Decomposing primary and secondary effects

Karlson, Kristian Bernt; Holm, Anders

Published in:
Research in Social Stratification and Mobility

DOI:
10.1016/j.rssm.2010.12.005

Publication date:
2011

Document version
Early version, also known as pre-print

Citation for published version (APA):
Karlson, K. B., & Holm, A. (2011). Decomposing primary and secondary effects: A new decomposition method. *Research in Social Stratification and Mobility*, *29*(2), 221-237. https://doi.org/10.1016/j.rssm.2010.12.005

Decomposing primary and secondary effects:

A new decomposition method

THIS PAPER IS PUBLISHED IN

RESEARCH IN SOCIAL STRATIFICATION AND MOBILITY, 2013 VOL. 29, 221-237

Link: http://www.sciencedirect.com/science/article/pii/S0276562410000697

This is a post-print (i.e. final draft post-refereeing) version according to SHERPA/ROMEO

Karlson, Kristian Bernt *

Holm, Anders **

* Corresponding author. The Danish School of Education, Aarhus University, Tuborgvej 164, DK-2400 Copenhagen NV, Denmark. Telephone: 0045 23369285. Fax: 0045 88889001. Email: kbk@dpu.dk

** The Danish School of Education, Aarhus University, Tuborgvej 164, DK-2400 Copenhagen NV, Denmark. Email: aholm@dpu.dk

December 3, 2010

7874 words (including all text, notes, and references)
8 tables
1 appendix

Running head: Decomposing primary and secondary effects

Acknowledgements**:** We thank Wout Ultee, Lars H. Andersen, Thomas DiPrete, Robert Hauser, and participants at the ISA RC28 meeting in Haifa in May 2010 for helpful comments.

Decomposing primary and secondary effects:

A new decomposition method

Abstract

One strand of educational inequality research aims at decomposing the effect of social class origin on educational choices into primary and secondary effects. We formalize this distinction and present a new and simple method that allows empirical assessment of the relative magnitudes of primary and secondary effects. Contrary to other decomposition methods, this new method is unbiased, is more intuitive, and decomposes effects of both discrete and continuous measures of social origin. The method also provides analytically derived statistical tests and is easily calculated with standard statistical software. We give examples using the Danish Longitudinal Survey of Youth.

Keywords: primary effects; secondary effects; decomposition; logit; logistic

1. Introduction

With the gaining popularity of the relative risk aversion hypothesis set forward by Breen and Goldthorpe (1997), Boudon's (1974) distinction between primary and secondary effects has become a topic of interest for educational researchers (Erikson *et al.*, 2005; Jackson *et al.*, 2007; Erikson, 2007). This strand of research interprets primary effects as those cultural and genetic influences that are expressed through the uneven academic performance of students with differing social origins. Secondary effects are, on the other hand, interpreted as the "residual" influences of social origin that operate over and above academic performance, potentially reflecting an anticipatory dimension of educational decision making. Postulating the existence of a fear of downward social mobility, secondary effects are given interpretation in terms of the class-specific expected pay-offs from completing further education. The theoretical distinction between primary and secondary effects consequently lays out two basic mechanisms that explain observed inequality in educational attainment (Goldthorpe, 1996).

Since Mare's (1980, 1981) seminal work, educational researchers have preferred analyzing educational choices as discrete events. They therefore often use binary logit models for studying the determinants of educational decisions. However, in nonlinear probability models such as the logit model, one cannot simply compare an uncontrolled coefficient of a variable of interest with a controlled counterpart (controlling for a confounder of interest) (cf. Winship & Mare, 1984). The reason for this is that regression coefficients and the underlying residual variance of binary regression models are not separately identified (cf. Amemiya, 1975; Cramer, 2003). This restriction, known as the scale identification issue, makes simple decompositions of effects, such as the primary and secondary effects decomposition, challenging. In this paper we present a new method that enables researchers to decompose the logit effect of social origin on educational decisions into primary and secondary effects. One notable feature of this method is that it is unaffected by the scale identification issue, and it is therefore not prone to attenuation bias created by differences in underlying residual variances. The method is a by-product of the generalized decomposition technique for nonlinear probability models recently developed by Karlson, Holm, and Breen (2010), henceforward KHB. We thus continue the work set out by Erikson et al. (2005), whose method allows for formal decompositions of effects of discrete variables in logit models. However, contrary to their method, the KHB-method is unbiased, is more intuitive, and involves well-known interpretations from linear regression analysis. The KHB-method is also more general in the sense that it does not require the variable whose effect we want to decompose to be discrete, does not rely on distributional assumptions of the control variables, and is easily calculated with standard statistical software.

We proceed as follows. First, we formalize the distinction between primary and secondary effects using standard results from linear regression models, in particular the distinction between direct and indirect effects when controlling for a third, potentially confounding variable. Second, we give an exposition of the KHB-method and we link it to the primary and secondary effects distinction. We also conduct a Monte Carlo study to compare the performance of our method relative to the method by Erikson et al. (2005) and its generalization developed by Buis (2010). Third, we present an example using the Danish Longitudinal Survey of Youth (DLSY). Fourth, we conclude the paper by pointing out theoretical and methodological challenges related to empirical decompositions of primary and secondary effects.

2. Formalizing primary and secondary effects

In his classic work on social inequality, Boudon (1974) distinguishes between two kinds of influence of social origin on educational decisions, namely primary and secondary effects. In this section we present the distinction and we propose a simple formalization based on well-known properties of linear regression models. Boudon (1974) begins with the observation that social origins and educational decisions are related, and that this association may be explained by the unequal distribution of scholastic abilities across social class (an explanation Boudon linked to the literature focusing on the social class variation in cultural and linguistic endowments). Boudon coins these effects “primary effects”. However, Boudon finds that even net of abilities the association between social origins and educational attainment persists. Boudon coins these net effects of social origins on educational decisions “secondary effects”, and he develops a rational choice model that gives interpretation to these secondary effects.

Boudon’s model lays out a mechanism that operates through rational expectations about future returns to education (Holm & Jæger, 2008). His model stipulates that families, irrespective of the position they occupy in the occupational structure, fear downward social mobility for their offspring. They are risk averse in the terminology of Breen and Goldthorpe (1997). Although this social demotion motive operates in all families, it has different consequences for students of different social origins. Students with similar academic performance, but different social origins, face different pay-offs from completing further education (cf. Boudon, 1989, 1994). Higher class students will have to attain a relatively higher level of schooling than lower class students, if they are to enter the same or higher social class as their parents. In other words, the expected utility associated with completing further education differs across social classes, the reason being the social demotion motive. Given that the educational system in late-industrial societies is a main channel for distributing individuals into the occupational structure, Boudon's model explains why we observe secondary effects. It does so by relating class-specific evaluations of costs and benefits to educational decision making, rather than using explanations based on the notion of socialization (cf. Goldthorpe, 1996, 2007).

