Effects of Birth Cohorts and Parents' Education in Probabilistic Models of Schooling Outcomes

G. Rohwer (August 2016)

Abstract: In the framework of a simple model in which schooling outcomes depend only on birth cohorts and parents' education, illustrated with data from the German National Educational Panel Study, the paper considers definitions of effects which are derived from conditional probabilities without presupposing a particular parametric model (logit models are only used as tools for smoothing observations). Effects are interpreted as partial descriptions of a structure defined by the conditional distributions which relate explanatory and outcome variables. Due to the interaction between birth cohorts and parents' education, only counterfactual definitions of direct and indirect effects of birth cohorts are possible. Although not effects in an ordinary sense, they can provide a quantification of the contribution of changes of parents' education to changes in the distribution of schooling outcomes. Using this interpretation, it is shown that an increasing part of the tendency towards higher schooling outcomes in Germany was due to changes in the distribution of parents' education.

Keywords: Models of schooling outcomes; Logit models; Definitions of effects; Direct and indirect effects; Interaction.

1 Introduction

Researchers interested in schooling outcomes often use probabilistic models in order to investigate how the probability of leaving the school system with a particular outcome depends on explanatory variables. A methodological question concerns how to define effects of the explanatory variables and whether and how one can distinguish between direct and indirect effects. Definitions suggested in the literature often presuppose particular parametric models, e.g., logit models (e.g., Long, 1997; Wooldridge, 2002). In this note, following Kuha and Goldthorpe (2010), I start from definitions which only use notions of conditional probability. Parametric models are considered as possibly helpful tools for the estimation (instead of definition) of effects.

As an example, I consider the question of how probabilities of schooling outcomes depend on parents' educational level. Of course, one has to take into account historical changes. The most simple theoretical framework is then the following:

$$
\begin{tikzcd}
X \arrow[drr] & & \\
C \arrow[u] \arrow[rr] & & Y
\end{tikzcd}
$$

where the dependent variable, Y, represents possible schooling outcomes, X records parents' educational level, and C distinguishes between birth cohorts. The framework assumes two functional relationships. The first is

$$c \longrightarrow \Pr(X = x \mid C = c) \tag{1}$$

which assumes that the distribution of parents' education depends on birth cohorts. The second is

$$(c, x) \longrightarrow \Pr(Y = j \mid C = c, X = x) \tag{2}$$

which assumes that the distribution of schooling outcomes depends on both birth cohorts and parents' education.

To illustrate the discussion, I use data from the German National Educational Panel Study (NEPS). In the next section, I describe the data and consider estimates of the relationship (2) (estimates of (1) will be presented in Section 4). In Section 3, I consider effects of parents' educational level which are derived from the relationship (2). I show that differences between conditional probabilities are well suited to describe that effects are context-dependent. As a possible alternative, I consider odds ratios. I argue that both definitions of effects can be interpreted as representing structural relationships which do not depend on marginal distributions of explanatory variables.

In Section 4, I consider effects of birth cohorts and discuss proposals for distinguishing between indirect effects, which are mediated through parents' education, and direct (residual) effects. Due to the interaction of birth cohorts and parents' education, only counterfactual partitions of effects of birth cohorts into indirect and direct components are possible. I describe different versions and present estimates of one of these partitions.

In the final discussion, I first summarize the consideration of different definitions of effects. I then stress that effects derived from counterfactual partitions are not ordinary effects which compare outcomes between individuals which differ in values of explanatory variables. This then creates the problem of how to interpret the counterfactually defined components, which depends on the application. I suggest that in the present application, a counterfactual partition can provide an informative quantification of the contribution of changes of parents' education to changes in the distribution of schooling outcomes. Using this interpretation, I show that an increasing part of the tendency towards higher schooling outcomes in Germany was due to changes in the distribution of parents' education.

2 Conditional probabilities of schooling outcomes

2.1 Data from the NEPS project

I use data from the German National Educational Panel Study (NEPS), Starting Cohort Adults (Version SC6:5.1.0).[1] I consider persons born between 1944 and 1986 (born in Germany or at most six years old when migrating to Germany), who began schooling in an elementary school (Grundschule). The outcome variable, Y, recording the level of schooling when leaving the school system, is derived from the variable `Ts11209` in the NEPS spell data file `spSchool`. I distinguish the following values (categories of the variable `Ts11209` in square brackets):

$$Y = \begin{cases} 1 & \text{without a degree (ohne Abschluss) [-20, -21] or} \\ & \text{low degree (Grund- or Hauptschule) [1, 2]} \\ 2 & \text{intermediate degree (mittlere Reife) [3])} \\ 3 & \text{upper degree (Hoch- or Fachhochschulreife) [4, 5]} \\ 4 & \text{residual category [6, 7]} \end{cases}$$

Since there are very few cases with $Y = 4$ (less than 2.5 %), I subsequently only consider persons with outcomes 1, 2, or 3.

