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.

Put another way – As you add variables to an OLS regression, the Total Sum of Squares stays the same, but the allocation between the Model (Explained) and Residual (Unexplained) Sums of Squares shifts, i.e. adding more variables increases the Explained SS and decreases the unexplained SS by a corresponding amount.

```
. nestreg: reg health black (age sex height weight)

Block 1: black

      Source |       SS           df       MS      Number of obs   =    10,335
-------------+----------------------------------   F(1, 10333)     =    173.65
       Model |  248.486541         1  248.486541   Prob > F        =    0.0000
    Residual |  14786.5348    10,333  1.43100115   R-squared       =    0.0165
-------------+----------------------------------   Adj R-squared   =    0.0164
       Total |  15035.0214    10,334   1.4549082   Root MSE        =    1.1962

[Output deleted]

Block 2: age sex height weight

      Source |       SS           df       MS      Number of obs   =    10,335
-------------+----------------------------------   F(5, 10329)     =    410.86
       Model |  2494.19977         5  498.839954   Prob > F        =    0.0000
    Residual |  12540.8216    10,329  1.21413705   R-squared       =    0.1659
-------------+----------------------------------   Adj R-squared   =    0.1655
       Total |  15035.0214    10,334   1.4549082   Root MSE        =    1.1019
```

Recall too that MS Total is the variance of y. It stays the same regardless of what variables are added or dropped from the model. In this case it equals 1.4549. In other words, v(y) is a fixed quantity and does not depend on the variables in the model.

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

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(εy*) 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 https://www3.nd.edu/~rwilliam/statafiles/standardized.dta, clear

. logit ybinary x1, nolog

Logistic regression                             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 |   .7388677   .0729611    10.13   0.000     .5958667    .8818688
       _cons |  -.0529777    .105911    -0.50   0.617    -.2605594     .154604
------------------------------------------------------------------------------

. logit ybinary x2, nolog

Logistic regression                             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      .394164    .5831861
       _cons |  -.0723833   .1058261    -0.68   0.494    -.2797987    .1350321
------------------------------------------------------------------------------

. 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. For example, his Table 7 somewhat understates how much the effect of race declines as controls are added. (We semi-replicate his work later in this handout.)

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).

The Karlson/Holm/Breen (KHB) method (Papers are available in Sociological Methodology and The Stata Journal) shows great 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.

For example, a researcher might want to know how much of the effect of race is direct and how much is indirect (e.g. race affects education and education in turn affects the dependent variable.)

If some of the effect of race is indirect, then the coefficient for race should decline as more variables are added to the model.

The KHB method provides a parsimonious and more accurate way of depicting such changes.

We'll now present a few examples using khb, starting with the hypothetical example we had earlier.

khb Example 1: Hypothetical Data

```
. khb logit ybinary x1 || x2

Decomposition using the KHB-Method

Model-Type:  logit                                Number of obs     =       500
Variables of Interest: x1                         Pseudo R2         =      0.64
Z-variable(s): x2
------------------------------------------------------------------------------
     ybinary | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
x1           |
     Reduced |    1.78923   .1823053     9.81   0.000     1.431918    2.146541
        Full |    1.78923   .1823053     9.81   0.000     1.431918    2.146541
        Diff |   1.05e-08   .0011743     0.00   1.000    -.0023016    .0023016
------------------------------------------------------------------------------
```

khb shows that, in reality, the effect of x1 doesn't change at all. The change shown earlier was entirely due to the rescaling of Y*.

Khb Example 2: Matthew Replication

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).

For unclear reasons, our replication results differ slightly from those presented in the paper.

In the first model (see next slide), race (0 = white, 1 = black) is the only independent variable. The estimated effect of race is -.507.