The formalization of Boudon's model by Breen and Goldthorpe (1997) has spurred a literature that focuses on empirical decompositions of primary and secondary effects (Erikson *et al.*, 2005; Jackson *et al.*, 2007; Erikson, 2007). The main purpose of this literature is to assess, empirically, the extent to which primary or secondary effects govern educational choices. Primary effects are the part of the social class effect on an educational decision that is mediated by academic performance, while secondary effects is the residual, or direct, effect of social class net of academic performance. Following Boudon's model, the literature then interprets the primary effects component as the part of the social class effect that exists because of cultural and genetic endowment differences (operating through academic performance), while the literature interprets secondary effects component as a choice-based part of the social class effect. In this article we adopt the same interpretation, but we recognize that at least one further assumption is to be maintained if we are to identify the underlying parameters of interest: Investments in abilities and performance are not a consequence of expectations about future returns to education. If the social demotion motive does not lead families to invest in their offspring's academic abilities and performance, then the motive is not reflected in primary effects, but only in secondary effects. Given this assumption (that ability and performance does not arise as a consequence of the social demotion motive) the empirical decomposition of primary and secondary effects can be said to identify the theoretical parameters of interest (i.e., cultural and genetic endowments, and the social demotion motive).[1] In this paper we maintain this assumption.

We now turn to a simple empirical strategy that translates the distinction between primary and secondary effects into empirical quantities. The strategy can be used for a formal assessment of the relative magnitudes of primary and secondary effects. We use the well-known properties of linear regression models, namely the rules for decomposing a total effect into the simple sum of direct and indirect effects (Duncan, 1966; Alwin & Hauser, 1975; Clogg *et al.*, 1995a). This decomposition is known from the early literature on path-models in sociology, but we do not imply the causal relationships usually assumed in these models (cf. Freedman, 1987, 2005). In other words, we use the decomposability of linear models descriptively to evaluate to which degree academic performance mediates the association between social origin and an educational decision. In the next section we show how to implement this method into the framework for binary logit models for educational transitions. As we will see, this implementation is not straightforward in binary logit models, because evaluating the extent to which academic performance mediates the effect of social origins on educational decisions is hampered by the scale identification of these models: comparing uncontrolled and controlled (controlling for academic performance) logit coefficients does not reflect confounding, but a mixture of both confounding and rescaling.

---

[1] We thank an anonymous reviewer for pointing out that giving reduced form regression coefficients a structural interpretation is difficult and often dubious. We also recognize that other mechanisms than the one stressed by Boudon (1974) may be responsible for bringing about secondary effects. However, this paper is not concerned with extending Boudon's model, but rather with introducing a new decomposition technique, and we consequently do not delve into these theoretical issues here.

2.1. Total, direct, and indirect effects in linear models

Let *SO* denote a measure of social origin and *EC* an educational choice (e.g., passing a certain educational level). Let *AP* denote the student's academic performance, and *e* a random error term. We write the following linear model:

$$EC = \alpha + \beta \cdot SO + \gamma \cdot AP + e \qquad (1)$$

The model in (1) relates social origins and academic performance to the educational choice. Now relate academic performance to social origin:

$$AP = \mu + \theta \cdot SO + v, \qquad (2)$$

where *v* is a random error term. From this simple system of equations we may derive the total, direct, and indirect effect of *SO* on *EC*, where the indirect effect is the effect of *SO* on *EC* mediated by *AP*. β is the direct effect and can be interpreted as the residual effect of social origins on the educational decision. In terms of primary and secondary effects, β may be interpreted as secondary effects, because it captures the social origin influences net of academic performance. The product, $\gamma\theta$, is the indirect effect, which can be interpreted as the degree to which academic performance mediates the effect of social origin on the educational choice. In other words, $\gamma\theta$ may be interpreted as primary effects. The indirect effect may also be obtained by estimating the effects of *SO* in a model without *AP* and in a model with *AP*, and then compute the simple difference between the effects of *SO* from each model. This difference equals the product, $\gamma\theta$ (Clogg et al., 1995a). Finally, the simple sum of the direct and indirect effects is the total effect: $\delta = \beta + \gamma\theta$. δ summarizes the total influence that social origin has on the educational choice. We note that the total effect is equal to effect of *SO* in a model without *AP* (Clogg et al., 1995a). Because the direct and indirect effect can be interpreted as secondary and primary effects, we may recover how much of the total effect each of these effects accounts for:

$$P = \frac{\theta\gamma}{\beta + \theta\gamma} = \frac{\theta\gamma}{\delta} \quad ; \quad S = \frac{\beta}{\beta + \theta\gamma} = \frac{\beta}{\delta}. \tag{3}$$

P and S denote the relative of magnitude of primary and secondary effects, respectively. Multiplying P or S with 100 returns the percentage that each effect accounts for of the total effect. Note that, because S is defined as the residual effect of social origins net of academic performance, it holds that S = 1 – P. Whenever we are interested in assessing the relative magnitude of primary and secondary effects, we can use the ratios in (3).[2]

3. The decomposition method

While the linear system expressed in Equations (1) and (2) is a convenient way of describing primary and secondary effects, we face the problem that educational decisions are measured as discrete variables (cf. Mare, 1980). Thus, linear models, however convenient they may be, do not adequately describe the relationships we are interested in. The problem somewhat relates to the belief that there is "no calculus of path coefficients" for nonlinear models such as logit models (Fienberg, 1980; Hagenaars, 1998). Therefore, Erikson et al. (2005) have recently proposed a method that enables decompositions of effects of discrete measures of social origin (e.g. social classes) on discrete educational choices. However, while this method solves the problem, we point out some shortcomings that the general method by Karlson, Holm, and Breen (2010) overcomes. In this section we therefore present the KHB-method and we compare it with the technique developed by Erikson et al. (2005) and its generalization by Buis (2010).

---

[2] Researchers acquainted with the consequences of omitted variable bias may note that the simple linear decomposition of social origin influences on educational choice may be *biased*, because unobserved variables (such as family structure or peer group structures) may influence the parameters in Equation (1). We return to this point in the discussion, where we also discuss the consequence of measurement error, in the academic performance measure, for the decomposition.

