Department of Statistics
London School of Economics

Research Report
Number 144

Path analysis for discrete variables: Education as mediator of social mobility in Britain

Jouni Kuha*
Department of Statistics
London School of Economics

John H. Goldthorpe
Nuffield College, Oxford

December 4, 2007

Abstract

A new general method for path analysis is proposed, and used to estimate direct and indirect effects in systems where some of the variables are categorical. The effects are defined in terms of expected values of a response variable averaged over distributions of explanatory variables which have the characteristics of direct and indirect effects. Because this separates the definition of the effects from the specification of models for the variables, the approach can be applied to any types of models and variables. It provides an exact additive decomposition of total effects in terms of mean differences, and one which is typically very accurate also for other measures of association such as log odds ratios. The case of a three-variable system with one intervening variable is described in detail, and extensions to more general cases are outlined. Estimates of the effects and their standard errors can be calculated easily, using standard output for fitted models. The method is applied to the problem of assessing the impact of educational attainment on intergenerational social mobility, and illustrated by an analysis of British survey data on mobility.

*Department of Statistics, London School of Economics, Houghton St, London WC2A 2AE, UK; email: j.kuha@lse.ac.uk

1 Introduction

Classical path analysis provides a convenient tool for analysing relationships between continuous variables. In its simplest case, we consider three observed variables, as represented in Figure 1. These are regarded as ordered on some substantive grounds, so that Y is preceded by X and Z, and Z is preceded by X. This is represented by statistical models specified for Z given X and for Y given X and Z.

One aim of path analysis is to try to decompose the *total effect* of X on Y into different parts. For the system of three variables, this may be divided between an *indirect effect* via the intervening (mediating) variable Z, and a *direct effect* of X on Y. These can be operationalised in terms of the models for the variables. In particular, suppose that linear models are specified for both Z and Y, with expected values $\mathrm{E}(Z|X) = \alpha_0 + \alpha_x X$ and $\mathrm{E}(Y|X, Z) = \beta_0 + \beta_x X + \beta_z Z$. Then the model for the mean of Y given X, obtained by integrating over Z, is also linear, with the coefficient

$$\gamma_x = \beta_x + \beta_z \alpha_x. \tag{1}$$

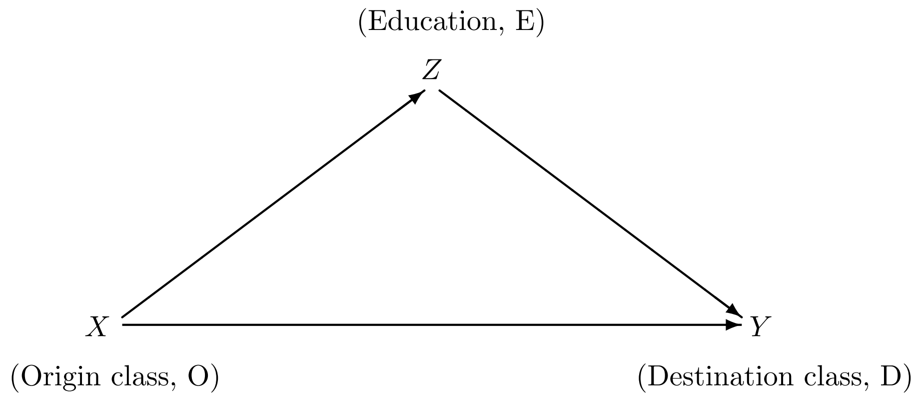
Here γ_x, β_x and $\beta_z \alpha_x$ are identified with the total, direct and indirect effects of X on Y respectively. This interpretation is intuitively obvious when we associate the coefficients with the corresponding paths in Figure 1. It also follows from more careful definitions of the effects, such as that discussed in Section 3.1 of this article. The result generalises in a straightforward way to systems of more than three variables, if linear models without interactions are used for all of them; the total effect is then again a sum of all the possible direct and indirect effects, each of which is given by the product of the coefficients along the corresponding path. These results hold also when applied to structural equation models where some or all of the variables may be latent variables measured by sets of observed indicators. (For early references to linear path analysis, see e.g. Wright 1921, 1934, 1960; Tukey 1954; Duncan 1966; and the bibliographies in Wolfle 1999, 2003; for a contemporary overview and further references, see Bollen 1989.)

It would often be convenient to have similar decompositions of effects for systems involving discrete variables for which linear models are less appropriate. For example, in this article we examine associations between a person's father's and his or her own social classes, with education in the role of the intervening variable Z. Some or all of these are typically treated as categorical variables with ordered or unordered categories.

Unfortunately, much of the simplicity of derivation and interpretation of linear path analysis is lost when the analysis involves categorical variables modelled using nonlinear models such as logistic regression models. As Fienberg (1980) notes, "there is no calculus of path coefficients" in this case, in the sense that there is no way to associate paths with regression coefficients in a way which provides an exact decomposition analogous to (1). It is not obvious in what direction the results for the linear case might best be generalised to achieve something even approximately similar for discrete variables. A number of different approaches have been proposed, as described in Section 3.6. However, none of these are fully general or satisfactory in all cases where they might be required.

A new method of path analysis for discrete and continuous variables is proposed here, and used in particular to obtain decompositions of total effects. The approach is quite general, in that it can be applied to any types of models and variables. It is motivated by, and applied to, the problem of quantifying the mediating effects of education on intergenerational social

Figure 1: The setup of basic path analysis of three variables. The names and labels of variables in parentheses refer to the social mobility example considered in the article.



mobility. This application is introduced in Section 2. The method of path analysis is described in Section 3, and applied to data on class mobility in Section 4. The article concludes with a discussion in Section 5.

Following conventional terminology, we talk of total, direct and indirect *effects* of X on Y. However, possible *causal* interpretations of these effects are not emphasised. Any such interpretation is conditional on the appropriateness of the proposed system of variables under consideration. In our mobility example, the ordering of the three variables is fairly straightforward, but the set of variables itself is not comprehensive: it is clear that class background and education are not the only factors which may affect a person's social class (indeed, even talking of the causal effect of a father's class on that of his children is somewhat contrived). The results presented here are thus best regarded in a rather descriptive spirit, as decompositions of *associations* between the variables. The indirect effect is then interpreted as that part of the total association between fathers' and their children's classes which can be accounted for by the association of education with the class variables, and the direct effect is the part of the total association which cannot be thus explained by education. An examination of the conditions under which these associations might be given firm causal interpretations is not attempted here. Such questions are examined in the literature on causal inference, where definitions of direct and indirect effects quiet similar to ours have also been proposed. This work is reviewed in Section 3.6.

2 Path analysis for education and class mobility

Path analysis with categorical variables is here used to examine the part played by education in intergenerational social mobility. In other words, how important is individuals' educational attainment in mediating the association between their own social class, or social status, and that of their parents? As well as being of academic interest in sociology, this is a question that is today of central political concern.

Consider two measures related to an individual's social status, the class of his or her father

("origin class", often denoted O below) and the individual's own current class ("destination class", D). Associations between these describe the level of intergenerational social mobility, with strong associations indicating relative lack of mobility. Suppose further that each individual's level education (E) is also recorded. We could then try to partition the total association between father's status and child's status into two additive components: (i) the indirect part that went via child's education; and (ii) the direct part that went straight from father's status to child's status without involving the intervening variable education. The variables thus have the roles of the three variables in simple path analysis as shown in Figure 1.

In the rest of this section, we first provide a brief historical review of the methods of analysis applied to this problem, to motivate the need for the kinds of methods proposed here. We then describe the data set used to illustrate the methods later in the article.

2.1 Approaches to the analysis: a historical review

In the context of intergenerational class mobility, the problem of path analysis for categorical variables has come to prominence through the long-term development of research and conceptualisation, following a four-stage process. Stage 1 is associated with the pioneering study of social mobility in Britain carried by David Glass, the first such study to be based on a probability sample of a national population (Glass, 1954). The main results of this were presented in contingency tables, initially in the form of "standard", bivariate mobility tables in which child's position is crossed with father's position (see Table 3 below for examples). Glass related child's social status to father's social status on the basis of 7 discrete, but ordered, occupational status categories, using, for the most part, simply percentaging by row or column. Education was then introduced into the analysis by creating three-way tables in which distributions of child's status were shown given combinations of father's status and of child's education, with education being also treated categorically by type of school attended. However, these three-way tables proved hard to interpret sociologically, chiefly because of the difficulty of distinguishing between two kinds of effect on counts in the tables: (i) effects of the marginal distributions of the three variables involved and (ii) effects of the net associations existing between these variables. This difficulty was a major factor prompting the next stage of development.

Stage 2 might be described as the "Duncan revolution" in mobility research, after O. D. Duncan, the American demographer and sociologist (Duncan and Hodge 1963; Blau and Duncan 1967). In this revolution, which began in the 1960s, there were three key elements. First, Duncan proposed measuring both status and education as continuous variables, status by a "socioeconomic status' index" of occupations that he himself developed, and education simply by the number of years completed. Second, he proposed a linear regression rather than a contingency table approach, in which child's occupational status was the dependent variable and father's status and child's education were the main explanatory variables. Third, Duncan took over from the geneticist Sewall Wright the method of path analysis with linear regression, which allowed the total association between father's status and child's status to be partitioned into direct and indirect components as described above. This approach proved very popular for a decade or more, but it then came under increasing criticism. The main objection was to the underlying assumption that in the "status attainment process", as it was called, the same linear regression rules applied everywhere, as e.g. in the assumption that father's status or child's education had the same effect on child's status at all levels of these two explanatory

variables. More detailed studies in fact indicated that important non-linearities and interaction effects occurred. Moreover, there was a reaction against measurement by continuous variables. On the one hand, there was a wish to shift the basis of mobility research from social status to social class, with class being treated as a categorical variable, and one that was not necessarily of a fully ordered kind. On the other hand, it was thought important to treat education in terms of qualifications achieved, rather than just years completed, and also to recognise type as well as level of qualification, so that again a categorical treatment seemed appropriate.

Stage 3 is characterised by a return to contingency table analysis in mobility research — now, typically, of bivariate tables of child's class by father's class. The analysis of such tables now moved beyond simple percentaging (or the calculation of merely ad hoc indices of mobility) to loglinear modelling. It was the introduction of this technique into sociology that provided the main methodological impetus for work in this stage (Goodman 1972; Hauser 1978). Most importantly, through loglinear modelling, the problem was effectively solved of separating marginal effects on counts in standard mobility tables from the effects of the net association between father's position and child's position. Under loglinear models, this association is expressed by odds ratios, which are not affected by changes in the marginal distributions of the two variables. Sociologists could thus compare mobility tables, over time or over place, and determine how far any differences in absolute, percentage rates were the result either of structural change, expressed through marginal effects, or of changes in the underlying set of odds ratios - that is, of changes in what became known as the "endogenous mobility regime". In addition, loglinear modelling allowed for the nature of this regime to be explored in detail, as, say, through so-called "topological" models. In other words, stage 3 amounted to a reversion to the Glass approach of stage 1, although with a more powerful statistical technology. However, in the context of the loglinear modelling of mobility tables, the stage 1 problem of how to incorporate education into the analysis re-presented itself. A method of path analysis for such tables was proposed by Goodman (1973a, 1973b; discussed further in Section 3.6 below), but this approach did not provide a way of quantifying direct and indirect components of the associations.

Stage 4 represents another shift back from contingency table to regression analysis. This involves, in effect, the rewriting of loglinear models for the grouped data of mobility tables as multinomial logistic regression models for individual-level data. Child's class is the dependent variable, and father's class and child's education are the explanatory variables, perhaps along with various others. Through this approach, a much more refined understanding than before has been gained of how class background influences children's educational attainment - one more responsive to nonlinearities and interaction effects; and, in turn, of how their educational attainment influences the class positions that children eventually achieve. However, it still remains the case that if we want to work with categorical variables and logistic regression - as most sociologists in the field would wish to do - we still have no established way, analogous to that which exists with continuous variables and linear regression, of partitioning direct and indirect effects in the "class attainment" process. This situation was noted by Winship and Mare (1983), and remains little changed today. Winship and Mare themselves described some methods for carrying out path analysis in this context, and several other approaches have been proposed elsewhere, not just for models of class mobility. These methods are reviewed in Section 3.6, after the approach proposed here has been introduced.

2.2 Example: mobility and education in Britain

Analysis of intergenerational class mobility is illustrated here using data from the British General Household Survey (GHS). The GHS is an official survey carried out continuously since 1971, using a multi-stage stratified sample of the British population living in private households. GHS data were used for an analysis of class mobility by Goldthorpe and Mills (2004), and we will mostly follow their choices on definitions of the variables and other settings. The data are analysed separately for men and women, and for surveys from 1973 and 1992, the two extremes of the range of years considered by Goldthorpe and Mills. Only respondents aged 25 or over are considered here, so that almost all of them will have completed their-full time education. The upper age limit is 59, because information on father's class was only available for respondents up to this age.

Social class is here classified using the occupation-based Goldthorpe class schema (also known as the CASMIN schema, see e.g. Erikson and Goldthorpe 1992; Goldthorpe 2007). This is typically used with a fairly detailed level of grouping, such as the nine-class version of the schema listed in Table 1 (the Roman numerals and letters in parentheses are standard labels for the CASMIN classes). For simplicity of presentation, however, we will consider only a three-class version of the schema, also shown in Table 1 (this differs slightly from the three-class schema employed by Goldthorpe and Mills, who deviated from the more common grouping used here in order to facilitate crossnational comparisons). The three classes are referred to as "Salariat and employers", "Intermediate class" and "Working class" (often abbreviated S, I and W respectively). Their distributions in the GHS data are shown in Table 1, separately for each combination of sex and year.

No ordering of the three classes is assumed, so the class variables are treated as nominal throughout. Here it would not be wholly unreasonable to consider them as ordinal instead, regarding the three classes as being in approximate order of decreasing social advantage. In general, however, classifications of social class (especially ones with larger numbers of levels) are typically best treated as unordered, as imposing an ordering on them along a single dimension would be theoretically unsatisfactory (c.f. Erikson and Goldthorpe 1992).

Each respondent's destination class is his or her own class position at the time of the interview, and origin class is the class position of the respondent's father or other family head. The destination class of respondents who were unemployed or not in the labour force is allocated based on their last employment, and the class position of married and cohabiting women is based on their own employment, without reference to that of their partners.

