THE STATA JOURNAL

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go "beyond the Stata manual" in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

For more information on the *Stata Journal*, including information for authors, see the web page

http://www.stata-journal.com

The *Stata Journal* is indexed and abstracted in the following:

- CompuMath Citation Index®
- Current Contents/Social and Behavioral Sciences®
- RePEc: Research Papers in Economics
- Science Citation Index Expanded (also known as SciSearch®)
- Social Sciences Citation Index®

Direct and indirect effects in a logit model

Maarten L. Buis
Department of Sociology
Tübingen University
Tübingen, Germany
maarten.buis@uni-tuebingen.de

Abstract. In this article, I discuss a method by Erikson et al. (2005, *Proceedings of the National Academy of Science* 102: 9730–9733) for decomposing a total effect in a logit model into direct and indirect effects. Moreover, I extend this method in three ways. First, in the original method the variable through which the indirect effect occurs is assumed to be normally distributed. In this article, the method is generalized by allowing this variable to have any distribution. Second, the original method did not provide standard errors for the estimates. In this article, the bootstrap is proposed as a method of providing those. Third, I show how to include control variables in this decomposition, which was not allowed in the original method. The original method and these extensions are implemented in the `ldecomp` command.

Keywords: st0182, ldecomp, mediation, intervening variable, logit, direct effect, indirect effect

In this article, I aim to study direct, indirect, and total effects in a logit model and, in particular, a generalization of a method by Erikson et al. (2005) for computing those effects when the variable whose effect we want to decompose is a categorical variable. Direct, indirect, and total effects are studied to reveal a mechanism through which one variable affects another variable. The idea is illustrated using figure 1: There is a variable X that has an effect on a variable Y, but part of this effect occurs through another variable, Z. This indirect effect occurs because X influences Z, which in turn influences Y. Within this framework, the effect of X on Y while controlling for Z is called the direct effect. The indirect and direct effect together form the total effect of X on Y. The indirect effect is the part of the effect of X on Y that can be explained, while the direct effect is the residual, or unexplained, part of the effect. The aim of such an analysis is to try to explain why a variable X influences a variable Y by specifying a mechanism: the effect occurs through the third variable, Z. Z is called an intervening variable. It differs from a confounding variable in the direction of the effect between X and Z. If Z was a confounding variable, then it would affect X rather than the other way around.
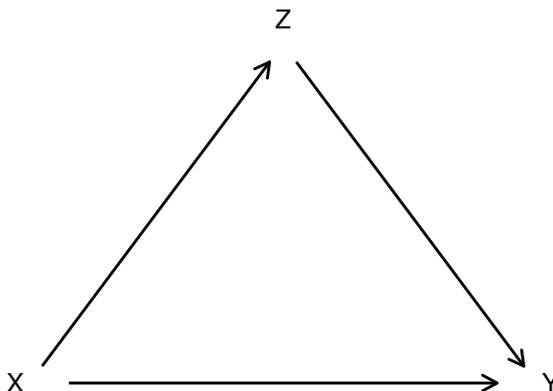
Figure 1. Direct, indirect, and total effects.

X, Y, and Z could be many things. For example, within political science there is literature (e.g., Campbell et al. [1960]) that explains the effect of party identification (X) on voting for a party (Y) in part through how one perceives the issues and the parties (Z). The idea is that someone's party identification is a relatively stable characteristic of a person, almost like a personality trait, which can color/distort the perceptions of the issues and the positions of the candidates or parties,[1] which in turn influences voting behavior. This indirect effect thus represents a mechanism through which party identification influences voting behavior, and we want to know how important that mechanism is relative to the total effect of party identification.

Alternatively, within sociology there is literature (e.g., Boudon [1974]) that looks at the degree to which children from different social backgrounds (X) have different probabilities of attending college (Y) and the part of this effect that can be ascribed to an indirect effect through academic performance in high school (Z): Children from a higher class background do well in high school, and those children who do well in high school are more likely to attend college. Possible explanations for this could be that higher class parents have better genes (nature) or are able to provide a more intellectually stimulating environment for their children (nurture) or both. Again, the aim is to estimate how important this mechanism is relative to the total effect of family background.

---

1. A striking example of this is reported by Bartels (2002), who found in a survey held in 1988 that more than 50% of the respondents who identified themselves as a "strong Democrat" thought that the inflation got worse or a lot worse during the Reagan presidency, while only 13% of the "strong Republicans" thought that this was the case. In actual fact, the inflation rate in consumer prices fell from 13.5% in 1980 to 4.1% in 1988.

More generally, the common practice of comparing regression coefficients before and after controlling for a set of other covariates is implicitly an attempt at decomposing a total effect into a direct and an indirect effect. Within linear regression ([R] **regress**), this is an easy and convenient way of estimating the direct, indirect and total effect. In this case, the total effect of X is estimated using a regression that does not control for Z. The direct effect of X is the effect of X in a regression with both X and Z as explanatory variables, and the indirect effect is the difference between these two, because the total effect is the sum of the direct and indirect effect. However, this method will not work when dealing with nonlinear models like logistic regression ([R] **logit**). In this article, I will show why this is the case. I will also discuss a method proposed by Erikson et al. (2005) for solving this problem, I will propose a generalization of this method, and I will introduce the `ldecomp` command, which implements these methods in Stata.

1 The problem of indirect effects in a logit model

The key problem when estimating the direct, indirect, and total effects is that the standard method of estimating them—comparing estimates from models that do and do not control for Z—will not work in nonlinear models like logistic regression. The easiest way to see that is in an example where there is no indirect effect.

