NBER WORKING PAPER SERIES

LOG ODDS AND ENDS

Edward C. Norton

Working Paper 18252
http://www.nber.org/papers/w18252

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2012

The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2012 by Edward C. Norton. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Log Odds and Ends
Edward C. Norton
NBER Working Paper No. 18252
July 2012
JEL No. C25,I19

ABSTRACT

Although independent unobserved heterogeneity—variables that affect the dependent variable but are independent from the other explanatory variables of interest—do not affect the point estimates or marginal effects in least squares regression, they do affect point estimates in nonlinear models such as logit and probit models. In these nonlinear models, independent unobserved heterogeneity changes the arbitrary normalization of the coefficients through the error variance. Therefore, any statistics derived from the estimated coefficients change when additional, seemingly irrelevant, variables are added to the model. Odds ratios must be interpreted as conditional on the data and model. There is no one odds ratio; each odds ratio estimated in a multivariate model is conditional on the data and model in a way that makes comparisons with other results difficult or impossible. This paper provides new Monte Carlo and graphical insights into why this is true, and new understanding of how to interpret fixed effects models, including case control studies. Marginal effects are largely unaffected by unobserved heterogeneity in both linear regression and nonlinear models, including logit and probit and their multinomial and ordered extensions.

Edward C. Norton
Department of Health Management and Policy
Department of Economics
University of Michigan
School of Public Health
1415 Washington Heights, M3108 SPHII
Ann Arbor, MI 48109-2029
and NBER
ecnorton@umich.edu

1. INTRODUCTION

Researchers often struggle with how to present the results from a model with a dichotomous dependent variable in a meaningful way. The most common ways to report results tend to fall along disciplinary lines. Epidemiologists, clinical researchers, and health services researchers often report odds ratios after estimating a logit model. Economists tend to report marginal effects, and are split on estimating logit, probit, and linear probability models. There is an increasing recognition that independent unobserved (or neglected) heterogeneity—variables that affect the dependent variable but are independent from the other explanatory variables of interest—affect the interpretation of the results (Mroz and Zayats 2008; Mood 2010). This is in contrast to linear regression models, where independent unobserved heterogeneity does not affect the magnitude of the coefficients or the corresponding marginal effects.

The fundamental issue is that including or excluding independent unobserved heterogeneity changes the information that the results are conditioned on. While conditioning on independent unobserved heterogeneity does not substantively affect the results in linear models, conditioning matters greatly in nonlinear models. In nonlinear models, conditioning on the mean of a variable leads to a different prediction than averaging predictions over all possible values. The difference leads to predictable differences in the magnitude of the reported results due to normalization of the error variance—and should lead to a different interpretation of statistics, such as the odds ratio, that are based only on the estimated coefficient. Unfortunately, the importance of conditioning to the interpretation is often neglected. The odds ratio must be reported conditional on the model specification because other model specifications will lead to different, often dramatically different, estimated odds ratios. The odds ratio is not invariant to the model specification involving variables that are independent of the variables of interest. In

contrast, marginal effects are largely unaffected by independent unobserved heterogeneity in nonlinear models.

Omitted variable bias in linear regression is a well-known problem (Griliches 1957). (Omitted variables that are correlated with the variables of interest cause omitted variable bias due to endogeneity; that challenging topic is not the topic of this paper.) Later statistical work focused on what conditions were needed for independent unobserved heterogeneity to matter for logit and probit models and other nonlinear models (Lee 1982; Gail, Wieand, Piantadosi 1984; Yatchew and Griliches 1985). While the odds ratio interpretation of coefficients from logit models increasingly became the norm, some researchers raised concerns about the understandability of odds ratios (e.g., Greenland 1987; Sackett, Deeks, Altman 1996; Altman, Deeks, Sackett 1998; Schwartz et al. 1999; Walter 2000; Kleinman and Norton 2009; Tajeu et al. 2012). More recently, Allison (1999) explained why odds ratios cannot be compared across samples. Mood (2010) extended this work nicely to show how odds ratios cannot be interpreted as substantive effects, nor can they be compared across models or across groups within models. Mroz and Zayats (2008) also pointed out the importance of unobserved heterogeneity in several kinds of models.

This paper builds on the prior literature and adds to it in several important ways. I provide a comprehensive review of independent unobserved heterogeneity in nonlinear models and how independent unobserved heterogeneity affects interpretation in meaningful ways. I start with linear regression to show the intuition for why independent unobserved heterogeneity in linear models does not affect the interpretation of the results. I review what is known from econometric theory about independent unobserved heterogeneity in nonlinear models. I show simple Monte Carlo results from probit and logit models to compare and contrast with the linear

regression results. I provide intuition from math, graphs, and simple stories, to explain why unobserved heterogeneity matters more in nonlinear models. The graphs are new and provide novel insights about the interpretation. I discuss results from other Monte Carlo simulations that do not adhere to the assumptions required for nice results, but reflect patterns more commonly found in real data, and discuss under what circumstances the independent unobserved heterogeneity seems to matter most. This paper also has a number of new results. I provide new insights from how unobserved heterogeneity matters in the interpretation of conditional fixed effects models. The discussion in not restricted to the logit model, but I explain how independent unobserved heterogeneity affects probit, linear probability, multinomial, and ordered models. Because risk ratios are an alternative to odds ratios, I present some evidence on how they are affected by unobserved heterogeneity. Most importantly, I argue that the long-standing argument in the literature against odds ratios (because they are hard to understand) is actually much more powerful. They are much harder to understand than previously thought—it is not only that people generally think risk ratio when they see "odds ratio" but more importantly that the odds ratio has an extremely narrow, precise, and non-generalizable meaning because it is conditioned on the data and the model. Correctly explaining odds ratios is now even harder. Finally, I have recommendations for best practice.

Many prior papers have criticized the use of odds ratios (or log odds) in multivariate analysis (Greenland S 1987; Sackett DL, Deeks JJ, Altman DG 1996; Altman DG, Deeks JJ, Sackett DL 1998; Schwartz et al. 1999; Kleinman LC, Norton EC 2009; Tajeu et al. 2012). While odds ratios have nice properties—they are simple to calculate and invariant to which risk is in the numerator and which is in the denominator—many researchers have complained that odds ratios are hard to understand. The specific complaint has been that odds ratios are often

misinterpreted as risk ratios, yet mathematically they diverge significantly from risk ratios when the baseline risk exceeds around 10 percent. This paper explains another reason why odds ratios are hard to understand. No odds ratio calculated from a multivariable model applies generally to all populations. Odds ratios are conditional on the model specification and conditioning is hard to explain. A different model will lead to a different calculated odds ratio, even when adding or excluding seemingly irrelevant variables or controlling for fixed effects. Odds ratios should only be reported when the information that the results are conditioned on is clear.