Level of education is characterised by the respondent's highest educational qualification. This is recorded in seven categories, as shown in Table 2. These will be treated either as unordered categories or as a seven-point interval-level scale, with the scores 1–7 shown in the first column of the table. In a few subsidiary analyses education is used instead with three categories, formed by levels 1–4, 5, and 6–7 of the seven-point scale (corresponding roughly to primary or incomplete secondary education, completed secondary education and tertiary education respectively), or as a dichotomous variable defined as levels 1–4 ("low education") vs. 5–7 ("high education").

Table 3 shows the joint distributions of the class variables, in the form of standard origin–

Table 1: Percentage distributions of the respondents' social classes of origin (O) and destination (D) in the GHS data sets. The main class variable used in the analyses is the three-category variable (with levels "Salariat", "Intermediate" and "Working"), obtained by combining categories of the nine-class version of the CASMIN class schema.

	1973				1992			
	Men		Women		Men		Women	
Class	O	D	O	D	O	D	O	D
Salariat and employers *(I+II+IVa)*	17	28	18	16	29	42	28	35
Intermediate class	21	20	21	39	25	24	26	36
Routine non-manual workers, higher grade (IIIa)	*7*	*10*	*8*	*36*	*5*	*6*	*5*	*31*
Self-employed workers (IVb)	*4*	*4*	*4*	*2*	*6*	*10*	*5*	*3*
Farmers (IVc)	*5*	*2*	*5*	*0*	*3*	*1*	*3*	*0*
Foremen and technicians (V)	*5*	*5*	*4*	*1*	*12*	*8*	*12*	*2*
Working class	62	52	61	45	46	34	46	29
Skilled manual workers (VI)	*37*	*32*	*36*	*6*	*26*	*18*	*26*	*3*
Routine non-manual workers, lower grade (IIIb)	*1*	*0*	*1*	*14*	*1*	*1*	*1*	*8*
Non-skilled manual workers (VIIa)	*20*	*18*	*20*	*24*	*18*	*14*	*17*	*18*
Agricultural workers (VIIb)	*4*	*2*	*4*	*1*	*2*	*1*	*3*	*1*
Sample size	6276		6882		4835		5284	

by-destination mobility tables. Associations in these tables are the "total effects of origin on destination" which we will try to decompose into their direct and indirect (education) components. Odds ratios or log odds ratios will be used as measures of the associations. All odds ratios in the table can be derived from any set of four nonredundant ratios. Here we consider the four local odds ratios, defined and labelled as shown in Table 4. For example, local odds ratio (1) is

$$\frac{P(\text{Destin.} = \text{Intermediate}|\text{Origin} = \text{Intermediate})/P(\text{Destin.} = \text{Salariat}|\text{Origin} = \text{Intermediate})}{P(\text{Destin.} = \text{Intermediate}|\text{Origin} = \text{Salariat})/P(\text{Destin.} = \text{Salariat}|\text{Origin} = \text{Salariat})}.$$

These ratios for the GHS data are also shown in Table 4. The association between origin and destination classes is positive for both men and women and in both of the years considered here. It is expected that this is partly due to the intervening effect of education, which is associated with both origin and destination, as shown in Table 5 (for simplicity, the three-level education variable is used here). The associations are in predictable directions and substantively strong, especially for destination class given education. The aim of the path analysis is then to examine how much of the total origin–destination association is accounted for by these associations involving education, and how much remains as an unexplained "direct effect" of origin on destination. We will return to such questions for these data in Section 4, after the methods of analysis have been introduced in Section 3.

Table 2: Percentage distributions of the respondents' levels of education in the GHS data sets. Education is used in most of the analyses as a seven-level variable.

			1973		1992	
Education			Men	Women	Men	Women
Low			84	91	59	73
	1	*None of the qualifications listed below*	*63*	*73*	*29*	*33*
	2	*CSE or lower-grade GCE O-level*	*0*	*0*	*5*	*4*
	3	*Clerical or commercial qualification, apprenticeship*	*10*	*7*	*6*	*9*
	4	*Some higher-grade GCE O-level*	*11*	*11*	*19*	*26*
Medium						
	5	*Some GCE A-level. ONC, OND.*	6	2	14	8
High			10	7	27	20
	6	*HNC, HND, non-graduate teaching or nursing qualification*	*5*	*5*	*14*	*12*
	7	*University degree or diploma, higher degree*	*5*	*1*	*13*	*8*
		Sample size	6276	6882	4835	5284

3 Path analysis for general types of variables

This section describes in general terms the method of path analysis proposed here. Returning to general notation, we will consider models for three variables X, Z, and Y, ordered as shown in Figure 1. For the moment, the models thus involve only one intervening variable Z, and no control variables on an equal footing with X. Extensions to these situations are discussed in Section 3.7.

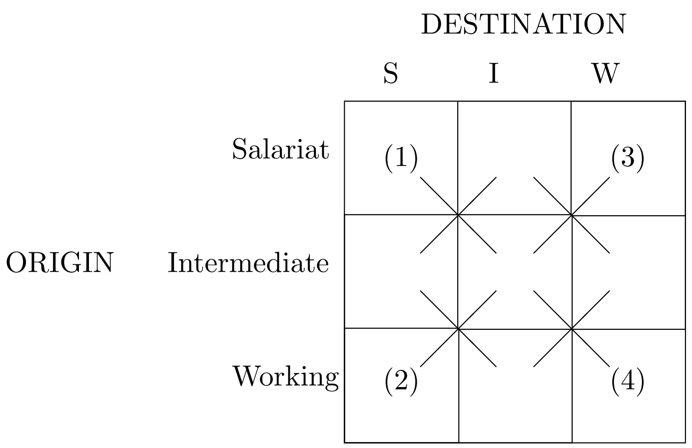
The variables can be of any type. For Y, the case where it is a categorical variable is of most interest here. For Z, the details of the calculations depend on whether the variable is categorical or continuous, so these cases are often discussed separately below. For X, only a finite (but arbitrary) set of values is considered at once. This raises some additional questions when X is continuous; these are also discussed in Section 3.7.

The numbers of possible values of categorical variables are denoted J, K, and L for X, Z, and Y respectively (J is also used for a continuous X, to denote the number of distinct values being considered). Subscripts in parentheses are used to indicate specific values; for example, the values of X being considered are $X_{(j)}$, $j = 1, \dots, J$. Dummy variables for the values of categorical variables are also used, and indicated by subscripts without parentheses. For example, $Y_l = 1$ when $Y = Y_{(l)}$ and $Y_l = 0$ otherwise. When Y is continuous, it is convenient to set $L = 1$ and redefine $Y_1 = Y$.

Table 3: Destination class-by-origin class mobility tables for the respondents in the GHS data (with row percentages in parentheses).

Year: 1973						
	Men			Women		
	Destination			Destination		
Origin	S	I	W	S	I	W
Salariat (S)	606	210	271	397	615	238
	(56)	(19)	(25)	(32)	(49)	(19)
Intermediate (I)	413	342	553	267	632	524
	(32)	(26)	(42)	(19)	(44)	(37)
Working (W)	733	718	2430	433	1463	2313
	(19)	(19)	(63)	(10)	(35)	(55)
Year: 1992						
Salariat	919	263	240	759	508	228
	(65)	(19)	(17)	(51)	(34)	(15)
Intermediate	490	361	334	519	503	342
	(41)	(30)	(28)	(38)	(37)	(25)
Working	633	542	1053	558	893	974
	(28)	(24)	(47)	(23)	(37)	(40)

Table 4: Values of the four local odds ratios for the origin class-by-destination class mobility tables for the respondents in the GHS data (c.f. Table 3). The odds ratios are labelled (1)–(4) as indicated by the picture above the table.



		Local odds ratio			
Year	Sex	(1)	(2)	(3)	(4)
1973	Men	2.39	1.18	1.25	2.09
	Women	1.53	1.43	2.14	1.91
1992	Men	2.57	1.16	1.01	2.10
	Women	1.45	1.65	1.51	1.60

Table 5: Tables of origin class vs. education and education vs. destination class for the respondents in the GHS data. The numbers in the table are row percentages (except for the row totals, shown in parentheses).

Education level given class of origin

Year: 1973

	Men				Women			
	Education				Education			
Origin	Low	Middle	High	(n)	Low	Middle	High	(n)
Salariat	64	11	25	(1087)	79	5	16	(1250)
Intermediate	81	7	12	(1308)	88	3	9	(1423)
Working	90	4	6	(3881)	96	1	3	(4209)
Year: 1992								
Salariat	39	17	44	(1422)	54	12	34	(1495)
Intermediate	61	13	26	(1185)	71	8	21	(1364)
Working	71	12	16	(2228)	85	5	10	(2425)

Class of destination given education level

Year: 1973

	Men				Women			
	Destination				Destination			
Education	S	I	W	(n)	S	I	W	(n)
Low	19	22	60	(5243)	10	41	49	(6267)
Middle	51	23	26	(377)	55	38	7	(165)
High	89	8	3	(656)	85	12	3	(450)
Year: 1992								
Low	25	28	47	(2871)	21	42	37	(3837)
Middle	47	28	25	(663)	41	45	14	(413)
High	78	14	8	(1301)	84	11	5	(1034)

The quantities defined below depend mainly on the expected values $\mathrm{E}(Y_l|X,Z)$ of Y_l ($l = 1,\ldots,L$) given X and Z, and the conditional probability or probability density function $p(Z|X)$ of Z given X. Similar notation is also used for other expectations and density functions where needed. The subscript notation defined above is used to shorten such expressions, so that, for example, $p(Z_{(k)}|X_{(j)}) = p(Z = Z_{(k)}|X = X_{(j)})$ and $\mathrm{E}(Y_l|X_{(j)}, Z_{(k)}) = p(Y_{(l)}|X_{(j)}, Z_{(k)}) = \mathrm{E}(Y_l|X = X_{(j)}, Z = Z_{(k)})$. The general ideas of the path analysis are independent of the specific types of models used for $\mathrm{E}(Y_l|X,Z)$ and $p(Z|X)$. The models which will be used for the social mobility example are defined in Section 3.2.1, and used to illustrate the details of the calculations subsequently.

Definitions of the effects considered here are given in Section 3.1 below. The calculation of their point estimates is described in Section 3.2, and of their standard errors in Section 3.3. Sections 3.4 and 3.5 describe approximate expressions for the effects in terms of regression coefficients and the accuracy of the estimates, respectively. Comparisons to other other approaches to path analysis for discrete variables are discussed in Section 3.6, and extensions to cases with more than three variables are outlined in Section 3.7.

3.1 Definitions of direct and indirect effects

As discussed earlier, the term *effect* is used here essentially to mean *association*. Thus X is said to have an effect on Y if the conditional distribution of Y is different at different values of X. Instead of the whole conditional distributions, the effects will be defined throughout in terms of such comparisons of their expected values.

Let $X_{(r)}$ and $X_{(s)}$ denote an arbitrary pair of values of X, between which the means of Y are to be compared. The *total effect* on Y of changing from $X_{(r)}$ to $X_{(s)}$ is here defined as a change in the marginal expected value

$$\mathrm{E}_r(Y_l) \equiv \mathrm{E}(Y_l|X_{(r)}) = \int \mathrm{E}(Y_l|X_{(r)}, Z)p(Z|X_{(r)})\, dZ = \mathrm{E}[\mathrm{E}(Y_l|X_{(r)}, Z)|X_{(r)}], \tag{2}$$

obtained by integrating over the conditional distribution of Z. One possible measure of total effect is then the mean difference $\mathrm{E}_s(Y_l) - \mathrm{E}_r(Y_l)$. This is an obvious choice if Y is continuous. For categorical Y, we will instead focus on the odds ratios

$$\theta(Y_l, Y_m|X_{(r)}, X_{(s)}) = \left[\frac{\mathrm{E}_s(Y_l)}{\mathrm{E}_s(Y_m)}\right] \left[\frac{\mathrm{E}_r(Y_l)}{\mathrm{E}_r(Y_m)}\right]^{-1} = \frac{p(Y_{(l)}|X_{(s)})\, p(Y_{(m)}|X_{(r)})}{p(Y_{(m)}|X_{(s)})\, p(Y_{(l)}|X_{(r)})} \tag{3}$$

or the log odds ratios $\log \theta(Y_l, Y_m|X_{(r)}, X_{(s)})$, for any pair $l, m = 1,\ldots,L$ of categories of Y. For example, the local odds ratios in Table 4 are of this form.

To motivate the definitions of direct and indirect effects considered here, it is useful to express (2) in a slightly different way, by defining

$$p_r(X) = 1 \text{ if } X = X_{(r)}, \text{ 0 otherwise, and} \tag{4}$$

$$p_r(X, Z) = p_r(X)p(Z_{(r)}|X), \tag{5}$$

so that we can write

$$\mathrm{E}_r(Y_l) = \int \mathrm{E}(Y_l|X, Z)p_r(X, Z)\, dXdZ. \tag{6}$$

In other words, the expected value of Y_l given $X = X_{(r)}$ can be thought of as $\mathrm{E}(Y_l|X,Z)$ integrated over a joint distribution of X and Z where the marginal distribution of X has a point mass at $X_{(r)}$, and the conditional distribution of Z is that of Z given $X_{(r)}$ in the observed data. Both of these thus depend on the value $X_{(r)}$ being conditioned on, and not on the value $X_{(s)}$ with which a comparison is to be made. In defining direct and indirect effects, this is changed so that only one of these distributions depends on $X_{(r)}$, while the other is independent of the choice between $X_{(r)}$ and $X_{(s)}$. For direct effects, define

$$\bar{p}_{rs}(Z|X) = [p(Z|X_{(r)}) + p(Z|X_{(s)})]/2, \tag{7}$$

$$p^D_{r(s)}(X,Z) = p_r(X)\bar{p}_{rs}(Z|X) \quad \text{and} \tag{8}$$

$$\mathrm{E}^D_{r(s)}(Y_l) = \int \mathrm{E}(Y_l|X,Z)p^D_{r(s)}(X,Z)\,dX dZ \tag{9}$$

$$= \int \mathrm{E}(Y_l|X_{(r)},Z)\left[\frac{p(Z|X_{(r)}) + p_s(Z|X_{(s)})}{2}\right] dZ \tag{10}$$

and, for indirect effects,

$$\bar{p}_{rs}(X) = 1/2 \quad \text{if } X = X_{(r)} \text{ or } X = X_{(s)}, \; 0 \text{ otherwise}$$

$$p^I_{r(s)}(X,Z) = \bar{p}_{rs}(X)p(Z|X_{(r)})$$

$$\mathrm{E}^I_{r(s)}(Y_l) = \int \mathrm{E}(Y_l|X,Z)p^I_{r(s)}(X,Z)\,dX dZ. \tag{11}$$

$$= \int \left[\frac{\mathrm{E}(Y_l|X_{(r)},Z) + \mathrm{E}(Y_l|X_{(s)},Z)}{2}\right] p(Z|X_{(r)})\,dZ \tag{12}$$

and define $\mathrm{E}^D_{s(r)}(Y_l)$ and $\mathrm{E}^I_{s(r)}(Y_l)$ similarly. The double subscripts here highlight the fact that these quantities are specific to the particular pair of values of X whose effects on Y are being compared, and each quantity depends on both of these values. Here, for example, $\mathrm{E}^D_{r(s)}(Y_l)$ denotes a quantity which will be used, when defining a direct effect, in the role of an expected value of Y_l given $X_{(r)}$, when the comparison is between $X_{(r)}$ and $X_{(s)}$.