In the final model controls are added for GPA, SES, and others. The effect of race declines to -.483, an apparent -.024 drop.

```
. use https://www3.nd.edu/~rwilliam/statafiles/soe2011, clear
. nestreg: logit jobenjoy i.race (gpa ses sex educjob educimportant luckimportant sbprevent

Logistic regression                                     Number of obs =  6,731
                                                        LR chi2(1)    =  22.74
                                                        Prob > chi2   = 0.0000
Log likelihood = -2740.9172                             Pseudo R2     = 0.0041

------------------------------------------------------------------------------
    jobenjoy | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        race |
       Black |  -.5071732   .1023392    -4.96   0.000    -.7077544    -.306592
       _cons |   1.856833   .0375953    49.39   0.000     1.783147    1.930518
------------------------------------------------------------------------------

Logistic regression                                     Number of obs =  6,731
                                                        LR chi2(8)    = 423.84
                                                        Prob > chi2   = 0.0000
Log likelihood = -2540.3718                             Pseudo R2     = 0.0770

------------------------------------------------------------------------------
     jobenjoy | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
--------------+---------------------------------------------------------------
         race |
        Black |  -.4833004   .1095584    -4.41   0.000    -.6980309     -.26857
          gpa |   .2896685   .0548232     5.28   0.000     .1822171      .39712
          ses |   .0984206   .0527711     1.87   0.062    -.0050088      .20185
          sex |   .1113716    .073942     1.51   0.132    -.0335521     .2562953
      educjob |   .2508703   .1794196     1.40   0.162    -.1007857     .6025263
educimportant |   .9474587   .0860938    11.00   0.000      .778718     1.116199
luckimportant |  -.2183139   .1183016    -1.85   0.065    -.4501808     .0135529
    sbprevent |  -.8611827    .079129   -10.88   0.000    -1.016273    -.7060928
        _cons |   .1883665    .233478     0.81   0.420     -.269242      .645975
------------------------------------------------------------------------------
```

The khb method (shown below) shows that the decline is actually about four times as great, -.089. Again this is at least partly because the variance of y* becomes greater as more variables are added, causing coefficients to increase.

Put another way, the effect of race in model 1, −.507, is adjusted upwards to -.5772, to reflect the increased variance of Y* as more variables are added.

Putting it yet another way, the indirect effect of race is underestimated if we don't make the khb correction. It appears that the indirect effect of race is only .024 when it is really .089. Without the KHB correction, you would underestimate the importance of the indirect effects race has by influencing other variables which in turn affect whether or not a person enjoys their job.

```
. 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 | Coefficient  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
------------------------------------------------------------------------------
```

Marginal Effects

Both Mize, Doan, & Long (2019) and Karlson, Holm, & Breen (2012) argue that it may be better to look at changes in marginal effects across nested models, rather than changes in coefficients.

KHB note that marginal effects have more intuitive appeal than do coefficients.

MDL further note that rescaling is not an issue with marginal effects

The KHB and MDL methods differ though

The Mize/Doan/Long Approach:

Report average marginal effects of variables

In Example 1,

```
. 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 marginal effects changed far less than the coefficients did (although it may seem odd that they changed at all, given that x1 and x2 are uncorrelated).

Mize, Doan, & Long (2019) demonstrate how to do formal tests of whether marginal effects significantly differ across nested models. (In this case, they don't.) See example 6.2 in https://journals.sagepub.com/doi/full/10.1177/0081175019852763

The coding is a little complicated (Mize is working on do files to simplify it) but the code and data used in Mize's paper is (as of February 26, 2022) at https://drive.google.com/drive/folders/18RS5C47b_ddGaRRXILfmxmfOB41AyA7D

The MDL approach includes output like the following. The main thing it adds to the margins commands just shown is a formal test of whether the change in marginal effects is statistically significant.

Average Discrete Changes for x1 and cross-model differences

	lincom	se	pvalue
ADC x1			
Model 1	**0.132**	0.007	0.000
Model 2	**0.139**	0.004	0.000
Diff in ADCs			
M1 - M2	**-0.007**	**0.005**	**0.168**

KHB provides what may be a better way to compare marginal effects across nested models. They argue that estimating marginal effects model by model (like MDL do and we just did in Example 1) will give at least slightly erroneous results. They show how to correct for this.

Using their ape (Average Partial Effect) option, even the small differences in marginal effects we saw in Example 1 across nested models go away.