3.1 The KHB-method

Let y^* be a continuous latent propensity to choose or pass a certain educational level or track (e.g., upper secondary or university education).[3] Let *x* be a continuous or discrete measure of social origin, and let *z* be a continuous or discrete measure of academic performance.[4] Now define an underlying linear model such that

$$y^* = \alpha + \beta x + \gamma z + u, \quad \text{where} \quad sd(u) = \sigma_u \tag{4}$$

and define the effect of *x* on *z* as in (2)

$$z = \mu + \theta x + v, \quad \text{where} \quad sd(v) = \sigma_v, \tag{5}$$

where *u* and *v* are random and independent error terms, and α and μ are intercepts. σ_u and σ_v are the residual standard deviations, i.e., the unexplained part in each model. Models (4) and (5) correspond to the system defined in (2) and (3), except that y^* is unobserved (and consequently we cannot estimate β, γ, or σ_u). We nevertheless observe a binary discrete realization of y^* such that

$$\begin{aligned} y &= 1 \ \text{ if } \ y^* > \tau \\ y &= 0 \ \text{ if } \ y^* \leq \tau, \end{aligned} \tag{6}$$

where τ denotes a threshold on the latent propensity. For example, τ may be the threshold for entering university. For convenience we set $\tau = 0$.[5] We now rewrite the error term in (4) such that $u = \sigma_e \cdot \omega$, where ω is a standard logistic random variable with mean zero and

---

[3] We assume that this latent propensity exists in order to motivate the methodology. To fix ideas, imagine two identical students born in two different cohorts. Both students have the same latent propensity to enroll in university education, but because of the educational expansion, going to university is harder for the student born in the oldest cohort than for the student born in the youngest cohort. A result may be that the student born in the oldest cohort did not go to university, while the other student did.

[4] Whenever *x* has more than two levels, e.g. a discrete social class measure, we may think of *x* as a vector of binary indicator variables that indicate the different levels of *x*. β are then the corresponding vector of regression coefficients. In the following description of the KHB-method we therefore find it easiest to think of *x* as binary whenever *x* is discrete with more than two categories.

[5] Whenever τ is nonzero it is absorbed in the intercept of the logit model (Long, 1997; Cramer, 2003).

variance $\pi^2/3$, and σ_e is the scale parameter of the logistic distribution capturing the effect of omitted variables independent of *x* and *z*, yielding a standard deviation of $\sigma_u = \sigma_e \cdot (\pi \ /\sqrt{3})$ for the error term in (4). We may then specify the probability of choosing *y = 1* as a function of *x*:

$$\Pr(y=1) = \Pr(y^* > 0) = \Pr\left(u > -\left[\frac{\alpha}{\sigma_e} + \frac{\beta}{\sigma_e}x + \frac{\gamma}{\sigma_e}z\right]\right)$$
$$= F\left(\frac{\alpha}{\sigma_e} + \frac{\beta}{\sigma_e}x + \frac{\gamma}{\sigma_e}z\right) = \frac{\exp\left(\frac{\alpha}{\sigma_e} + \frac{\beta}{\sigma_e}x + \frac{\gamma}{\sigma_e}z\right)}{1+\exp\left(\frac{\alpha}{\sigma_e} + \frac{\beta}{\sigma_e}x + \frac{\gamma}{\sigma_e}z\right)} \quad ,$$

where *F(.)* is the cumulative logit function and where the third equality holds because we assume ω to be a standard logistic random variable. Taking the logarithm to the odds of the probability, we obtain a logistic regression model:

$$\operatorname{logit}(\Pr(y=1)) = a + bx + cz = \frac{\alpha}{\sigma_e} + \frac{\beta}{\sigma_e}x + \frac{\gamma}{\sigma_e}z \,. \tag{7}$$

Thus, the parameters of the logit model in (7) are in fact the parameters from the underlying linear model *divided* by the residual deviation of that model multiplied by a constant (i.e., $\sigma_e = \sigma_u \cdot (\sqrt{3}/\pi)$). In logit models we therefore only identify the underlying coefficients relative to a scale, which is a function of the residual standard deviation of the underlying linear model (cf. Amemiya, 1975; Cramer, 2003). In other words, we can only estimate the logit coefficients (log odds-ratios) *b* and *c* such that:

$$b = \frac{\beta}{\sigma_e}; \quad \mathrm{c} = \frac{\gamma}{\sigma_e} \ . \tag{8}$$

However, this scale sensitivity of logit coefficients does not prevent a linear decomposition of the effect of *x* on *y* into a part attributable to *z* and a part not attributable to *z*. Karlson, Holm, and Breen (2010) develop a formal method for making such decomposition. They begin with the observation that comparing coefficients across nested logit models successively including potentially confounding variables does not have the same interpretation as in linear models. In logit models, changes in coefficients across models are a result of two underlying mechanisms. First, similar to linear models, confounding alters the coefficient magnitude (and, in some situations, the sign of the coefficient). Second, unlike linear models, a rescaling of the model increases the coefficient magnitude (Winship & Mare, 1984). This latter mechanism arises as a consequence of differences in the scale parameter across models. A model excluding the confounder has a larger scale parameter (i.e., larger error dispersion) than a model including the confounder. This rescaling carries over to the estimated logit coefficients, because, as is evident from (8), the scale parameter enters the denominator. In the Appendix we formally show why this is so, and we also present the solution to this problem by Karlson, Holm, and Breen (2010). Their method allows coefficient comparisons across logit models that are unaffected by rescaling, and their method consequently captures confounding net of rescaling. From the derivations in the Appendix, it follows that we can decompose the total effect of x on y measured in logits (log odds-ratios) into a direct and an indirect part measured on the same scale:

Direct: $$b = \frac{\beta}{\sigma_e} \tag{9a}$$

Indirect: $$c \cdot \theta = \frac{\gamma}{\sigma_e} \cdot \theta = \frac{\gamma\theta}{\sigma_e} \tag{9b}$$

Total: $$\frac{\delta}{\sigma_e} = \frac{\beta}{\sigma_e} + \frac{\gamma\theta}{\sigma_e} = \frac{\beta + \gamma\theta}{\sigma_e}, \tag{9c}$$

where (9a) is the direct effect of x controlled for z, (9b) is the part of the effect of x mediated by z, and (9c) is the total effect. All effects are measured on the same scale, making them directly comparable. Had we instead estimated the total effect in a logit model *without z*, as is possible in the linear case, it will generally not equal the expression in (9c). The reason for this is that a model excluding *z* has a different residual variance and the logit coefficient of *x* from that model is consequently identified up to a different scale than σ_e in (9c) (see the Appendix; cf. Winship & Mare, 1984; Wooldridge 2002). Had we followed the principles in linear models and computed the indirect effect as the difference between the effect of *x* estimated in a logit model excluding *z* and the effect of *x* estimated in a logit model including *z*, this difference would be conflated by differences in residual variation across models. What we consequently achieve by using the *product* of coefficients to compute the indirect effect is that the all three effects in (9) are measured on the same scale (i.e., divided by the same scale parameter, σ_e). Compared to the expression of these effects in linear models, the only difference is that the coefficients (β, γ, θ, and δ) in (9) are divided with a scale parameter (σ_e).