A *direct effect* of X on Y, comparing distributions given $X_{(r)}$ and $X_{(s)}$, is here quantified in terms of comparisons of (functions of) $\mathrm{E}^D_{r(s)}(Y_l)$ and $\mathrm{E}^D_{s(r)}(Y_l)$. As shown in (7)–(10), these are expected values of Y_l averaged over two joint distributions of (X,Z), in which the marginal distributions of X are different (with probability 1 for $X_{(r)}$ or $X_{(s)}$) but the distribution of Z is the same in both, specifically the equally weighted average of its true conditional distributions given $X_{(r)}$ and $X_{(s)}$. In other words, X varies in the same way as in the quantities (4)–(6) used to define the total effect, but the distribution of Z does not change accordingly.

An *indirect effect* of X on Y is formulated as a comparison of $\mathrm{E}^I_{r(s)}(Y_l)$ and $\mathrm{E}^I_{s(r)}(Y_l)$. In these, conversely, the marginal distributions of X are the same (with equal probabilities for $X_{(r)}$ and $X_{(s)}$) but the distributions of Z differ, being the same conditional distributions as in the definition of the total effect. In other words, the distribution of Z changes as if in response to changes in X, but the marginal distribution of X itself is unchanged.

Each of these calculations involves, in the role of the distribution of either X or Z, an equally weighted average of quantities corresponding to $X_{(r)}$ and $X_{(s)}$ in the marginal expectations (6). These weights have nothing to do with the relative frequencies of $X_{(r)}$ of $X_{(s)}$ in the actual data. Such proportions do not enter into any of the calculations, which are conditional on X

throughout. Instead, the equal weights reflect the fact that the comparisons are symmetric in $X_{(r)}$ and $X_{(s)}$.

These definitions have some face validity, in that they have characteristics which we might intuitively associate with direct and indirect effects. A more formal justification for considering them is that they provide an exact additive decomposition of the total effect when it is quantified as the mean difference, i.e.

$$\begin{aligned}
\mathrm{E}_s(Y_l) - \mathrm{E}_r(Y_l) &= [\mathrm{E}^D_{s(r)}(Y_l) - \mathrm{E}^D_{r(s)}(Y_l)] + [\mathrm{E}^I_{s(r)}(Y_l) - \mathrm{E}^I_{r(s)}(Y_l)] && (13)\\
&= \int [\mathrm{E}(Y_l|X_{(s)}, Z) - \mathrm{E}(Y_l|X_{(r)}, Z)] \left[\frac{p(Z|X_{(r)}) + p(Z|X_{(s)})}{2}\right] dZ && (14)\\
&\quad + \int \left[\frac{\mathrm{E}(Y_l|X_{(r)}, Z) + \mathrm{E}(Y_l|X_{(s)}, Z)}{2}\right] [p(Z|X_{(s)}) - p(Z|X_{(r)})]\, dZ. && (15)
\end{aligned}$$

This can be shown by substituting (10) and (12) into (13), together with similar forms for $\mathrm{E}^D_{s(r)}$ and $\mathrm{E}^I_{s(r)}$.

The decomposition (13) reduces to the familiar results of classical path analysis when $\mathrm{E}(Y_l|X, Z) = \mathrm{E}(Y|X, Z)$ and $\mathrm{E}(Z|X)$ are both linear. In particular, if $\mathrm{E}(Y|X, Z) = \beta_0 + \beta_x X + \beta_Z Z$ and $\mathrm{E}(Z|X) = \alpha_0 + \alpha_x X$, we get

$$\begin{aligned}
&\text{Total effect:} & \mathrm{E}_s(Y) - \mathrm{E}_r(Y) &= (\beta_x + \beta_z\alpha_x)(X_{(s)} - X_{(r)}), && (16)\\
&\text{Direct effect:} & \mathrm{E}^D_{s(r)}(Y) - \mathrm{E}^D_{r(s)}(Y) &= \beta_x(X_{(s)} - X_{(r)}), && (17)\\
&\text{Indirect effect:} & \mathrm{E}^I_{s(r)}(Y) - \mathrm{E}^I_{r(s)}(Y) &= \beta_z\alpha_x(X_{(s)} - X_{(r)}), && (18)
\end{aligned}$$

the same result as from (1). It is important to note, however, that decomposition (13) is not limited to the linear case, but holds for any forms of $\mathrm{E}(Y_l|X, Z)$ and $p(Z|X)$. For nonlinear models, the effects are not in general equally simple expressions of regression coefficients, but they can still be easily calculated, as described in Section 3.2 below. Note also that these definitions give zero effects when expected; for example, $\mathrm{E}^I_{r(s)}(Y_l) = \mathrm{E}^I_{s(r)}(Y_l)$ and $\mathrm{E}^D_{r(s)}(Y_l) = \mathrm{E}_r(Y_l)$ when $\mathrm{E}(Y_l|X, Z) = \mathrm{E}(Y_l|X)$ or $p(Z|X_{(r)}) = p(Z|X_{(s)})$.

We propose to use the quantities defined by (7)–(12) to define estimated direct and indirect effects even when the effect of X on Y is described by some other measure of association than the mean difference. In particular, we define direct and indirect log odds ratios of the form

$$\begin{aligned}
\log\,\theta^D(Y_l, Y_m|X_{(r)}, X_{(s)}) &= \log\left[\frac{\mathrm{E}^D_{s(r)}(Y_l)}{\mathrm{E}^D_{s(r)}(Y_m)}\right] - \log\left[\frac{\mathrm{E}^D_{r(s)}(Y_l)}{\mathrm{E}^D_{r(s)}(Y_m)}\right] \quad \text{and} && (19)\\
\log\,\theta^I(Y_l, Y_m|X_{(r)}, X_{(s)}) &= \log\left[\frac{\mathrm{E}^I_{s(r)}(Y_l)}{\mathrm{E}^I_{s(r)}(Y_m)}\right] - \log\left[\frac{\mathrm{E}^I_{r(s)}(Y_l)}{\mathrm{E}^I_{r(s)}(Y_m)}\right] && (20)
\end{aligned}$$

and use these to define an approximate decomposition of the total odds ratio (3) into the sum of direct and indirect effects, i.e.

$$\log\theta(Y_l, Y_m|X_{(r)}, X_{(s)}) \approx \log\theta^D(Y_l, Y_m|X_{(r)}, X_{(s)}) + \log\theta^I(Y_l, Y_m|X_{(r)}, X_{(s)}). \qquad (21)$$

This also gives an approximate multiplicative decomposition

$$\theta(Y_l, Y_m|X_{(r)}, X_{(s)}) \approx \theta^D(Y_l, Y_m|X_{(r)}, X_{(s)}) \times \theta^I(Y_l, Y_m|X_{(r)}, X_{(s)})$$

for the total odds ratio.

The total effect can be represented either (as is mostly done here) by the sum of the direct and indirect effects, or by the corresponding bivariate sample association between X and Y. For log odds ratios, these are the right and left-hand sides of (21) respectively. Unlike (13) for the mean difference, the additive decomposition (21) is not exact, in that the two estimated total effects are not exactly equal. However, the agreement between them is typically very good, as discussed further in Section 3.5. Note also that, as log odds ratios, the terms of (21) are on the same scale as the regression coefficients of logistic models for categorical Y. This raises the possibility of obtaining for such models explicit effect decompositions in terms of the regression coefficients, comparable to (16)–(18) for linear models. This possibility, which requires further approximations, is discussed in Section 3.4.

3.2 Estimation of the effects

To obtain estimated values for these direct and indirect effects, we need first to specify and estimate models for $\mathrm{E}(Y_l|X,Z)$ and $p(Z|X)$, and then use these to calculate estimates of the quantities $\mathrm{E}^D_{r(s)}(Y_l)$ and $\mathrm{E}^I_{r(s)}(Y_l)$ defined by (10) and (12) for different r, s and l. These two steps are discussed separately below. From their results, we can calculate appropriate measures of direct and indirect effects, here in particular the log odds ratios (19) and (20). The total effect is here estimated by the sum of the direct and indirect effects. We will also use the ratio of, say, the indirect to the total effect to describe the relative sizes of the effects.

3.2.1 Models for the variables

The specification of models for Z given X and for Y given X and Z is independent of the general expressions of the previous section. For illustration, we consider in more detail models which are appropriate for the social mobility data introduced in Section 2.2. There we have categorical X and Y with $J = L = 3$ categories, which are treated as unordered. The intervening variable Z has $K = 7$ values which may be treated as discrete or as scores of a continuous variable. For Y given X and Z, we consider a multinomial logistic model

$$\begin{aligned}&\mathrm{E}(Y_l|X,Z)\\ &= p(Y_{(l)}|X,Z) = \frac{\exp(\beta_{l0}+\beta_{lx2}X_2+\cdots+\beta_{lxJ}X_J+\beta_{lz}Z)}{\sum_{l=1}^{L}\exp(\beta_{l0}+\beta_{lx2}X_2+\cdots+\beta_{lxJ}X_J+\beta_{lz}Z)} = \frac{\exp(\boldsymbol{\beta}_l'\mathbf{w})}{\sum_{l=1}^{L}\exp(\boldsymbol{\beta}_l'\mathbf{w})}\end{aligned} \tag{22}$$

where $\mathbf{w} = (1, X_2, \ldots, X_J, Z)' = (\mathbf{w}_x', Z)'$, $\boldsymbol{\beta}_l = (\beta_{l0}, \beta_{lx2}, \ldots, \beta_{lxJ}, \beta_{1z})'$ for $l = 1, \ldots, L$, and $\boldsymbol{\beta}_1 = \mathbf{0}$ for identifiability. If Z were treated as categorical instead of continuous, $\beta_{lz}Z$ would be replaced by a set of terms for the dummy variables $Z_2, \ldots, Z_K$.

For Z given X, we consider two formulations. The first is a linear model where Z given X is normally distributed with mean

$$\mathrm{E}(Z|X) = \alpha_0 + \alpha_{x2}X_2 + \cdots + \alpha_{xJ}X_J = \boldsymbol{\alpha}'\mathbf{w}_x. \tag{23}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_{x2}, \ldots, \alpha_{xJ})'$. We will later consider both heteroscedastic models where the conditional variance $\mathrm{var}(Z|X_{(j)}) = \sigma_j^2$ depends on X, and homoscedastic models where it is

constant. The former might be used because in these analyses, where the main interest is on associations involving Y, it seems desirable to keep the model for Z as unconstrained as possible. In a similar spirit, whenever Z can have only a small number of values, we might ignore their possible ordering and model their conditional probabilities simply as

$$p(Z_{(k)}|X_{(j)}) = \pi_{jk} \quad \text{for } j = 1, \dots, J;\ k = 1, \dots, K. \tag{24}$$

This is equivalent to a saturated multinomial logistic model for Z given X.

Other models than these can be considered if necessary. For example, if Y were treated as an ordinal categorical variable, a proportional odds logistic model (see e.g. Agresti 2002) could be used instead of (22). Subsequent effect calculations would then be most conveniently carried out with Y_l redefined as an indicator variable for $Y \leq l$.

Estimates of the model parameters are obtained in standard ways, and then used to calculate fitted values for $\mathrm{E}(Y_l|X, Z)$ and $p(Z|X)$. Standard (maximum likelihood) estimates for multinomial logistic and linear models are used for (22) and (23) respectively, and the estimated probabilities for (24) are

$$\hat{\pi}_{jk} = \frac{n_{jk}}{n_{j\cdot}} \tag{25}$$

where n_{jk} is observed frequency of $(X = X_{(j)}, Z = Z_{(k)})$ and $n_{j\cdot} = \sum_k n_{jk}$. Estimated variances and covariances of the parameter estimates are also obtained in the usual way. These will be used in the calculation of standard errors of the estimated effects in Section 3.3.

Subsequent calculations of the effects require from this step only parameter estimates for the models and their estimated variances and covariances. They are thus not affected by *how* these estimates were obtained. This means that the effect calculations proceed in the same way even if the parameter estimation is modified to account for particular features of the data, such as allowing for clustering of observations or incorporation of survey weights.

3.2.2 Calculating the estimated effects

The estimated direct and indirect effects are based on the quantities defined by (10) and (12) above. These involve integration over a distribution of the intervening variable Z. The details of this depend on whether Z is treated as categorical or continuous, so these two cases are discussed separately here.

