Prelude to Comparing Coefficients Between Nested Models

Richard Williams, University of Notre Dame, https://www3.nd.edu/~rwilliam/
Last revised October 10, 2024

Following are results of two bivariate and one multivariate logistic regressions using a data set I constructed:

```
. use https://www3.nd.edu/~rwilliam/statafiles/standardized.dta, clear
. 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 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 go 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? ☺