Estimating Primary and Secondary Effects on Entering College

Maarten L. Buis
Vrije Universiteit Amsterdam

1. The Problem

Primary effect higher class students → higher probability of attending college because they perform better at school.

Secondary effect higher class students → higher probability of attending college, even if equal performance at school.

Total effect The sum of these two effects.

The problem is that these effects can **not** be estimated using logistic regressions using only class (for the total effect) and class and performance (for the secondary effect).

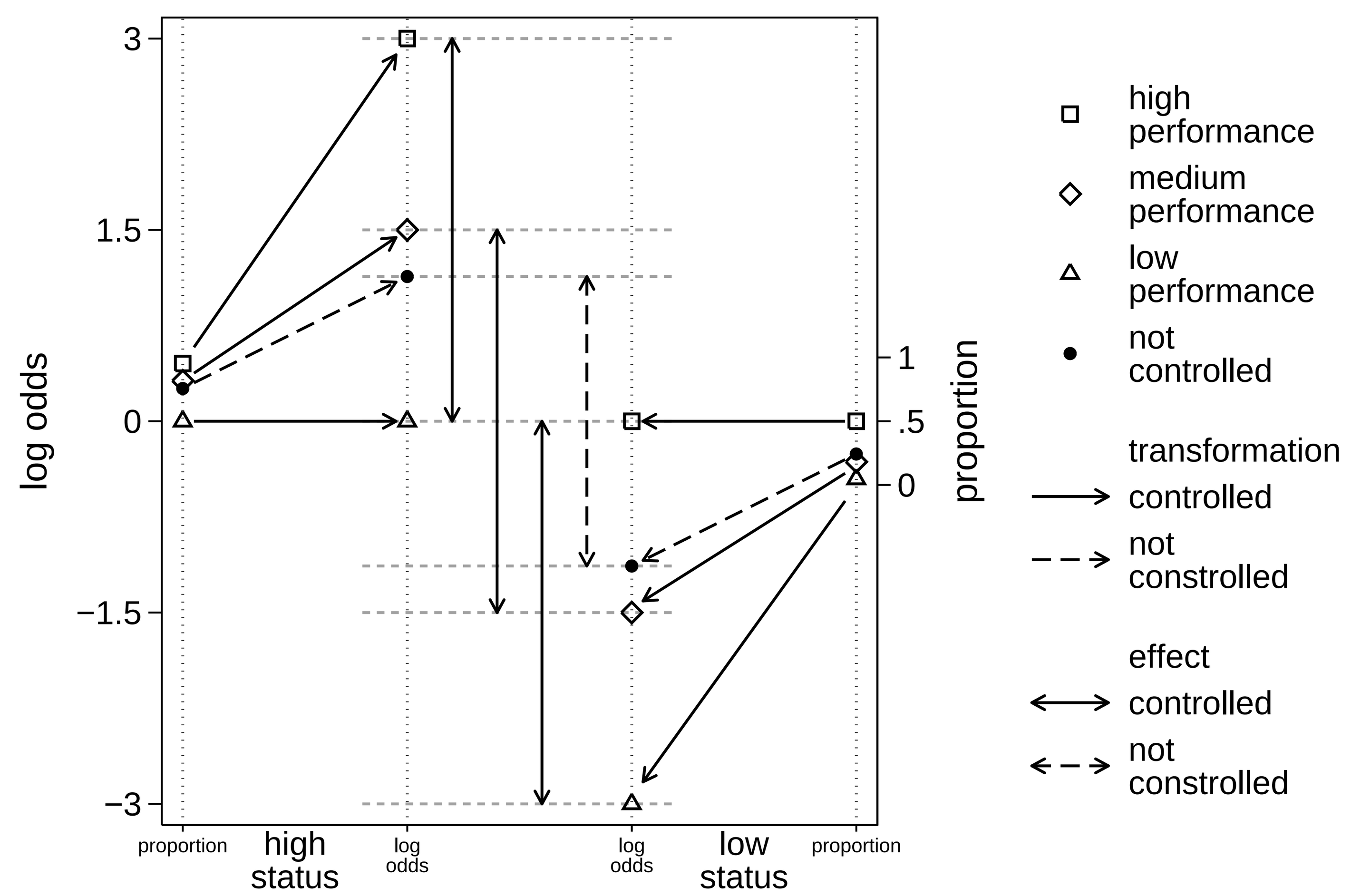
Example: there is no primary effect, so:
the secondary effect = the total effect.

A logistic regression with both class and performance (for the secondary effect) can be thought of as:

- compute proportions entering college for all combinations of class and performance, and transform these to log odds
- for each level of performance compare the log odds

A logistic regression with only class (for the total effect) can be though of in the same way, except that the proportions are first averaged within each class.