2. REVIEW OF OLS

In least squares regression, adding covariates that are independent of all the covariates already in the model has no effect on the magnitude of the coefficients on those covariates of interest (Griliches 1957). Therefore, marginal and incremental effects are also unaffected.

However, adding independent covariates reduces the standard errors on the coefficients of the covariates already in the model and thereby raises the *t*-statistics and lowers the *p*-values. The reason is obvious from the formula for the variance-covariance matrix of the estimated coefficients: $V(\hat{\beta}) = \sigma^2(X'X)^{-1}$. The standard error for an estimated coefficient is proportional to the variance of the model error σ^2, so when additional important covariates are added, the error variance shrinks. Additional independent covariates leave $(X'X)^{-1}$ unaffected on the main diagonal for the coefficients of interest.

I demonstrate these properties in two ways using Monte Carlo data. Both examples use linear regression; the first has a continuous dependent variable and the second has a dichotomous dependent variable. Although these examples are trivial, they serve as good comparisons for the more complicated nonlinear models that are the focus of this paper.

The continuous dependent variable *y* is a linear function of a dummy variable x_d and four continuous variables x_1 through x_4. For these illustrative examples, the variables of interest are the dummy variable (x_d) and the first two continuous covariates (x_1 and x_2). In equation (1), the third and fourth continuous covariates are in parentheses to indicate that they are the unobserved heterogeneity and can sometimes be part of the error term.

$$y = 0.5x_d + x_1 + 2x_2 + (x_3 + 3x_4) + \varepsilon \qquad (1)$$

The mean of the continuous dependent variable is about 0.25 (see Table I). The dummy covariate has a mean of one-half and the continuous covariates all have zero mean.

To make these examples as clear as possible, it is important to have covariates that are perfectly independent, not just having correlation close to zero. After creating normally distributed variables (*N* = 10,000) and a dummy independent variable, I ensure independence by projecting each covariate on the others and replacing the variable with the residual. The correlation matrix shows that the covariates are all independent of each other (see Table II).

The main parameters of interest are the first three coefficients in the model (0.5, 1.0, and 2.0). In a least squares regression model the estimated coefficients on the three variables of interest are exactly 0.5, 1.0, and 2.0—the joys of the perfectly controlled world of Monte Carlo data (see Table III). What happens to the coefficients of interest in equation (1) when other independent variables x_3 and x_4 are added? The coefficients remain exactly the same. The first column shows the regression of *y* on just the dummy variable and the first two continuous variables. In the second column, the other two continuous variables are added. The coefficients on the three variables of interest remain unchanged. The standard errors for these coefficients predictably decline by a ratio of about 4.53, which is the ratio of root mean squared errors in these two models. The *t*-statistics increase and the *p*-values decrease as expected.

The corresponding linear probability model, with a dichotomous dependent variable, is instructive for later comparisons to probit and logit models. The dichotomous dependent variable y_d is constructed from the continuous dependent variable. If *y* is a latent continuous variable, then y_d is the corresponding dichotomous variable, equal to one if *y* is positive.

$$y_d = 1(y > 0) \equiv 1(0.5x_d + x_1 + 2x_2 + (x_3 + 3x_4) + \varepsilon > 0) \quad (2)$$

The mean of the dichotomous dependent variable is just over one-half.

The results for the linear probability model, with the dependent variable y_d, are qualitatively similar in terms of adding in the unobserved heterogeneity. Adding independent covariates does not affect the point estimates of the coefficients of interest (see the third and fourth columns of Table III). The coefficients are, of course, different than in the linear regression with *y* as the dependent variable. The coefficients of interest are now about 0.04, 0.08, and 0.16 (same proportions as before). These are the numbers that will be most comparable to the marginal effects calculated later from logit and probit models on the same data. The decline in the standard errors from the simpler model to the model including the unobserved heterogeneity is proportional to the decline in the root mean square error, as before. There are no surprises in Table III.

The results from these illustrative examples for OLS apply to other situations. As long as the unobserved heterogeneity is independent of the other covariates, then the inclusion or exclusion of these variables has no effect on the point estimates of the covariates of interest. Marginal and incremental effects are unchanged. The statistical significance, however, does depend on whether the unobserved heterogeneity is included in the model. These results do not depend on the distribution of the dependent variable—in addition to normal and dichotomous, the dependent variable can be uniform, skewed, log transformed, count, or any other distribution.

The nuisance variables can be correlated with each other, just not with the variables of interest. In the interest of space, I have only included the simple examples in order to focus on less well known results for nonlinear models.

3. BASIC PROBIT AND LOGIT

Proving what happens to estimated coefficients in probit and logit models is more complicated than in linear models. The crux of the problem is that these models do not estimate the coefficients β, instead they estimate β/σ. The normalization is clear by rewriting equation (2), the probability that $y_d = 1$ conditional on covariates *x*, and showing that it is the same as the probability that $y > 0$, conditional on covariates *x* and then rewriting the expression in terms of a normalized error. Start by showing the relationship between the dichotomous dependent variable to its latent continuous dependent variable.

$$Pr(y_d = 1|x) = Pr(y > 0|x) \qquad (3)$$

Write this in terms of the linear index and then put the mean-zero error on one side.

$$Pr(y_d = 1|x) = Pr(0.5x_d + x_1 + 2x_2 + (x_3 + 3x_4) + \varepsilon > 0|x)$$

$$= Pr(\varepsilon < 0.5x_d + x_1 + 2x_2 + (x_3 + 3x_4)|x)$$

If the mean-zero error term is normalized by its standard error, then one can make statistical statements about the probability that a standardized error takes on a range of values, after assuming a particular distribution for the error (e.g., normal or logistic).

$$Pr(y_d = 1|x) = Pr\left(\frac{\varepsilon}{\sigma} < \frac{0.5x_d + x_1 + 2x_2 + (x_3 + 3x_4)}{\sigma}|x\right) \qquad (4)$$

The root mean squared error σ cannot be estimated in a probit or logit model. The root mean squared error depends on the unit of the dependent variable. For example, in a model of wages (unlogged) the magnitude of σ depends on whether wages are measured in dollars,

pennies, euros, pounds, yen, yuans, or other currency. In the corresponding probit or logit, the dependent variable measures whether wages exceed some absolute monetary threshold—the currency is irrelevant.

