# Comparing Logit and Probit Coefficients between Models

Richard Williams (with assistance from Cheng Wang)
Notre Dame Sociology
rwilliam@ND.Edu
https://www3.nd.edu/~rwilliam
August 2012 Annual Meetings of the American Sociological Association

Last revised March 28, 2020

# Introduction

- We are used to estimating models where an observed, continuous independent variable, Y, is regressed on one or more independent variables, i.e.

$$Y = \alpha + \sum X\beta + \varepsilon,\ \ \varepsilon \sim N(0, \sigma^2)$$

- Since the residuals are uncorrelated with the Xs, it follows that

$$V(Y) = V(\alpha + \sum X\beta) + V(\varepsilon)$$

$$= \text{Explained Variance} + \text{Residual Variance}$$

- As you add explanatory variables to a model, the variance of the observed variable Y stays the same in OLS regression. As the explained variance goes up, the residual variance goes down by a corresponding amount.

- But suppose the observed Y is not continuous – instead, it is a collapsed version of an underlying unobserved variable, Y*

- Examples:
  - Do you approve or disapprove of the President's health care plan? 1 = Approve, 2 = Disapprove
  - Income, coded in categories like $0 = 1, $1- $10,000 = 2, $10,001-$30,000 = 3, $30,001-$60,000 = 4, $60,001 or higher = 5

- For such variables, also known as limited dependent variables, we know the interval that the underlying Y* falls in, but not its exact value

- Binary & Ordinal regression techniques allow us to estimate the effects of the Xs on the underlying Y*. They can also be used to see how the Xs affect the probability of being in one category of the observed Y as opposed to another.

- The latent variable model in binary logistic regression can be written as

$$y^* = \alpha + \sum X\beta + \varepsilon, \ \ \varepsilon \sim \text{Standard Logistic}$$

If y* >= 0, y = 1
If y* < 0, y = 0

In logistic regression, the errors are assumed to have a standard logistic distribution. A *standard logistic distribution* has a mean of 0 and a variance of $\pi^2/3$, or about 3.29.

- Since the residuals are uncorrelated with the Xs, it follows that

$$V(y^*) = V(\alpha + x\beta) + V(\varepsilon_{y^*}) = V(\alpha + x\beta) + \pi^2 / 3 = V(\alpha + x\beta) + 3.29$$

- Notice an important difference between OLS and Logistic Regression.
  - In OLS regression with an observed variable Y, V(Y) is fixed and the explained and unexplained variances change as variables are added to the model.
  - But in logistic regression with an unobserved variable y*, $V(\varepsilon_{y^*})$ is fixed so the explained variance and total variance change as you add variables to the model.
  - This difference has important implications. Comparisons of coefficients between nested models and across groups do not work the same way in logistic regression as they do in OLS.

# Comparing Logit and Probit Coefficients across Models

```
. use http://www.nd.edu/~rwilliam/xsoc73994/statafiles/standardized.dta
. logit  ybinary x1, nolog

Logit estimates                                   Number of obs   =        500
                                                  LR chi2(1)      =     161.77
                                                  Prob > chi2     =     0.0000
Log likelihood = -265.54468                       Pseudo R2       =     0.2335

------------------------------------------------------------------------------
     ybinary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .7388678    .072961    10.13   0.000     .5958668    .8818687
       _cons |  -.0529777    .105911    -0.50   0.617    -.2605593     .154604
------------------------------------------------------------------------------

. logit  ybinary x2, nolog

Logit estimates                                   Number of obs   =        500
                                                  LR chi2(1)      =     160.35
                                                  Prob > chi2     =     0.0000
Log likelihood = -266.25298                       Pseudo R2       =     0.2314

------------------------------------------------------------------------------
     ybinary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x2 |   .4886751   .0482208    10.13   0.000     .3941641    .5831861
       _cons |  -.0723833   .1058261    -0.68   0.494    -.2797986     .135032
------------------------------------------------------------------------------

. logit  ybinary x1 x2, nolog
```

```
. logit  ybinary x1 x2, nolog

Logit estimates                                   Number of obs   =        500
                                                  LR chi2(2)      =     443.39
                                                  Prob > chi2     =     0.0000
Log likelihood = -124.73508                       Pseudo R2       =     0.6399

------------------------------------------------------------------------------
     ybinary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |    1.78923   .1823005     9.81   0.000     1.431927    2.146532
          x2 |   1.173144   .1207712     9.71   0.000     .9364369    1.409851
       _cons |  -.2144856   .1626906    -1.32   0.187    -.5333532    .1043821
------------------------------------------------------------------------------
```