To illustrate the problem, I create artificial data where I know that there cannot be an indirect effect of X through Z on Y; I show that the naïve method finds a substantial indirect effect. The variable Z is created so that it can take three values (0, 1, and 2), and the variable X is created so that it can take two values (0 and 1). There is no relationship between X and Z (a low X individual is as likely to be a high value on Z as a high X individual). So in this example, there is no indirect effect of X on Y through Z. We start by creating the variables X and Z:

```
. drop _all
. set obs 60000
obs was 0, now 60000
. generate z = ceil(_n/20000) - 1
. bysort z: generate x = ceil(_n/10000) - 1
. tab x z
```

		z		
x	0	1	2	Total
0	10,000	10,000	10,000	30,000
1	10,000	10,000	10,000	30,000
Total	20,000	20,000	20,000	60,000

Next we create the dependent variable Y according to the logistic regression equation $P(y=1)/1 - P(y=1) = -4 + 4X + 2Z$, as discussed in Buis (2007b).

```
. set seed 12345
. generate y = runiform() < invlogit(-4 + 4*x + 2*z)
```

Next we compute the naïve estimates of the indirect effect:

```
. quietly logit y x z
. estimates store direct
. local direct = _b[x]
. quietly logit y x
. estimates store total
. local total = _b[x]
. estimates tab direct total
```

Variable	direct	total
x	4.0391332	2.6256242
z	2.026339	
_cons	-4.0452305	-1.3123133

```
. display "naive indirect effect = " `total´ - `direct´
naive indirect effect = -1.413509
```

Someone who did not know that the data were created such that the indirect effect is exactly zero and so used the naïve estimate of the indirect effect would conclude that the indirect effect is about 54% of the total effect; the fact that it has the opposite sign from the total effect would suggest that this indirect effect has a noticeable dampening influence on the effect of X on Y. This is not caused by sampling variation but instead caused by a structural bias in the naïve estimator.

The reason for this bias is that a logistic regression is comparison of proportions that have first been transformed into log odds-ratios;[2] this is illustrated in figure 2. It consists of four vertical lines: on the two outer lines are plotted the probabilities of success on Y, while the equivalent log odds are plotted on the two inner lines. The two left lines represent the high X group, while the two right lines represent the low X group. The open symbols and the solid arrows show how the probabilities are transformed into log odds and how, within each category of Z, the log odds of the high X group are compared with the log odds of the low X group.

This represents what happens when we estimate a `logit` model with both X and Z as explanatory variables. When we leave the variable Z out of the model—for example, because we want to estimate the total effect—we are in effect first computing the average of the proportions and then transforming them into log odds. This is represented by the closed circles and the dashed arrows. However, the more extreme values are less extreme in the probability metric than in the log odds metric; that is, the probabilities close to either 0 or 1 are more bunched together than their log odds counterparts. Therefore, computing the average proportion before transforming the proportions into log odds means that the extreme values are less influential than they would have been if the

---

2. The odds and the log odds contain exactly the same information as the proportion, just presented differently. The proportions times a hundred tell you how many people out of a hundred are expected to attend college, while the odds tell you how many people are expected to attend college for every person who does not attend college. The odds (O) can be derived from the proportion (p) in the following way: $O = p/1 - p$.

means were computed in the log odds metric, so the average is being pulled toward the less extreme categories. As a consequence, the effect in terms of log odds will be less when Z is left out of the model, even if there is no indirect effect. This problem is very similar to the problems discussed in Bartus (2005) and Buis (2007a).
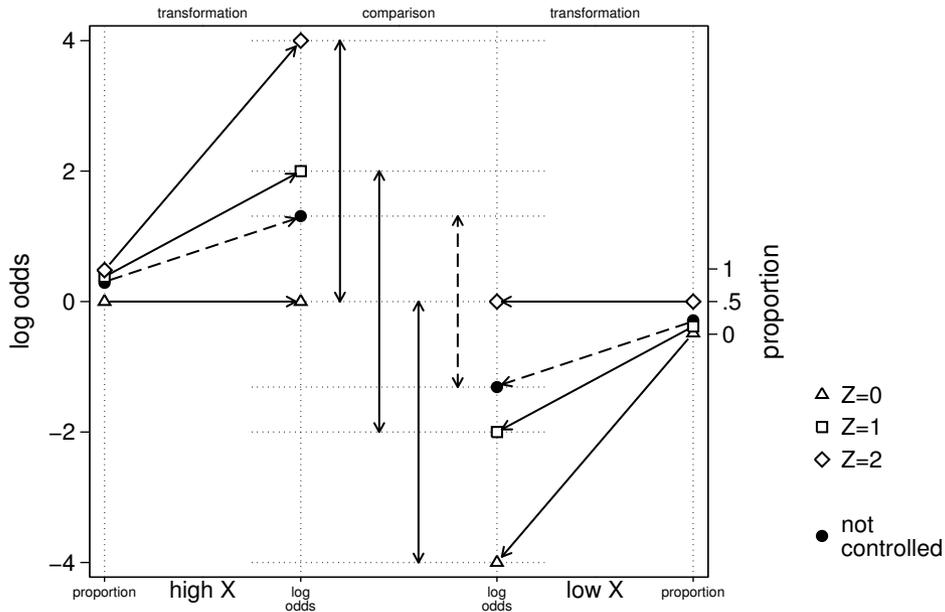
Figure 2. The effect of X with and without controlling for Z.

