Københavns Universitet

Comparing Regression Coefficients Between Same-sample Nested Models Using Logit and Probit

Karlson, Kristian Bernt; Holm, Anders; Richard, Breen

Comparing Regression Coefficients Between Same-sample Nested Models Using Logit and Probit: A New Method

Authors: Kristian Bernt Karlson (kbk@soc.ku.dk), Anders Holm, and Richard Breen

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.

Comparing Regression Coefficients Between Same-Sample Nested Models using

Logit and Probit: A New Method*

Kristian Bernt Karlson

The Danish National Centre for Social Research and

Department of Education, Aarhus University

Anders Holm

Department of Education, Aarhus University

and

Richard Breen

Center for Research on Inequality and the Life Course,

Department of Sociology, Yale University

10653 words (including all text, notes, and references)

Running head: Comparing logit and probit regression coefficients

Comparing Regression Coefficients Between Same-Sample Nested Models using Logit and Probit: A New Method

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 scaling 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 Educational Longitudinal Survey.

Comparing Regression Coefficients Between Same-Sample Nested Models using Logit and Probit: A New Method

Introduction

Nonlinear probability models such as binary logit and probit models are widely used in quantitative sociological research. One of their most common applications is to estimate the effect of a particular variable of interest on a binary outcome when potentially confounding variables are controlled. Interest in such genuine or "true" coefficients has a long history in the social sciences and is usually associated with the elaboration procedure in cross-tabulation suggested by Lazarsfeld (1955, 1958; Kendall and Lazarsfeld 1950; cf. Simon 1954 for partial product moment correlations). Nevertheless, controlled logit or probit coefficients do not have the same straightforward interpretation as controlled coefficients in linear regression. In fact, comparing uncontrolled and controlled coefficients across nested logit models is not directly feasible, but this appears to have gone unrecognized in much applied social research, despite an early statement of the problem by Winship and Mare (1984).

In this paper we offer a solution. We develop a method that gives unbiased comparisons of logit or probit coefficients of the same variable (x) across same-sample nested models successively including control variables (z). The method decomposes the difference in the logit or probit coefficient of x between a model excluding z and a model including z, into a part attributable to confounding (i.e., the part mediated or explained by z) and a part attributable to rescaling of the coefficient of x. Our method also deals with a further problem in making comparisons, namely that,

in most cases, the error distribution, and not just its variance, will differ across models excluding and including z. Under the assumption that the full model (i.e., the model including both x and z) is correctly specified, Monte Carlo analyses indicate that while other methods suggested for resolving the rescaling problem, y-standardization, average partial effects derived from logit or probit models, and the linear probability model effectively deal with rescaling, they can lead to mistaken inferences about confounding if the error distributions are very different across models. In contrast, in all scenarios studied, our approach performs at least as well, and in some cases better than, others when faced with differences in the error distributions across models.

We proceed as follows. We first review the problem of comparing coefficients across nested logit or probit models, then we present our solution and show how to assess the magnitude of confounding relative to the magnitude of rescaling. Next we develop a test statistic that enables a formal test of confounding. Using Monte Carlo simulations, we show that our method is preferable to other methods for comparing coefficients across nested models. We illustrate our approach using data from the National Educational Longitudinal Survey. A short concluding section summarizes our arguments and suggests further applications.

Comparing coefficients across logit and probit models

In linear regression, the concept of controlling for confounding variables is well understood and has great practical value. For example, a researcher might want to assess how much the effect of years of education on log annual income changes when holding constant academic ability and gender by comparing the uncontrolled coefficient for years of education with its counterpart controlling for ability and

gender. The difference between the two coefficients reflects the degree to which the impact of years of education is mediated or confounded by ability and gender.[1] This kind of design is straightforward within the OLS modeling framework and is probably one of the most widespread practices in empirical social research (Clogg, Petkova, and Haritou 1995).

In logit and probit models, however, uncontrolled and controlled coefficients can differ not only because of confounding but also because of a rescaling of the model. By rescaling we refer to the situation where the size of the estimated coefficient of the variable of interest depends on the error variance of the model and, consequently, on which other variables are included in the model. Including a control variable, z, in a logit or probit model will alter the coefficient of x whether or not z is correlated with x because, if z explains any of the variation in the outcome variable, its inclusion will reduce the error variance. Consequently, logit or probit coefficients from different nested models are not measured on the same scale and are therefore not directly comparable. This is because, unlike linear models, in nonlinear probability models, the error variance is not independently identified and is fixed at a given value (Cramer 2003:22). This identification restriction is well known in the literature on limited dependent variable models (Amemiya 1975; Maddala 1983; Winship and Mare 1984; Yatchew and Griliches 1985; Long 1997; Powers and Xie 2000; Agresti 2002; Cramer 2003), but the consequences of rescaling for the interpretation of logit or probit coefficients are seldom recognized in applied social research (for similar statements, see Allison 1999, Hoetker 2004; 2007, Williams 2009, or Mood 2010).