Therefore, while textbooks often present this as an arbitrary normalization (set $\sigma = 1$), it is better to think of this instead as the models estimate the ratio of β/σ. The distinction between β and β/σ is central to this paper, and the ratio β/σ depends on which variables *x* are being conditioned on. As nuisance variables are added to the model, they are removed from the error term. This makes the error variance smaller and the estimate of β/σ larger.

The other issue for formal proofs is that taking expectations depends on functional form. Showing nice results is easy only for certain functional forms. Wooldridge (2002) proves that if an omitted variable is normally distributed and independent of the other variables in the model, then the probit will consistently estimate β/σ.

Is an estimate of β/σ useful? In section 4, where I discuss interpretation, I argue that the answer is *yes*—if the results are presented in terms of quantities that do not depend on β but instead depend on β/σ. Average partial effects, the marginal effect of a change in a variable averaged over the distribution of omitted variables, depend on β/σ and are unaffected (given the strong assumptions of normality and independence). In contrast, the odds ratio (from a logit model) depends on β, so using β/σ as an estimate of β will give different magnitude of results depending on whether independent variables are included or omitted from the model. There is more explanation in section 4.

Before showing results for the logit, I return to the probit model, where it is well known what will happen to the estimated β/σ when independent variables are added to the model (Wooldridge 2002). Using the same data as for the linear probability model, the results for

probit at first appear quite different (see Table IV). In contrast to the linear regression where coefficients are identical across the model specifications, for the probit the estimated β/σ increase when independent variables are added. For example, the estimated coefficient (actually β/σ) on the dummy variable increases from 0.1122, to 0.1193, to 0.5033 when the two independent covariates are added sequentially. The estimated coefficients on the two other variables of interest also increase, by roughly similar proportions.

The standard errors increase overall, but it is hard to make specific statements about the magnitude of the change because of two opposing effects. One effect is that the standard error shrinks as the model error gets smaller, just as in linear regression models. The other effect is that the coefficient is rescaled by σ, which should increase the absolute magnitude of the standard error in proportion to σ.

If the estimated coefficients change, then do the marginal and incremental effects also change in the probit model? No, these statistics are basically unaffected (see Table V). The row marked "IE" indicates the incremental effect of a change in the dummy variable. The rows marked "ME" indicate the marginal effects of a change in a continuous variable. These marginal and incremental effects correspond to the coefficients in the linear probability model (see Table III) which are also marginal or incremental effects. The estimated marginal and incremental effects are virtually the same across model specifications, and are essentially the same for both the linear probability model and the probit. The differences in the estimated marginal effects are well within the standard errors; these results confirm what Wooldridge proved (2002). The standard errors are similar in the probit compared to the linear probability model, although a bit smaller in the probit, especially in column 3.

While the incremental effect is a difference between two probabilities, the risk ratio is the ratio of two probabilities. Because the risk ratio is calculated from β/σ and not from β, it too is largely unaffected by independent unobserved heterogeneity. The rows in Table V marked "RR" show that these quantities are essentially unchanged across model specifications.

Lee (1982) discusses the situations in which omitted variables do not affect the estimated coefficients of the included variables for logit and multinomial logit models. He proves that the sufficient condition is that, conditional on the dependent variable, the omitted and included variables must be independent. The additional conditioning on the dependent variable is stronger than in the linear regression model. The Monte Carlo data for this paper were constructed so that the additional variables are independent conditional on the dependent variable as well, so this property is satisfied.

The logit results for the same Monte Carlo data are similar to those of the probit, with one important exception. As before, the estimated coefficients (again, really β/σ) increase when omitted independent variables are added back into the model (see columns 4 through 6 of Table IV). The difference is that the magnitudes of the logit coefficients are larger than those of the corresponding probit, by the well-known amount of roughly 1.6.

Like the probit, the estimated incremental and marginal effects and the risk ratios are largely unaffected by the inclusion or exclusion of the independent unobserved heterogeneity. The magnitudes of the marginal effects are nearly identical to those of the probit and linear probability model, certainly the same within the confidence intervals.

For the logit, many researchers like to report odds ratios, which are exponentiated coefficients. Actually—and this distinction is critical—exponentiated β/σ conditional on the variables in the model. Odds ratios increase when independent omitted variables are added to

the model (see Table VIII). The increase can be substantial. Adding in x_4 to the model increases the odds ratios of the other variables by several orders of magnitude.

In one specific Monte Carlo example I have shown that the marginal effects in probit and logit models are nearly constant. I have also run many other Monte Carlo experiments to show that this holds empirically in many other situations (results not reported in tables). It holds for dependent variables whose mean is close to zero or close to one, not just for a mean near one-half. It holds for many different distributions of included and omitted variables: normal, logistic, and uniform. However, it does not seem to hold for highly skewed data, such as the exponential of a random normal. More research could be done to understand exactly which distributions matter.

4. INTERPRETATION

Next I provide interpretation of the various statistics produced by probit and logit models. The interpretation is fundamentally linked to the variables in the model (or those excluded), even when those variables are independent of the variables of interest. This is an important difference from interpreting linear regression models, and again shows how intuition from the familiar linear regression model fails in nonlinear models. The differences in results across the different model specifications are due to differences in what is being conditioned on. The simpler models, with independent unobserved heterogeneity as part of the error term, are conditioned on fewer variables. The richer models, which include the independent variables, are conditioned on more variables. This conditioning affects both the magnitudes of statistics like odds ratios and their interpretation.

One way to understand the intuition for why the estimated coefficients β/σ get larger as independent variables are added to the model is to return to the formulas. The mathematical way to see this is through equation (4) (see also Mroz and Zayats (2008) for a similar discussion of the probit model). When x_3 and x_4 are excluded from the model and therefore part of the error term, then the model error σ is larger than when x_3 and x_4 are included in the model.

$$Pr(y_d = 1|x_d, x_1, x_2) = Pr\left(\frac{\varepsilon}{\sigma_{d12}} < \frac{0.5x_d + x_1 + 2x_2}{\sigma_{d12}}|x\right) \qquad (5)$$

$$Pr(y_d = 1|x_d, x_1, x_2, x_3, x_4) = Pr\left(\frac{\varepsilon}{\sigma_{d1234}} < \frac{0.5x_d + x_1 + 2x_2 + x_3 + 3x_4}{\sigma_{d1234}}|x\right) \qquad (6)$$

The logit model estimates β/σ_{d12} and β/σ_{d1234}.