2 A solution

2.1 Outline

For there to be an indirect effect of X on Y through Z, X needs to have an effect on Z and Z has to have an effect on Y. For example, children from higher classes (X) are likely to do better in high school (Z) than children from lower classes, and those who have done better in high school (Z) are more likely to enter college (Y). In figure 2, there was no indirect effect because the distribution of Z was the same for both categories of X. In figure 3, this has changed to include an indirect effect. The distribution of Z is represented by the size of the symbols. So in this new example, there are within the high X group more high Z individuals than medium Z individuals and more medium Z individuals than low Z individuals. The distribution of Z for the low X group is exactly the opposite. Now there are two reasons why high X individuals are more likely to belong to the high Y group: 1) They are more likely to belong to the high Z group, and

those who belong to the high Z group are more likely to belong to the high Y group. This is the indirect effect of X through Z on Y. Figure 3 shows this in the following way: for the high X group, the high Z group is more influential when computing the average probability because the high Zs are more numerous, while for the low X group, the low Zs are more influential. This leads to a larger difference between the high and low status group. 2) They are more likely to belong to the high Y group even when compared with a low X group that has the same value on Z. Figure 3 shows this in the following way: for each level of Z, the high X group has a higher proportion belonging to the high Y group than does the low X group. This is the direct effect.
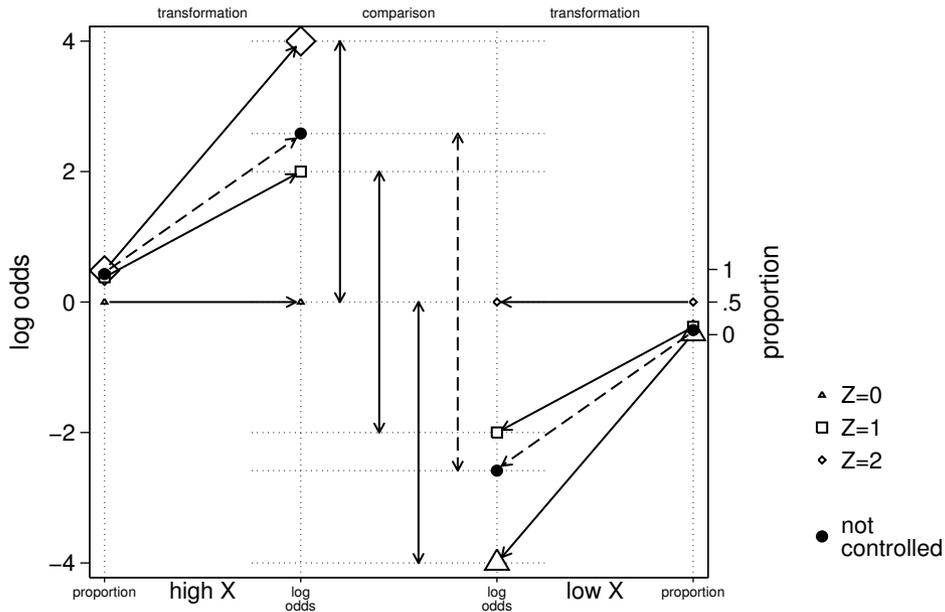
Figure 3. Indirect and direct effects.

Erikson et al. (2005) propose two related methods for estimating a direct, indirect, and total effect such that the direct and indirect effects add up to the total effect. The first method is illustrated in figure 4 and consists of the following steps:

1. Estimate a logistic regression using both X and Z, and optionally, the interactions between X and Z.

2. Predict for each individual the log odds of success on Y (this is the linear predictor) and transform these to probabilities.

3. Compute within each group of X the average predicted probability and transform these to log odds. The difference in these log odds between the groups of X represents the total effect.

4. Assume that the low X individuals are actually high X individuals, predict for each low X individual the log odds in this counterfactual scenario, and transform these to probabilities.

5. Compute the average of these probabilities. This is the counterfactual probability of success on Y for high X individuals if they had the distribution of Z of the low X individuals. These are then transformed into log odds.

6. The high X individuals and this counterfactual group differ with respect to the distribution of Z, but the probabilities conditional on X and Z are kept constant. Therefore, comparing these groups gives the effect of X caused by the differences in the distribution of Z, that is, the indirect effect.

7. The low X individuals and the counterfactual group differ with respect to the probabilities conditional on X and Z, but the distribution of Z is kept constant. Therefore, comparing these groups gives the effect of X while controlling for the distribution of Z, that is, the direct effect.
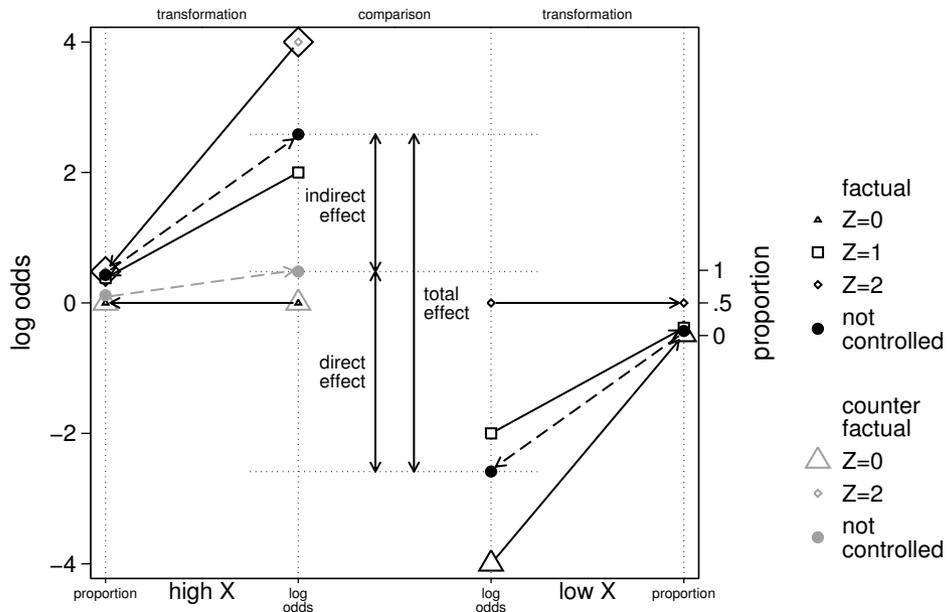
Figure 4. Method 1.