Because the effects in (9) are measured on the same scale, we may obtain knowledge about the relative magnitude of primary and secondary effects similar to the linear case (stated in (3)):

$$P = \frac{c \cdot \theta}{b + c \cdot \theta} = \frac{\frac{\gamma\theta}{\sigma_e}}{\frac{\delta}{\sigma_e}} = \frac{\gamma\theta}{\beta + \gamma\theta} \; ; \; \mathrm{S} = \frac{b}{b + c \cdot \theta} = \frac{\frac{\beta}{\sigma_e}}{\frac{\delta}{\sigma_e}} = \frac{\beta}{\beta + \gamma\theta}, \tag{10}$$

which is the same definitions of primary and secondary effects as in the linear model, except that the parameters in (10) stem from the underlying linear model. While we cannot estimate these parameters directly, we can compute their *ratios* as in (10). By computing their ratios we obtain measures of the relative magnitude of primary and secondary effects that are *unaffected by the scale parameter*. In other words, the ratios in (10) overcome the scale identification issue, because the scale parameter cancels out. Moreover, it is precisely these ratios that have our interest, whenever we are to decompose primary and secondary effects in a logit model. Multiplying the quantities in (10) with 100 returns the relative percentage that each component, P and S, accounts for of the total effect of social origin on educational choice. The expressions in (10) show why the KHB-method extends the decomposition features of linear models to nonlinear probability models such as the logit model.

A further consequence of this property is that the KHB-method allows researchers to control the decomposition defined in (10) for possibly confounding variables. Following Sobel (1998:32), we name these variables *concomitants*, w_j, $j = 1, 2, \ldots, J$, where J denotes the number of variables. We may control for the potential confounding influence of these concomitants *on the decomposition*. Such control is unaffected by the scale identification of logit coefficients, because the ratios in (10) do not involve the scale parameters. We define the following logit model as in (4) and (7):

$$y^* = \tilde{\alpha} + \tilde{\beta}x + \tilde{\gamma}z + \tilde{\lambda}_j w_j + l, \quad \text{where } sd(l) = \sigma_l \text{ and } \sigma_l = \sigma_k \cdot (\pi / \sqrt{3}) \qquad (11a)$$

$$\operatorname{logit}(\Pr(y = 1)) = \tilde{a} + \tilde{b}x + \tilde{c}z + \tilde{d}_j w_j = \frac{\tilde{\alpha}}{\sigma_k} + \frac{\tilde{\beta}}{\sigma_k}x + \frac{\tilde{\gamma}}{\sigma_k}z + \frac{\tilde{\lambda}_j}{\sigma_k}w_j, \qquad (11b)$$

where σ_k is smaller than σ_e in (7), because the added concomitants, insofar they explain variation in the latent propensity, reduce the residual variation. Replacing b and c in (9) and (10) with $\tilde{b}$ and $\tilde{c}$ makes it possible to control the decomposition for important concomitants. We reiterate that this replacement does not entail any comparison problems brought about by the scaling of the coefficients in equation (11b). The scale parameter, σ_k, cancels out whenever we replace b and c with $\tilde{b}$ and $\tilde{c}$ in (10). Thus, the KHB-method also allows for controlling the decomposition for concomitants. For example, the decomposition of primary and secondary effects may be influenced by family factors or peer group structures.

Including concomitant variables measuring these attributes ensures that the relative magnitude of primary and secondary effects is not distorted by these attributes.

3.2 Statistical tests of primary and secondary effects using the KHB-method

Often researchers are not only interested in the relative magnitude of primary and secondary effects, but also whether each of them can be said to be statistically significant. On the one hand, testing secondary effects is easy, because it amounts to testing the direct effect of b in (7) or $\breve{b}$ in (11b), depending on whether or not concomitants are included in the model. Standard statistical software reports test statistics that in large samples are asymptotically normally distributed. For continuous x, z-statistics may be reported, while for discrete x, joint Wald tests may be computed.

On the other hand, testing primary effects, which amounts to testing the indirect effect, may at first seem as a complicated affair, because it involves a product of two coefficients. However, whenever z is a single variable (e.g., academic performance), it can be shown that the indirect effect is statistically significantly different from zero whenever both c and θ are significantly different from zero.[6] We may therefore use the statistical tests for each of the two coefficients. For the first coefficient, c, we report the z-statistic from the logit model (cf. (7) or (11b)). For the second coefficient, θ, we may report the z-statistic from the linear model (cf. (5)). If x is a discrete variable with more than two categories, we may use an F-test in the linear model (jointly testing θ for each category relative to the reference). In the Appendix we carry out a Monte Carlo study that shows that our test-statistic performs

[6] In more formal terms, if either $c = 0$ or $\theta = 0$ holds, then the null hypothesis of no confounding is true (cf. Clogg *et al.*, 1995b). In case of more than a single z, we refer to the test-statistic derived in Karlson, Holm, and Breen (2010).

satisfactorily when it comes to rejecting the null hypothesis of no indirect effect. Thus, for decomposing primary and secondary effects, the KHB-method also provides researchers with easily accessible statistical tests, because they only need the output of the logit model and the linear model to do the testing. In the empirical examples we show how these test-statistics can be reported.

3.3. Comparing the KHB-method with the methods by Erikson et al. and Buis

The KHB-method provides researchers with practicable and easily accessible tools for decomposing primary and secondary effects in a logit model. Erikson et al. (2005) propose another solution that involves counterfactual analyses of log-odds ratios. We briefly outline their solution and then compare it with the KHB-method using a Monte Carlo study. The study reveals that the method by Erikson et al. (2005) and the generalization by Buis (2010) return biased estimates of the true decomposition percentages in a range of scenarios encountered in real applications, while the KHB-method returns unbiased estimates. We explain this discrepancy in terms of a possible distributional sensitivity of the methods by Erikson et al. and Buis.

As in the KHB-method Erikson et al. (2005) want to assess the direct effect of x (social origin) on y (educational choice) and the indirect effect trough z (academic performance). First, they identify the direct effect of x by calculating counterfactual log-odds-ratios of x, a discrete social class measure. They compare the log-odds of success (y=1) for one social class with another social class given that latter social class have the distribution of academic performance (z) of the former social class. Second, they find the indirect effect by comparing the log-odds of success within one social class with the counterfactual log-odds within the same social class with a distribution of academic performance of another social class. The method by Erikson et al. (2005) makes comparisons of the factual and counterfactual groups possible. For the direct effect the factual and counterfactual social classes only differ with regards to social class, *x*, and not the academic performance distribution, *z*. For the indirect effect the factual and counterfactual social classes only differ with respect to the academic performance, *z*, and not social class, *x*. Consequently, the method makes it possible to compare the relative magnitude of primary and secondary effects, because the direct and indirect effects are directly comparable (Erikson et al., 2005:9733). The method by Erikson et al. (2005) integrates numerically over the distribution of academic performance (*z*) to generate the counterfactuals. While Erikson et al. assume normality of *z* in order to make this numerical integration, Buis (2010) recently generalized the method to accommodate any distribution of *z*, making the method applicable to a range of situations met in real applications.