But it is easier to understand the importance of conditioning using a different example, one that explicitly shows the importance of conditioning on values of the independent variable and allows for simple graphs. For this second Monte Carlo example, there are two explanatory variables. The first variable w_1 takes on discrete values of one through seven (think of this as answers on a seven-item Likert scale). The second variable w_2 has three discrete values: –1, 0, and 1. There are 21 unique combinations of these two variables; I generated 10 observations for each unique combination, for a total of 210 observations. The two variables are independent from each other by construction. The data generating process has the dichotomous dependent variable *y* equal to one if the sum of w_1 and $2w_2$ plus a random normal error ω (mean zero) exceed 4, giving the mean value of *y* about one-half.

$$y = 1(w_1 + 2w_2 + \omega > 4) \qquad (7)$$

Before showing the logit results, it is useful to discuss the results from the linear probability model. The results from the simple and the richer linear probability models (*N* = 210) are

$$y = -0.1 + 0.146w_1 \qquad R^2 = 0.34 \quad \hat{\sigma} = 0.41$$

$$y = -0.1 + 0.146w_1 + 0.307w_2 \qquad R^2 = 0.60 \quad \hat{\sigma} = 0.32$$

The graph of the predicted probability that $y = 1$ is linear in w_1 and w_2 (see Figure 1). A one-unit increase in w_1 raises the predicted probability by 14.6 percentage points. A one-unit increase in w_2 shifts the predicted probability by about 31 percentage points for any value of w_1. More importantly for comparison with the logit model, for a given value of w_1, the following are identical: the average of the predictions over all values of w_2 and the prediction of the average of w_2 (i.e., $w_2 = 0$). Of course, this graph also illustrates the main problem with linear probability models—some predicted probabilities are outside the feasible range of [0,1]. Next we estimate a logit model, which constrains predictions to be within the feasible range, on the same data.

The logit model shows that, not surprisingly, the estimated coefficient (again, β/σ) on w_1 is positive, and increases when w_2 is added to the model (see Table IX). The odds ratio also increases, while the average marginal effect stays about the same. This is consistent with the prior results when adding a variable that is independent of the other covariates to the model.

The advantage of this simple discrete design is that it allows for easy graphing of the results (see Figure 2). The dots connected by the solid line represent predicted values of the dependent variable from the simple model that only includes w_1. These predictions follow the familiar *S*-curve, showing that an increase in w_1 increases the probability that the dependent variable equals one, but in a nonlinear way. The other dots, and lines connecting them, represent various calculations from the model that also includes w_2. The predicted probability that the dependent variable equals one, conditional on $w_2 = 1$ (or $w_2 = -1$), is represented by the upper (or lower) *S*-curve, connected by dots. In between the outer lines are the predictions for $w_2 = 0$,

connected by long dashes. Lying close to the first set of predictions are the average predictions for the second model, conditional only on w_1. This graph illustrates the importance of conditioning.

Consider the predicted probability that $y = 1$ conditional on w_1 for the richer model, which includes w_2. Of particular interest is the difference between averaging over all the three possible values of w_2 compared to conditioning on $w_2 = 0$. Jensen's Inequality, which states that the mean of a nonlinear function does not equal the nonlinear function of the mean, applies. Computing predictions for all observations and then taking the average (given w_1) yields predicted probabilities conditional on w_1 that are quite close to those of the simple model. However, first taking the mean of w_2 (i.e., condition on $w_2 = 0$) and then computing predicted probabilities yields the curve connected by the long dashes. Predicted probabilities conditional on $w_2 = 0$ are further away from 0.5 than averaging over all values of w_2. Why? Conditional probabilities are based on β/σ with a smaller σ, making them larger in absolute value. More information (conditioning on more variables) improves the ability to sort the data into those that are likely to have $y = 1$ and those that are likely to have $y = 0$. The observations that continue to have predictions in the mid-range, even after conditioning on lots of variables, are the observations that are more likely to be affected by the variables of interest.

Here we can visualize the main intuition. When conditioning on more variables, the coefficient on w_1 increases, which makes the slope of the CDF with respect to that variable steeper in the middle range. Consider what happens for a particular value of w_1. Let $w_1 = 2$. Without conditioning on w_2, the mean predicted probability is about 0.15 (from the simple logit model). Also conditioning on w_2 leads to three different predictions. The predicted probabilities for $w_2 = 0$ and $w_2 = -1$ are close together because of the floor effect—probabilities are bound

by zero; this is in contrast to the linear probability model that allows nonsensical negative predicted probabilities. Therefore, the average predicted probability over the three values of w_2 lies between the values for $w_2 = 0$ and $w_2 = 1$, and is about the same as the unconditional amount. The nonlinearity of the logit function for this calculation is clear from the graph.

Next consider how conditioning affects the odds ratio. Part of the appeal of the odds ratio is that it simplifies a complicated nonlinear equation to a single number. However, even though the formula does not appear to depend directly on any other covariate (everything falls out), the interpretation does fundamentally depend on the covariates in the model. The simplicity of the formula belies the complexity of the interpretation. Here is the derivation of the odds ratio for the dummy variable in the first Monte Carlo example, where the odds are the ratio of the probability to one minus the probability.

$$Odds(y_d = 1|x_d = 1, x) = e^{\frac{0.5+x_1+2x_2+x_3+3x_4}{\sigma_{d1234}}}|x_d = 1, x \quad (8)$$

$$Odds(y_d = 1|x_d = 0, x) = e^{\frac{x_1+2x_2+x_3+3x_4}{\sigma_{d1234}}}|x_d = 0, x \quad (9)$$

This example includes both x_3 and x_4 in the model specification, so the error variance σ_{d1234} is explicitly conditional on all those variables. The odds ratio that $y_d = 1$ comparing $x_d = 1$ to $x_d = 0$ and conditional on the other *x* variables is

$$OddsRatio(y_d = 1|x) = e^{\frac{0.5}{\sigma_{d1234}}}|x \quad (10)$$

The explicit dependence of this odds ratio on the error variance and also on the other variables *x* is rarely acknowledged, even in textbooks. Yet, clearly the model specification will affect σ and hence the estimated coefficient β/σ, and therefore the reported odds ratio (Mood 2010).