Figure 4 above and (1) on the following page both show that this way the total effect is always the sum of the direct and indirect effect. In (1), the O is the odds of success on Y, the first subscript represents the logistic regression coefficients, and the second subscript represents the distribution of Z.[3] So $O_{x=1,z\,|\,x=1}$ is the odds of success for the high X group, while $O_{x=1,z\,|\,x=0}$ is the counterfactual odds of success for the high X group if it had the distribution of Z of the low X group.

$$\underbrace{\ln(O_{x=1,z\,|\,x=1}) - \ln(O_{x=0,z\,|\,x=0})}_{\text{total}} = \underbrace{\ln(O_{x=0,z\,|\,x=1}) - \ln(O_{x=0,z\,|\,x=0})}_{\text{indirect}} + \\ \underbrace{\ln(O_{x=1,z\,|\,x=1}) - \ln(O_{x=0,z\,|\,x=1})}_{\text{direct}} \qquad (1)$$

Using the rule that $\ln(a) - \ln(b) = \ln(a/b)$, it can be shown that these effects are actually log odds-ratios:

$$\underbrace{\ln\left(\frac{O_{x=1,z\,|\,x=1}}{O_{x=0,z\,|\,x=0}}\right)}_{\text{total}} = \underbrace{\ln\left(\frac{O_{x=0,z\,|\,x=1}}{O_{x=0,z\,|\,x=0}}\right)}_{\text{indirect}} + \underbrace{\ln\left(\frac{O_{x=1,z\,|\,x=1}}{O_{x=0,z\,|\,x=1}}\right)}_{\text{direct}} \qquad (2)$$

This means that this decomposition can also be presented in terms of odds ratios, by exponentiating both sides of (2). Because of the rule that $\exp(a+b) = \exp(a) \times \exp(b)$, the total effect is now the product of the direct and indirect effects:

$$\underbrace{\frac{O_{x=1,z\,|\,x=1}}{O_{x=0,z\,|\,x=0}}}_{\text{total}} = \underbrace{\frac{O_{x=0,z\,|\,x=1}}{O_{x=0,z\,|\,x=0}}}_{\text{indirect}} \times \underbrace{\frac{O_{x=1,z\,|\,x=1}}{O_{x=0,z\,|\,x=1}}}_{\text{direct}} \qquad (3)$$

The second method of decomposing the total effect into a direct and indirect effect proposed by Erikson et al. (2005) is exactly the same as the first method, except that it uses the counterfactual probability of success on Y for the low X individuals assuming that they have the distribution of Z of the high X individuals, as is illustrated in figure 5. The logic behind these two methods is exactly the same, but they do not have to result in exactly the same estimates for the direct and indirect effects, though they are often very close. Jackson et al. (2007) propose to solve this problem by computing the size of the indirect effect relative to the total effect using both methods and report the average of the two.

---

3. In fact, as can be seen in the description of this technique, these are averaged estimated log odds, so some people would prefer to incorporate that in the notation by adding hats and bars. However, I think that this notation is complicated enough as is, and hats and bars are not crucial for the points that these equations make, so I chose not to include them.
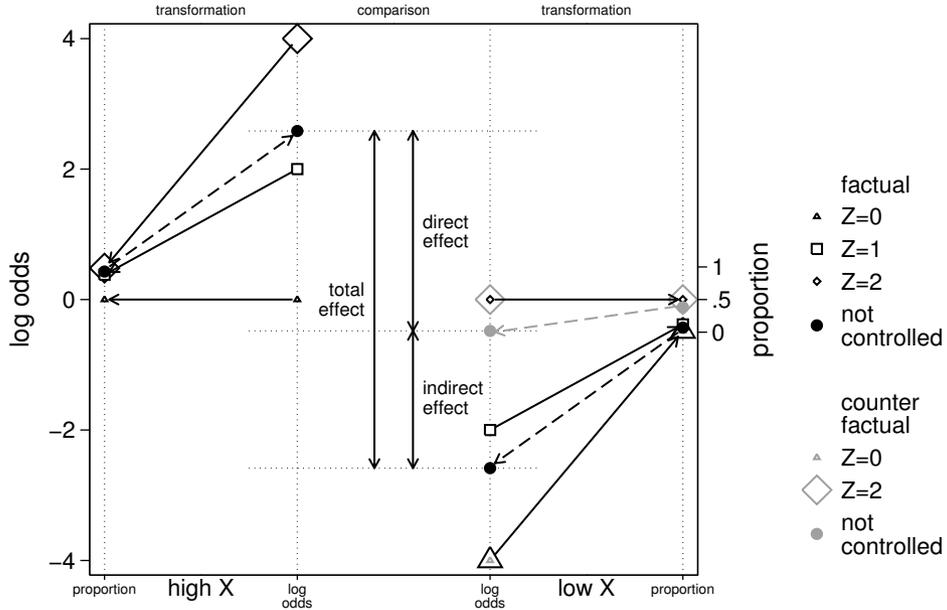
Figure 5. Method 2.

Another difficulty with this method is that the variable whose effect we want to decompose (X) must be categorical and the number of effects that must be decomposed increases very quickly with the number of categories in X. The reason is that the decomposition will be different for all pairwise comparisons of categories. Normally, a categorical explanatory variable consisting of C categories can be entered as $C-1$ dummies, with each dummy representing a comparison between the reference category and one of the other categories, leading to $C-1$ effects. All other pairwise comparisons can be directly derived from those effects. Consider a logit model with a categorical variable with three categories—low ($x = 0$), middle ($x = 1$), and high ($x = 2$)—and the low category was chosen to be the reference category. This will result in two odds ratios: the odds ratio comparing the medium and the low group, and the odds ratio comparing the high and the low group, or $O_{x=1}/O_{x=0}$ and $O_{x=2}/O_{x=0}$, respectively. This excludes the third possible comparison: the medium versus the high group. This comparison can be derived directly from the other two odds ratios by dividing the two odds ratios:

$$\frac{O_{x=1}/O_{x=0}}{O_{x=2}/O_{x=0}} = \frac{O_{x=1}}{O_{x=0}} \times \frac{O_{x=0}}{O_{x=2}} = \frac{O_{x=1}}{O_{x=2}}$$

This same reasoning, however, will not work for the decomposition. Consider the three indirect effects we get when using the first method: $O_{x=0,z\,|\,x=1}/O_{x=0,z\,|\,x=0}$, $O_{x=0,z\,|\,x=2}/O_{x=0,z\,|\,x=0}$, and $O_{x=1,z\,|\,x=2}/O_{x=1,z\,|\,x=1}$.

$$\frac{O_{x=0,z\,|\,x=2}/O_{x=0,z\,|\,x=0}}{O_{x=0,z\,|\,x=1}/O_{x=0,z\,|\,x=0}} \neq \frac{O_{x=1,z\,|\,x=2}}{O_{x=1,z\,|\,x=1}}$$

So with this decomposition, we cannot get away with only presenting all comparisons with a single reference category (leading to $C-1$ effects), but we will have to display the decomposition for all pairwise comparisons (leading to $\binom{C}{2}$ effects).