To learn about the performance of the KHB-method relative to the methods by Erikson et al. and Buis, we conducted a Monte Carlo study. The results, which are reported in the Appendix, reveal that the method by Erikson et al. (2005) and the generalization by Buis (2010) do not consistently recover the true fraction of the indirect effect relative to the total effect in a range of situations relevant for real applications. The bias arises whenever the control variable, *z*, is skew (lognormal in the Monte Carlo study). This bias appears to vary with the degree of true confounding and the distribution of *y*: The more confounding and the more skew y, the larger the bias. For example, in the situation of strong confounding (defined as a situation in which the indirect effect accounts for two-thirds of the total effect), a 75/25 distribution of *y*, and a 50/50 distribution of *x*, the method by Erikson et al. is around 6 percentage points off the true confounding percentage, and the method by Buis is around 5 percentage points off. These biases are substantial. However, the KHB-method is unbiased, being less around 0.2 percentage point off the true fraction. In passing, we note that the Monte Carlo study also shows that whenever the control variable, *z*, is not skew, all three methods consistently recover the true fraction of the indirect effect. This finding suggests that the methods by Erikson et al. and Buis work well with a symmetrically distributed control variable, but not an asymmetrically distributed control variable. In other words, their methods appear to be distribution sensitive. Whenever *z* is skew, their methods do not recover the true fraction of the indirect effect relative to the total effect. This does not hold for the KHB-method, which is insensitive to the shape of the distribution of *z*.

Thus, the Monte Carlo study suggests that the KHB-method performs better than the methods by Erikson et al. (2005) and Buis (2010), because it consistently recovers the true fraction in all scenarios. In addition to this, we also raise four other shortcomings of their methods that do not exist for the KHB-method. First, their methods use numerical integration to estimate the counterfactual log odds-ratios, while the KHB-method is an analytical solution. Second, compared to the KHB-method that produces one set of estimates, their methods produce two sets of estimates of direct and indirect effects, based on two different, though highly similar, counterfactual analyses (Erikson et al., 2005; Buis, 2010). Although Jackson et al. (2007) suggest to report the average of the two estimates for each set of direct and indirect effects, as Buis (2010:28) notes, such ambiguity makes the method “less than elegant.” Third, contrary to the KHB-method, the method by Erikson et al. requires the variable whose effect we want to decompose to be categorical. Fourth, the method by Erikson et al. involves several separate computational steps (Buis, 2010:16-17). Although the *ldecomp* Stata command by Buis (2010) calculates these steps, the method returns much output that researchers need to filter. By way of contrast, the KHB-method uses standard statistical output from logit and linear models and provides readily interpretable statistical tests.

We consequently suggest that researchers interested in decomposing primary and secondary effects use the KHB-method. It is easy to understand and use, because it extents the decomposition features of linear models to logit models. Because the properties of linear models are well-known to most researchers, we believe that the KHB-method is a practicable way forward in decomposing primary and secondary effects. In some situations, the method developed by Erikson et al. (2005) and the generalization by Buis (2010) will produce identical results to those of the KHB-method, but in some situations this will not be the case. As we will see in the example, all three methods return more or less identical results, the reason being that the measure of academic performance is normally distributed (and consequently does not have a skew distribution).

4. Examples

We have presented a simple formalization of primary and secondary effects, shown why the KHB-method is useful for decompositions of primary and secondary effects, and argued that the KHB-method is an unbiased and more practicable alternative than the method developed by Erikson et al. (2005) or its generalization by Buis (2010). In this section we turn to two examples based on the Danish Longitudinal Survey of Youth (Hansen, 1995; Jæger & Holm, 2007). This survey follows the life course of some 3,150 children born in 1954.[7] The final sample consists of 1,896 individuals. Because the examples presented here work as illustrations of the KHB-method, we do not go into the missing data patterns and the possible biases they may produce.

---

[7] The data is cluster-sampled with school classes as clusters. In this analysis, we ignore the clustering, thereby reporting slightly incorrect standard errors. However, performing analyses with cluster-robust standard errors does not alter the substantial conclusions drawn in this article.

We give a brief outline of the data used. The binary dependent variable, *y*, is whether the individual finished university (=1) or not (=0). We have two measures of social origin, *x*. The first is a discrete five-category measure of father's social class that was constructed by the authors of the original study (see Hansen, 1995). High values denote higher classes, while low values denote lower classes. The second is a continuous composite measure constructed as the principal component of a principal component analysis involving father's social class, educational attainments, and income. This measure is supposed to approximate the socioeconomic status index developed by Duncan (1961).[8] We standardize this variable to have expectation zero and variance of unity. We have a single continuous measure of academic performance, *z*, namely a composite measure of three low-stake tests (verbal, inductive and spatial) extracted from a principal component analysis.[9] While this composite variable may measure academic "ability" more than academic "performance", we take it as a good proxy for academic performance in the sense that students condition their educational choices on their perceived abilities. We standardize the variable to have mean zero and unit variance. Finally, we have two control variables that may affect the decomposition: whether the child comes from an intact family (=1) or not (=0), and the gender of the individual (1 = male; 0 = female). Table 1 describes the variables.

-- TABLE 1 HERE --

4.1 Decomposition with a discrete measure of social origin

We begin with a decomposition of the effect of a discrete measure of social origin on whether the child completes university or not, and we thereafter compare results from the KHB-

[8] We use polychoric correlations for the principal components analyses, because the three indicators are ordered discrete variables. The first component explains around 57 percent of the total variation in the three indicators.
[9] The first component explains around 89.6 percent of the total variation in the three items.

method and the methods by Erikson et al. (2005) and Buis (2010). Tables 2 and 3 report regression estimates from a logit model (cf. (7)) and a linear model (cf. (5)), respectively. The direct effect (*b*) is the effect of the discrete social class measure in Table 2. The pattern of this effect is somewhat linear, i.e., the higher the social class, the higher the propensity to complete university. The indirect effect is product of the effect of academic performance (*c*) in the logit model in Table 2 *and* the effect of social class (θ) on academic performance in the linear model in Table 3. The effects in both tables are, as expected, highly significant. We use formula (9b) to calculate the indirect effects for *each* social class. Because social class is a discrete measure, the indirect effect for each social class has to be interpreted relative to the reference group (the lowest social class). For example, for the highest social class relative to the lowest social class we calculate the indirect effect as: $c \cdot \theta = 1.054 \cdot 0.930 = 0.980$.