Consider again a comparison between distributions of Y given two different values $X_{(r)}$ and $X_{(s)}$ of X. If Z is a categopical variable, the calculations involve only summation over its K possible values. The required quantities, for any Y_l, $l = 1, \dots, L$, are then given by

$$\mathrm{E}^{D}_{r(s)}(Y_l) = \frac{1}{2}\sum_{t=r,s}\sum_{k=1}^{K} \mathrm{E}(Y_l|X_{(r)}, Z_{(k)})p(Z_{(k)}|X_{(t)}) \tag{26}$$

$$\mathrm{E}^{I}_{r(s)}(Y_l) = \frac{1}{2}\sum_{t=r,s}\sum_{k=1}^{K} \mathrm{E}(Y_l|X_{(t)}, Z_{(k)})p(Z_{(k)}|X_{(r)}) \tag{27}$$

and similarly for $\mathrm{E}^{D}_{s(r)}(Y_l)$ and $\mathrm{E}^{I}_{s(r)}(Y_l)$. In general, we may require these quantities for several categories of Y, and for many pairs of values of X. To facilitate simultaneous calculation of

them and their estimated variances (and computer programming of such calculations), define first the $(JL) \times K$ matrix

$$\mathbf{A} = \begin{bmatrix} \mathrm{E}(Y_1|X_{(1)}, Z_{(1)}) & \ldots & \mathrm{E}(Y_1|X_{(1)}, Z_{(K)}) \\ \vdots & \ddots & \vdots \\ \mathrm{E}(Y_L|X_{(1)}, Z_{(1)}) & \ldots & \mathrm{E}(Y_L|X_{(1)}, Z_{(K)}) \\ \mathrm{E}(Y_1|X_{(2)}, Z_{(1)}) & \ldots & \mathrm{E}(Y_1|X_{(2)}, Z_{(K)}) \\ \vdots & \ddots & \vdots \\ \mathrm{E}(Y_L|X_{(J)}, Z_{(1)}) & \ldots & \mathrm{E}(Y_L|X_{(J)}, Z_{(K)}) \end{bmatrix} \tag{28}$$

(i.e. one where $\mathrm{E}(Y_l|X_{(j)}, Z_{(k)})$ is in the kth column of the $[(j-1)L + l]$th row), the $K \times J$ matrix

$$\mathbf{B} = \begin{bmatrix} p(Z_{(1)}|X_{(1)}) & \ldots & p(Z_{(1)}|X_{(J)}) \\ \vdots & \ddots & \vdots \\ p(Z_{(K)}|X_{(1)}) & \ldots & p(Z_{(K)}|X_{(J)}) \end{bmatrix} \tag{29}$$

and $\mathbf{C} = \mathrm{vec}(\mathbf{AB})$ where $\mathrm{vec}(\cdot)$ denotes the vectorization operator which creates a column matrix by stacking the columns of a matrix. Defining $p(l, j, j') = (j'-1)JL + (j-1)L + l$, $\mathbf{C}$ is a $J^2L \times 1$ vector whose $p(l, j, j')$th element is

$$\mathbf{C}[p(l, j, j')] = \sum_{k=1}^{K} \mathrm{E}(Y_l|X_{(j)}, Z_{(k)})\, p(Z_{(k)}|X_{(j')}). \tag{30}$$

With this notation, (26) and (27) can be expressed as

$$\begin{aligned} \mathrm{E}^D_{r(s)}(Y_l) &= \frac{1}{2}\,\mathbf{C}[p(l, r, r)] + \frac{1}{2}\,\mathbf{C}[p(l, r, s)] \qquad \text{and} \\ \mathrm{E}^I_{r(s)}(Y_l) &= \frac{1}{2}\,\mathbf{C}[p(l, r, r)] + \frac{1}{2}\,\mathbf{C}[p(l, s, r)]. \end{aligned}$$

All the building blocks of our direct and indirect effects are thus linear combinations of elements of $\mathbf{C}$, and any set of them we might require is obtained as $\mathbf{E} = \mathbf{MC}$ for some known matrix $\mathbf{M}$. This is estimated by $\widehat{\mathbf{E}} = \mathbf{M}\widehat{\mathbf{C}}$, where $\widehat{\mathbf{C}} = \mathrm{vec}(\widehat{\mathbf{A}}\widehat{\mathbf{B}})$ and $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$ are estimates of $\mathbf{A}$ and $\mathbf{B}$ obtained by substituting into their elements estimates for the parameters of $\mathrm{E}(Y_l|X, Z)$ and $p(Z|X)$.

If, instead, Z is continuous (and Y is categorical), the integrals involved in the calculations cannot in general be expressed in a closed form, so numerical methods or closed-form approximations are needed to evaluate them. Here we apply straightforward Monte Carlo integration (for an overview of this and other numerical methods, see e.g. Tanner 1993). Let $Z^{(r)}_1, \ldots, Z^{(r)}_M$ be IID random variates generated from $p(Z|X_{(r)})$ given estimated values of its parameters, and let $Z^{(s)}_1, \ldots, Z^{(s)}_M$ be IID draws from an estimate of $p(Z|X_{(s)})$ similarly. The required quantities are then approximated by

$$\hat{\mathrm{E}}^D_{r(s)}(Y_l) = \frac{1}{2M} \sum_{t=r,s} \sum_{m=1}^{M} \hat{\mathrm{E}}(Y_l|X_{(r)}, Z^{(t)}_m) \tag{31}$$

$$\hat{\mathrm{E}}^I_{r(s)}(Y_l) = \frac{1}{2M} \sum_{t=r,s} \sum_{m=1}^{M} \hat{\mathrm{E}}(Y_l|X_{(t)}, Z^{(r)}_m) \tag{32}$$

and similarly for $\hat{\mathrm{E}}^{D}_{s(r)}(Y_l)$ and $\hat{\mathrm{E}}^{I}_{s(r)}(Y_l)$, for all $l = 1, \ldots, L$. Here $\hat{\mathrm{E}}(Y_l|X, Z)$ denotes an estimate of $\mathrm{E}(Y_l|X, Z)$ obtained by substituting estimates for its parameters. Let $\hat{\mathbf{E}}$ denote the vector of all of these quantities. Analogously with (30), its elements are again linear combinations of the elements of a vector $\hat{\mathbf{C}} = M^{-1}\sum_m \hat{\mathbf{C}}_m$, which are of the form

$$\hat{\mathbf{C}}[p(l, j, j')] = \sum_{m=1}^{M} \hat{\mathrm{E}}(Y_l|X_{(j)}, Z_m^{(j')})\, \frac{1}{M}. \tag{33}$$

3.3 Standard errors of the estimated effects

The estimated effects are functions of $\hat{\mathbf{E}} = \mathbf{M}\hat{\mathbf{C}}$, where $\mathbf{M}$ is a known matrix and $\hat{\mathbf{C}}$ is an estimate of a vector $\mathbf{C}$ defined by (30) or (33), obtained by substituting estimates for the parameters of the models for $\mathrm{E}(Y_l|X, Z)$ and $p(Z|X)$. It is assumed throughout that these parameter estimates are asymptotically normally distributed, and that standard asymptotic variance estimates are available for them. Estimated standard errors for the estimated effects are then obtained through repeated application of the multivariate delta method (see e.g. Agresti 2002, Ch. 14, for a summary).

First, variance matrices of estimates of $\mathrm{E}(Y_l|X, Z)$ and $p(Z|X)$ are obtained. The details of this step depend on the specific models being used; as an example, they are described below for the models discussed in Section 3.2.1. Next, the variance matrix of $\hat{\mathbf{C}}$ is obtained. This depends on whether Z is continuous or categorical, so these cases will be discussed separately below. Then $\mathrm{var}(\hat{\mathbf{E}}) = \mathbf{M}\mathrm{var}(\hat{\mathbf{C}})\mathbf{M}'$ and, finally, another application of the delta method gives the variance matrix of whichever functions of $\hat{\mathbf{E}}$ are used to describe the direct and indirect effects. For example, suppose this is done using the direct and indirect log odds ratios (19) and (20) and ratios of them. To simplify the notation temporarily, let $\hat{D}$ and $\hat{I}$ denote estimates of a direct and an indirect log odds ratio respectively, and let $\hat{\mathbf{F}} = (\hat{D}, \hat{I})'$. Furthermore, reexpress these estimates as $\hat{D} = \log(\hat{E}_1/\hat{E}_2) - \log(\hat{E}_3/\hat{E}_4)$ and $\hat{I} = \log(\hat{E}_5/\hat{E}_6) - \log(\hat{E}_7/\hat{E}_8)$, where $\hat{\mathbf{E}} = (\hat{E}_1, \ldots, \hat{E}_8)$ are estimates of the matching quantities in (19) and (20). The partial derivatives of these are

$$\frac{\partial \hat{D}}{\partial \hat{E}_1} = \frac{1}{\hat{E}_1}, \quad \frac{\partial \hat{D}}{\partial \hat{E}_2} = -\frac{1}{\hat{E}_2}, \quad \frac{\partial \hat{D}}{\partial \hat{E}_3} = -\frac{1}{\hat{E}_3}, \quad \text{and} \quad \frac{\partial \hat{D}}{\partial \hat{E}_4} = \frac{1}{\hat{E}_4},$$

and $\partial\hat{I}/\partial\hat{E}_5, \ldots, \partial\hat{I}/\partial\hat{E}_8$ defined analogously. Denoting the vector of these by $\partial\hat{\mathbf{F}}/\partial\hat{\mathbf{E}}$, we then have $\mathrm{var}(\hat{\mathbf{F}}) = (\partial\hat{\mathbf{F}}/\partial\hat{\mathbf{E}})\,\mathrm{var}(\hat{\mathbf{E}})\,(\partial\hat{\mathbf{F}}/\partial\hat{\mathbf{E}})'$. Finally, consider, for example, the ratio $\hat{R} = \hat{D}/(\hat{D} + \hat{I})$. Another application of the delta method gives a formula for the variance of this, noting that $\partial\hat{R}/\partial\hat{D} = \hat{I}/(\hat{D}+\hat{I})^2$ and $\partial\hat{R}/\partial\hat{I} = -\hat{D}/(\hat{D}+\hat{I})^2$. Estimates of all of these variance matrices are then obtained by substituting estimates of the model parameters.

When the intervening variable Z is categorical, $\hat{\mathbf{C}} = \mathrm{vec}(\hat{\mathbf{A}}\hat{\mathbf{B}})$ where $\hat{\mathbf{B}}$ and $\hat{\mathbf{A}}$ are estimates of (28) and (29) respectively. Then

$$\mathrm{var}(\hat{\mathbf{C}}) = [\mathbf{B}' \otimes \mathbf{I}_{JL}]\,\mathrm{var}[\mathrm{vec}(\hat{\mathbf{A}})]\,[\mathbf{B}' \otimes \mathbf{I}_{JL}]' + [\mathbf{I}_J \otimes \mathbf{A}]\,\mathrm{var}[\mathrm{vec}(\hat{\mathbf{B}})]\,[\mathbf{I}_J \otimes \mathbf{A}]' \tag{34}$$

where $\otimes$ denotes the Kronecker product and $\mathbf{I}_m$ is the $m \times m$ identity matrix (obtaining this requires partial derivatives of $\mathrm{vec}(\hat{\mathbf{A}}\hat{\mathbf{B}})$, see e.g. Lütkepohl 1996). Note that (34) assumes that $\mathrm{cov}[\mathrm{vec}(\hat{\mathbf{A}}), \mathrm{vec}(\hat{\mathbf{B}})] = \mathbf{0}$, i.e. that estimates of the parameters of $\mathrm{E}(Y_l|X, Z)$ and $p(Z|X)$ are

asymptotically uncorrelated. This condition is satisfied for the specific models considered here, because we assume that the parameters of the two models are distinct, and use maximum likelihood estimates for them. If $\text{cov}[\text{vec}(\hat{\mathbf{A}}), \text{vec}(\hat{\mathbf{B}})] \neq \mathbf{0}$, and additional term involving it is added to (34).

In the case of a continuous Z, we have $\hat{\mathbf{C}} = M^{-1} \sum_{m=1}^{M} \hat{\mathbf{C}}_m$, where $\hat{\mathbf{C}}_m$ are vectors of estimates of $\text{E}(Y_l|X_{(j)}, Z_m^{(j')})$ over $l = 1, \ldots, L$; $j, j' = 1, \ldots, J$, and $Z_m^{(j')}$ are independent draws from an estimate of $p(Z|X_{(j')})$. Here we consider in detail only the case where the estimated $p(Z|X)$ is specified by a normal linear model, so that

$$Z_m^{(j')} = \hat{\boldsymbol{\alpha}}' \mathbf{w}_{xj'} + \epsilon_{j'm} \tag{35}$$

where $\mathbf{w}_x$ is a vector of explanatory variables (including a constant term) which depend on X (c.f. equation 23), $\hat{\boldsymbol{\alpha}}$ is a corresponding vector of estimated regression coefficients, and $\epsilon_{j'm}$ are random variates from a normal distribution with mean 0 and variance $\hat{\sigma}^2_{j'} = \hat{\text{var}}(Z|X_{(j')})$, independent across all m and j'.

Let $\hat{\boldsymbol{\theta}}$ denote estimates of the parameters of $\text{E}(Y|X, Z)$ and $\text{E}(Z|X)$, thus excluding $\hat{\sigma}^2_{j'}$ which will be treated as fixed here. The variability in $\hat{\mathbf{C}}_m$ arises from two sources, the sampling variability of $\hat{\boldsymbol{\theta}}$ and the simulation variation in $\boldsymbol{\epsilon}_m = (\epsilon_{m1}, \ldots, \epsilon_{mJ})'$. Taking both of these into account, and noting that the $\hat{\mathbf{C}}_m$ are exchangeable random variables, the variance matrix of $\hat{\mathbf{C}}$ is given by

$$\begin{aligned} \text{var}(\hat{\mathbf{C}}) &= M^{-1} \left[\sum_m \text{var}(\hat{\mathbf{C}}_m) + \sum_{t \neq u} \text{cov}(\hat{\mathbf{C}}_t, \hat{\mathbf{C}}_u) \right] \\ &= M^{-1} \text{var}(\hat{\mathbf{C}}_m) + (1 - M^{-1}) \, \text{cov}(\hat{\mathbf{C}}_t, \hat{\mathbf{C}}_u) \\ &= M^{-1} \left\{ \text{var}_\epsilon[\hat{\mathbf{C}}_m] + \text{E}_\epsilon[\text{var}_\theta(\hat{\mathbf{C}}_m)] \right\} + (1 - M^{-1}) \left\{ \text{E}_\epsilon[\text{cov}_\theta(\hat{\mathbf{C}}_t, \hat{\mathbf{C}}_u)] \right\}. \end{aligned} \tag{36}$$

Here the expressions with the subscript θ denote the asymptotic variances and covariances of $\hat{\mathbf{C}}_m$ with respect to $\hat{\boldsymbol{\theta}}$ only, treating $\boldsymbol{\epsilon}_m$ as fixed. The variances and expectations with the subscript ϵ are with respect to the distribution of $\boldsymbol{\epsilon}_m$. These can be evaluated with Monte Carlo integration, using the same simulated values of $\boldsymbol{\epsilon}$ that are used to calculate $\hat{\mathbf{C}}$ itself; for the covariance term, we have evaluated it over M pairs of distinct $\hat{\mathbf{C}}_t$ and $\hat{\mathbf{C}}_u$, for example obtained by pairing $\boldsymbol{\epsilon}_m$ with $\boldsymbol{\epsilon}_{m+1}$.