2.2 Generalization

In the original formulation of this method, the variable through which the indirect effect occurs (Z) has to be normally distributed. This is because of the way Erikson et al. (2005) propose to compute the average predicted probabilities. As discussed in steps 3 and 5 in the previous section, these averaged predicted probabilities play a key role in this decomposition. Erikson et al. (2005) compute these averages by assuming that Z follows a normal distribution, and then they integrate over this distribution. This method is generalized in this article to allow Z to follow any distribution by computing the average predicted probability by using the arithmetic mean of the predicted and counterfactual probabilities, thereby integrating over the empirical distribution of Z instead of over a normal distribution. As an added bonus, this method is also easier to compute because the integration over the normal distribution has to be done numerically since there is no closed-form solution for this integral. For these reasons, the generalized method is the default method in the `ldecomp` command, which implements both decompositions in Stata.

2.3 Standard errors

To get an idea about the degree of uncertainty around these estimates, one would usually also estimate standard errors. These are not provided by Erikson et al. (2005) and Jackson et al. (2007), but can be easily computed using the bootstrap (Efron and Tibshirani 1993). The bootstrap uses the idea that the standard error is the result of the following thought experiment: Assume we could draw many samples from the population and compute a statistic in each of these samples. Because these samples will slightly differ from one another, so will the estimates of the statistic. The standard error is the standard deviation of these different estimates and indicates how much variation one can expect due to the estimate being based on a random sample. Drawing many random samples from the population is not practical, but we do have a good estimate of the population—the sample—and we can without any difficulty draw (with replacement) many samples from this "estimate of the population". So when using the bootstrap, many samples are drawn with replacement from the observed sample, the statistic is computed within each sample, and the estimate of the standard error is the

standard deviation of these statistics. Within Stata, this process has been automated with the `bootstrap` command; see [R] **bootstrap**.

2.4 Control variables

Control variables can be included in this method by estimating the logit model from step 1 (in section 2.1) including these control variables, but afterward, at steps 2 and 4, fixing their value at one specific value, for example their mean.

3   Implementation in Stata

Both the original method proposed by Erikson et al. (2005) and the generalizations proposed in this article have been implemented in Stata as the `ldecomp` command.

3.1 Syntax

`ldecomp` *depvar* $\big[$ *control_var1* $\big[$ ... $\big]\big]$ $\big[$ *if* $\big]$ $\big[$ *in* $\big]$ $\big[$ *weight* $\big]$, <u>direct</u>(*varname*) <u>indirect</u>(*varlist*) $\big[$ at(*control_var1* # $\big[$; *control_var2* # $\big]$ $\big[$ ... $\big]$) <u>obspr</u> <u>predpr</u> <u>predodds</u> or <u>rind</u>irect <u>norm</u>al range(# #) nip(#) <u>interactions</u> <u>nolegend</u> <u>nodec</u>omp <u>noboot</u>strap *bootstrap_options* $\big]$

`fweight`s, `pweight`s, and `iweight`s are allowed when the `nobootstrap` option is specified.

3.2 Options

<u>direct</u>(*varname*) specifies the variable whose direct effect we want to decompose into an indirect and total effect. This has to be a categorical variable, and each value of *varname* is assumed to represent a group.

<u>indirect</u>(*varlist*) specifies the variable(s) through which the indirect effect occurs. By default, multiple variables are allowed and these can follow any distribution. If the `normal` option is specified, only one variable may be entered, and this variable is assumed to be normally distributed.

at(*control_var1* # $\big[$; *control_var2* # $\big]$ $\big[$ ... $\big]$) specifies the values at which the control variables are to be fixed. The default is to fix the control variables at their mean value.

`obspr` specifies that a table of the observed proportions be displayed.

`predpr` specifies that a table of predicted and counterfactual proportions be displayed. If the `normal` option is not specified and there are no control variables, then the diagonal elements of this table will be exactly the same as the observed proportions.

`predodds` specifies that a table of predicted and counterfactual odds be displayed.

`or` specifies that the decomposition be displayed in terms of odds ratios instead of log odds-ratios.

`rindirect` specifies that the relative contributions of the indirect effects to the total effect (in terms of log odds-ratios) be displayed.

`normal` specifies that the predicted and counterfactual proportions be computed according to the method specified by Erikson et al. (2005). This means that the variable specified in `indirect()` is assumed to be normally distributed. This option was primarily added for compatibility with Erikson et al. (2005). By default, `ldecomp` uses the more flexible method described in this article.

`range(# #)` specifies the range over which the numerical integration of the variable specified in `indirect()` is to be performed. The default is the minimum of that variable minus 10% of the range of the variable and the maximum of the variable plus 10% of the range of the variable. This option may only be specified with the `normal` option because in the default method there is no need for numerical integration.

`nip(#)` specifies the number of integration points used in the numerical integration of the variable specified in `indirect()`. The default is `nip(1000)`. This option may only be specified with the `normal` option because in the default method there is no need for numerical integration.