For representing birth cohorts, I use a variable, C = `birthyear` − 1900, varying between 44 and 86. The variable X, recording the educational level of the parents, is based on CASMIN categories.[2] I distinguish four groups:

$$X = \begin{cases} 1 & \text{inadequately completed [1a] or general elementary school [1b]} \\ 2 & \text{basic vocational qualification [1c]} \\ 3 & \text{intermediate degree [2b, 2a], or general maturity cert. [2c\_gen]} \\ 4 & \text{lower or higher tertiary education [3a, 3b], or maturity cert.} \\ & \text{with vocational training [2c\_voc]} \end{cases}$$

There are 11468 cases with known values of all three variables (Y, C, and X).

2.2 Estimation of conditional probabilities

Conditional probabilities of schooling outcomes, $\Pr(Y = j \mid C = c, X = x)$, can be estimated by the corresponding observed proportions. Figure 1 shows how these proportions depend on values of C and X.

Instead of nonparametrically smoothing observed proportions, one could use different kinds of parametric and semi-parametric regression models for estimating the conditional probabilities. As an example, I use a multinomial logit model and interpret it as an alternative way of smoothing the observed proportions.

[1] From 2008 to 2013, NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network. For general information, see Blossfeld et al. (2011).

[2] The variables for the mother and the father are `T731301_g2` and `T731351_g2`, respectively. If values are different, I use the higher level.

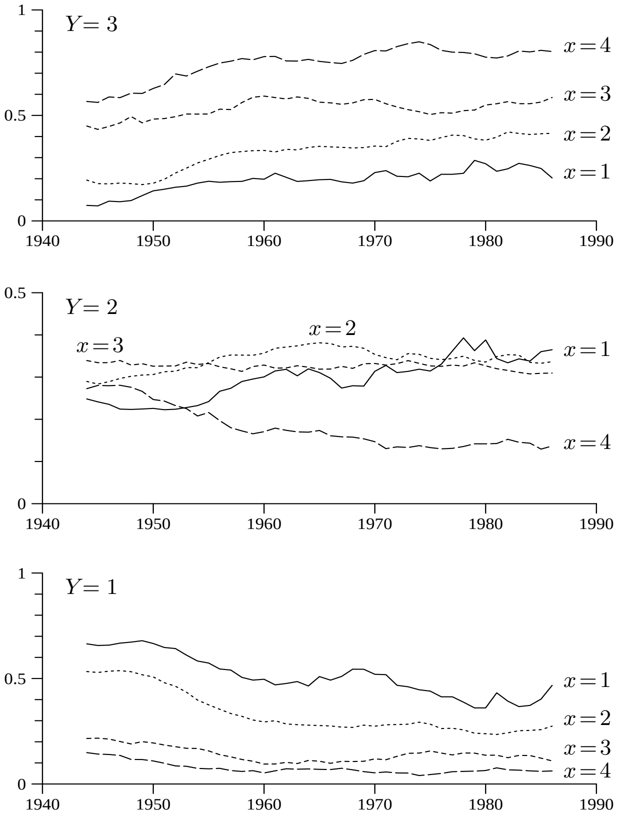

Figure 1 Observed proportions corresponding to $\Pr(Y{=}j \mid C{=}c, X{=}x)$, smoothed with running means of length 7.

From inspection of Figure 1 it is clear that one cannot treat X as just one quantitative variable, and one needs to specify interactions between C and X. I therefore use dummy variables $X_l = I[X = l]$, with values x_l, and consider X_1 as a reference. In order to improve the fit, I also include the square of C (divided by 100). The resulting specification is

$$\Pr(Y{=}j \mid C{=}c, X = x) \approx \qquad (3)$$
$$\frac{\exp(\alpha_j + c\,\gamma_j + c^2\,\gamma'_j + \sum_{l=2}^{4} x_l\,\beta_{lj} + \sum_{l=2}^{4} c\,x_l\,\delta_{lj})}{\sum_{k=1}^{3} \exp(\alpha_k + c\,\gamma_k + c^2\,\gamma'_k + \sum_{l=2}^{4} x_l\,\beta_{lk} + \sum_{l=2}^{4} c\,x_l\,\delta_{lk})}$$

Table 1 shows the estimated parameters (based on the constraint that the pa-

Table 1 Estimated parameters; standard errors in brackets.