To illustrate those elements of these calculations which depend on specific models, consider first $\text{var}[\text{vec}(\hat{\mathbf{A}})]$ and $\text{var}[\text{vec}(\hat{\mathbf{B}})]$ in (34), when $\text{E}(Y|X, Z)$ and $p(Z|X)$ are defined by (22) and (24) respectively. Then $\text{vec}(\hat{\mathbf{B}}) = (\hat{\boldsymbol{\pi}}'_1, \ldots, \hat{\boldsymbol{\pi}}'_J)'$ where $\hat{\boldsymbol{\pi}}_j = (\hat{\pi}_{j1}, \ldots, \hat{\pi}_{jK})'$ and $\hat{\pi}_{jk}$ are defined by (25), and $\text{var}[\text{vec}(\hat{\mathbf{B}})]$ is block-diagonal with diagonal blocks $\text{var}(\hat{\boldsymbol{\pi}}_j) = n_{j\cdot}^{-1}[\text{diag}(\hat{\boldsymbol{\pi}}_j) - \hat{\boldsymbol{\pi}}\hat{\boldsymbol{\pi}}']$, $j = 1, \ldots, J$. For the multinomial logistic model (22), we have $\text{vec}(\hat{\mathbf{A}}) = (\hat{\boldsymbol{\tau}}'_1, \ldots, \hat{\boldsymbol{\tau}}'_T)'$ where the $\hat{\boldsymbol{\tau}}_t$ are estimates of $[\text{E}(Y_1|X_{(j)}, Z_{(k)}), \ldots, \text{E}(Y_L|X_{(j)}, Z_{(k)})]'$ for all $T = JK$ combinations of $X_{(j)}$ and $Z_{(k)}$. Let $\hat{\boldsymbol{\Theta}} = [\hat{\boldsymbol{\beta}}_1 \ldots \hat{\boldsymbol{\beta}}_L]$ where the $\hat{\boldsymbol{\beta}}_l$ are vectors of estimates of the regression coefficients in the numerator of (24) for each $l = 1, \ldots, L$ (with $\hat{\boldsymbol{\beta}}_1 = \mathbf{0}$), let $\mathbf{w}_t$ denote the vector explanatory variables, corresponding to $(X_{(j)}, Z_{(k)})$, which appear in the model, and define $\hat{\boldsymbol{\eta}}_t = \hat{\boldsymbol{\Theta}}' \mathbf{w}_t$. Then $\hat{\boldsymbol{\tau}}_t = \exp(\hat{\boldsymbol{\eta}}_t) / \mathbf{1}' \exp(\hat{\boldsymbol{\eta}}_t)$ where $\mathbf{1}$ is a unit vector and $\exp(\cdot)$ is used as an elementwise operator. Define further $\mathbf{W} = [\mathbf{w}_1 \ldots \mathbf{w}_T]$ and $\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\eta}}'_1, \ldots, \hat{\boldsymbol{\eta}}'_T)' = \text{vec}[\hat{\boldsymbol{\Theta}}' \mathbf{W}]$, and let $\mathbf{D}$ be the $LT \times LT$ permutation matrix with elements equal to 1 at locations $[(t-1)L+$

$l, (l-1)T+t]$ for $t = 1, \dots, T$, $l = 1, \dots, L$, and 0 elsewhere. Then $\mathbf{V} = \partial \text{vec}(\hat{\mathbf{A}})/\partial \hat{\boldsymbol{\eta}}'$ is a block-diagonal matrix with blocks $\partial \text{vec}(\hat{\mathbf{A}})/\partial \hat{\boldsymbol{\eta}}_t' = \text{diag}(\hat{\boldsymbol{\tau}}_t) - \hat{\boldsymbol{\tau}}_t \hat{\boldsymbol{\tau}}_t'$, $\mathbf{G} = \partial \hat{\boldsymbol{\eta}}/\partial \text{vec}(\hat{\boldsymbol{\Theta}})' = \mathbf{D}(\mathbf{I}_L \otimes \mathbf{W}')$ and, finally, $\text{var}[\text{vec}(\hat{\mathbf{A}})] = (\mathbf{VG})\text{var}[\text{vec}(\hat{\boldsymbol{\Theta}})](\mathbf{VG})'$ where $\text{var}[\text{vec}(\hat{\boldsymbol{\Theta}})]$ is the standard asymptotic variance matrix for the estimated coefficients of the multinomial logistic model.

For an example of the calculation of $\text{var}_\theta(\hat{\mathbf{C}}_m)$ and $\text{cov}_\theta(\hat{\mathbf{C}}_t, \hat{\mathbf{C}}_u)$ in (36), suppose that $\text{E}(Y|X, Z)$ is again given by (22), and that $p(Z|X)$ is specified by (35). Only a modification of the results in the discrete-Z case is then needed. First, denote $\hat{\mathbf{C}}_m = \text{vec}(\hat{\mathbf{A}}_m)$, where $\mathbf{A}_m$ is similar to matrix $\mathbf{A}$ in (28), except that its columns depend on $Z_m^{(1)}, \dots, Z_m^{(J)}$ instead of $Z_{(1)}, \dots, Z_{(K)}$. Redefining $T = J^2$ and inserting the subscript m throughout, the results of the previous paragraph up to $\mathbf{G}_m = \partial \hat{\boldsymbol{\eta}}_m/\partial \text{vec}(\hat{\boldsymbol{\Theta}}) = \mathbf{D}(\mathbf{I}_L \otimes \mathbf{W}_m')$ then hold also in this case. Now, however, $\hat{\boldsymbol{\eta}}_m$ also depends on $\hat{\boldsymbol{\alpha}}$, in that the columns of $\mathbf{W}_m$ are of the form $\mathbf{w}_{tm} = (\mathbf{w}_{xj}', \hat{\boldsymbol{\alpha}}'\mathbf{w}_{xj'} + \epsilon_{j'm})'$ (c.f. equation 35). Denoting the length of $\mathbf{w}_x$ by p_x and $p = p_x + 1$, then $\partial \hat{\boldsymbol{\eta}}_m/\partial \text{vec}(\mathbf{W}_m)' = \mathbf{I}_T \otimes \hat{\boldsymbol{\Theta}}'$, and $\partial \text{vec}(\mathbf{W}_m)/\partial \hat{\boldsymbol{\alpha}}' = \mathbf{H}$ is a $Tp \times p_x$ matrix with $\mathbf{w}_{xj}'$ in its $[(j-1)Jp + j'p]$th rows ($j' = 1, \dots, J$) for each $j = 1, \dots, J$, and zeros elsewhere. Then

$$\text{var}_\theta(\hat{\mathbf{C}}_m) = \mathbf{V}_m\{\mathbf{G}_m \text{var}[\text{vec}(\hat{\boldsymbol{\Theta}})]\mathbf{G}_m' + (\mathbf{I}_T \otimes \hat{\boldsymbol{\Theta}}')\mathbf{H}\text{var}(\hat{\boldsymbol{\alpha}})\mathbf{H}'(\mathbf{I}_T \otimes \hat{\boldsymbol{\Theta}}')'\}\mathbf{V}_m'$$

and $\text{cov}_\theta(\hat{\mathbf{C}}_t, \hat{\mathbf{C}}_u)$ is obtained similarly.

3.4 Approximate expressions

The general effect calculations outlined above do not produce expressions of direct and indirect effects in terms of the coefficients of the regression models, comparable to (17) and (18) for standard linear path analysis. Such expressions can be obtained also for nonlinear models, by employing further approximations. Although not proposed here as alternatives to the methods of Section 3.2.2 for the actual calculations, these approximations provide some further insight into the definitions of the effects, as well as links to some of the other approaches to path analysis discussed in Section 3.6.

Only the case where Y is modelled with the multinomial logistic model (22) is considered here. It will be convenient to deviate temporarily from previous notation, by allowing Z and X to be vectors, and denoting their coefficients in (22) by β_{lx} and β_{lz} respectively ($l = 1, \dots, L$). Suppose we want to decompose the log of a particular odds ratio (3), which involves two values $Y_{(l)}$ and $Y_{(m)}$ of Y. Conditioning on Y having one of these two values, it follows from (22) that

$$\begin{aligned} &\text{E}(Y_l|X, Z;\ Y = Y_{(l)} \text{ or } \ Y = Y_{(m)}) \\ &= \frac{\exp[(\beta_{l0} - \beta_{m0}) + (\beta_{lx} - \beta_{mx})X + (\beta_{lz} - \beta_{mz})Z]}{1 + \exp[(\beta_{l0} - \beta_{m0}) + (\beta_{lx} - \beta_{mx})X + (\beta_{lz} - \beta_{mz})Z]} \equiv \frac{\exp(\beta_0 + \beta_x X + \beta_z Z)}{1 + \exp(\beta_0 + \beta_x X + \beta_z Z)} \end{aligned} \tag{37}$$

which is of the binary logistic form. Suppose further that Z given X *and* ($Y = Y_{(l)}$ or $Y = Y_{(m)}$) is normally distributed, with mean $\alpha_0 + \alpha_x X$ and variance matrix Σ_z. Integrating (37) over Z we then get, approximately,

$$\text{E}(Y_l|X;\ Y = Y_{(l)} \text{ or } \ Y = Y_{(m)}) \approx \frac{\exp[\gamma_0 + a(\beta_x + \beta_z\alpha_x)X]}{1 + \exp[\gamma_0 + a(\beta_x + \beta_z\alpha_x)X]} \tag{38}$$

where $a = [1 + c^2\beta_z\Sigma_z\beta_z']^{-1/2}$, $c^2 \approx 0.346$ and $\gamma_0 = a(\beta_0 + \beta_z\alpha_0)$. This approximation, which is based on the similarity between the logistic and normal distributions (see e.g. Johnson et al.

1995, p. 119), is commonly used for such logistic-normal integrals for binary logistic models in various contexts (see e.g. Zeger et al. 1988 and Liang and Liu 1991).

Applying (38) for each $Y_{(l)}$, $l = 2, \ldots, L$, in turn against $Y_{(1)}$, and recombining the results, shows that the model for $\mathrm{E}(Y_l|X)$ is approximately of the multinomial logistic form, with parameters as in (38) (with different values for different Y_l). Expression (38) also indicates that the log of the total odds ratio (3) is here approximately $(a\beta_x + a\beta_z\alpha_x)(X_{(s)} - X_{(r)})$. Arguing by obvious analogy to (16)–(18) in the linear case, we might then further identify $a\beta_x(X_{(r)} - X_{(s)})$ with the direct log odds ratio (19) and $a\beta_z\alpha_x(X_{(r)} - X_{(s)})$ with the indirect log odds ratio (20). These are of the same form as the effects in the linear case, apart from the attenuation factor a which emerges from integrating the nonlinear expression (37) over Z.

A slightly different set of approximate expressions is obtained if we apply a similar approximation to (9) and (11) separately. These calculations are again conditional on ($Y = Y_{(l)}$ or $Y = Y_{(m)}$), although for simplicity this is not shown in the notation. Here the distributions involved can no longer be taken to be normal, but the same approximation is applied regardless to obtain simple closed-form expressions. For (9), the approximate integral is calculated over a distribution of Z with mean $\alpha_0 + \alpha_x\bar{X}_{rs}$, where $\bar{X}_{rs} = (X_{(r)} + X_{(s)})/2$, and variance $\Sigma_z + \alpha_x V_{rs}\alpha_x'$, where $V_{rs} = (X_{(r)} - X_{(s)})(X_{(r)} - X_{(s)})'/4$ for both $X_{(r)}$ and $X_{(s)}$. This gives $\mathrm{E}^D_{r(s)}(Y_l) \approx \{1 + \exp[-(\gamma^D_{0(rs)} + a^D_{rs}\beta_x X_{(r)} + a^D_{rs}\beta_z\alpha_x\bar{X}_{rs})]\}^{-1}$ where $a^D_{rs} = [1 + c^2(\beta_z\Sigma_z\beta_z' + \beta_z\alpha_x V_{rs}\alpha_x'\beta_z')]^{-1/2}$ and $\gamma^D_{0(rs)} = a^D_{rs}(\beta_0 + \beta_z\alpha_0)$. For (11), consider an integral over a joint distribution of independent X and Z where the distribution of Z is the same as in the calculation of (38), and X has mean $\bar{X}_{rs}$ and variance V_{rs}. This gives $\mathrm{E}^I_{r(s)}(Y_l) \approx \{1 + \exp[-(\gamma^I_{0(rs)} + a^I_{rs}\beta_x\bar{X}_{rs} + a^I_{rs}\beta_z\alpha_x X_{(r)})]\}^{-1}$ where $a^I_{rs} = [1 + c^2(\beta_z\Sigma_z\beta_z' + \beta_x V_{rs}\beta_x')]^{-1/2}$ and $\gamma^I_{0(rs)} = a^I_{rs}(\beta_0 + \beta_z\alpha_0)$. Substituting these to (19) and (20) gives the approximations

$$\log\theta^D(Y_l, Y_m|X_{(r)}, X_{(s)}) \approx a^D_{rs}\beta_x(X_{(s)} - X_{(r)}) \quad \text{and} \tag{39}$$

$$\log\theta^I(Y_l, Y_m|X_{(r)}, X_{(s)}) \approx a^I_{rs}\beta_z\alpha_x(X_{(s)} - X_{(r)}) \tag{40}$$

for the direct and indirect log odds ratios respectively. These are again of the same form as the effects in linear path analysis. Unlike the expressions obtained from (38), (39) and (40) involve slightly different attenuation factors (a^D_{rs} and a^I_{rs}) for the two effects.

These expressions could in principle be used to obtain estimated effect decompositions, by substituting sample estimates for the parameters. An example of this is given in the next section. In general, however, the more general and less approximate formulas of Section 3.2 are preferable for the actual calculations.

3.5 Accuracy of the effect decompositions

We have argued that the quantities proposed here seem appropriate to the extent that their forms correspond to what we would expect of direct and indirect effects. In addition, there should be practical criteria for assessing the adequacy of the effect decompositions. Here we do this by comparing the two definitions of the total effect, i.e. as the bivariate association between X and Y (called "marginal total effect" below) or as the sum of the measures of direct and indirect effects ("estimated total effect"). These are equal for effects defined as mean differences, a result which includes that of standard linear path analysis. The same is not true

for other measures of association, such as for the decomposition (21) of log odds ratios. Even then, however, we would want the estimated total effect to be fairly close to the marginal one. While exact agreement may not be essential, a large disparity would clearly compromise the interpretability of the proposed effect decomposition.

We carried out a brief investigation of these quantities for binary variables over a range of parameter values. This suggested that largest disparities between marginal and estimated total log odds ratios occur when the models involve very strong nonlinearities, i.e. large associations or interactions, or probabilities close to 0 or 1. For any observed set of data, it is of course easy to compare the two estimates of the total effect. In the analyses considered here their differences are very small, as shown below and in Chapter 4. This provides reassurance of the adequacy of the results of the path analyses for these data.