`interactions` specifies that interactions between the categories of the variable specified in `direct()` and the variable(s) specified in `indirect()` be included. In other words, the effects on the dependent variable of the variables specified in `indirect()` are allowed to differ from one another for each category of the variable specified in `direct()`. This option was added primarily for compatibility with Erikson et al. (2005).

`nolegend` suppresses the legend that is by default displayed at the bottom of the main table.

`nodecomp` prevents `ldecomp` from displaying the table of decompositions, which can be useful in combination with the `obspr`, `predpr`, or `predodds` options.

`nobootstrap` prevents `ldecomp` from using `bootstrap` to calculate standard errors.

bootstrap_options are allowable options of `bootstrap`. The following options are allowed: <u>reps</u>(#), <u>strata</u>(*varlist*), <u>size</u>(#), <u>cluster</u>(*varlist*), <u>id</u>cluster(*newvar*), bca, <u>saving</u>(*filename*[, *suboptions*]), <u>jackknifeopts</u>(*jkopts*), mse, seed(#), nodots, and <u>level</u>(#). See [R] **bootstrap**.

3.3 Example

The use of `ldecomp` is illustrated using data from the Wisconsin Longitudinal Study (Hauser and Sewell 1957–1977), which contain data on a random sample of 10,317 men and women who graduated from Wisconsin high schools in 1957. In this example, we want to study the part of the effect of social class on the probability of entering college that can be explained by performance during high school. The dependent variable is whether a respondent ever attended college (`college`). Class is measured by the type of occupation of the father (`ocf57`). The original data contained five classes, which would lead to $\binom{5}{2} = 10$ effects that are to be decomposed. To keep the number of effects manageable, the number of classes has been reduced to three: a lower class (unskilled workers and farmers), a middle class (skilled workers), and a higher class (white collar workers, professionals, and executives). The performance during high school is measured with the percentile rank of high school grades (`hsrankq`). This means that computing the counterfactual proportions using the method proposed by Erikson et al. (2005) will be problematic, because percentile rank scores will follow a uniform distribution instead of a normal distribution. However, the default method can be used without problem because it does not make any assumption about the distribution of performance. Relevant descriptive statistics are shown below:

```
. use wisconsin, clear
(Wisconsin Longitudinal Study)
. recode ocf57 2=1 3=2 4=3 5=3
(ocf57: 6196 changes made)
. label define ocf57 1 "lower" 2 "middle" 3 "higher", modify
. label value ocf57 ocf57
. table ocf57 if !missing(hsrankq, college),
> contents(mean college mean hsrankq freq) format(%9.3g) stubwidth(15)
```

occupation of r father in 1957	mean(college)	mean(hsrankq)	Freq.
lower	.284	48.2	5218
middle	.38	50.6	868
higher	.619	56.2	2837

There are big differences between the classes in the proportion of respondents that attend college. Moreover, those classes with a low proportion attending college also tend to have done worse during high school. So part of the differences in proportions attending college could be due to differences in performance. `ldecomp` is used to estimate these direct, indirect, and total effects. In the example below, the effects are presented as odds ratios, so the total effect is the product of the indirect and direct effect. Consider the decomposition according to the first method of the difference between high class (denoted as 3) and low class (denoted as 1) students, that is, the first three coefficients of the panel labeled "`3/1`". Overall, the odds of attending college for high class students is 4.10 times as large as the odds for low class students (the total effect). Low class students would have 1.23 times higher odds of attending college if they had the same performance as high class students (indirect effect according to method 1), while the

high class students would have 3.34 times higher odds of attending college than low
class students when we keep the performance constant at the level of high class students
(direct effect according to method 1).

```
. ldecomp college, direct(ocf57) indirect(hsrankq) or
(running _ldecomp on estimation sample)
Bootstrap replications (50)
─────┼─── 1 ───┼─── 2 ───┼─── 3 ───┼─── 4 ───┼─── 5
..................................................    50
Bootstrap results                        Number of obs    =    8923
                                         Replications     =      50
```

	Observed Odds Ratio	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf. Interval]	
2/1						
total	1.547746	.1087768	6.22	0.000	1.34858	1.776327
indirect1	1.061166	.0297615	2.12	0.034	1.004408	1.12113
direct1	1.458534	.0889065	6.19	0.000	1.294287	1.643624
indirect2	1.060416	.0292343	2.13	0.033	1.004638	1.11929
direct2	1.459565	.0889226	6.21	0.000	1.295284	1.644683
3/1						
total	4.098896	.1563335	36.99	0.000	3.80366	4.417047
indirect1	1.228616	.0205526	12.31	0.000	1.188987	1.269566
direct1	3.33619	.1231675	32.63	0.000	3.103313	3.586542
indirect2	1.22293	.0212813	11.56	0.000	1.181922	1.26536
direct2	3.351702	.1256571	32.26	0.000	3.114249	3.607259
3/2						
total	2.6483	.1803249	14.30	0.000	2.317438	3.026399
indirect1	1.157325	.03452	4.90	0.000	1.091606	1.226999
direct1	2.288295	.1361168	13.92	0.000	2.036475	2.571253
indirect2	1.153977	.0331169	4.99	0.000	1.090861	1.220745
direct2	2.294933	.1384071	13.77	0.000	2.039079	2.582889

```
in equation i/j (comparing groups i and j)
let the fist subscript of Odds be the distribution of the the indirect variable
let the second subscript of Odds be the conditional probabilities
Method 1: Indirect effect = Odds_ij/Odds_jj
          Direct effect = Odds_ii/Odds_ij
Method 2: Indirect effect = Odds_ii/Odds_ji
          Direct effect = Odds_ji/Odds_jj
value labels
    1 lower
    2 middle
    3 higher
```