	Model 3			Model 4	
	$Y = 2$		$Y = 3$		$Y = 3$
α_2	−6.768 (0.895)	α_3	−9.756 (0.948)	α	−7.118 (0.806)
γ_2	0.170 (0.027)	γ_3	0.242 (0.027)	γ	0.152 (0.022)
γ'_2	−0.110 (0.020)	γ'_3	−0.159 (0.020)	γ'	−0.097 (0.016)
β_{22}	0.759 (0.473)	β_{23}	0.583 (0.562)	β_2	0.509 (0.524)
β_{32}	2.529 (0.568)	β_{33}	3.546 (0.627)	β_3	2.585 (0.549)
β_{42}	3.123 (0.623)	β_{43}	3.859 (0.654)	β_4	2.513 (0.551)
δ_{22}	−0.002 (0.008)	δ_{23}	0.006 (0.009)	δ_2	0.003 (0.008)
δ_{32}	−0.017 (0.009)	δ_{33}	−0.018 (0.010)	δ_3	−0.015 (0.009)
δ_{42}	−0.026 (0.010)	δ_{43}	−0.008 (0.010)	δ_4	0.001 (0.009)

rameters for $j = 1$ are zero). Figure 2 compares the estimated probabilities with the observed proportions. Obviously, the model entails a heavy smoothing of the observed proportions. I assume that this is acceptable when being concerned with trends regarding the outcomes $Y{=}1$ and $Y{=}3$.

Model (3) is one example of a broad variety of parametric models of conditional probabilities (e.g., Long, 1997; Wooldridge, 2002). For the present application, a general notation is

$$\Pr(Y{=}j \mid C{=}c, X{=}x) \approx h(c, x; \theta_j)$$

where $h(c, x; \theta)$ denotes a mathematical function with arguments c and x, which depends on a parameter vector θ_j. There are three requirements.

a) It should be possible to find a parameter vector, $\hat{\theta}_j$, such that $h(c, x; \hat{\theta}_j)$ is a sensible representation (with regard to the model's purpose) of the observations in terms of conditional probabilities.

b) Values of the function $h(c, x; \hat{\theta}_j)$ should lie between zero and one. This is required for all possible values of C and X for which the model is intended to provide conditional probabilities. In the present application it suffices that the requirement is met for all observed values of C and X.

c) $\sum_j h(c, x; \hat{\theta}_j) \approx 1$. When using a multinomial logit model, the sum of estimated probabilities exactly equals 1. An approximate equality suffices

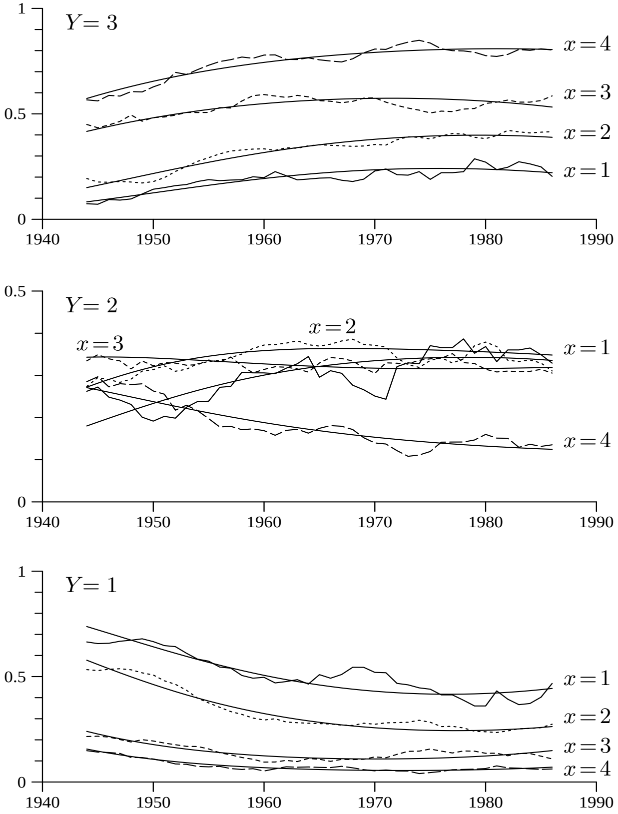

Figure 2 Observed proportions corresponding to $\Pr(Y = j \,|\, C = c, X = x)$, taken from Figure 1, and estimates calculated with Model (3).

when the estimated probabilities are considered as approximations. This view also allows one to use models which focus on particular outcomes.

For example, if the focus is on a single outcome, say $Y = j$, one can use a binary logit model for a corresponding dummy variable, $Y_j = I[Y = j]$. With a specification analogous to (3), the model is

$$\Pr(Y_j\!=\!1 \,|\, C\!=\!c, X = x) \approx \qquad (4)$$
$$L\Big(\alpha_j + c\,\gamma_j + c^2\,\gamma'_j + \sum_{l=2}^{4} x_l\,\beta_{lj} + \sum_{l=2}^{4} c\,x_l\,\delta_{lj}\Big)$$

where $L(u) = \exp(u)/(1 + \exp(u))$ is the inverse of the logit link function.