---

[1] We use 'confounding' as a general term to cover all cases in which additional variables, z, are correlated with the original predictor variable, x, and also affect the outcome variable, y. This includes cases where the additional variables are believed to 'mediate' the original x-y relationship. In terms of path analyses, the part of the x-y relationship mediated or confounded by z is the indirect effect.

There are at least two important reasons why we should be concerned with these problems. The first arises when we have a policy variable, z, which we believe will mediate the relationship between x and a binary outcome, y. Typically, we might first use a logit model to regress y on x to gauge the magnitude of the relationship, and then we might add z as a control to find out how much of the relationship is mediated via z. This would seem to tell us how much we could affect the x-y relationship by manipulating z. But, as we show below, in general such a strategy will tend to underestimate the mediating role of z, increasing the likelihood of our concluding, incorrectly, that changing z would have little or no impact on the x-y relationship. The second reason arises when z is a common cause of x and y. In this situation we would control for z to assess the extent to which the x-y relationship is spurious. However, similar to the first example, because of rescaling, this strategy will underestimate the confounding role of z, thereby increasing the risk of drawing erroneous conclusions about the x-y relationship.

Separating confounding and rescaling

In this section we first show how coefficient rescaling operates in binary non-linear probability models such as the logit and probit. Then we introduce a novel method that decomposes the change in logit or probit coefficients across nested models into a confounding component and a rescaling component, and we also explain why our method is unaffected not only by rescaling (i.e., changes in variance of the error distribution), but also by the changes in the shape of the error distribution across nested models. We also derive standard errors and Z-statistics for formally assessing

the degree of confounding and we present results from a Monte Carlo study showing that the statistical test performs well.

Latent linear formulation of a logit model

Let y^* be a continuous latent variable, x a predictor variable of interest, and z one or more control variables. In this general formulation, y^* may be truly latent in the sense of being a construct not directly measurable as a continuous variable (for example an attitude or a propensity to do something) but it may also be directly measurable in principle but treated as latent because, in the particular case at hand, it was only measured partially (for example we might only know whether someone's income exceeded a particular amount or not). In both cases we would like to estimate the relationship between the predictor variables and the underlying latent y^*.

In this exposition we assume that x and z are correlated and that z may thus confound the x-y^* relationship. Omitting individual subscripts and centering x and z on their respective means (i.e., omitting the intercepts), we follow the notation for linear models in Clogg, Petkova, and Haritou (1995; cf. Blalock 1979) and specify two latent variable models:

$$H_R : y^* = \beta_{yx}x + e \tag{1}$$

$$H_F : y^* = \beta_{yx.z}x + \beta_{yz.x}z + v \tag{2}$$

e and v are error terms. H_R and H_F denote the reduced and full model respectively and we take H_F to be the true model.[2] However, we do not observe y^*: instead we observe y, a dichotomized version of the latent propensity such that:

---

[2] Following Clogg et al. (1995) we use the term "full model" to denote the model including controls, while we denote the model without controls the "reduced model". Since both models cannot be true

$$y = 1 \quad \text{if } y^* > \tau$$
$$y = 0 \quad \text{otherwise,}$$

(3)

where τ is a threshold, normally set to zero.[3] The expected outcome of this binary indicator is the probability of choosing $y = 1$, i.e., $E(y = 1) = \Pr(y = 1)$. Assuming that the error term in (2) follows a logistic distribution we can write $v = \sigma_F u$ where σ_F is a scale parameter and u is a standard logistic random variable with mean 0 and standard deviation $\pi / \sqrt{3}$ (Cramer 2003:22; Long 1997:119). The scale parameter rescales the variance of the error term relative to the variance of the standard logistic distribution and so $\sigma_v = \sigma_F \pi / \sqrt{3}$. We can also write $e = \sigma_R \zeta$, where $\sigma_R = \dfrac{\sigma_e \times \sqrt{3}}{\pi}$ and we define $\zeta = e / \sigma_R$. That is, the standard deviation of the error in the reduced model is written as equal to the standard deviation of the standard logistic multiplied by a scale factor (cf. Yatchew and Griliches 1985).

However, although we can express σ_e in these terms, e is unlikely to follow the same distribution as v. The distribution of the error of the reduced model depends on the distribution of v, the error of the full model, and on the distribution of the omitted confounder, z. Thus, the two models will differ not only in their scaling but also in the distribution of their error terms.[4] This can have consequences for our ability to make cross-model comparisons when the logistic functional form (or whichever

---

simultaneously we hold the full model to be the "true" model, i.e., the model on which we will base our inferences. In this context, "reduced model" should not be confused with "reduced form model".

[3] Whenever the threshold constant is nonzero, it is absorbed in the intercept of the logit model. Because it does not affect the effect estimates setting it to zero, as we do here, is innocuous.

[4] An exception occurs when both v and z are normally distributed (and we would fit a probit to the data). In this case e will also be normally distibuted and so the two models will differ in scale parameters but not in the distribution of their errors. We exploit this property later in our simulations.

functional form we believe appropriate) is a much poorer approximation to the actual

error distribution in one model than in the other.