-- TABLE 2 AND 3 HERE --

Using this strategy, we summarize the direct, indirect, and total effects in Table 4 for each social class, where the total effect is given by the sum of direct and indirect effects (see (9c)). The two last columns in Table 4 also report the relative magnitudes of primary and secondary effects, using the formulae in (10). For example, 51.2 percent of the effect of the low social class on the educational outcome is mediated by academic performance, whereas 38.4 percent is so for the highest social class. A joint test of these primary effects shows that they are statistically significant. The bottom columns of Table 4 provide the two test statistics necessary for testing the indirect effect. The *z*-value, 9.78, of academic performance (*c*) on the educational decision in the logit model is much larger than a critical value of 1.64, and the F-value, 28.18, for the joint test of the effect of father's social class (θ) on academic performance in the linear model is much larger than the critical value obtained from an F(4,1891) distribution. Because both tests reject the null hypothesis, we conclude that the indirect effect, i.e., the primary effect, is statistically significant.

-- TABLE 4 HERE --

Because we define secondary effects as residual primary effects (S = 1 – P), the estimated secondary effects in Table 4 necessarily reflect the estimated primary effects. However, we may want to ascertain whether these secondary effects are statistical significant. To do this we simply evaluate the joint Wald test of the effect of social class (b) on the educational decision in the logit model in Table 2. We report this test-statistic, 33.34, in the third final row of Table 4. We find that it is much larger than the critical value obtained from a chi-square distribution with four degrees of freedom. Thus, the joint social class effect on university enrollment is statistically significant, indicating statistically significant secondary effects.

In Table 4 we also report a simple (i.e. unweighed) average of primary and secondary effects across social classes. These percentages give an overall assessment of the extent to which academic performance can be said to mediate the effect of social class on university completion. On average primary effects account for 42 percent of the effect of social class on university completion, while secondary effects account for 100-42 = 58 percent. Secondary effects thus appear to account for a slightly larger proportion than primary effects of the total influence of social origin on the educational decision.

The final part of this analysis is a comparison of the KHB-method and the method suggested by Erikson et al. (2005) and the generalization developed by Buis (2010). Because the two latter methods each reports two sets of estimates of direct and indirect effects, we follow Jackson et al. (2007) and report their average. Table 5 compares results generated by the three methods (the Appendix contains the output behind the Erikson et al. and Buis results). The immediate impression is that the results are highly similar for all three methods. Only minor deviations occur. On average, the method by Erikson et al. (2005) and its generalization by Buis (2010) return a fairly smaller part attributable to primary effects compared to the method by KHB. Around one percentage point distinguishes the results produced by the methods. This similarity, however, is not a surprise, since both methods use the same data and the measure of academic performance is normally distributed. Consequently, the distributional sensitivity of the methods by Erikson et al. and Buis does not appear to distort the results in this example.

-- TABLE 5 HERE --

4.2 Decomposition with continuous measure of social origin

We now perform a decomposition analysis using the continuous measure of social origin, interpreted as father's socioeconomic status (SES). Table 6 provides the decomposition results, while we have placed the output behind the table in the Appendix. We find both the direct and indirect effect of SES on university completion to be statistically significant. Around 32 percent of the effect of SES is mediated by academic performance, leaving around 68 unexplained. Because the test statistics in the bottom of Table 6 reject the null hypotheses of no direct or indirect effects, we conclude that both primary and secondary effects are statistically significant. We therefore confirm the results from the previous analysis with the discrete social class measure, namely that secondary effects are comparatively larger than primary effects. Because SES is a continuous measure, we cannot compare it with the method by Erikson et al. (2005) or the generalization by Buis (2010). This fact suggests that using the KHB-method may be the preferred alternative, because it does not require any assumptions with regards to the distribution of the variable whose effect we want to decompose.

-- TABLE 6 HERE --

4.3 Decomposition with concomitants

The final two analyses control the decompositions reported in the two first analyses for the influence of two possibly confounding variables, namely intact family and gender (cf. the logit model described in (11)). Table 7 compares the uncontrolled and controlled primary and secondary effects for the discrete measure of social origin, while Table 8 compares the continuous counterparts.[10] From Table 7 we see that controlling the decomposition for intact family and gender reduces the average primary effects from 42 to 39.5 percent, and from Table 8 we see a decline from 31.5 to 30.3. Both reductions are small, and they therefore add to the overall impression that secondary effects are more important than primary effects. Moreover, net of concomitants, both primary and secondary effects are statistically significant, thereby corroborating the conclusions drawn in the previous analyses. We thus conclude that the reported magnitudes of primary and secondary effects are robust to two concomitants that potentially could affect the decomposition results.

-- TABLE 7 AND 8 HERE --

[10] The output behind the results in Tables 7 and 8 are available from the authors.

5. Discussion

The distinction between primary and secondary effects has become a topic of interest for educational researchers. How much of the influence of social origin on educational choices is explained by differences in academic performance between students of different social origins, and how much is explained by differences in expected returns to education, triggered by a fear of downward social mobility? In this paper we present a new method that allows researchers to decompose the effect of social origin on discrete educational decisions into primary and secondary effects. The method is based on a general decomposition method recently developed by Karlson, Holm, and Breen (2010). We show that this method is more effective than other methods currently available for decompositions in nonlinear probability models such as logit models. In contrast to the method developed by Erikson et al. (2005) and its generalization by Buis (2010), the KHB-method gives unbiased decompositions, is more intuitive, and is simpler to compute.

Using the KHB-method we find that around 40 and 60 percent of the effect of a discrete measure of social origin on educational choice is attributable to primary and secondary effects, respectively, while the decomposed counterpart of a continuous social origin measure is 30 and 70 percent, respectively. Using a measure of academic ability (an IQ test) similar to one used in this paper, Jackson et al. (2007) report similar findings for the UK. However, they do not investigate whether their results are statistically significant. We do so in this paper and we also find that our results are robust to two possibly confounding variables, namely intact family and gender. However, because other generally unobserved variables may confound the decomposition, these results are only suggestive. If families invest in their offspring's academic abilities and performance, and these investments are driven by the social demotion motive, then a part of the primary effect may be attributed to the motive, rather than to cultural or genetic endowments. In other words, investments in academic performance may be a result of a choice-based mechanism, thereby hampering a simple linear decomposition of primary and secondary effects. If this logic holds, primary effects will be overestimated, and the reported secondary effects in this article are lower bound estimates.

However, another—somewhat overlooked—issue related to the estimation of effects may complicate these matters. Measurement error in the academic performance measure will deprive its effect on educational decisions, and primary effects will therefore generally be underestimated. A consequence is that measurement error in academic performance measures establish estimated secondary effects as upper bound estimates. Thus, anticipatory decision making influencing academic performance and measurement error in academic performance measures may counteract each other, the result being that true decompositions lie somewhere in between. Future research will need to develop new methods if it is to overcome these issues. The method developed in this paper may aid such future research.

References

Alwin, D. F. & Hauser, R. M. (1975). The Decompositions of Effects in Path Analysis. *American Sociological Review*, 40, 37-47.

Amemiya, T. (1975). Qualitative Response Models. *Annals of Economic and Social Measurement*, 4, 363-388.