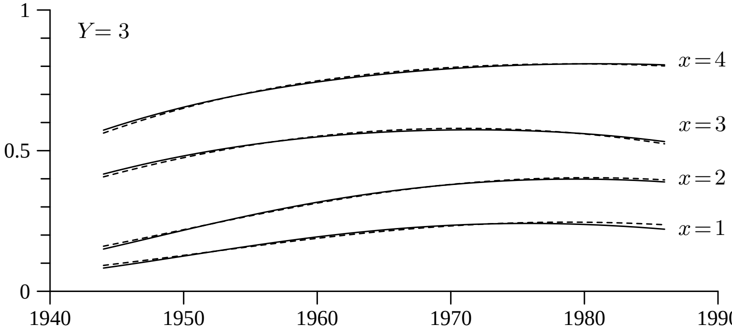

Figure 3 Comparison of probabilities $\Pr(Y_3 = 1 \,|\, C = c, X = x)$, estimated with Model (3) (solid) and Model (4) (dashed).

Table 1 shows the parameter estimates for Y_3.

Of course, Model (4) is not mathematically equivalent to Model (3). As illustrated in Figure 3, one nevertheless gets very much the same estimates of conditional probabilities. The same is true when the focus is on the outcome $Y = 1$. Therefore, as long as the focus is on a single outcome, and one is using a parametric model only for smoothing the observed proportions, one can alternatively use the multinomial or the binary logit model as a framework for linking covariates to outcome probabilities. In both cases, the model's flexibility is due to the specification of the linear predictor.

3 Effects of parents' educational level

3.1 Comparing conditional probabilities

What is the effect of X on Y? A general leading idea is that the distribution of Y depends on values of X.[3] In order to define an effect one therefore needs a reference to at least two values of X, say x' and x'', and the effect describes 'the difference' between the corresponding probabilities of $Y = j$ (separately for each j). This can be done in different ways. An often used simple and easily

[3]Thinking of a probabilistic dependency in this general sense does not require to assume also a 'causal relation' in any specific sense. Questions of possible causal interpretations will not be considered in this paper.

understandable definition of ‘the difference’ is

$$\Pr(Y = j \,|\, C = c, X = x'') - \Pr(Y = j \,|\, C = c, X = x') \tag{5}$$

which shows how a difference between values of X is connected with a corresponding numerical difference in probabilities of the outcome. So these effects are defined on a probability scale. They are visible as the differences between the curves in Figure 2.

This approach to understanding, and defining, effects has two important implications. First, in general, there is a different effect for each pair of values selected for the comparison. In the present example, this is directly seen in Figure 2.

Second, effects of a variable almost always depend on a context given by values of other variables. Subsequently, I use this as a definition of ‘interaction’: Two variables interact if the effect of one variable depends on values of the other variable (and vice versa). In our example, X and C interact, that is, effects of X depend on values of C.

The definition (5) must be distinguished from the notion of a ‘marginal effect’ which presupposes a parametric model and the possibility to consider derivatives. Using the notation $h(c, x; \hat{\theta}_j)$, the marginal effect of X would be $\partial h(c, x; \hat{\theta}_j)/\partial x$. In general, except for a linear model, the marginal effect does not show how a difference between two values of X relates to a difference in outcome probabilities (Petersen, 1985; Long, 1997).

3.2 Odds ratios

Instead of differences between probabilities, one could use odds ratios (or their logarithms). Corresponding to (5), the odds ratio is

$$\frac{\Pr(Y = j \,|\, C = c, X = x'')}{\Pr(Y \neq j \,|\, C = c, X = x'')} \Big/ \frac{\Pr(Y = j \,|\, C = c, X = x')}{\Pr(Y \neq j \,|\, C = c, X = x')} \tag{6}$$

Like differences between conditional probabilities, also odds ratios are derived from conditional probabilities. They cannot, however, be derived from each other. Both entail a reduction of the information given by two separate conditional probabilities.

There also is a difference when using the definition of ‘interaction’ given above which presupposes a notion of effects. In the present application, whether

the odds ratio (6) depends on values of C also depends on the model used to estimate the conditional probabilities. For example, when using Model (4), the log odds ratio is $(\beta_{x''j} - \beta_{x'j}) + c\,(\delta_{x''j} - \delta_{x'j})$. Without the explicitly specified interaction terms, the dependence on values of C would vanish.

3.3 Effects and marginal distributions

It has been argued that odds ratios are particularly well suited for representing effects because they ‘do not depend on marginal distributions’ (e.g., Mare, 1981; Marshall and Swift, 1999). As shown by Hellevik (2007), the claim is not immediately clear.

(a) I begin with the marginal distribution of the explanatory variable, X. In fact, all measures of effects which are derived from the conditional probabilities, $\Pr(Y = j \,|\, C = c, X = x)$, do not depend on the distribution of X. The conditioning on values of X entails the separation of a structural relationship from the distribution of X:

$$\text{Distr. of } X \rightarrow \boxed{\Pr(Y = j \,|\, C = c, X = x)} \rightarrow \text{distr. of } Y$$

The conditional distribution is considered as the structure which links the two marginal distributions. Taking this view, also differences between conditional probabilities do not depend on the marginal distribution of X.