The accuracy of the effect decompositions in any particular analysis depends partly on how the quantities $\mathrm{E}^{D}_{r(s)}(Y_l)$ and $\mathrm{E}^{I}_{r(s)}(Y_l)$ from (10) and (12) are specified and estimated. A fairly minor part of this is the accuracy of their estimates given specific models for $\mathrm{E}(Y|X, Z)$ and $p(Z|X)$. If the intervening variable Z is categorical, the estimates (26) and (27) introduce no new sources of systematic or random errors, in addition to those of the estimates of the model parameters themselves. With continuous Z, the Monte Carlo integration in (31) and (32) adds some simulation variation, but this can be reduced by increasing the number of simulations. For example, its contribution was negligible in the the estimates reported below, which used $M = 100,000$ simulations. Finally, larger and potentially systematic errors would be incurred if we used the kinds of first-order approximations discussed in Section 3.4. This is done here only for illustrative purposes.

The estimated effects also depend on how the models for $\mathrm{E}(Y|X, Z)$ and $p(Z|X)$ are specified. This is is examined in Table 6, using for illustration the 1992 samples of the mobility data introduced in Section 2.2 (these are analysed in more detail in Section 4). Here the effects are logs of the local odds ratios defined in Table 4. The upper part of Table 6 shows estimates of the total effect of X (origin class) on Y (destination class), comparing marginal total effects (i.e. logs of the sample odds ratios in Table 4) to estimated total effects obtained under various model specifications. Least restrictive of these is the saturated model (numbered [1] in the table), which specifies $\mathrm{E}(Y|X, Z)$ as a multinomial logistic model which includes an interaction between X and Z, and $p(Z|X)$ as (24) for the seven-category education variable Z given origin class X. The other estimates in the table are obtained using more parsimonious models. In model [2], the interaction of Z and X (which is in fact not significant for these data) is omitted from the model for Y (this is used also for Table 8 of Section 4). Models [3] and [4] simplify this further by using a linear model for Z, with the variance of Z depending on X in [3] but not in [4] (the latter is also used for Table 9). Finally, results [5] are obtained by applying the first-order approximations of Section 3.4.

The estimated total effects from the saturated model [1] are very close to the marginal total effects. While the differences are slightly larger for some of the more parsimonious models, they are still very small for all of them, even for the rather rough approximations used for [5]. Estimates of the indirect effects, which are shown in the lower part of Table 6, are also fairly similar for all model specifications, although some of these differences are slightly larger than for the total effects. In particular, the approximate estimates [5] are consistently smaller than the others; the greater accuracy of these for the total effects is in fact due to a corresponding slight overestimation of the direct effects.

Table 6: Estimated total and indirect effects from path analysis decompositions of local log odds ratios (c.f. the labels in Table 4) between origin and destination classes in the 1992 GHS data, based on different model specifications. See the text for more details.

	Local odds ratio							
	Men				Women			
Total effect	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Marginal	.95	.15	.01	.74	.37	.50	.42	.47
Estimated								
[1] Saturated model	.93	.15	.03	.73	.37	.50	.42	.47
[2] No interaction of X and Z	.93	.15	.01	.74	.37	.50	.41	.47
[3] Heteroscedastic linear model for Z	.92	.15	.03	.73	.36	.50	.42	.46
[4] Homoscedastic linear model for Z	.92	.15	.01	.74	.37	.46	.41	.49
[5] Approximation of S. 3.4	.91	.18	.01	.74	.36	.51	.43	.47
Indirect effect								
Estimated								
[1] Saturated model	.37	.20	.18	.10	.31	.33	.22	.24
[2] No interaction of X and Z	.36	.20	.18	.08	.30	.34	.24	.22
[3] Heteroscedastic linear model for Z	.34	.19	.18	.08	.28	.34	.24	.22
[4] Homoscedastic linear model for Z	.34	.19	.17	.09	.30	.30	.23	.24
[5] Approximation of S. 3.4	.31	.14	.12	.06	.26	.31	.23	.19

3.6 Other approaches

This section reviews other methods of path analysis of discrete variables which have been proposed in the literature. The focus is on approaches which also provide estimates of direct and indirect effects. A prominent method which does not do this is that proposed by Goodman (1973a; 1973b). It defines a joint model for the variables as the union of log-linear models for a sequence of marginal tables corresponding to the proposed ordering of the variables. For the system in Figure 1, this would involve one model for the $X \times Z$ marginal table and another for the full $X \times Z \times Y$ table, the latter being saturated for the $X \times Z$ margin. These can also be expressed as logistic models like (22), for Z given X and for Y given (X, Z). The approach proposed here can thus be seen as a way of associating effect estimates with Goodman's path-analytic models.

One way of retaining the convenient results of linear path analysis is simply to use linear models also for probabilities of categorical response variables. These may be estimated using weighted or unweighted least squares estimation (Davis 1975, 1980; Hellevik 1983, 1984), or with maximum likelihood estimation for multinomial responses. Direct and indirect effects are then defined by the coefficients of the models in the usual way, and interpreted in terms of differences in the response probabilities. The usefullness of the approach thus depends partly on whether such differences are regarded as appropriate measures of association for the categorical responses. Even when they are, however, the method proposed here also provides a way (c.f. equation 13) of decomposing such differences, even when the probabilities themselves are specified through standard nonlinear models.

Another apparently obvious approach would be to define effects as regression coefficients and their differences, analogously with the linear case. For instance, we could define the total effect as the coefficient of X in a logistic model for Y given X, the direct effect as its coefficient in a similar model given both X and Z, and the indirect effect as their difference. For example, this approach was used for the analysis of social mobility data by Breen and Luijkx (2004). The results, however, will be misleading for nonlinear models. This can be seen by considering the approximate model (38). There the total and direct effects from the differencing approach would estimate (approximately) $a(\beta_x + \beta_z \alpha_x)$ and β_x respectively. The proposed total effect then involves an attenuation factor a, while the direct effect β_x is unattenuated. The two are thus expressed on different scales, the distinction of which is essentially that of marginal (population-averaged) and conditional (within-subject) coefficients of an explanatory variable in models for clustered categorical response data (see e.g. Agresti 2002). The relative size of the direct effect is then inevitably overestimated.

A yet another way of translating a discrete-variable analysis into one for continuous variables is to use a latent-variable motivation of models for categorical responses (see e.g. Agresti 2002). This was proposed for systems of simultaneous equations by Heckman (1978), and applied to path analysis of recursive systems by Winship and Mare (1983) and Xie (1989). In this approach, each categorical response variable is regarded as a single indicator of a continuous latent variable, related to it by a known (deterministic) measurement model. A set of linear models is then specified for the latent variables (given X) as in standard path analysis, and effects are defined by the coefficients of these linear structural models. A limitation of this approach is that it can only be used when the categorical response variables are binary or ordinal. This makes it inapplicable to problems where the categories of a variable are unordered,

or where they may be treated as ordered but an interpretation in terms of an underlying latent continuum is substantively unappealing. The social mobility example considered here is one such application.

Winship and Mare (1983) also consider the situation where a (binary) discrete intervening variable Z itself rather than a corresponding latent variable affects a continuous outcome variable Y^*, which may be a latent variable corresponding to a binary observed outcome Y. In this case they consider $\partial \mathrm{E}[Y^*|X, \mathrm{E}(Z|X)]/\partial X$, identifying its two parts with the direct and indirect effects. This gives effectively the same result for the effects of X on Y^* as the approximations of Section 3.4, when we linearise $\mathrm{E}(Z|X)$ using a first-order Taylor series approximation. For the effects of X on Y, on the other hand, Winship and Mare's results omit the attenuation factor arising from the integration over Z in a nonlinear model (this, however, does not affect the relative sizes of the estimated direct and indirect effects).

General formulations of direct and indirect effects which are most closely related to those described here have been proposed in the literature on causal inference for experimental and observational studies, especially in epidemiology. In addition to the ones discussed below, key references include Robins and Greenland (1992), Pearl (1998), Ten Have et al. (2005), van der Laan and Petersen (2005), and Petersen et al. (2006). All of these go beyond the aims of this paper in that they consider the effects in causal terms, and examine conditions for when they may be interpreted as true causal effects.

Working in a decision-theoretic causal inference framework, Didelez et al. (2006) and Geneletti (2007) define effects for general path analysis as follows. In the notation used here, they first define the total effect as $\mathrm{E}_s(Y_l) - \mathrm{E}_r(Y_l)$, as in (13) above. A *standardised direct effect* (called "generated direct effect" in Geneletti 2007) is then defined as

$$\int [\mathrm{E}(Y_l|X_{(s)}, Z) - \mathrm{E}(Y_l|X_{(r)}, Z)]\, p^*_{rs}(Z)\, dZ \tag{41}$$

where $p^*_{rs}(Z)$ is a distribution of Z which is the same for both values of X being compared. Our definition of a direct effect, which for mean differences is given by (14), is clearly a special case of this, with $p^*_{rs}(Z) = [p(Z|X_{(r)}) + p(Z|X_{(s)})]/2$. The special case of (41) considered in most detail by Didelez et al. (2006) and Geneletti (2007), and previously, in a counterfactual framework for causal inference, by Pearl (2001) and Robins (2003), is the *natural direct effect*. It is obtained with $p^*_{rs}(Z) = p(Z|X_{(r)})$, i.e.

$$\int [\mathrm{E}(Y_l|X_{(s)}, Z) - \mathrm{E}(Y_l|X_{(r)}, Z)]\, p(Z|X_{(r)})\, dZ. \tag{42}$$

This thus treats $X_{(r)}$ (say) as a reference value, in a way which makes the definition of the effect asymmetric in the values of X. This is most natural in contexts where the reference value corresponds to a control condition of absence of treatment or exposure. It is less appropriate when, as in our social mobility example, many pairs of values of X may be compared, and there is no obvious reference value in any strong substantive sense.

For indirect effects, Geneletti (2007) defines them as the difference between the total effect and the direct effect defined by (41). With the natural direct effect (42), this gives the "natural indirect effect"

$$\int \mathrm{E}(Y_l|X_{(s)}, Z)\, [p(Z|X_{(s)}) - p(Z|X_{(r)})]\, dZ. \tag{43}$$

Note that this fixes the value of X in $\mathrm{E}(Y_l|X, Z)$ (corresponding to removing the direct effect) at $X_{(s)}$, while in the corresponding direct effect (42) it is fixed at $X_{(r)}$. An exact additive decomposition of the total effect into the sum of (42) and (43) is thus achieved with a definition which is asymmetric in X in two ways, in that the result depends on the choice of the reference value, and that X is fixed at different values in the direct and indirect effects. The latter is avoided if the indirect effect corresponding to (42) is defined with $X_{(r)}$ instead of $X_{(s)}$ in the first term of (43); this is referred to as the "natural indirect effect" by Pearl (2001) and as the "pure indirect effect" by Robins (2003) (who calls (43) the "total indirect effect"). With this definition, on the other hand, the sum of the direct and indirect effects is no longer exactly equal to the total effect (a property which, however, is arguably not crucial). By comparison, the approach proposed here defines the indirect effect for mean differences as (15). Paired with (14), this gives a definition of the effects which both provides an exact partitioning of the total mean difference and is unaffected by the choice of a reference level for X. This makes these effect quantities particularly convenient for the descriptive purposes considered here.

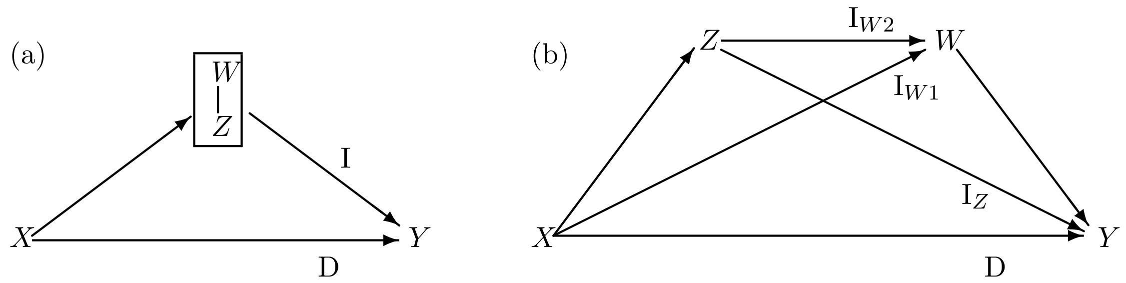
3.7 Extensions to more variables

So far, only problems with the three variables X, Z and Y have been considered. This can be extended in a number of ways. First, suppose there are variables W which are considered on an equal footing with X, so that they should be controlled for in models for Z and Y. The easiest way to extend the analysis to this case is to calculate the path analysis decompositions of the effect of X on Y using estimates of $p(Z|X, W)$ and $\mathrm{E}(Y|X, Z, W)$, conditional on specific values of W. If the models include interactions involving W, this should produce different results at different values of W; if not, we would except W to have little or no influence on the effects of X on Y. For linear models without such interactions, all terms involving W cancel out in the effect calculations. The same is true for decompositions of log odds ratios for categorical Y, if we use the approximate logistic models discussed in Section 3.4. Our analyses of the social mobility data suggest that the results of the more accurate effect calculations of Section 3.2 are also insensitive to W in this respect. Specifically, we fitted models similar to those in Table 9 of Section 4 (here with $M = 10,000$), but with the respondent's age as an additional control variable. All the results of the analyses were essentially unaffected by varying the value at which age was controlled across its observed range.

Similar considerations apply when X itself is continuous. The estimated effects described above are calculated for a comparison of specific pairs of values of X, so several of these would in principle be needed to cover its whole range. Doing so is essential if the model for Y includes and interaction of X and Z, so that the effects will be different at different values of X. Without an interaction, the results might again be insensitive to the specific values of X considered, at least as long as the association between X and Y is not extremely strong. This is again supported by examples with the mobility data, treating a nine-level version of the origin class variable as a continuous X. It is straightforward to check whether this is the case in a particular application, by calculating and comparing effects at selected values over the range of X.

A more important extension is to cases with more than one intervening variable. Suppose there are two of them, denoted Z and W. We may distinguish two ways of treating the intervening variables, as represented in Figure 2. In the first case, corresponding to diagram (a), Z and W are considered on an equal footing, and we do not aim to distinguish their contributions to

Figure 2: Graphical representations of two cases of path analysis with two intervening variables. See the text for more details.



the indirect effect. The previous results then hold unchanged, using the joint variable (Z, W) in the role of Z throughout. The joint distribution of (Z, W) given X may be specified in any convenient manner, for example through $p(Z, W|X) = p(Z|X)p(W|X, Z)$. The indirect effect then represents the effect due to the associations of X and Y with (Z, W) jointly.