The common complaint about reporting only estimated coefficients from a logit or probit model is that the coefficients on their own are hard to interpret, without further manipulation.

The same can almost be said about odds ratios. To be clear, an estimated odds ratio does have a specific interpretation, but the correct interpretation is far more complex than commonly believed or reported (Mood 2010). Without stating the proper conditioning, a statement like, "The estimated odds ratio is 1.5." is factually incorrect. A more accurate, but imprecise, statement would be "An estimated odds ratio is 1.5." A correct precise interpretation might be, "The estimated odds ratio is 1.5 conditional on age, gender, race, and income, but a different odds ratio would be found if the model also controlled for other important factors not included in the data set. This estimated odds ratio cannot be directly compared to odds ratios estimated from other data sets, or even from the same data set with different model specification." The odds ratio is primarily useful to show the sign and statistical significance of an effect, but the same can be said about the estimated coefficient β/σ. Other statistics derived from β/σ have the same problem. Odds, log odds, and log odds ratios are all conditional on the model specification.

In contrast, the average marginal effect does not have this problem. This has been proved rigorously for the case of independent omitted variables for the probit model and for omitted variables that are independent of both the dependent variable and other explanatory variables for the logit (and multinomial logit) (Lee 1982; Yatchew and Griliches 1985; Wooldridge 2002). As Wooldridge explains the intuition, the formula for the marginal effect in a probit model has two parts $\frac{\beta}{\sigma}\phi\left(\frac{x\beta}{\sigma}\right)$, and an increase in σ lowers the first term β/σ and raises the second term $\phi\left(\frac{x\beta}{\sigma}\right)$.

Because incremental effects are similar to marginal effects—they are a discrete difference instead of a derivative—incremental effects and risk differences empirically tend to be unaffected by the inclusion or exclusion of independent variables. I confirmed this with Monte Carlo experiments on a variety of data sets with different distributions. When the incremental change was small, for most distributions the incremental effect was essentially unchanged. As

with marginal effects, however, when the distribution of the omitted variable was highly skewed, then the incremental effect changed.

The risk ratio is in between the marginal effect and the odds ratio in terms of being affected by unobserved heterogeneity. In my Monte Carlo simulations, the risk ratio seemed to shift more than marginal effects, but still far less than odds ratios. More research could be done to understand exactly which distributions matter.

5. FIXED EFFECTS MODELS

Hoes does unobserved heterogeneity inform the interpretation of fixed effects models? With panel data, the researcher can often control for unobserved heterogeneity that is common to a group of observations. For example, in repeated observations on individuals, an individual fixed effect controls for characteristics of the person that are time invariant, such as gender, eye color, and perhaps risk preferences. Individual fixed effects would not control for time-varying characteristics such as age, education (for a young sample), or health status. Controlling for fixed effects in linear regression, which has fixed effects independent of the variables of interest, has exactly the same effect on the coefficient found before—none.

The logit model with fixed effects is more interesting, however, than just adding independent variables to a logit, for three reasons. One reason is that marginal effects are not defined in a Chamberlain conditional fixed effect logit model (Chamberlain 1980). The fixed effects are never estimated. There is no constant term, and the only way to interpret the estimated coefficients is as an odds ratio (unless one makes strong additional assumptions, discussed below). The second reason is that the results are conditional on the data in another way. Observations with no within-group variation in the dependent variable are dropped from

the model. This does more than change the sample size—it changes the interpretation and the generalizability of the results. The third reason is that we can now better understand how to interpret case-control studies, which are a common study design in epidemiology where subjects are matched on some common characteristics but differ on one important variable of interest (e.g., treatment).

Returning to the first reason, consider analyzing one large data set, with clusters at the state, zip code, household, and individual level. How will the results change if one adds fixed effects at progressively finer levels, starting at the state level and moving to the individual level? In linear regression, if all the fixed effects are independent of the variables of interest, then the marginal effects will remain the same no matter the level of the fixed effects. In a logit model with fixed effects, as the fixed effects are applied at a finer level, then the coefficients (again, β/σ) get larger because of the renormalization. So far, this is just the same as with other independent variables.

One major difference between a fixed effects logit and a regular logit is that one cannot directly compute predicted probabilities from the conditional fixed effects model. The lack of a constant term means that only relative statements about risks can be made. Without further assumptions, the odds ratio is the only possible interpretation. Statistical packages like Stata allow the user to make additional assumptions and thereby calculate predicted probabilities. For example, one can assume that all the fixed effects are zero. But it is odd to go to the trouble of modeling fixed effect heterogeneity carefully and then assume homogeneity when making predictions. Therefore, in this model with no constant term, the odds ratio interpretation can be useful, as long as proper conditioning is understood.

In a conditional logit model with fixed effects some observations may be dropped. Any observation from a group that does not have variation in the dependent variable within the group will be dropped. This is consistent with the graph of predicted values from the simple logit model. When the variable w_2 was added to the model, then more of the conditional predicted probabilities were close to either zero or one. More information allows for better sorting, and fixed effects at a fine level add much more information. The remaining observations are those that have at least one observation with $y = 1$ and at least one with $y = 0$ within the same cluster. The name *conditional* logit is because the results are conditional on the sum of the dependent variable within a group. Analogous to the local average treatment effect (LATE) interpretation of instrumental variables, in a fixed effects logit model when observations are dropped from the model, the interpretation of the effect applies to the remaining marginal observations. In a short panel in which many observations are dropped, the remaining observations are not representative of the entire data set, and the interpretation is only about those whose clusters are not dropped.

Another way to think about conditioning in the fixed effects logit model is for the case-control study. Consider a simple hypothetical study design; let there be one treatment person matched to each control person. With no conditioning, the matching could be strange. For example, just picking two human beings at random from anywhere in the world could result in a 4-year old girl from Indonesia matched to 82-year old sheep farmer from the Australian Outback. With no matching on controls, the results from pairs like this would not be very informative. Yes, there may be a difference in outcome, but if there is, is the difference due to the treatment or any of the other innumerable differences between these two people? This corresponds to having a relatively small β/σ because σ is large because there is so much unaccounted for in the model.