(b) All measures of effects which are defined in terms of conditional probabilities do not depend on the distribution of X, but only the odds ratio also does not depend on the distribution of Y. How to understand the latter kind of ‘independence’? A purely formal explanation can be given by referring to a crosstabulation of cases, for example:

	$Y_3 = 0$	$Y_3 = 1$
x'	a	b
x''	c	d

All measures which only depend on conditional probabilities do not change if one multiplies the rows by arbitrary constants. The odds ratio, that is, $(d/c)/(b/a)$, also does not change if one multiplies the columns by arbitrary constants. The suggested symmetry hides, however, that the relation between X and Y (as defined by the theoretical framework) is not symmetric. The requirement that a measure of effects should not depend on the distribution of X can be satisfied by conditioning on values of X without a reference to Y. On the other hand, conditioning on values of Y is not compatible with the idea that the distribution of Y is the result of a given distribution of X and a structural relationship (see the diagram above).

Taking this idea seriously, possible marginal distributions of Y are determined by the marginal distribution of X and the conditional probabilities. In the example, assuming that x' and x'' are the only categories of X:

$$\Pr(Y_3 = 1) = \frac{b}{a+b}\Pr(X = x') + \frac{d}{c+d}\Pr(X = x'')$$

Consequently, if $b/(a+b) \leq d/(c+d)$, the range of possible distributions is given by

$$b/(a+b) \leq \Pr(Y_3 = 1) \leq d/(c+d)$$

This shows that the idea that a measure of effects should be compatible with all conceivable distributions of outcomes, independent of restrictions due to the presupposed structure, is not warranted. On the other hand, the property that the measure does not depend on outcome distributions which are compatible with the structure, is automatically satisfied if the measure does not depend on the distribution of X.

4 Effects of birth cohorts

Analogous to effects of parents' education, X, one can investigate effects of birth cohorts, C. Given our theoretical framework, there is no further context and one can simply compare probabilities of schooling outcomes between birth cohorts. On the other hand, since the distribution of X depends on C, one can now consider X as a variable which mediates effects of birth cohorts. In fact, as can be seen in Figure 4, the distribution of parents' educational levels heavily changed across birth cohorts in our observation period.

In this section, I use this example to discuss proposals to distinguish between direct and indirect effects. The example is well suited because of the interaction of X and C which entails that one cannot define direct and indirect effects in an ordinary sense. I begin with a brief description of effects of birth cohorts, and then discuss proposals for decompositions of these effects into direct and

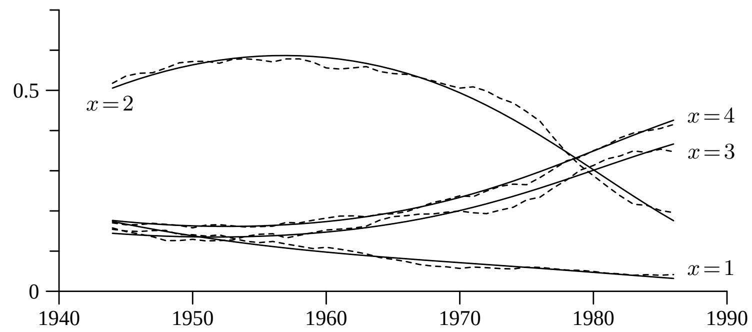

Figure 4 Dashed: observed proportions $\Pr(X = x \mid C = c)$, smoothed with running means of length 7. Solid: estimates calculated with a multinomial logit model.

indirect components which are based on various counterfactual assumptions. Finally, I show estimation results of one of these decompositions.

4.1 Total effects of birth cohorts

I consider effects of birth cohorts defined by

$$\Pr(Y_j = 1 \mid C = c'') - \Pr(Y_j = 1 \mid C = c') \tag{7}$$

for $j = 1$ and $j = 3$. For example, if $c' = c$ and $c'' = c + 1$ these effects compare adjacent birth cohorts. For estimating the conditional probabilities, I use binary logit models

$$\Pr(Y_j = 1 \mid C = c) \approx L(\alpha_j^* + c\,\gamma_j^* + c^2\,\gamma_j^{*'}) \tag{8}$$

(as before, values of C^2 are divided by 100). The estimated parameters are:

j	α_j^*	γ_j^*	$\gamma_j^{*'}$
1	4.410 (0.656)	−0.132 (0.021)	0.069 (0.016)
3	−4.064 (0.610)	0.083 (0.019)	−0.035 (0.014)

Figure 5 compares the estimated probabilities with corresponding observed proportions. The comparison illustrates again how the parametric model provides a smoothing of the observations.