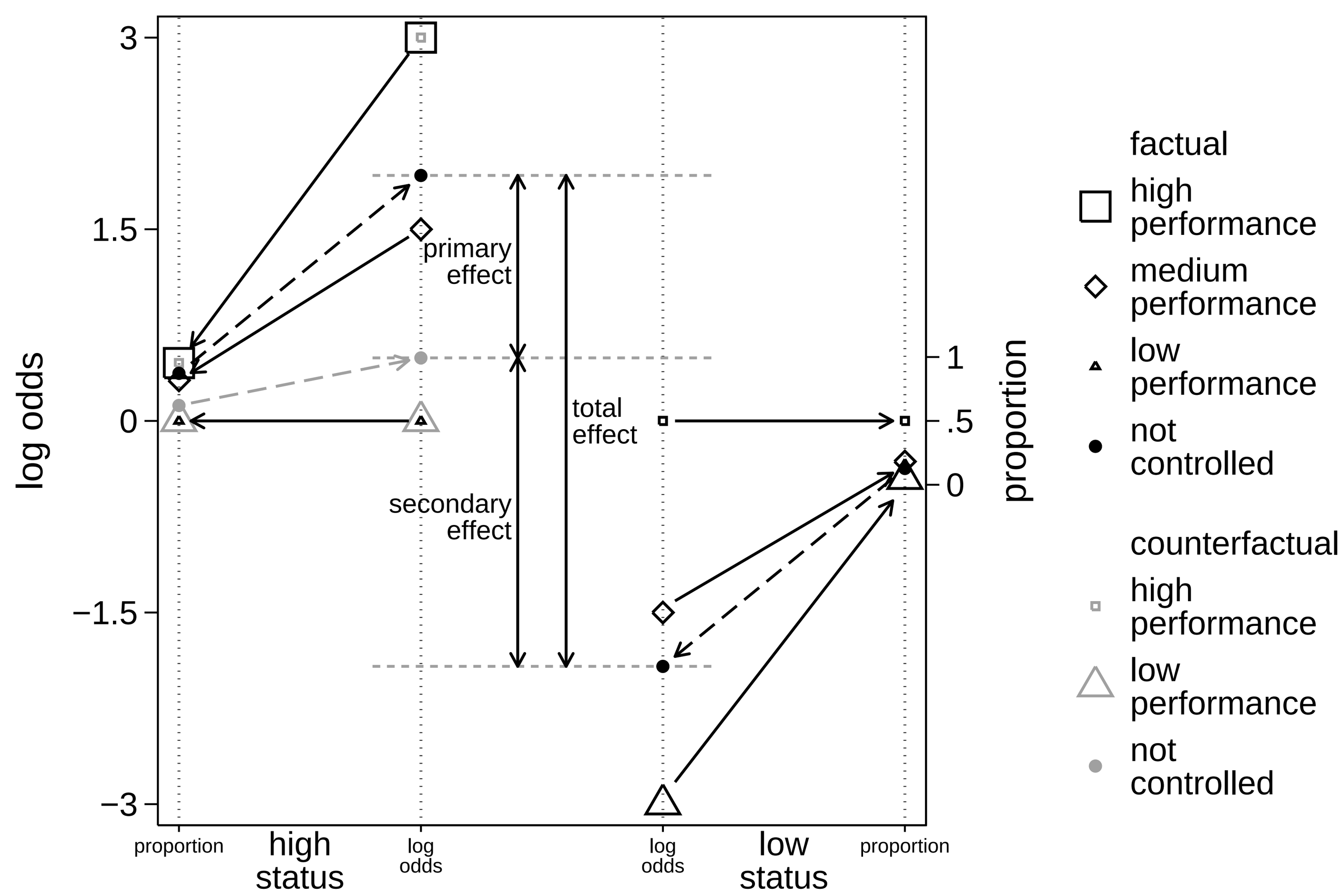
The more extreme proportions are less influential because they are more 'squashed', thus leading to a smaller effect of class, even though there is no primary effect.



2. A Solution

Erikson et al. (2005) propose the following solution:

- Estimate a logistic regression with both class and performance.
- Predict the log odds for each respondent and transform these to proportions (the sizes of the black symbols represents the number of respondents).
- Compute the average proportion within class, and transform back to log odds: the difference between classes is the total effect.
- Compute the average proportion for high class student, assuming they have the distribution of performance of the low class students (represented by the size of the grey symbols).
- The only difference between the high class and the counterfactual group is the distribution of performance, so this difference represents the primary effect.
- The distribution of performance remains constant when comparing the counterfactual group with the low class, so this difference represents the secondary effect.



An alternative decomposition is possible by computing the counterfactual proportion for low class students assuming they have the distribution of performance of the high class.

This alternative decomposition will lead to similar but not exactly the same results.

3. Two Extensions

1. Erikson et al. (2005) propose to compute the average proportions given the observed and counterfactual distribution of performance by assuming that performance is normally distributed, and then integrate over this normal distribution.

Alternatively these averages can be computed by predicting the observed and counterfactual proportions, add them up and divide by the number of respondents in that group.

The latter method has the advantage of making less assumptions about the distribution of performance, as it integrates over the empirical distribution of performance instead of over a normal distribution.

2. Standard errors of these estimates can be computed using the bootstrap.

The bootstrap is based on the following logic:

If we could draw many samples from the population and compute a statistic in each sample then the standard error is the standard deviation of these statistics.

The sample is an estimate of the population, so the standard error can be estimated by drawing many samples (with replacement) from the observed sample, compute the statistic in each new sample, and compute the standard deviation of the statistics.

Both the original method and the extensions proposed in this poster are implemented as the Stata package `ldecomp` (Buis 2008). To install type in Stata: `ssc install ldecomp`

References

M.L. Buis (2008), "Direct and Indirect Effects in a Logit Model", `http://home.fsw.vu.nl/m.buis/wp/ldecomp.html`.

R. Erikson, 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(27): 9730–3.