At the other extreme, consider matching pairs of monozygotic twins, or a person to themselves in different time periods (pre-post trial). When holding nearly all relevant factors constant (e.g., age, gender, genetics, current health, health history, smoking status, risk preferences) the effect of a treatment relative to a control can be isolated. In a conditional fixed effects logit model, many of the paired observations will fall out of the model because both have the same outcome—conditional on all the matching characteristics, the treatment did not matter. However, for those pairs of observations with different outcomes, the effect of treatment is strong. This corresponds to having a relatively large β/σ because σ is small because there is so much accounted for in the model.

The results from a conditional fixed effects logit model can only be interpreted as conditional on the fixed effects. Returning to example with clusters at the state, zip code, household, and individual level: Fixed effects at a different levels lead to different magnitudes of the estimated β/σ conditioned on those other fixed effects. The results are not directly comparable; the interpretation is completely different.

The bottom line is that in a conditional fixed effects logit model, the exponentiated estimated coefficient (β/σ) can be interpreted as the odds ratio in a precise and meaningful way. But the correct precise interpretation (conditional on the data and model specification) means that the magnitude cannot be directly comparable to other results. In contrast, marginal effects cannot be calculated directly from the conditional fixed effects logit model. An understanding of what conditioning means in this model is useful in understanding what conditioning means in regular logit models.

6. EXTENSIONS

All the points that apply to probit and logit models also apply to their related multinomial and ordered models. Those models also have an arbitrary normalization and so the estimated coefficeints β/σ again depend on independent unobserved heterogeneity.

The same issues also matter in two-stage residual inclusion, the control function instrumental variables method for dealing with endogeneity when the main dependent variable is dichotomous and the endogenous variable is continuous (Newey, Powell, Vella 1999; Terza, Basu, Rathouz 2008). Independent unobserved heterogeneity in the main probit equation changes the point estimate of the endogenous variable of interest in the usual way. The local average treatment effect interpretation is conditional on the model specification.

Mroz and Zayats (2008) discuss several of these issues in detail. They make many of the same basic arguments made in this paper. They show how excluding an independent normally distributed variable from a probit model will change the error variance and hence the arbitrary scaling of the estimated coefficients β/σ. This, they argue, has led to bias in several important nonlinear models. Mroz and Zayats specifically show problems with tests of endogeneity in a conditional logit model, with the interpretation of multilevel models with binary outcomes, and for hazard models with heterogenetiy controls. They point out that many problems of interpretation of arbitrarily scaled coefficients could be avoided by estimating the substantive effect of interest.

7. IMPLICATIONS

There is no one odds ratio. Each odds ratio estimated in a multivariate model is conditional on the data and model specification in a way that makes comparisons with results

from other models difficult or impossible (Allison 1999; Mood 2010). A study that aims to estimate *the* odds ratio is misguided. Given the voluminous literature in epidemiology, clinical research, health services research, and other social sciences that estimates and reports odds ratios without proper discussion of conditioning, arbitrary normalization of parameters, or independent unobserved heterogeneity, there is a long way to go to improve best practice and translation of results.

Another implication is that the importance of conditioning to the odds ratio interpretation complements and enhances the already strong criticism of reporting odds ratios on the basis of misunderstanding by others (Greenland 1987; Sackett, Deeks, Altman 1996; Altman, Deeks, Sackett 1998; Kleinman and Norton 2009; Tajeu et al. 2012). Most prior arguments have focused on the difference between risk ratios and odds ratios, and how people mistakenly interpret odds ratios as risk ratios (Sackett, Deeks, and Altman (1996) also discuss other points). This paper adds to that debate by pointing out that the correct interpretation of odds ratios is even harder and more subtle to explain. It is not enough understand odds ratios; an understanding of what the odds ratio is conditioned on is also necessary. So while the odds ratio in multivariate analysis does have a precise interpretation, that interpretation is even harder to explain than previously thought.

What should researchers report when the dependent variable is dichotomous? The answer depends on the research question. There is no single right way for all studies. Nonlinear models are inherently complicated, and predictions based on estimated models will depend on whether the predictions are conditional on a particular set of covariates or averaged over a distribution of covariates. A linear probability model can be useful if the goal is an overall average marginal effect (Angrist 2001), although considerable caution is needed because of the

problems of predicted probabilities that lie outside of [0,1] and the assumption of linearity of effects. In a conditional fixed effects model (or case-control study design), interpreting results as odds ratios may be preferable to making additional assumptions needed to compute other summary statistics, although considerable caution is needed because of the importance of reporting the results as conditional on the fixed effects (or matched sample characteristics) and being applicable only to the subset of observations that are not dropped from the estimation sample. Risk ratios may be preferable in many cases for reasons of interpretation (Kleinman and Norton 2009), but they are still conditional on the model specification. More work is needed to understand how sensitive the adjusted risk ratio is to unobserved heterogeneity. But in most cases, I believe that some version of a marginal or incremental effect, perhaps averaged over the sample or subsample, will be both most understandable to the reader and least sensitive to independent unobserved heterogeneity. Mood (2010) has a comprehensive discussion of alternatives.

In summary, the most important things are to answer the research question and to communicate the results clearly to the intended audience.

8. CONCLUSION

The correct interpretation of odds ratios acknowledges that the magnitude is conditional on the model. When more independent variables are included in the model, the error variance is reduced and the estimated coefficient β/σ increases. Yes, the odds ratio has meaning, but a narrow conditional meaning that cannot be directly compared to other models. A vast literature in epidemiology, clinical medicine, health services research, and other social sciences reports odds ratios without the correct interpretation, and apparently without understanding that odds

ratios are highly sensitive to unobserved heterogeneity. Given the documented difficulties that most people have in understanding odds ratios, this places further burden on researchers to correctly interpret the results to readers in a meaningful way.

There are no conflicts of interest in this research. No committee or institution had to approve this manuscript.

REFERENCES

Allison PD. 1999. Comparing logit and probit coefficients across groups. *Sociological Methods and Research* 28: 186–208.

Altman DG, Deeks JJ, Sackett DL. 1998. Odds ratios should be avoided when events are common [Letter]. *British Medical Journal* 317: 1318.

Angrist JD. 2001. Estimation of limited dependent variable models with dummy endogenous regressors: Simple strategies for empirical practice. *Journal of Business and Economic Statistics* 19(1): 2–28.

Chamberlain G. 1980. Analysis of Covariance with Qualitative Data. *The Review of Economic Studies* 47(1): 225–238.

Gail MH, Wieand S, Piantadosi S. 1984. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* 71(3): 431–444.