Accepting this smoothing, the estimated probabilities can immediately be used to calculate the effects of birth cohorts defined in (7). Figure 6 shows

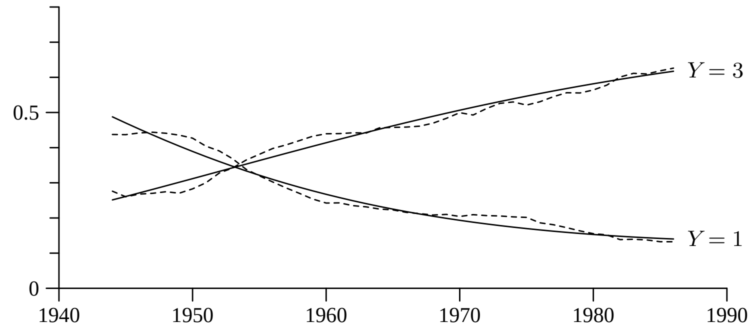

Figure 5 Solid lines: probabilities $\Pr(Y_j = 1 \mid C = c)$ estimated with Model (8). Dashed lines: corresponding observed proportions smoothed with running means of length 7.

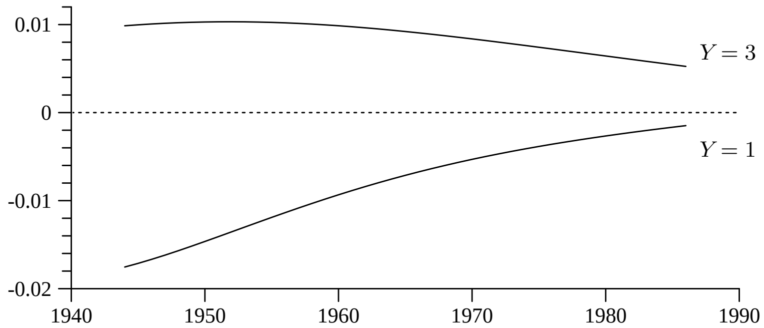

Figure 6 Effects of C on the probability of $Y = 1$ and $Y = 3$, as defined in (7) with $c' = c$ and $c'' = c + 1$; derived from Model (8).

these effects for adjacent birth cohorts. They decreased slightly for $Y = 3$ and substantially for $Y = 1$.

4.2 Direct and indirect effects

The direct effect of birth cohorts is conditional on values of the mediating variable, X. One can start from the definition

$$\Pr(Y_j = 1 \mid C = c'', X = x) - \Pr(Y_j = 1 \mid C = c', X = x)$$

If C and X interact, as it is the case in our application, this effect is not independent of values of X. Consequently, a unique direct effect does not exist

(see also Kaufman et al., 2004). Note that the argument does not depend on assuming a particular nonlinear model; the same consequence would result when using a linear regression model with an interaction term.

It might seem possible to consider a mean direct effect with respect to a distribution of X:

$$\sum\nolimits_x \left[\Pr(Y_j = 1 \mid C = c'', X = x) - \Pr(Y_j = 1 \mid C = c', X = x)\right] \Pr(X = x)$$

One has to take into account, however, that the actual distribution of X depends on values of C; and due to the interaction between X and C, this definition of a mean direct effect depends on the chosen distribution of X.

Several choices are possible. One could use $\Pr(X = x \mid C = c')$ which leads to what has been called a 'natural' mean direct effect by Pearl (2001). Alternatively, one could use the distribution $\Pr(X = x \mid C = c'')$, or, as proposed by Kuha and Goldthorpe (2010), a mean of the two distributions.

Assuming a temporal ordering, $c' < c''$, allowing one to think of a change in the value of C, it seems plausible to fix the distribution at the beginning, $C = c'$. In the present application, however, it seems preferable to use the distribution at $C = c''$ because this allows a more straightforward interpretation of the indirect effect which is of primary interest. In fact, only the indirect effect contributes to the understanding of how an effect of C is generated. The mean direct effect (in whatever definition) only represents a residual; and in order to better understand this residual one would have to think of further mediating variables.

The choice of $\Pr(X = x \mid C = c'')$ leads to the following partition of the total effect of C.

$$\begin{aligned} &\Pr(Y_j = 1 \mid C = c'') - \Pr(Y_j = 1 \mid C = c') = \qquad (9)\\ &\sum\nolimits_x \left[\Pr(Y_j = 1 \mid C = c'', X = x) - \Pr(Y_j = 1 \mid C = c', X = x)\right] \Pr(X = x \mid C = c'') + \\ &\sum\nolimits_x \left[\Pr(X = x \mid C = c'') - \Pr(X = x \mid C = c')\right] \Pr(Y_j = 1 \mid C = c', X = x) \end{aligned}$$

The first term on the right-hand side defines a mean direct effect which can be attributed to the difference between c' and c'', assuming counterfactually a constant distribution of X. The second term on the right-hand side defines an indirect effect which shows the contribution of the change in the distribution of X, assuming counterfactually that the dependence of $Y_j = 1$ on C and X had not changed.

Note that the partition (9) does not presuppose a particular parametric model. In contrast, for example, the proposal made by Breen, Karlson and Holm (2013) relates to a linear regression model for a latent variable behind Y_j. Moreover, their proposal presupposes that there is no interaction between C and X, and further assumes a linear regression model for the dependence of X on C. Both assumptions are not met in the present application, and I therefore do not consider this proposal as a possible alternative.