Boudon, R. (1974). *Education, Opportunity and Social Inequality*. New York: Wiley.

Boudon, R. (1989). Subjective Rationality and the Explanation of Social Behavior. *Rationality and Society*, 1, 173-196.

Boudon, R. (1994). *The Art of Self-Persuasion. The Social Explanation of False Beliefs*. Cambridge: Polity Press.

Breen, R. & Goldthorpe, J. (1997). Explaining Educational Differentials: Towards a Formal Rational Action Theory. *Rationality and Society*, 9, 275-305.

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

Cameron, S. V. & Heckman, J. J. (1998): Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males. *Journal of Political Economy*, 106, 262-333.

Clogg, C.C., Petkova, E., & Haritou, A. (1995a). Statistical Methods for Comparing Regression Coefficients Between Models. *The American Journal of Sociology*, 100, 1261-1293.

Clogg, C.C., Petkova, E., & Haritou, A. (1995b). Reply to Allison: More on Comparing Regression Coefficients. *The American Journal of Sociology*, 100, 1305-1312.

Cramer, J.S. (2003). *Logit Models*. *From Economics and Other Fields.* Cambridge: Cambridge University Press.

Duncan, O. D. (1961). A Socioeconomic Index for All Occupations. In A.J. Reiss (Ed.), *Occupations and Social Status* (pp. 109-138). New York: The Free Press of Glencoe, Inc.

Duncan, O. D. (1966). Path Analysis: Sociological Examples. *The American Journal of Sociology*, 72, 1-16.

Erikson, R. (2007). Social selection in Stockholm schools: primary and secondary effects on the transition to upper secondary education. In: Scherer, S., Pollak, R., Otte, G., Gangl, M. (eds.): *From Origin to Destination: Trends and Mechanisms in Social Stratification Research*. Chicago: The University of Chicago Press.

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 Science (PNAS) of the USA*, 102, 9730-9733.

Fienberg, S.E. (1980). *The Analysis of Cross-classified Categorical Data.* 2nd ed. Cambridge, Massachusetts: MIT Press.

Freedman, D.A. (1987). As Others See Us: A Case Study in Path Analysis. *Journal of Educational Statistics*, 12, 101-128.

Freedman, D.A. (2005). *Statistical Models: Theory and Practice*. Cambridge: Cambridge University Press.

Goldthorpe, J.H. (2007). *On Sociology*. (2nd ed., vol. 2). Stanford: Stanford University Press, (Chapter 7).

Goldthorpe, J. H. (1996). Class analysis and the reorientation of class theory: the case of persisting differentials in educational attainment. *British Journal of Sociology*, 47, 481-505.

Hagenaars, J. A. (1998). Categorical Causal Modeling: Latent Class Analysis and Directed Log-Linear Models with Latent Variables. *Sociological Methods Research*, 26, 436-486.

Hansen, E. J. (1995). *En generation blev voksen*. [A Generation Grew Up]. Copenhagen: SFI – Danish Institute of Social Research.

Holm, A. & Jæger, M. M. (2008). Does Relative Risk Aversion explain educational inequality? A dynamic choice approach, *Research social stratification and mobility*, 26, 199-219.

Jackson, M., Erikson, R., Goldthorpe, J.H., & Yaish, M. (2007). Primary and Secondary Effects in Class Differentials in Educational Attainment: The Transition to A-Level Courses in England and Wales. *Acta Sociologica*, 50, 211-229.

Jæger, M. M. & Holm, A. (2007). Does parent's economic, cultural, and social capital explain the social class effect on educational attainment in the Scandinavian mobility regime? *Social Science Research*, 36, 719-744.

Karlson, K. B., Holm, A., & Breen, R. (2010). Comparing Regression Coefficients Between Logit and Probit Models: A New Method. *Unpublished paper*.

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

Mare, R. D. (1980). Social Background and School Continuation Decisions. *Journal of the American Statistical Association*, 75, 295-305.

Mare, R.D. (1981). Change and Stability in Educational Stratification. *American Sociological Review*, 46, 72-87.

Sobel, M. E. (1998). Causal Inference in Statistical Models of the Process of Socioeconomic Achievement. *Sociological Methods & Research*, 27, 318-48.

Winship, C. & Mare, R. D. (1984). Regression Models with Ordinal Variables. *American Sociological Review*, 49, 512-525.

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

Appendix

Deriving the decomposition

Karlson, Holm, and Breen (2010) develop a formal method for making unbiased comparisons of logit coefficients across models without and with potentially confounding variables. Using definitions similar to those in (7), we may write a logit model with only *x* as:

$$\text{logit}(\Pr(y=1)) = \frac{\alpha^*}{\sigma_e^*} + \frac{\beta^*}{\sigma_e^*}x, \qquad \text{(A1)}$$

The star (*) indicates that the parameters differs from those in (7). We note that the scale parameter, σ_e^*, is greater than the scale parameter of the model in (7), σ_e (see Yatchew & Griliches, 1985). Consequently, comparing the logit coefficient of *x* between (A1) and (7) conflates true confounding ($\beta^* - \beta$) with rescaling ($\sigma_e^* - \sigma_e$). However, Karlson, Holm, and Breen (2010) show how to bring the difference, $\beta - \beta^*$, on the same scale, namely the scale parameter of the model in (7), σ_e. Replacing the *z*-variable in (7) with the residual of the model in (5), *v*, we obtain a reparametrization of the model in (7). (In the parlance of Karlson, Holm, and Breen (2010), the residual of the model in (5) is the *x*-residualized control variable):

$$\text{logit}(\Pr(y=1)) = \frac{\alpha^\dagger}{\sigma_e} + \frac{\beta *}{\sigma_e}x + \frac{\gamma}{\sigma_e}v, \qquad \text{(A2)}$$

which has the same scale parameter as the model in (7), σ_e, because the two models have the same fit to the data. Karlson, Breen, and Holm (2010) prove this, and they also prove that the underlying beta coefficient, $\beta *$, in (A2) equals the beta coefficient in (A1). Thus, the logit coefficient of *x* in the model in (A2) is the total effect of *x* measured on the scale of the model in (7). Taking the difference between two coefficients, we obtain

$$\frac{\beta *}{\sigma_e}-\frac{\beta}{\sigma_e}=\frac{\beta *-\beta}{\sigma_e}. \tag{A3}$$

This difference is a measure of confounding due to the inclusion of *z*, measured on the scale of the model including *z*. This measure does not conflate true confounding with rescaling, because both coefficients are measured on the same scale. It is now straightforward to prove that

$$\frac{\beta *-\beta}{\sigma_e}=\frac{\gamma\theta}{\sigma_e}\Rightarrow \beta *-\beta=\gamma\theta. \tag{A4}$$