To get an idea of the relative importance of the indirect effect compared with the
total effect, one can add the `rindirect` option, like in the example below. This means
that the sizes of the indirect effects relative to total effects are shown at the bottom of the
decomposition table. These are labeled #/#r. So if we look again at the comparison
between children from higher and lower class fathers (the rows labeled `3/1` and `3/1r`),
we see that according to method 1 the indirect effect is $0.206/1.41 \times 100\% = 14.6\%$
of the total effect. According to method 2, this is $0.201/1.41 \times 100\% = 14.3\%$. So,
on average, the indirect effect is 14.5% of the total effect. In general, this table shows

that the size of the indirect effect is about 14 percent of the total effects. Additionally, the example below illustrates that by leaving out the `or` option, `ldecomp` will show the direct, indirect, and total effects in terms of log odds-ratios, which means that the total effect is now the sum of the direct and indirect effects.

```
. ldecomp college, direct(ocf57) indirect(hsrankq) rind nolegend
(running _ldecomp on estimation sample)

Bootstrap replications (50)
────┼── 1 ──┼── 2 ──┼── 3 ──┼── 4 ──┼── 5
.................................................  50

Bootstrap results                          Number of obs    =      8923
                                           Replications     =        50
```

	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf. Interval]	
2/1						
total	.4367997	.0681297	6.41	0.000	.303268	.5703314
indirect1	.0593679	.0244276	2.43	0.015	.0114906	.1072451
direct1	.3774319	.0684413	5.51	0.000	.2432894	.5115743
indirect2	.0586611	.0240686	2.44	0.015	.0114875	.1058347
direct2	.3781386	.0684292	5.53	0.000	.2440198	.5122575
3/1						
total	1.410718	.0421018	33.51	0.000	1.3282	1.493236
indirect1	.2058881	.014635	14.07	0.000	.177204	.2345723
direct1	1.204829	.0390007	30.89	0.000	1.12839	1.281269
indirect2	.2012494	.014978	13.44	0.000	.1718931	.2306057
direct2	1.209468	.0386461	31.30	0.000	1.133723	1.285213
3/2						
total	.9739179	.075011	12.98	0.000	.8268989	1.120937
indirect1	.1461109	.0234312	6.24	0.000	.1001865	.1920353
direct1	.8278069	.0759825	10.89	0.000	.678884	.9767298
indirect2	.1432144	.0237856	6.02	0.000	.0965955	.1898332
direct2	.8307035	.0760533	10.92	0.000	.6816418	.9797652
2/1r						
method1	.1359155	.0574128	2.37	0.018	.0233885	.2484425
method2	.1342975	.0568445	2.36	0.018	.0228844	.2457105
average	.1351065	.0571076	2.37	0.018	.0231776	.2470354
3/1r						
method1	.1459457	.0096034	15.20	0.000	.1271234	.164768
method2	.1426575	.0097455	14.64	0.000	.1235567	.1617583
average	.1443016	.0095878	15.05	0.000	.1255099	.1630933
3/2r						
method1	.1500239	.0256887	5.84	0.000	.0996749	.2003729
method2	.1470497	.0258689	5.68	0.000	.0963475	.1977519
average	.1485368	.0256919	5.78	0.000	.0981816	.198892

Notice that the size of the indirect effect relative to the total effect can be larger than 100%, negative, or both. This might puzzle people who think of these relative effects as the proportion of the total effect that can be explained by the indirect effect. However, there is no reason why the direct and indirect effect cannot have opposite signs, and in

those cases, these "weird" proportions can occur. Thinking of these numbers as the size of the indirect effect relative to the size of the total effect can help avoid confusion.

In addition to the decomposition itself, `ldecomp` can produce a number of tables that together illustrate step by step how it builds this decomposition. The first of these tables is the table of the predicted and counterfactual proportions, shown below.

```
. ldecomp college, direct(ocf57) indirect(hsrankq) predpr nodecomp
predicted and counterfactual proportions
```

distribution	association		
	lower	middle	higher
lower	.284	.366	.571
middle	.296	.38	.585
higher	.327	.415	.619

The diagonal elements in this table represent the predicted proportions, both factual distribution of performance (the rows) and factual conditional probabilities (the columns), while the off-diagonal elements represent the counterfactual proportions. For example, 28.4% of the children from lower class fathers will attend college. If these children had the same performance as the children of higher class fathers, then 32.7% of them would attend college. If they had the same conditional probabilities as the children of higher class fathers, then 57.1% would attend college. This indicates that the direct effect is stronger than the indirect effect. The next step in the computation is to transform these proportions into odds:

```
. ldecomp college, direct(ocf57) indirect(hsrankq) predodds nodecomp
predicted and counterfactual odds
```

distribution	association		
	lower	middle	higher
lower	.396	.578	1.33
middle	.421	.613	1.41
higher	.487	.71	1.62

The proportions are transformed into odds by dividing the proportion by one minus the proportion, so the odds of attending college for children from lower class fathers is $0.284/1 - 0.284 = 0.397$. The difference between this number and the number in the table is caused by rounding. This odds is interpreted as follows: for every child of a lower class father who does not go to college, there are only 0.397 children of lower class fathers who do go to college.

The results that were obtained at the beginning of this example can be computed using the predicted and counterfactual odds from the previous table. For example, if we return to the decomposition of the total effect of having a higher class father rather than a father from the lower class using method 1, we can compute it by filling in (3) with the predicted and counterfactual odds from the previous table:

$$\underbrace{\frac{1.62}{0.396}}_{\text{total}} = \underbrace{\frac{0.487}{0.396}}_{\text{indirect}} \times \underbrace{\frac{1.62}{0.487}}_{\text{direct}}$$

$$\underbrace{4.09}_{\text{total}} = \underbrace{1.23}_{\text{indirect}} \times \underbrace{3.33}_{\text{direct}}$$

This example also makes it possible to see if the naïve method is really as bad as I claim, and if the generalization that I proposed in this article is really an improvement. We know that `hsrankq` deviates considerably from a normal/Gaussian distribution,[4] so if the generalized method is an improvement on the original method, then in this example it should yield noticeably different estimates. The estimates using the original method by Erikson et al. (2005) are obtained by specifying the `normal` option in `ldecomp`.