4.3 Estimation of decompositions

Estimation of the partition (9), or alternative versions, requires estimates of the conditional probabilities involved. Practical approaches depend, in particular, on the representation of conditional distributions of X. In the present application, X is a categorical variable with only four categories; so one can use the multinomial logit model illustrated in Figure 4.[4] In addition, I use Model (4) for estimating probabilities of $Y_j = 1$ conditional on C and X.

For the illustration, I consider the total effect as defined in (7) with $c' = c$ and $c'' = c + 1$. Results are shown in Figure 7 for $Y = 1$ and $Y = 3$. The labels D and I of the curves denote, respectively, the first and the second term on the right-hand side of (9).

The dashed curves show the sum of the two terms and represent an approximation to the total effect of the birth cohorts. Of course, one cannot expect that these curves equal the total effects shown in Figure 5 which are derived from Model (8). However, notwithstanding the limited accuracy of the estimated partitions, they suggest some conclusions.

With regard to $Y = 3$, the partition suggests that the rising probability of this schooling outcome (as indicated in Figure 4) was increasingly a result of a change in the distribution of parents' educational levels. With regard to $Y = 1$, one has to consider that the effect of birth cohorts on this schooling outcome was negative, approaching zero at the end of the observation period. Taking this into account, also this partition shows that a change in parents' educational levels increasingly contributed to the effect of birth cohorts.

[4] In other applications, it could be preferable to use a parametric continuous distribution. See, for example, Erikson et al. (2005) who used counterfactual decompositions for defining and estimating primary and secondary effects of the family background. For further discussion see also Buis (2010).

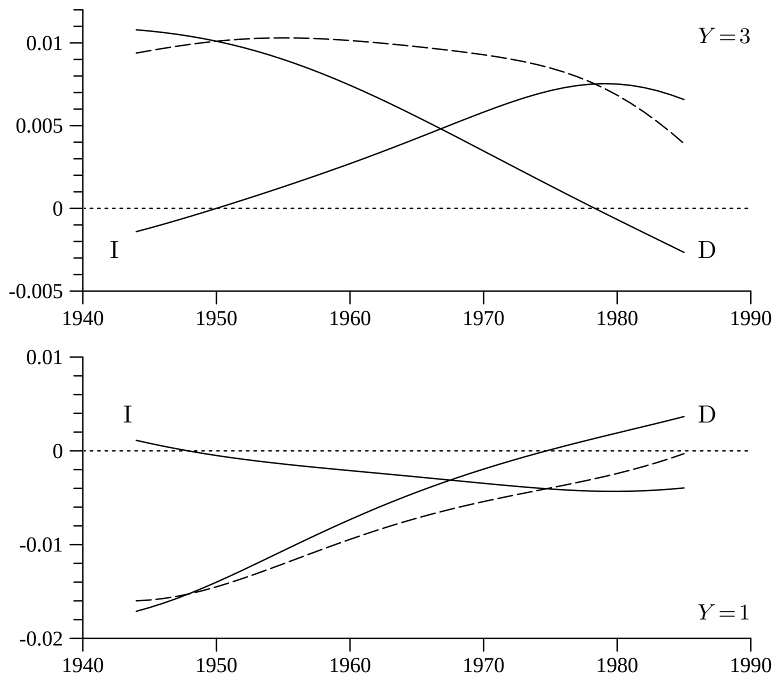

Figure 7 Estimation of the partition (9) for $Y = 3$ (upper plot) and $Y = 1$ (lower plot). The labels D and I denote, respectively, the first and the second term on the right-hand side of (9). The dashed curves show the sum of D and I.

5 Discussion

I have considered a simple model in which schooling outcomes probabilistically depend on birth cohorts and parents' educational level. In this framework, I have compared two definitions of effects of parents' education which can be derived from conditional probabilities of schooling outcomes without requiring the reference to a parametric model: simple differences between conditional probabilities and odds ratios.

For comparing the two definitions of effects, I have considered them as (partial) characterizations of a structure defined by the conditional distributions which relate explanatory and outcome variables. It is seen, then, that both definitions provide characterizations which do not depend on the marginal distributions of parents' education. A given odds ratio is also compatible with all conceivable distributions of the outcome variable. However, given that a measure of effects should characterize the structure relating explanatory and outcome variables, it suffices that the measure is compatible with all outcome distributions which can be generated by the structure;[5] and this property is common to all measures of effects which are derived from the conditional probability distributions.

Using simple differences between conditional probabilities as a measure of effects is also useful when being interested in direct and indirect effects of an explanatory variable. In the present paper, I have considered effects of birth cohorts which represent relevant conditions for the schooling outcomes of children. As a consequence of the interaction of birth cohorts and parents' education, only counterfactual partitions of effects of birth cohorts into direct and indirect components are possible. This then creates the further problem of how to interpret the counterfactually defined components.