```
. khb logit ybinary x1 || x2, ape

Decomposition using the APE-Method

Model-Type:  logit                                Number of obs     =     500
Variables of Interest: x1                         Pseudo R2         =    0.64
Z-variable(s): x2
------------------------------------------------------------------------------
     ybinary | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
x1           |
     Reduced |   .1386605   .0043668    31.75   0.000     .1301018    .1472192
        Full |   .1386605   .0043668    31.75   0.000     .1301018    .1472192
        Diff |   3.42e-10          .        .       .            .           .
------------------------------------------------------------------------------
 Note: Standard errors of difference not known for APE method
```

Applying the MDL method to Example 2,

Average Discrete Changes for i.race and cross-model differences

		lincom	se	pvalue
ADC i.race				
	Model 1	**-0.071**	0.016	0.000
	Model 2	**-0.061**	0.016	0.000
Diff in ADCs				
	M1 - M2	**-0.010**	**0.006**	**0.103**

Applying khb with the ape option to Example 2,

```
. khb logit jobenjoy i.race || gpa ses sex educjob educimportant luckimportant s
> bprevent, ape

Decomposition using the APE-Method

Model-Type:  logit                                Number of obs     =    6731
Variables of Interest: i.race                     Pseudo R2         =    0.08
Z-variable(s): gpa ses sex educjob educimportant luckimportant sbprevent
------------------------------------------------------------------------------
    jobenjoy | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
0.race       |  (base outcome)
-------------+----------------------------------------------------------------
1.race       |
     Reduced |  -.0742168   .0154848    -4.79   0.000    -.1045665    -.043867
        Full |  -.0613298   .0154254    -3.98   0.000    -.0915631   -.0310966
        Diff |  -.0128869          .        .       .            .           .
------------------------------------------------------------------------------
 Note: Standard errors of difference not known for APE method
```

For Example 2 – Estimating marginal effects the MDL way, the marginal effect of race declines by about .010 between nested models. The change is NOT statistically significant at even the .10 level.

However, using khb, the change is slightly larger, about .013. Further, khb does NOT provide a test of the statistical significance of the change.

It may not matter that much whether or not you just estimate the marginal effects model by model like MDL do or if you use khb, as the differences seem to be minor in practice.

Summary

When you estimate a series of nested models using logit or probit, comparisons of coefficients across models may be problematic, because Y* is scaled differently in each model.

You may just want to not even present the results from nested models. Often people do so but ignore everything but the final model. So why waste space on something you aren't using and which could mislead people?

If you do want to present sequences of nested models and see how coefficients change (e.g. you want to see how the effect of race declines as more variables are added to the model) you probably want to use the khb method so results across models are directly comparable.

Rather than focusing on how coefficients change, you may prefer to focus on how marginal effects change. This may be more intuitively meaningful.

Mize, Doan, & Long (2019), as well as Karlson, Holm, and Breen (2012), have suggested ways to validly compare marginal effects across nested models.

I'm currently not sure which is best. According to KHB the MDL approach seems slightly off. On the other hand MDL provides a formal statistical test of the differences between marginal effects but KHB does not. The KHB software is currently much easier to use but easier-to-use software for MDL may be coming. In any event, it may not matter that much as differences between the two approaches seem to be small.

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. https://journals.sagepub.com/doi/10.1177/0081175012444861. Sociological Methodology August 2012 vol. 42 no. 1 286-313

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.

Mize, Trenton, Long Doan, & J. Scott Long. 2019. A General Framework for Comparing Predictions and Marginal Effects across Models. Sociological Methodology vol 49 no 1: pp. 152-189. https://journals.sagepub.com/doi/full/10.1177/0081175019852763

Matthew, Ervin (Maliq). 2011. Effort Optimism in the Classroom: Attitudes of Black and White Students on Education, Social Structure, and Causes of Life Opportunities. Sociology of Education Volume: 84 issue: 3, page(s): 225-245. https://journals.sagepub.com/doi/10.1177/0038040711402360