Usually, when we add variables to a model (at least in OLS regression), the effects of variables added earlier goes down. However, in this case, we see that the coefficients for x1 and x2 increase (seemingly) dramatically when both variables are in the model, i.e. in the separate bivariate regressions the effects of x1 and x2 are .7388678 and .4886751, but in the multivariate regressions the effects are 1.78923 and 1.173144, more than twice as large as before. This leads to two questions:

1. If we saw something similar in an OLS regression, what would we suspect was going on? In other words, in an OLS regression, what can cause coefficients to get bigger rather than smaller as more variables are added?
2. In a logistic regression, why might such an interpretation be totally wrong?

```
. corr, means

(obs=500)

    Variable |         Mean    Std. Dev.          Min          Max
-------------+--------------------------------------------------------
           y |     5.51e-07     3.000001    -8.508021     7.981196
     ybinary |         .488     .5003566            0            1
          x1 |    -2.19e-08            2     -6.32646     6.401608
          x2 |     3.57e-08            3    -10.56658     9.646875


             |        y  ybinary       x1       x2
-------------+------------------------------------
           y |   1.0000
     ybinary |   0.7923   1.0000
          x1 |   0.6667   0.5248   1.0000
          x2 |   0.6667   0.5225   0.0000   1.0000
```

- x1 and x2 are uncorrelated! So suppressor effects cannot account for the changes in coefficients.
- Long & Freese's listcoef command can add some insights.

```
. quietly logit ybinary x1
. listcoef, std

logit (N=500): Unstandardized and Standardized Estimates

 Observed SD: .50035659
   Latent SD: 2.3395663

  Odds of: 1 vs 0

------------------------------------------------------------------------------
     ybinary |      b         z     P>|z|    bStdX    bStdY   bStdXY      SDofX
-------------+----------------------------------------------------------------
          x1 |   0.73887   10.127   0.000   1.4777   0.3158   0.6316     2.0000
------------------------------------------------------------------------------

. quietly logit  ybinary x2
. listcoef, std

logit (N=500): Unstandardized and Standardized Estimates

 Observed SD: .50035659
   Latent SD: 2.3321875

  Odds of: 1 vs 0

------------------------------------------------------------------------------
     ybinary |      b         z     P>|z|    bStdX    bStdY   bStdXY      SDofX
-------------+----------------------------------------------------------------
          x2 |   0.48868   10.134   0.000   1.4660   0.2095   0.6286     3.0000
------------------------------------------------------------------------------
```

```
. quietly logit  ybinary x1 x2
. listcoef, std

logit (N=500): Unstandardized and Standardized Estimates

 Observed SD: .50035659
   Latent SD: 5.3368197

  Odds of: 1 vs 0

------------------------------------------------------------------------------
     ybinary |      b         z     P>|z|    bStdX    bStdY   bStdXY      SDofX
-------------+----------------------------------------------------------------
          x1 |   1.78923     9.815   0.000   3.5785   0.3353   0.6705     2.0000
          x2 |   1.17314     9.714   0.000   3.5194   0.2198   0.6595     3.0000
------------------------------------------------------------------------------
```

- Note how the standard deviation of y* fluctuates from one logistic regression to the next; it is about 2.34 in each of the bivariate logistic regressions and 5.34 in the multivariate logistic regression.
- It is because the variance of y* changes that the coefficients change so much when you go from one model to the next. In effect, the scaling of Y* is different in each model. By way of analogy, if in one OLS regression income was measured in dollars, and in another it was measured in thousands of dollars, the coefficients would be very different.

- Why does the variance of y* go up? Because it has to. The residual variance is fixed at 3.29, so improvements in model fit result in increases in explained variance which in turn result in increases in total variance.
- Hence, comparisons of coefficients across nested models can be misleading because the dependent variable is scaled differently in each model.

