold that of the direct effect, and the direct effect is small. The total effect of Z on Y is about 2.7-fold that of X. The effect of Z on Y is more pronounced for women than for men.

In a comparison of men and women, the effect of Z on Y for women is about 1.3-fold the effect for men, and, contrarily, the effect of X on Y for men is about 1.4-fold the effect for women. For both men and women, the direct effects of X on Y are mostly very small, and this decomposition of effects shows that education plays an important role in determining social class as an adult.

Table 5. Summary Direct, Indirect, and Total Effects of X and Z on Y .

Sex	Explanatory Variables	Direct Effect	Indirect Effect	Total Effect
Men	Parental Class X	0.033 (0.004) *	0.076 (0.07)	0.109 (0.008)
	Education Z	0.168 (0.010)	---	0.168 (0.010)
	(X,Z)	---	---	0.276 (0.013)
Women	Parental Class X	0.011 (0.002)	0.068 (0.06)	0.079 (0.007)
	Education Z	0.210 (0.011)	---	0.210 (0.011)
	(X,Z)	---	---	0.289 (0.012)

* The numbers in parentheses are the standard errors.

7. Discussion

In the usual path analysis of continuous variable systems, use of the regression coefficients allows straightforward calculation of total, direct and indirect effects, and the total effect can be expressed by the sum of the direct and indirect effects. However such techniques cannot be applied to structural GLMs with categorical variables or variables that are not normally distributed. Moreover, multiple variable categories make the problem more complicated in comparison with linear equation models for continuous variables. In the present paper, a path analysis approach for structural GLM models was proposed, and calculation of the direct and indirect effects was discussed. Although the analysis of effects of explanatory variables on response variables can be discussed in detail by using log odds ratios, and the effects can be interpreted as changes of relative information, the results are generally quite complicated as demonstrated in Tables 2–4. The present path analysis summarizes the effects, as measured by log odds ratios, and the standardized summary total, direct, and indirect effects are interpreted in the framework of entropy. The present path analysis approach has potential for wide application in practical data analyses of causal systems represented as GLMs, and is particularly well suited to categorical data analysis. The present study has provided a basic idea for path analysis of recursive systems with GLMs, where all the variables concerned are causally ordered, and further studies are needed for performing path analysis of more complicated recursive GLM systems and assessing spurious effects.

Acknowledgments

The authors would like to thank the two referees and the editor for their useful comments and suggestions for improving the first version of this paper. This research was supported by Grant-in-aid for Scientific Research 26330045, Ministry of Education, Culture, Sports, Science and Technology of Japan.

Author Contributions

Nobuoki Eshima and Claudio Giovanni Borroni wrote the manuscript. Nobuoki Eshima and Minoru Tabata designed the research. Nobuoki Eshima, Claudio Giovanni Borroni and Yutaka Kano carried out the present path analysis for a real dataset. All authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Asher, H.B. *Causal Modelling*; Sage Publications: Beverly Hills, CA, USA, 1976.
2. Bentler, P.M.; Weeks, D.B. Linear structural equations with latent variables. *Psychometrika* **1980**, *45*, 289–308.
3. Jöreskog, K.G.; Sörbom, D. *LISREL8: User's Reference Guide*, 2nd ed.; Scientific Software International: Chicago, IL, USA, 1996.
4. Goodman, L.A. Causal analysis of data from panel studies and other kinds of surveys. *Am. J. Sociol.* **1973**, *78*, 1135–1191.
5. Goodman, L.A. The analysis of multidimensional contingency tables when some variables are posterior to others: A modified path analysis approach. *Biometrika* **1973**, *60*, 179–192.
6. Goodman, L.A. The analysis of systems of qualitative variables when some of the variables are unidentifiable: Part I. A modified latent structure approach. *Am. J. Sociol.* **1974**, *79*, 1179–1259.
7. Goodman, L.A. Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **1974**, *61*, 215–231.
8. Hagenaars, J.A. Categorical causal modeling: Latent class analysis and directed loglinear models with latent variables. *Sociol. Methods Res.* **1998**, *26*, 436–489.
9. Eshima, N.; Tabata, M.; Geng, Z. Path analysis with logistic regression models: Effect analysis of fully recursive causal systems of categorical variables. *J. Jpn. Stat. Soc.* **2001**, *31*, 1–14.
10. Kuha, J.; Goldthorpe, J.H. Path analysis for discrete variables: The role of education in social mobility. *J. R. Stat. Soc. A* **2010**, *173*, 351–369.
11. Albert, J.M.; Nelson, S. Generalized causal mediation analysis. *Biometrics* **2011**, *67*, 1028–1038.
12. Nelder, J.A.; Wedderburn, R.W.M. Generalized linear model. *J. R. Stat. Soc. A* **1972**, *135*, 370–384.
13. McCullagh, P.; Nelder, J.A. *Generalized Linear Models*, 2nd ed.; Chapman and Hall: London, UK, 1989.
14. Eshima, N.; Tabata, M. Entropy correlation coefficient for measuring predictive power of generalized linear models. *Stat. Probab. Lett.* **2007**, *77*, 588–593.
15. Eshima, N.; Tabata, M. Entropy coefficient of determination for generalized linear models. *Comput. Stat. Data Anal.* **2010**, *54*, 1381–1389.
16. Eshima, N.; Tabata, M. Three predictive power measures for generalized linear models: Entropy coefficient of determination, entropy correlation coefficient and regression correlation coefficient. *Comput. Stat. Data Anal.* **2011**, *55*, 3049–3058.

17. Patnaik, P.B. The non-central χ^2 and F-distributions and their applications. *Biometrika* **1949**, *36*, 202–232.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).