Greenland S. 1987. Interpretation and choice of effect measures in epidemiologic analyses. *American Journal of Epidemiology*.125(5): 761–768.

Griliches Z. 1957. Specification bias in Estimates of Production Functions. *Journal of Farm Economics* 39(1): 8–20.

Kleinman LC, Norton EC. 2009. What's the risk? A simple approach for estimating adjusted risk ratios from nonlinear models including logistic regression. *Health Services Research* 44(1): 288–302.

Lee L-F. 1982. Specification error in multinomial logit models, analysis of omitted variable bias. *Journal of Econometrics* 20: 197-209.

Mood C. 2010. Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review* 26(1): 67–82.

Mroz TA, Zayats YV. 2008 Arbitrarily normalized coefficients, information sets, and false reports of "biases" in binary outcome models. *Review of Economics and Statistics* 90(3): 406–413.

Newey WK, Powell JL, Vella F. 1999. Nonparametric estimation of triangular simultaneous estimation models. *Econometrica* 67(3): 565–603.

Sackett DL, Deeks JJ, Altman DG. 1996. Down with odds ratios! *Evidence Based Medicine* 1: 164-166.

Schwartz LM, Woloshin S, Welch HG. 1999. Misunderstandings about the effects of race and sex on physicians' referrals for cardiac catheterization. *New England Journal of Medicine* 341(4): 279–283.

Tajeu GS, Sen B, Allison DB, Menachemi N. Forthcoming 2012. Misuse of odds ratios in obesity literature: An empirical analysis of published studies. *Obesity*.

Terza JV, Basu A, Rathouz PJ. 2008 Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. Journal of Health Economics 27(3): 531–543.

Walter SD. 2000. Choice of effect measure for epidemiological data. *Journal of Clinical Epidemiology* 53: 931–939.

Wooldridge JM. 2002. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press: Cambridge, MA.

Yatchew A, Griliches Z. 1985. Specification error in probit models. *Review of Economics and Statistics* 67(1): 134-139.

Table I. Summary Statistics for the first Monte Carlo data set.

Variable	Mean	Std. Dev.	Minimum	Maximum
y	.250	4.64	-19.159	17.894
y_d	.529	.499	0.0	1.0
x_d	.500	.500	0.0	1.0
x_1	0.0	1.00	-3.558	3.324
x_2	0.0	1.00	-3.840	4.640
x_3	0.0	1.41	-4.793	5.223
x_4	0.0	1.22	-4.776	5.082

N = 10,000

Table II. Correlations of the variables in the first Monte Carlo data set.

	y	y_d	x_d	x_1	x_2	x_3	x_4
y	1.0000						
y_d	0.7997	1.0000					
x_d	0.0538	0.0391	1.0000				
x_1	0.2154	0.1775	0.0000	1.0000			
x_2	0.4307	0.344	0.0000	0.0000	1.0000		
x_3	0.3046	0.239	0.0000	0.0000	0.0000	1.0000	
x_4	0.7913	0.634	0.0000	0.0000	0.0000	0.0000	1.0000

Table III: OLS and LPM results for the simple and full model specifications.

	OLS		LPM	
Variables	Simple	Full	Simple	Full
Constant	.000	.000	.5092	.5092
	(.057)	(.014)	(.0065)	(.0044)
x_d	.500	.500	.0390	.0390
	(.081)	(.020)	(.0092)	(.0062)
x_1	1.000	1.000	.0886	.0886
	(.041)	(.010)	(.0046)	(.0031)
x_2	2.000	2.000	.1717	.1717
	(.041)	(.010)	(.0046)	(.0031)
x_3		1.0000		.0844
		(.0071)		(.0022)
x_4		3.0000		.2584
		(.0082)		(.0025)
R-squared	0.23	0.95	0.15	0.61
RMSE	4.06	1.00	0.46	0.31

Table IV. Multivariate results for the probit model using the first Monte Carlo data set.

	Probit			Logit		
Variables	Simple	Add x_3	Full	Simple	Add x_3	Full
Constant	0.025 (0.019)	0.027 (0.019)	0.044 (0.037)	0.042 (0.031)	0.045 (0.032)	0.072 (0.066)
x_d	0.112 (0.026)	0.119 (0.027)	0.503 (0.053)	0.186 (0.044)	0.198 (0.045)	0.913 (0.097)
x_1	0.255 (0.014)	0.271 (0.014)	1.023 (0.035)	0.420 (0.022)	0.451 (0.024)	1.837 (0.066)
x_2	0.495 (0.014)	0.527 (0.015)	2.078 (0.055)	0.817 (0.025)	0.878 (0.026)	3.74 (0.11)
x_3		0.260 (0.010)	1.017 (0.030)		0.433 (0.017)	1.824 (0.058)
x_4			3.073 (0.075)			5.51 (0.15)
Pseudo *R2*	0.12	0.17	0.78	0.12	0.17	0.78

Table V. Incremental effects, marginal effects, and risk ratios for the probit model using the first Monte Carlo data set.

		Probit		
Variables		Simple	Add x_3	Full
x_d	IE	0.0390	0.0389	0.0414
		(0.0092)	(0.0088)	(0.0043)
	RR	1.0766	1.0764	1.0814
		(0.0187)	(0.0180)	(0.0088)
x_1	ME	0.0886	0.0884	0.0842
		(0.0044)	(0.0043)	(0.0020)
	RR	1.1688	1.1682	1.1632
		(0.0089)	(0.0085)	(0.0042)
x_2	ME	0.1721	0.1719	0.1708
		(0.0041)	(0.0039)	(0.0019)
	RR	1.3392	1.3390	1.3416
		(0.0099)	(0.0096)	(0.0047)

Notes: IE = incremental effect; ME = marginal effect; RR = risk ratio. Std. errors in parentheses.

Table VII. Incremental effects, marginal effects, and risk ratios for the logit model using the first Monte Carlo data set.

		Logit		
Variables		Simple	Add x_3	Full
x_d	IE	0.0392	0.0388	0.0418
		(0.0092)	(0.0088)	(0.0043)
	RR	1.0770	1.0762	1.0822
		(0.0188)	(0.0181)	(0.0088)
x_1	ME	0.0887	0.0884	0.0840
		(0.0045)	(0.0043)	(0.0020)
	RR	1.1689	1.1685	1.1636
		(0.0089)	(0.0086)	(0.0042)
x_2	ME	0.1723	0.1720	0.1708
		(0.0041)	(0.0040)	(0.0020)
	RR	1.3409	1.3406	1.3424
		(0.0102)	(0.0099)	(0.0047)

Notes: IE = incremental effect; ME = marginal effect; RR = risk ratio. Standard errors in parentheses.