In the second case, represented by diagram (b), an ordering is specified for the intervening variables, here that Z precedes W. It then becomes meaningful to try to divide the indirect effect of X on Y via Z and W into three parts: one (indicated by I_Z in the figure) via Z alone and two via W, corresponding to the dependence of the distribution of W on X directly (I_{W1}) and via Z (I_{W2}). Consider again two values $X_{(r)}$ and $X_{(s)}$ of X, and denote $\mu_r = \mathrm{E}(Y_l|X_{(r)}, Z, W)$, $\alpha_r(Z) = p(Z|X_{(r)})$ and $\beta_r(Z) = p(W|X_{(r)}, Z)$, and define

$$\gamma_r = p(W|X_{(r)}) = \int p(W|X_{(r)}, Z)p(Z|X_{(r)})\, dZ \equiv \int \beta_r(Z)\alpha_r(Z)\, dZ. \tag{44}$$

A convenient effect decomposition for this case is obtained by specifying the total effect in terms of comparisons of

$$\mathrm{E}_r^T(Y_l) = \int \mu_r\gamma_r\alpha_r(Z)\, dZdW, \tag{45}$$

rather than $\mathrm{E}_r(Y)$, i.e. replacing $\beta_r(Z) = p(W|X_{(r)}, Z)$ in $\mathrm{E}_r(Y_l)$ (c.f. equation 2) with $\gamma_r = p(W|X_{(r)})$. Similarly, let us define

$$\begin{aligned}
\mathrm{E}_{r(s)}^{D*}(Y_l) &= \int \mu_r \frac{1}{2}[\gamma_r\alpha_r(Z) + \gamma_s\alpha_s(Z)]\, dZdW, \\
\mathrm{E}_{r(s)}^{Z}(Y_l) &= \int \bar{\mu}\,\bar{\gamma}\,\alpha_r(Z)\, dZdW, \\
\mathrm{E}_{r(s)}^{W1}(Y_l) &= \int \bar{\mu}\,[\int \beta_r(Z^*)\bar{\alpha}(Z^*)\, dZ^*]\,\bar{\alpha}(Z)\, dZdW, \\
\mathrm{E}_{r(s)}^{W2}(Y_l) &= \int \bar{\mu}\,[\int \bar{\beta}(Z^*)\alpha_r(Z^*)\, dZ^*]\,\bar{\alpha}(Z)\, dZdW,
\end{aligned}$$

where a bar indicates an average of quantities depending on $X_{(r)}$ and $X_{(s)}$; for example, $\bar{\alpha} = (\alpha_r + \alpha_s)/2$. Intuitively, these quantities again seem to have the characteristics which we could associate with specific direct and indirect effects. For example, in $\mathrm{E}_{r(s)}^{W1}$ only that part of the distribution of W which depends directly on X is conditional on a specific value of X, while everything else is averaged over the two values being compared. Defining similarly quantities

subtitled $s(r)$, we obtain the exact decomposition

$$\begin{aligned} \mathrm{E}_s^T(Y_l) - \mathrm{E}_r^T(Y_l) &= [\mathrm{E}_{s(r)}^{D*}(Y_l) - \mathrm{E}_{r(s)}^{D*}(Y_l)] + [\mathrm{E}_{s(r)}^{Z}(Y_l) - \mathrm{E}_{r(s)}^{Z}(Y_l)] \\ &\quad + [\mathrm{E}_{s(r)}^{W1}(Y_l) - \mathrm{E}_{r(s)}^{W1}(Y_l)] + [\mathrm{E}_{s(r)}^{W2}(Y_l) - \mathrm{E}_{r(s)}^{W2}(Y_l)] \end{aligned} \quad (46)$$

where the four mean differences on the right hand side of (46) correspond to the effects D, I_Z, I_{W1} and I_{W2} in Figure 2(b) respectively. The quantities can again be used to derive comparable decompositions for, say, log odds ratios. The calculations are a straightforward extension of those in Section 3.2, and involve only estimates of $\mathrm{E}(Y|X, Z, W)$, $p(Z|X)$ and $p(W|X, Z)$. The sum of the last two differences in (46) is the difference of quantities of the form $\mathrm{E}_{r(s)}^{W}(Y_l) = \int \bar{\mu}\gamma_r\bar{\alpha}(Z)\, dZdW$, which represents the combined indirect effect from the two paths via W in Figure 2(b). Similarly, the last three differences in (46) add up to the difference of quantities of the form $\mathrm{E}_{r(s)}^{I*}(Y_l) = \int \bar{\mu}\gamma_r\alpha_r(Z)\, dZdW$, which combines the indirect effects through all paths via Z and W.

The decomposition (46) is obtained by using $\gamma = p(W|X)$ instead of $\beta = p(W|X, Z)$ in the calculations. In other words, this involves ignoring the association of Z and W given X, and defining the total effects being decomposed in terms of the expectations (45) rather than the actual conditional expectations $\mathrm{E}(Y|X)$ defined by (2). A special case where both approaches give identical results, i.e. where the conditional association of Z and W can be ignored without loss, is one where the model for $\mathrm{E}(Y|X, Z, W)$ is linear in Z and W, and involves no interaction between them. More generally, the adequacy of decomposition (46) and its generalisations remains to be investigated. However, it seems plausible that it may be reasonable at least when the model for Y implies a model with no strong interaction between Z and W on the scale of $\mathrm{E}(Y|X, Z, W)$. This result is also obtained, for example, by considering the approximate expression (38) for a multinomial logistic model (37), letting Z in (37) stand for the vector (Z, W). Here their covariance given X contributes only to the attenuation coefficient a.

4 Application to British mobility data

Parameter estimates for models for the British GHS data introduced in Section 2.2 are shown in Table 7. The first column shows coefficients for linear regression models for education (treated as a seven-level continuous variable, c.f. Table 2) given origin class (in three categories, c.f. Table 1). Here "I vs. S" denotes the expected difference in education between Intermediate and Salariat origin classes, and "W vs. I" a contrast between working and intermediate classes similarly. These are thus estimates of the coefficients α_{xj} in (23), defined to correspond to the contrasts of origin classes in the local odds ratios shown in Table 4. When education is treated as a categorical seven-level variable, estimates of π_{jk} in (24) are its conditional probabilities given origin class (not shown here), comparable to those in the upper part of Table 5 (where education is shown with three levels).

The remaining columns of Table 7 show estimated coefficients for multinomial logistic models for destination given origin and education (treated as a continuous variable). These are estimates of β_{lxj} and β_{lz} in (22), again defined so that they correspond to the local log odds ratios. The estimates in the "I vs. S" column are log odds ratios for the destination class being Intermediate rather than Salariat (corresponding to odds ratios (1) and (2)), and those in the "W vs. I" column log odds ratios between Working and Intermediate destination classes

Table 7: Estimated coefficients (with standard errors in parentheses) for regression models for education given origin class and for destination class given origin and education in the GHS data. The categories of the class variables are abbreviated S, I and W for Salariat, Intermediate and Working classes respectively. See the text for an explanation of the models.

			Response variable					
		Explanatory			Destination class			
Year	Sex	variable	Education		I vs. S		W vs. I	
1973	Men	Origin class						
		I vs. S	-1.06	(.08)	.61	(.12)	.14	(.12)
		W vs. I	-.56	(.06)	-.01	(.10)	.70	(.08)
		Education	–		-.42	(.02)	-.21	(.02)
	Women	Origin class						
		I vs. S	-.80	(.06)	.19	(.11)	.58	(.10)
		W vs. I	-.67	(.05)	.02	(.10)	.47	(.07)
		Education	–		-.45	(.02)	-.58	(.03)
1992	Men	Origin class						
		I vs. S	-1.09	(.08)	.67	(.10)	-.12	(.12)
		W vs. I	-.59	(.07)	.03	(.10)	.69	(.09)
		Education	–		-.39	(.02)	-.19	(.02)
	Women	Origin class						
		I vs. S	-.84	(.07)	.11	(.09)	.21	(.11)
		W vs. I	-.89	(.06)	.23	(.09)	.30	(.09)
		Education	–		-.44	(.02)	-.35	(.02)

similarly (for odds ratios (3) and (4)). Several of the partial effects of origin on destination are not statistically significant, while those of education on destination (as well as of origin on education above) are all strongly significant.

Table 8 shows results of the estimated path analysis decompositions of the local log odds ratios between origin and destination classes. Here education is treated as a nominal categorical variable in the model for education given origin, and as continuous in the model for destination. The first line ("Marginal total effect") shows the estimated log odds ratios from a model for destination given origin, i.e. the logs of the odds ratios in Table 4. The "Estimated total effects" are sums of the estimated direct effect (19) and indirect effect (20), which are also shown in the table. The accuracy of the decomposition is examined with the difference between the marginal and estimated total effects and the exponential of this difference, expressed as a percentage difference of the estimated from the marginal odds ratio. These disparities are all small, the largest proportional difference between marginal and estimated odds ratios being 1.6%.

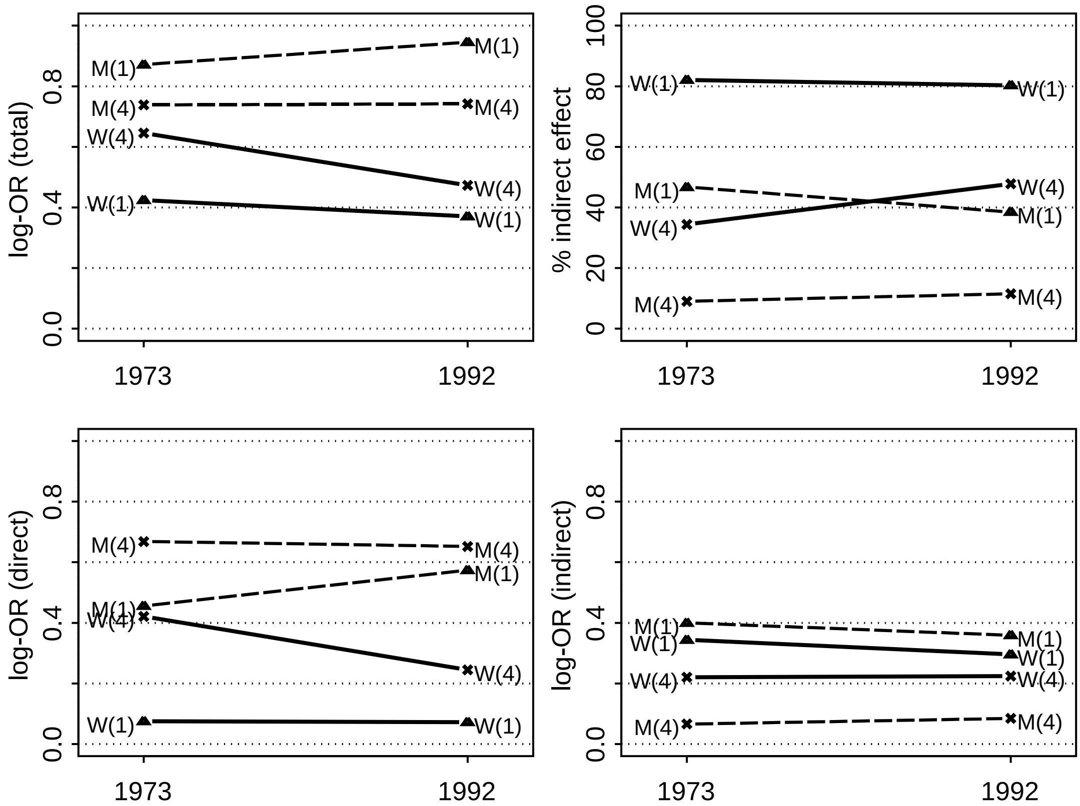
The estimated standard errors of the estimated total effects, obtained using the formulas of Section 3.3, are very similar to those of the marginal total effects. Both are comparable to the standard errors of the direct effects, and larger than those of the indirect ones (these two effects are negatively correlated). These patterns are identical to ones which are obtained if

Table 8: Estimated path analysis decompositions of local log odds ratios (c.f. the labels in Table 4) between origin and destination classes in the GHS data (with standard errors in parentheses). Here the intervening variable education has 7 levels (c.f. Table 2), and its conditional distribution given origin class was modelled using the conditional probabilities (24) for each level. The symbol ± indicates that the estimated indirect effect is positive and direct effect negative (standard errors for the percentage of indirect effect are not shown in these cases).

Year: 1973	Local odds ratio							
	Men				Women			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Marginal	.87	.17	.23	.74	.42	.36	.76	.65
total effect	(.11)	(.09)	(.11)	(.08)	(.10)	(.09)	(.10)	(.07)
Estimated	.86	.16	.22	.73	.42	.35	.76	.64
total effect	(.10)	(.09)	(.11)	(.08)	(.10)	(.09)	(.10)	(.07)
Difference	-.016	-.005	-.001	-.004	-.005	-.006	-.005	-.004
% difference in OR	-1.57	-.45	-.06	-.41	-.49	-.61	-.47	-.37
Direct	.46	-.12	.09	.67	.08	-.07	.49	.42
effect	(.10)	(.09)	(.12)	(.08)	(.09)	(.09)	(.09)	(.07)
Indirect	.40	.28	.14	.07	.34	.42	.27	.22
effect	(.04)	(.03)	(.02)	(.01)	(.04)	(.04)	(.03)	(.02)
% Indirect	47	±*	62*	9	82*	±*	35	34
effect	(6)	(–)	(33)	(2)	(18)	(–)	(5)	(4)
Year: 1992								
Marginal	.95	.15	.01	.74	.37	.50	.42	.47
total effect	(.10)	(.09)	(.12)	(.09)	(.08)	(.08)	(.11)	(.08)
Estimated	.93	.15	.01	.74	.37	.50	.41	.47
total effect	(.10)	(.09)	(.12)	(.09)	(.08)	(.08)	(.11)	(.08)
Difference	-.013	-.001	.001	-.005	-.002	-.005	-.002	-.003
% difference in OR	-1.28	-.09	-.08	-.50	-.15	-.47	-.19	-.32
Direct	.57	-.05	-.16	.65	.07	.15	.18	.25
effect	(.10)	(.09)	(.12)	(.09)	(.08)	(.08)	(.10)	(.08)
Indirect	.36	.20	.18	.08	.30	.34	.24	.22
effect	(.03)	(.03)	(.02)	(.02)	(.03)	(.03)	(.03)	(.02)
% Indirect	38	±*	±*	12	80*	69	57	48
effect	(5)	(–)	(–)	(2)	(18)	(11)	(15)	(9)

* The data are consistent (at 95% level of confidence) with 100% indirect effect.