The components are not ordinary effects which compare outcomes between individuals which differ in values of explanatory variables. Consider, for example, the direct effect as defined by the partition (9) which can be written

$$\Pr(Y_j = 1 \mid C = c'') - \sum\nolimits_x \Pr(Y_j = 1 \mid C = c', X = x) \Pr(X = x \mid C = c'')$$

While the first term is an empirically observable quantity, the second term describes a fictitious construct.

Interpretations of counterfactual partitions depend on the application. In the present application, one can make use of the fact that there is a temporal ordering of birth cohorts: $c' < c''$. So one can think of a change in the conditional distribution of parents' education, $\Pr(X = x \mid C = c') \longrightarrow \Pr(X = x \mid C = c'')$, and consider the question of how this change contributed to a corresponding change in the distribution of schooling outcomes. The indirect effect as defined by the partition (9) can then be interpreted as providing an answer to this question based on assuming that the structural relationships between the variables had not changed.

To illustrate, I consider the development of $\Pr(Y_3 = 1 \mid C = c)$ across birth cohorts from 1960 to 1986. As shown by the solid line in Figure 8, there was a substantial increase. For comparison, the dashed line shows a cumulated indirect effect defined by

[5] Other considerations might become relevant when being interested in measures of inequality, but see Marks (2004) for a critique.

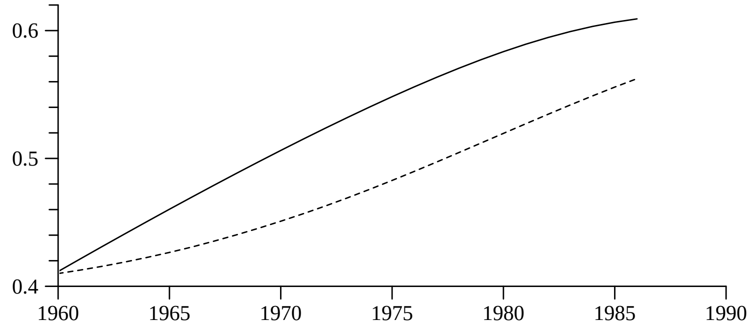

Figure 8 Solid line: $\Pr(Y_3 = 1 \mid C = c)$. Dashed line: the cumulated indirect effect defined in (10).

$$\sum_{k=61}^{c} \sum\nolimits_x \big[\Pr(X = x \mid C = k) - \Pr(X = x \mid C = k-1)\big] \times \Pr(Y_3 = 1 \mid C = k-1, X = x) + \Pr(Y_3 = 1 \mid C = 60) \tag{10}$$

Note that this definition uses a concatenation of 'temporally local' indirect effects. It is not counterfactually assumed that structural relationships between the variables involved remained constant since the birth cohort 1960, but only between adjacent birth cohorts. The comparison in Figure 8 can therefore be interpreted as demonstrating how an increasing part of the changes in the probability of the schooling outcome $Y = 3$ can be attributed to changes in the distribution of parents' education between adjacent birth cohorts.

References

Blossfeld, H.-P., Roßbach, H.-G, von Maurice, J. (Eds.), 2011. Education as a lifelong process – the German National Educational Panel Study (NEPS). Zeitschrift für Erziehungswissenschaft, 14.

Breen, R., Karlson, K.B., Holm, A., 2013. Total, direct, and indirect effects in logit and probit models. Sociological Methods and Research 42, 164–191.

Buis, M.L., 2010). Direct and indirect effects in a logit model. The Stata Journal 10, 11–29.

Erikson, R., Goldthorpe, J.H., Jackson, M., Yaish, M., Cox, D.R., 2005. On class differentials in educational attainment. Proceedings of the National Academy of Sciences 102, 9730–33.

Hellevik, O., 2007. 'Margin insensitivity' and the analysis of educational inequality. Czech Sociological Review 43, 1095–1119.

Kaufman, J.S., MacLehose, R.F., Kaufman, S., 2004. A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation. Epidemiologic Perspectives & Innovations 1:4, 1–13.

Kuha, J., Goldthorpe, J.H., 2010. Path analysis for discrete variables: the role of education in social mobility. Journal of the Royal Statistical Society A 173, 351–369.

Long, J.S. 1997. Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks: Sage.

Mare, R.D., 1981. Change and stability in educational stratification. American Sociological Review 46, 72–87.

Marshall, G., Swift, A., 1999. On the meaning and measurement of inequality. Acta Sociologica 42, 241–250.

Marks, G.N. (2004). The measurement of socio-economic inequalities in education. A further comment. *Acta Sociologica*, 47, 91–93.

Pearl, J., 2001. Direct and indirect effects. Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, 411–420. San Francisco, CA.

Petersen, T., 1985. A comment on presenting results from logit and probit models. American Sociological Review 50, 130–131.

Wooldridge, J.M., 2002. Econometric Analysis of Cross Section and Panel Data. Cambridge: MIT Press.