Table VIII. Odds ratios for the logit model using the first Monte Carlo data set.

	Logit		
Variables	Simple	Add x_3	Full
Constant	1.043	1.046	1.075
	(0.032)	(0.033)	(0.071)
x_d	1.204	1.219	2.49
	(0.053)	(0.055)	(0.24)
x_1	1.522	1.570	6.277
	(0.034)	(0.037)	(0.041)
x_2	2.263	2.405	41.8
	(0.056)	(0.063)	(4.5)
x_3		1.542	6.198
		(0.027)	(0.036)
x_4			248.3
			(36.7)
Pseudo *R2*	0.12	0.17	0.78

Table IX. Multivariate logit results for the second Monte Carlo data set.

	Logit Coeficients		Logit Odds Ratios		Marginal Effects	
Variables	Simple	Full	Simple	Full	Simple	Full
Constant	–3.49	–7.4	0.0304	0.0006		
	(.49)	(1.2)	(.015)	(0.0007)		
w_1	0.86	1.83	2.36	6.2	0.1296	0.1378
	(.11)	(.28)	(.26)	(1.7)	(.0051)	(.0031)
w_2		3.62		37.2		
		(.61)		(22.6)		
Pseudo *R2*	0.32	0.65				
N	210	210				

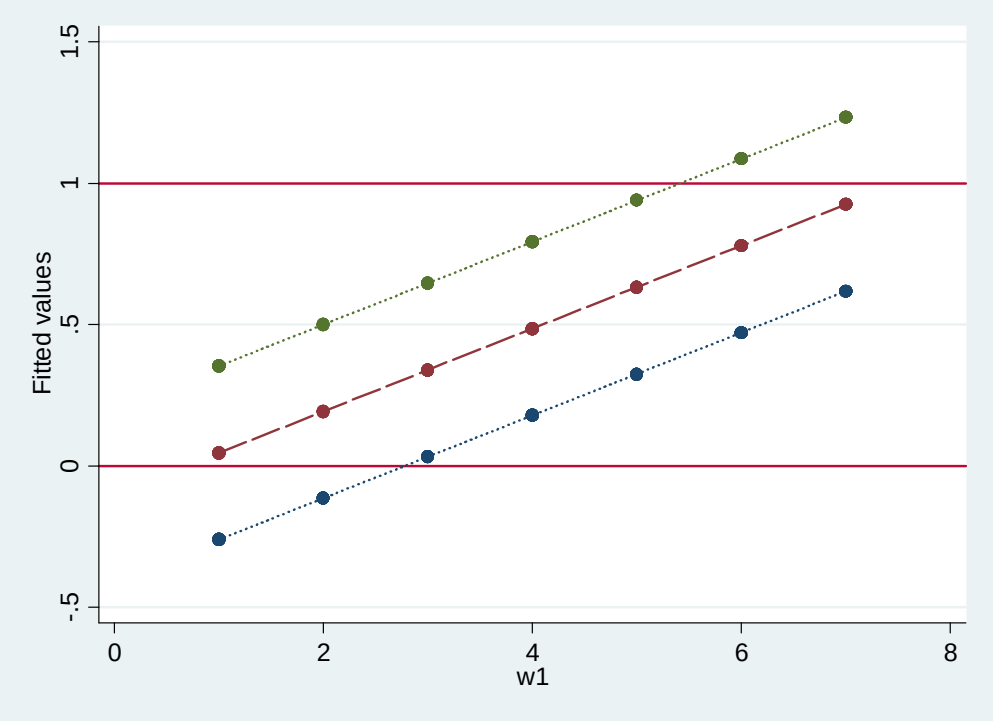


Figure 1. Predicted probabilities from a linear probability model for the second Monte Carlo data set, conditional on w_1 and w_2.

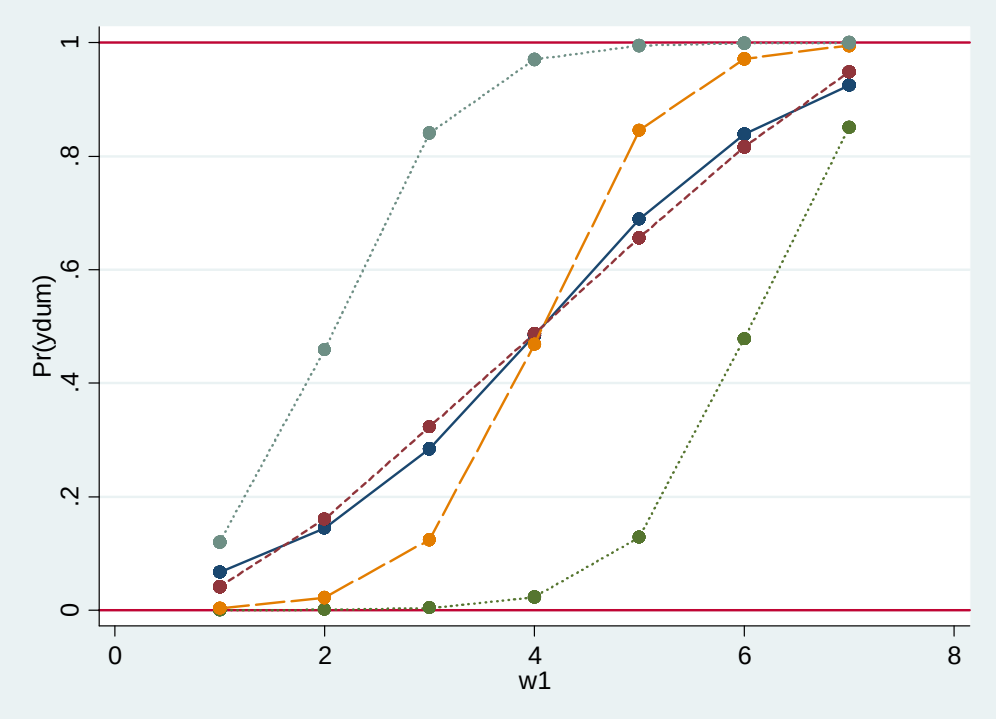


Figure 2. Predicted probabilities from a logit model for the second Monte Carlo data set, conditional on w_1 and w_2.