- How serious is the problem in practice?
  - Hard to say. We easily found dozens of recent papers that present sequences of nested models. Their numbers are at least a little off, but without re-analyzing the data you can't tell whether their conclusions are seriously distorted as a result.
  - Several attempts of our own using real world data have failed to raise major concerns with the comparisons
  - We asked several authors for copies of their data, but most were unwilling or unable to do so.

- One author, Ervin (Maliq) Matthew, did graciously provide us with the data used for his paper “Effort Optimism in the Classroom: Attitudes of Black and White Students on Education, Social Structure, and Causes of Life Opportunities” (Sociology of Education 2011 84:225-245)
- The paper contains potentially problematic statements such as “The effect of race on the dependent variable is even stronger once GPA, SES, and sex are controlled for (Model 2), indicating that when blacks and whites have equal GPAs and family SES, blacks are more likely to agree with this statement.”
- In practice, however, we found that any potential errors were modest, with estimates being only slightly affected by solutions we discuss later. For example, his Table 7 modestly understates how much the effect of race declines as controls are added.

- Nonetheless, researchers should realize that
  - Increases in the magnitudes of coefficients across models need not reflect suppressor effects
  - Declines in coefficients across models will actually be understated, i.e. you will be understating how much other variables account for the estimated direct effects of the variables in the early models.
  - Distortions are potentially more severe when added variables greatly increase the pseudo R^2 statistics, as the variance of Y* will increase more when that is the case.

- What are possible solutions?
  - Just don't present the coefficients for each model in the first place. Researchers often present chi-square contrasts to show how they picked their final model and then only present the coefficients for it.
  - Use y-standardization. With y-standardization, instead of fixing the residual variance, you fix the variance of y* at 1. This does not work perfectly, but it does greatly reduce rescaling of coefficients between models.
    - Listcoef gives the y-standardized coefficients in the column labeled bStdy, and they hardly changed at all between the bivariate and multivariate models (.3158 and .2095 in the bivariate models, .3353 and .2198 in the multivariate model).

- Report average marginal effects of variables. In our original example,

```
. use http://www3.nd.edu/~rwilliam/statafiles/standardized.dta, clear
. qui logit  ybinary x1, nolog
. qui margins, dydx(*) post
. est store m1
. qui logit  ybinary x2, nolog
. qui margins, dydx(*) post
. est store m2
. qui logit  ybinary x1 x2, nolog
. qui margins, dydx(*) post
. est store m3
. esttab m1 m2 m3, z
```

| | (1) | (2) | (3) |
|---|---|---|---|
| x1 | 0.132*** | | 0.139*** |
| | (18.74) | | (31.75) |
| x2 | | 0.0874*** | 0.0909*** |
| | | (18.77) | (27.49) |
| N | 500 | 500 | 500 |

z statistics in parentheses

- The Karlson/Holm/Breen (KHB) method (Papers are available in Sociological Methodology and Stata Journal) shows promise
  - According to KHB, their method separates changes in coefficients due to rescaling from true changes in coefficients that result from adding more variables to the model (and does a better job of doing so than y-standardization and other alternatives)
  - They further claim that with their method the total effect of a variable can be decomposed into its direct effect and its indirect effect.

- We would add that, when authors estimate sequences of models, it is often because they want to see how the effects of variables like race decline (or increase) after other variables are controlled for. The KHB method provides a parsimonious and more accurate way of depicting such changes.
- We'll first present a simple example showing the relationship between diabetes, race & weight.

# khb example 1

```
. webuse nhanes2f, clear
. khb logit diabetes black || weight

Decomposition using the KHB-Method

Model-Type:  logit                                Number of obs     =   10335
Variables of Interest: black                      Pseudo R2         =    0.02
Z-variable(s): weight
------------------------------------------------------------------------------
    diabetes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
black        |
     Reduced |   .6038012   .1236714     4.88   0.000     .3614098    .8461926
        Full |   .5387425   .1241889     4.34   0.000     .2953368    .7821483
        Diff |   .0650587   .0132239     4.92   0.000     .0391403    .0909771
------------------------------------------------------------------------------
```