Figure 3: Results of the path analysis decompositions for the GHS data, from Table 8. Here M and W denote men and women respectively. Only results for the local log odds ratios labelled (1) and (4) are shown.



standard linear path analysis is applied to these data (with both education and destination treated as continuous; these results are not shown here). The standard errors of the estimated proportions of indirect effect show that these proportions are reasonably precisely estimated when both effects are significantly different from 0, but rather larger when one of them (here always the direct effect) is not. In the latter case, the data are consistent with the claim that the total effect is entirely due to an indirect effect via education.

Figure 3 shows the main results from Table 8 in a graphical form. The plot omits the "off-diagonal" log odds ratios (2) and (3) which are somewhat less easy to interpret (and for which many of the estimates are consistent with no direct effect at all), and includes only (1) and (4). These describe "symmetric" associations between the same pair of origin and destination classes. High values of them indicate relative lack of intergenerational mobility between these classes. Thus log odds ratio (1) describes the tendency of respondents from Salariat and Intermediate origin classes to remain in these classes rather than move between them, and log odds ratio (4) describes a similar association involving Intermediate and Working classes.

These results show that the total association between origin and destination classes is in all cases stronger for men than for women; the difference is significant at the 5% level for all but association (4) in 1973. The disparity is particularly large for the immobility among the two non-working classes (association 1), where the total log odds ratio is in both years more than

twice as large for men as for women. Between 1973 and 1992, the estimated total associations have become somewhat stronger for men but weaker for women, but these changes are not significant. For direct effects, the patterns are roughly similar to those for total effects. The indirect (education) effects are fairly stable, in that there is no significant difference across years, or between men and women for association (1). The indirect effect for association (4) is significantly larger for women than for men.

The estimated effects imply much variation in the proportions of indirect effect, i.e. the proportion of the total association between origin and destination classes which can be explained by associations of education with origin and destination. This ranges from around 80% (which is not significantly different from 100%) for association (1) for women to around 12% for association (4) for men. In other words, most or all of the tendency of women from salariat or intermediate backgrounds to remain in the same class rather than move between them can be explained by the observed differences in educational attainment between women from these origin classes, together with the dependence of corresponding destination classes on educational qualifications. In contrast, only about 12% of the observed lack of mobility among men from intermediate or working-class background appears to be due to educational differences. There are no systematic or significant changes in such proportions from 1973 to 1992, but there are large and significant differences between men and women. In each case the proportion of dependence associated with education is larger for women than for men.

The remaining analyses of this section examine the effects on the results of changes in the model for the intervening education variable. Table 9 shows the results when education is still a seven-level variable, but a homoscedastic linear model is used for it given origin class. The calculations were carried out using Monte Carlo integration as in (31) and (32), with $M = 100,000$ simulated values from the estimated distribution of education given each origin class. The results are broadly similar to those in Table 8. The differences between marginal and estimated total effects are somewhat larger than there, mainly for the off-diagonal associations (2) and (3), presumably because the model for education is now somewhat less adequate than in Table 8 (c.f. the discussion in Section 3.5). The estimated standard errors are also very similar to those in Table 8. Here the Monte Carlo simulation variation adds little to the standard errors, because M is very large; the terms of order $O(M^{-1})$ contribute less than 1 per cent of the estimated variances obtained from (36).

The results of path analyses like these will depend on how the intervening variable is recorded. Table 10 illustrates the effects of this on the estimated proportions of the indirect effect for the diagonal log odds ratios (1) and (4). The first two models are those of Tables 8 and 9, where education has seven levels. In the other two models it is used (in the models both given origin and for destination) as a categorical variable with only three or two levels (c.f. Table 2 for the definitions of these). This leaves the relative size of the education effect roughly unchanged for some associations and reduced for others. The most noticeable effect is the substantial decrease of the education effect for the local odds ratio (4) for women. Clearly changing from seven to three or two levels of education has here resulted in substantial loss of information relevant to class mobility. A further examination of the data suggests that this is mainly the case for the association between education and class of destination. For women, there are substantial differences in proportions of Working and Intermediate destination classes between education levels 1–4, distinctions which are lost when these levels are combined into a single "Low" level of education. Similar strong associations do not occur for men, and for them the decrease in the proportions in Table 10 is much smaller.

Table 9: Estimated path analysis decompositions of local log odds ratios between origin and destination classes in the GHS data (with standard errors in parentheses). Here the intervening variable education has 7 levels, and its conditional distribution given origin class was modelled using a homoscedastic linear model. The calculations were carried out using Monte Carlo integration, with $M = 100,000$ simulated values. See the caption of Table 8 for further information.

Year: 1973	Local odds ratio							
	Men				Women			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Marginal	.87	.17	.23	.74	.42	.36	.76	.65
total effect	(.11)	(.09)	(.11)	(.08)	(.10)	(.09)	(.10)	(.06)
Estimated	.87	.10	.27	.75	.41	.20	.85	.70
total effect	(.11)	(.09)	(.11)	(.08)	(.10)	(.09)	(.09)	(.06)
Difference	-.003	-.067	.047	.015	-.009	-.159	.084	.052
% difference in OR	-.33	-6.48	4.81	1.51	-.91	-14.70	8.74	5.32
Direct	.51	-.08	.09	.66	.13	-.04	.48	.39
effect	(.11)	(.09)	(.11)	(.08)	(.10)	(.09)	(.09)	(.06)
Indirect	.36	.19	.18	.09	.29	.23	.37	.30
effect	(.03)	(.02)	(.02)	(.01)	(.02)	(.02)	(.03)	(.02)
% Indirect	42	±*	68*	12	70*	±	43	43
effect	(6)	(–)	(29)	(2)	(17)	(–)	(5)	(5)
Year: 1992								
Marginal	.95	.15	.01	.74	.37	.50	.42	.47
total effect	(.10)	(.09)	(.12)	(.09)	(.08)	(.08)	(.11)	(.08)
Estimated	.92	.15	.01	.74	.37	.46	.41	.49
total effect	(.10)	(.09)	(.12)	(.09)	(.08)	(.08)	(.10)	(.08)
Difference	-.021	-.004	.001	-.0002	.003	-.039	-.001	.013
% difference in OR	-2.13	-.41	.11	-.02	.29	-3.87	-.11	1.32
Direct	.58	-.04	-.15	.65	.08	.17	.18	.24
effect	(.10)	(.09)	(.12)	(.09)	(.08)	(.08)	(.10)	(.08)
Indirect	.34	.19	.17	.09	.30	.30	.23	.24
effect	(.03)	(.03)	(.02)	(.01)	(.03)	(.03)	(.02)	(.02)
% Indirect	37	±*	±*	12	79*	64	56	50
effect	(5)	(–)	(–)	(2)	(18)	(12)	(15)	(9)

* The data are consistent (at 95% level of confidence) with 100% indirect effect.

Table 10: Percentage of indirect (education) effect for local odds ratios (1) and (4), for models with different versions of the education variable. Here models [1] and [2] are those shown in Tables 8 and 9 respectively. In models [3] and [4], education is treated as a categorical variable (with three or two levels, as defined in Table 2) in the models for both education given origin class and destination given origin and education.

	Local odds ratio							
	Year: 1973				Year: 1992			
	Men		Women		Men		Women	
Education included as	(1)	(4)	(1)	(4)	(1)	(4)	(1)	(4)
[1] 7-level variable, categorical given Origin class	47	9	82	34	28	12	80	48
[2] 7-level variable, continuous given Origin class	42	12	70	43	37	12	79	50
[3] 3-level variable	36	6	76	5	33	7	86	10
[4] 2-level variable	34	6	75	5	28	7	82	12

5 Discussion

A key feature of the method of path analysis proposed here is that it separates the definition of the direct and indirect effects from the specification of the regression models for the variables. This means that the method can be applied in essentially the same way for any combination of types of models and variables. Models for continuous, nominal, ordinal or count outcomes can thus be handled in the same framework. Each model may include not only main effects of the explanatory variables but also interactions and nonlinear effects, as well as zero partial effects. The method also works the same way when some or all of the variables of interest are latent variables, as would be the case if we considered structural equation models involving categorical latent variables (e.g. as described in Hagenaars 1993). The models used to define and estimate the effects are then the structural models of such systems.

The focus here has been on the three-variable case involving only one intervening variable. A generalisation of this approach to problems with more variables was outlined in Section 3.7. The performance of this remains to be examined in future research.

In future work, we will use the method proposed here to examine in more detail the role of education in social mobility in Britain. This will require the use of data from other surveys than the GHS, for which information on the respondent's father's occupation is not available after 1992. Such analyses should also use a more detailed classification of social class than the 3-category version considered here. This, however, considerably increases the number of effect parameters. For example, associations for 7 unordered classes involve 36 odds ratios, each of which can be partitioned into direct and indirect effects. The question of how best to summarise such results needs to be considered carefully.

References

Agresti, A. (2002). *Categorical Data Analysis* (Second ed.). New York: Wiley.

Blau, P. M. and O. D. Duncan (1967). *The American Occupational Structure*. New York: Wiley.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Breen, R. and R. Luijkx (2004). Conclusions. In R. Breen (Ed.), *Social Mobility in Europe*, pp. 383–410. Oxford: Oxford University Press.

Davis, J. (1980). Contingency table analysis: Proportions and flow graphs. *Quality and Quantity 14*, 117–153.

Davis, J. A. (1975). Analyzing contingency tables with linear flow graphs: D systems. In D. R. Heise (Ed.), *Sociological Methodology 1976*, pp. 111–145. San Francisco: Jossey-Bass.

Didelez, V., A. P. Dawid, and S. Geneletti (2006). Direct and indirect effects of sequential treatments. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*, pp. 138–146. Arlington: Association for Uncertainty in Artifical Intelligence Press.

Duncan, O. D. (1966). Path analysis: Sociological examples. *American Journal of Sociology 72*, 1–16.

Duncan, O. D. and R. W. Hodge (1963). Education and occupational mobility: A regression analysis. *American Journal of Sociology 68*, 629–644.

Erikson, R. and J. H. Goldthorpe (1992). *The Constant Flux: A Study of Class Mobility in Industrial Societies*. Oxford: Clarendon Press.

Fienberg, S. E. (1980). *The Analysis of Cross-classified Categorical Data* (Second ed.). Cambridge, MA: MIT Press.

Geneletti, S. (2007). Identifying direct and indirect effects in a non-counterfactual framework. *Journal of the Royal Statistical Society B 69*, 199–215.

Glass, D. V. (Ed.) (1954). *Social Mobility in Britain*. London: Routledge.

Goldthorpe, J. H. (2007). *On Sociology* (Second ed.), Volume 2. Stanford, CA: Stanford University Press.

Goldthorpe, J. H. and C. Mills (2004). Trends in intergenerational class mobility in Britain in the late twentieth century. In R. Breen (Ed.), *Social Mobility in Europe*, pp. 195–224. Oxford: Oxford University Press.

Goodman, L. A. (1972). A general model for the analysis of surveys. *American Journal of Sociology 77*, 1035–1086.

Goodman, L. A. (1973a). The analysis of multidimensional contingency tables when some variables are posterior to others: A modified path analysis approach. *Biometrika 60*, 179–192.

Goodman, L. A. (1973b). Causal analysis of data from panel studies and other kinds of surveys. *American Journal of Sociology 78*, 1135–1191.

Hagenaars, J. A. (1993). *Loglinear Models with Latent Variables.* Newbury Park, CA: Sage Publications.

Hauser, R. M. (1978). A structural model of the mobility table. *Social Forces 56*, 919–953.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica 46*, 931–959.

Hellevik, O. (1983). Decomposing proportions and differences in proportions: Approaches to contingency-table analysis. *Quality and Quantity 18*, 79–111.

Hellevik, O. (1984). *Introduction to Causal Analysis.* London: George Allen & Unvin.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1995). *Continuous Univariate Distributions* (Second ed.), Volume 2. New York: Wiley.

Liang, K.-Y. and X.-H. Liu (1991). Estimating equations in generalized linear models with measurement error. In V. P. Godampe (Ed.), *Estimating Functions*, pp. 47–63. Oxford: Oxford University Press.

Lütkepohl, H. (1996). *Handbook of Matrices.* Chichester: Wiley.

Pearl, J. (1998). Graphs, causality, and structural equation models. *Sociological Methods & Research 27*, 226–284.

Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, pp. 411–420. San Francisco: Morgan Kaufmann.

Petersen, M. L., S. E. Sinisi, and M. J. van der Laan (2006). Estimation of direct causal effects. *Epidemiology 17*, 276–284.

Robins, J. M. (2003). Semantics of causal DAG models and the identification of direct and indirect effects. In P. Green, N. Hjort, and S. Richardson (Eds.), *Highly Structured Stochastic Systems*, pp. 70–81. Oxford: Oxford University Press.

Robins, J. M. and S. Greenland (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology 3*, 143–155.

Tanner, M. A. (1993). *Tools for Statistical Inference: Methods for Exploration of Posterior Distributions and Likelihood Functions* (Second ed.). New York: Springer.

Ten Have, T. R., M. Joffe, K. Lynch, G. Brown, and S. Maisto (2005). Causal mediation analyses with structural mean models. Biostatistics working paper, University of Pennsylvania.

Tukey, J. W. (1954). Causation, regression and path analysis. In O. Kempthorne, T. A. Bancroft, J. W. Gowen, and J. L. Lush (Eds.), *Statistics and Mathematics in Biology*, pp. 35–66. New York: Hafner.

van der Laan, M. J. and M. L. Petersen (2005). Direct effect models. Working Paper 187, U. C. Berkeley Division of Biostatistics.

Winship, C. and R. D. Mare (1983). Structural equations and path analysis for discrete data. *American Journal of Sociology 89*, 54–110.

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

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

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research 20*, 557–585.

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

Wright, S. (1960). Path coefficients and path regressions: Alternative or complementary concepts? *Biometrics 16*, 189–202.

Xie, Y. (1989). Structural equation models for ordinal variables: An analysis of occupational destination. *Sociological Methods & Research 17*, 325–352.

Zeger, S. L., K.-Y. Liang, and P. S. Albert (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics 44*, 1049–1060.