The naïve estimate was computed as follows: First, a `logit` model of `college` on `ocf57` was estimated. The effect of `ocf57` is the naïve estimate of the total effect. Then a `logit` model of `college` on `ocf57` and `hsrankq` was estimated. The effect of `ocf57` is the naïve estimate of the direct effect. The naïve estimate of the indirect effect relative to the total effect is (total effect − direct effect) / total effect. The results are shown in table 1. This table clearly illustrates the major underestimation of the indirect effect when the naïve method is used. Moreover, the method by Erikson et al. (2005) also leads to a considerable underestimation of about a quarter of the effect obtained using the generalized method. This underestimation is the result of the incorrect assumption made by Erikson et al. (2005) that `hsrankq` is normally distributed. Because the generalized method makes no such assumption, it seems to be the safest method of the three for computing the decomposition of total effects into direct and indirect effects after a logistic regression model.

Table 1. Comparing different estimates of the size of the (average) indirect effect relative to the size of the total effect.

	generalization	Erikson et al. (2005)	naïve
middle versus low	0.135	0.110	0.009
high versus low	0.144	0.105	0.014
high versus middle	0.149	0.102	0.017

4   Summary and discussion

In this article, I began by showing why getting estimates of the direct and indirect effects in a logistic regression is hard. I then presented a method by Erikson et al. (2005) to

---

4. As noted before, it is a percentile rank score, so it follows a uniform distribution.

estimate these effects, and I proposed a generalization of this method. The idea is that a categorical variable X can have a direct effect on a variable Y and an indirect effect through Z. One can find the direct effect of X by comparing the log odds of successes in one category of X with the counterfactual log odds of successes of another category of X given that they have the distribution of Z of the first category. This way, the factual and counterfactual group only differ with respect to X and not with respect to the distribution of Z, thus controlling for the distribution of Z. Similarly, the indirect effect can be found by comparing the log odds of success within one category of X with the counterfactual log odds of success within that same category of X with a distribution of Z of another category of X. This way, the factual and counterfactual groups differ only with respect to the distribution of Z. In its original form, this method assumed that the variable through which the indirect effect occurs has a normal distribution. In this article, this method was generalized by allowing the variable to have any distribution. Moreover, the use of the bootstrap is proposed to estimate standard errors, and I added the possibility to include control variables.

This is a relatively new methodology, and so there are still some loose ends to tie up. First, the fact that there are two different estimates of the direct effects and two different estimates of the indirect effects is less than elegant. Second, the fact that a separate decomposition needs to be estimated for all pairwise class comparisons instead of all comparisons with a single reference category can quickly lead to a very large number of estimates. Third, this method is not the only way of attaining estimates of direct and indirect effects; there are, for instance, the methods proposed by Gomulka and Stern (1990), Even and Macpherson (1990), Fairlie (2005), Yun (2004), and Bauer and Sinning (2008). Two of these have been implemented in Stata: the method by Fairlie (2005) in the `fairlie` command (Jann 2006), and the method by Bauer and Sinning (2008) in the `nldecompose` command (Sinning, Hahn, and Bauer 2008). How these alternatives compare with the method discussed in this article needs to be explored.

5   Acknowledgments

6   References

Bartels, L. M. 2002. Beyond the running tally: Partisan bias in political perceptions. *Political Behavior* 24: 117–150.

Bartus, T. 2005. Estimation of marginal effects using margeff. *Stata Journal* 5: 309–329.

Bauer, T. K., and M. Sinning. 2008. An extension of the Blinder–Oaxaca decomposition to nonlinear models. *Advances in Statistical Analysis* 92: 197–206.

Boudon, R. 1974. *Education, Opportunity and Social Inequality Prospects in Western Society*. New York: Wiley.

Buis, M. L. 2007a. Predict and adjust with logistic regression. *Stata Journal* 7: 221–226.

———. 2007b. Stata tip 48: Discrete uses for uniform(). *Stata Journal* 7: 434–435.

Campbell, A., P. E. Converse, W. E. Miller, and D. E. Stokes. 1960. *The American Voter*. New York: Wiley.

Efron, B., and R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. New York: Chapman & Hall/CRC.

Erikson, R., J. H. Goldthorpe, M. Jackson, M. Yaish, and D. R. Cox. 2005. On class differentials in educational attainment. *Proceedings of the National Academy of Science* 102: 9730–9733.

Even, W. E., and D. A. Macpherson. 1990. Plant size and the decline of unionism. *Economics Letters* 32: 393–398.

Fairlie, R. W. 2005. An extension of the Blinder–Oaxaca decomposition technique to logit and probit models. *Journal of Economic and Social Measurement* 30: 305–316.

Gomulka, J., and N. Stern. 1990. The employment of married women in the United Kingdom 1970–83. *Economica* 57: 171–199.

Hauser, R. M., and W. H. Sewell. 1957–1977. *Wisconsin Longitudinal Study (WLS) [graduates and siblings]: 1957–1977*. Madison, WI: University of Wisconsin–Madison. http://www.ssc.wisc.edu/~wls/documentation/.

Jackson, M., R. Erikson, J. H. Goldthorpe, and M. Yaish. 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.

Jann, B. 2006. fairlie: Stata module to generate nonlinear decomposition of binary outcome differentials. http://ideas.repec.org/c/boc/bocode/s456727.html.

Sinning, M., M. Hahn, and T. K. Bauer. 2008. The Blinder–Oaxaca decomposition for nonlinear regression models. *Stata Journal* 8: 480–492.

Yun, M.-S. 2004. Decomposing differences in the first moment. *Economics Letters* 82: 275–280.

About the author

Maarten L. Buis is affiliated with the Department of Sociology at the University Tübingen, where he studies the interaction between demographic processes and inequality of educational opportunity. He is also a frequent contributor to Statalist.