Applying basic principles of OLS to the latent beta coefficients, we may write the coefficients of *x* in the linear models underlying (A1) and (7) as:

$$\beta * = r_{yx}\frac{\sigma_y}{\sigma_x} \quad \text{and} \quad \beta=\frac{(r_{yx}-r_{xz}r_{yz})}{(1-r_{xz}^2)}\frac{\sigma_y}{\sigma_x},$$

where *r* denotes the correlation coefficient. Clogg *et al.* (1995b) show that the difference between these two coefficients equals

$$\beta *-\beta=\frac{r_{xz}(r_{yz}-r_{xz}r_{yx})}{(1-r_{xz}^2)}\frac{\sigma_y}{\sigma_x}. \tag{A5}$$

Now, the coefficients γ and θ from the models in (7) and (5) can be written as

$$\gamma=\frac{(r_{yz}-r_{xz}r_{yx})}{(1-r_{xz}^2)}\frac{\sigma_y}{\sigma_z} \quad \text{and} \quad \theta=r_{zx}\frac{\sigma_z}{\sigma_x}.$$

The product of these two coefficients is

$$\beta_{yz\cdot x}\theta_{zx}=\frac{(r_{yz}-r_{xz}r_{yx})}{(1-r_{xz}^2)}\frac{\sigma_y}{\sigma_z}r_{xz}\frac{\sigma_z}{\sigma_x}=\frac{r_{xz}(r_{yz}-r_{xz}r_{yx})}{(1-r_{xz}^2)}\frac{\sigma_y}{\sigma_x}, \tag{A6}$$

which equals (A5). We have consequently proven the equality in (A4). In other words, the product of coefficients equals the indirect effect of *x* that operates through *z*. In the situation of logit models, this indirect effect is measured on the scale defined by the model in (7). In this model, the effect of *x* is measured on the same scale as the indirect effect, thereby allowing them to be directly compared. Moreover, because the direct and indirect effects are measured on the same scale, their sum, the total effect, is also measured on the same scale. We note that this total effect equals the logit coefficient of x in model (A2), $\frac{\beta *}{\sigma_e}$, not the one in (A1). In summary, we have shown how to decompose, in logit models, the total effect of x into its direct and indirect effects. Karlson, Holm, and Breen (2010) provide a more thorough discussion of the issues of separating confounding and rescaling in cross-model comparisons of coefficients.

Monte Carlo study of performance statistical test

We conduct a Monte Carlo study to learn about the power of the statistical test. The true model is

$$y* = \beta x + \gamma z + u,$$

where x and z are normally distributed variables and u follows a standard logistic distribution with mean zero and a standard deviation of 1.81. We construct z such that

$$z = \theta x + 2v,$$

where v is a normally distributed random error term with mean zero and variance of unity. We divide $y*$ at its median such that y is distributed 50/50. The values of x are randomly drawn for each sample. We fix $\beta = 1$ and vary the parameters γ and θ, reflecting different degrees of confounding (i.e., relative magnitude of the indirect effect). We also vary the sample size. Results for each scenario are based on 1,000 replications. Because the indirect effect is a product of two coefficients, we want to test the following null hypothesis:

$$H_0: \quad \gamma = 0 \quad \vee \quad \theta = 0$$
$$H_1: \quad \gamma \neq 0 \quad \wedge \quad \theta \neq 0$$

We note that we reject the null whenever both γ and θ is statistically significant from zero; that is, to reject the null hypothesis, we carry out a statistical test for each parameter. The null is true if one of the two parameters cannot be rejected. We specify each test as a two-sided test and use a five percent significance level, leading us to reject the null hypothesis whenever the critical value exceeds ±1.96.

In Table A1 we report the probability of rejecting H_0 in a range of scenarios with varying magnitudes of the indirect effect relative to the total effect. We see that the test performs well. Even for moderate fractions of the indirect effect as in scenario D, where the indirect effect accounts for 20 percent of the total effect, the power is above 90 percent for sample sizes of N=200. In fact, whenever the sample size reaches N=1,000, irrespective of scenario, the power is high. In scenarios I, J, and K, in which there is the null hypothesis is true (i.e., no indirect effect), we also see that the tests rejects the null hypothesis around 5 percent of the times, as we would expect, when we work with a five percent significance level. Thus, combining the test statistics of γ and θ provides us with a means of testing the null hypothesis of no indirect effect.

-- TABLE A1 HERE –

Monte Carlo study comparing KHB-method with methods by Erikson et al. and Buis

We conduct a Monte Carlo study to learn about the performance of the KHB-method relative to the method by Erikson et al. (2005) and the generalization by Buis (2010). In the study we vary distributions of *x* and *z*, the degree of confounding, and the distribution of *y*. We draw 5,000 observations and run 500 replications. The true model is

$$y^* = \alpha + 1x + \gamma z + u,$$

where u follows a standard logistic distribution with mean zero and standard deviation 1.81. The observed y is defined as

$$\begin{aligned} y &= 0 \quad \text{if} \quad y^* \leq \tau \\ y &= 1 \quad \text{otherwise.} \end{aligned}$$

We vary the distribution of y by combining the values of α and τ such that they produce three different distributions: 50/50, 75/25, and 95/5. In all simulations, x is a discrete variable, and z is a continuous variable. We vary the distributions of x and z to create scenarios with skewed distributions. We also vary the degree of confounding. We define weak confounding as $r_{xz} = 0.25$ and $\gamma = 1$, yielding a true fraction of the indirect effect of the total effect of about one-fifth, while we define strong confounding as $r_{xz} = 0.50$ and $\gamma = 4$, yielding a true fraction of about two-thirds. We use logit models to assess the degree of confounding and we report the fractions recovered from the true, underlying model, and from the three techniques (KHB, Erikson et al., and Buis). Because the methods by Erikson et al. and Buis each give two fractions, we report, for each method, the average of these fractions (as suggested in Jackson *et al.*, 2007).

Table A2 reports the results of the Monte Carlo study. In the situation of 50/50-distributed x and normal z, we see that all three methods correctly recover the true fraction. Even for highly skewed y and strong confounding, the methods are never more than one percentage point off the true fraction. However, when z has a lognormal distribution and x is 50/50-distributed, the methods by Erikson et al. and Buis do not recover the true fraction. This bias appears to vary with the distribution of y and the degree of confounding. The more skew y and the more confounding, the larger the bias. This bias does not exist for the KHB-method. For example, in the situation of 75/25-distributed y and strong confounding, Buis' method is about 5 percentage points off the true fraction, and the method by Erikson et al. is about 6 percentage points off. The KHB-method is, on the other hand, 0.2 percentage points off, a negligible bias. To summarize, findings from Monte Carlo study reveal that the methods by Erikson et al. and Buis are biased in situations where the control variable, z, is skewed.

-- TABLE A2 HERE –

Output behind Table 5

-- TABLE A3 HERE –

-- TABLE A4 HERE –

Output behind Table 6

-- TABLE A5 HERE –

-- TABLE A6 HERE –