- Possible interpretation of results
  - In the line labeled Reduced, only black is in the model. .6038 is the total effect of black.
  - However, blacks may have higher rates of diabetes both because of a direct effect of race on diabetes, and because of an indirect effect: blacks tend to be heavier than whites, and heavier people have higher rates of diabetes.
  - Hence, the line labeled Full gives the direct effect of race (.5387) while the line labeled  Diff gives the indirect effect (.065)

# Khb Example 2

- Matthew (2011; see Table 7, p. 240) examines the determinants of how likely a student is to feel they will have a job he or she enjoys (0 = 50 percent or lower; 1 = better than 50 percent).
- In the first model, race (0 = white, 1 = black) is the only independent variable. The estimated effect of race is -.510.
- In the final model controls are added for GPA, SES, and others. The effect of race declines to -.471, an apparent -.039 drop.
- The khb method shows that the decline is actually about twice as large. Again this is at least partly because the variance of y* becomes greater as more variables are added, causing coefficients to increase.

```
. khb logit jobenjoy race || gpa ses sex educjob educimportant luckimportant sbprevent

Decomposition using the KHB-Method

Model-Type:  logit                                  Number of obs     =      6731
Variables of Interest: race                         Pseudo R2         =      0.08
Z-variable(s): gpa ses sex educjob educimportant luckimportant sbprevent
```

| jobenjoy | Coef. | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| race | | | | | | |
| Reduced | -.5727334 | .10607 | -5.40 | 0.000 | -.7806269 | -.3648399 |
| Full | -.4833004 | .1095584 | -4.41 | 0.000 | -.6980309 | -.26857 |
| Diff | -.089433 | .0349898 | -2.56 | 0.011 | -.1580117 | -.0208542 |

# Selected References


- Karlson, Kristian B., Anders Holm and Richard Breen. 2011. Comparing Regression Coefficients between Same-Sample Nested Models using Logit and Probit: A New Method. doi: 10.1177/0081175012444861. Sociological Methodology August 2012 vol. 42 no. 1 286-313

- Abstract: "Logit and probit models are widely used in empirical sociological research. However, the common practice of comparing the coefficients of a given variable across differently specified models fitted to the same sample does not warrant the same interpretation in logits and probits as in linear regression. Unlike linear models, the change in the coefficient of the variable of interest cannot be straightforwardly attributed to the inclusion of confounding variables. The reason for this is that the variance of the underlying latent variable is not identified and will differ between models. We refer to this as the problem of rescaling. We propose a solution that allows researchers to assess the influence of confounding relative to the influence of rescaling, and we develop a test to assess the statistical significance of confounding. A further problem in making comparisons is that, in most cases, the error distribution, and not just its variance, will differ across models. Monte Carlo analyses indicate that other methods that have been proposed for dealing with the rescaling problem can lead to mistaken inferences if the error distributions are very different. In contrast, in all scenarios studied, our approach performs as least as well as, and in some cases better than, others when faced with differences in the error distributions. We present an example of our method using data from the National Education Longitudinal Study"

- Kohler, Ulrich, Kristian B. Carlson and Anders Holm. 2011. Comparing Coefficients of nested nonlinear probability models. The Stata Journal Volume 11 Number 3: pp. 420-438. http://www.stata-journal.com/article.html?article=st0236.

- Abstract: "In a series of recent articles, Karlson, Holm, and Breen (Breen, Karlson, and Holm, 2011, http://papers.ssrn.com/sol3/papers.cfm?abstractid=1730065; Karlson and Holm, 2011, *Research in Stratification and Social Mobility* 29: 221–237; Karlson, Holm, and Breen, 2010, http://www.yale.edu/ciqle/Breen Scaling%20effects.pdf) have developed a method for comparing the estimated coefficients of two nested nonlinear probability models. In this article, we describe this method and the user-written program **khb**, which implements the method. The KHB method is a general decomposition method that is unaffected by the rescaling or attenuation bias that arises in cross-model comparisons in nonlinear models. It recovers the degree to which a control variable, *Z*, mediates or explains the relationship between *X* and a latent outcome variable, *Y**, underlying the nonlinear probability model. It also decomposes effects of both discrete and continuous variables, applies to average partial effects, and provides analytically derived statistical tests. The method can be extended to other models in the generalized linear model family.

- Long, J. Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata*, 3rd Edition. College Station, Texas: Stata Press.