Using this additional notation and setting $\tau = 0$, we obtain the following

two logit models, corresponding to (1) and (2) above:

H_R^{Logit} :

$$\Pr(Y=1) = \Pr(y^* > 0) = \Pr\left(u > -\frac{\beta_{yx}}{\sigma_R}x\right) = \frac{\exp\left(\frac{\beta_{yx}}{\sigma_R}x\right)}{1+\exp\left(\frac{\beta_{yx}}{\sigma_R}x\right)} = \qquad (4)$$

$$\frac{\exp(b_{yx}x)}{1+\exp(b_{yx}x)} \Leftrightarrow \text{logit}(\Pr(y=1)) = b_{yx}x = \frac{\beta_{yx}}{\sigma_R}x$$

and in a similar way we obtain the full model

H_F^{Logit} :

$$\Pr(y=1) = \frac{\exp\left(b_{yx\cdot z}x + b_{yz\cdot x}z\right)}{1+\exp\left(b_{yx\cdot z}x + b_{yz\cdot x}z\right)} \qquad (5)$$

$$\Leftrightarrow \text{logit}(\Pr(y=1)) = b_{yx\cdot z}x + b_{yz\cdot x}z = \frac{\beta_{yx\cdot z}}{\sigma_F}x + \frac{\beta_{yz\cdot x}}{\sigma_F}z$$

where the scale parameters, σ_R and σ_F, are defined as above. We can immediately

see that the coefficients for x from the two models, b_{yx} and $b_{yx\cdot z}$, are identified relative

to the equation-specific scale parameters, and are therefore influenced by the

magnitude of the residual variance. While the above holds for the logit case, in the

probit case we assume that v in (2) is normally distributed and, because the standard

normal has a standard deviation of unity, we have $\sigma_v = \sigma_F$. In (4) and (5) we would

replace the logistic function with the cumulative normal to yield (as the counterpart to

(5)):

$$\Pr(y = 1) = \Phi\left[\frac{\beta_{yx.z}}{\sigma_F}x + \frac{\beta_{yz.x}}{\sigma_F}z\right]$$

and so the coefficients from the probit are identified relative to the scale parameters and will therefore be influenced by the residual variance in the same way as coefficients from the logit. Because of this equivalence between the logit and probit case, in what follows we develop our method only for the logit, but our results extend readily to the probit.

Two sources of change: Confounding and rescaling

From the logit models in (4) and (5) we see that, because we cannot estimate the variance of y^* (i.e., we only observe y as in (3)), a restriction is necessary for identifying the model. This means that we cannot estimate the regression coefficients of x in the underlying models in (1) and (2), but only

$$b_{yx} = \frac{\beta_{yx}}{\sigma_R}; \quad b_{yx \cdot z} = \frac{\beta_{yx \cdot z}}{\sigma_F}. \tag{6}$$

The estimated logit coefficients are equal to the underlying coefficients divided by the scale parameter. Therefore, controlling a variable of interest (x) for a confounding variable (z) that explains variation in the outcome variable (y) will alter the coefficient of interest as a result of both confounding and rescaling. Confounding occurs whenever x and z are correlated *and* z has an independent effect on y^* in the full model. Rescaling occurs because the model without the confounding variable, z, has a different residual standard deviation than the model that includes the confounding variable; indeed, because we always explain at least as much variation in the full model as in the reduced model, it holds that $\sigma_R \geq \sigma_F$. The logit coefficients of x are

therefore measured on different scales, except for the situation where z has no effect on y^*.

When employing the logit model, we are interested in the difference between the underlying coefficients, β_{yx} and $\beta_{yx \cdot z}$ in (1) and (2), because this difference is the result of confounding only (and not rescaling). However, because we only observe the coefficients in (6) we cannot distinguish between changes in b_{yx} compared to $b_{yx \cdot z}$ due to rescaling and to confounding:

$$b_{yx} - b_{yx \cdot z} = \frac{\beta_{yx}}{\sigma_R} - \frac{\beta_{yx \cdot z}}{\sigma_F} \neq \beta_{yx} - \beta_{yx \cdot z}. \tag{7}$$

Researchers using the naïve difference on the left hand side of (7) as a proxy for the difference on the right hand side, will generally underestimate the role played by confounding, because $\sigma_R \geq \sigma_F$. In certain circumstances, rescaling and confounding will operate in opposite directions such that the difference on the left hand side of (7) is zero, which may lead to the (incorrect) impression that z does not mediate or confound the effect of x. Researchers may also incorrectly report a suppression effect, which is not a result of confounding (i.e., x and z are uncorrelated), but only a result of rescaling (i.e., z has an independent effect on y). The inequality stated in (7) may be known by most sociologists specializing in quantitative methods but the sociological literature is replete with examples in which the naïve comparison is made and interpreted as though it reflected pure confounding.[5]

---

[5] Sociologists sometimes prefer not to invoke a latent variable model to motivate the logit model. However, the problem in comparing coefficients across nested logit models stated in (7) also applies when the models are derived without assuming a latent outcome variable. We show this in the Appendix.

The solution

Now we present a method that overcomes the cross-model coefficient comparability problem. Let $\tilde{z}$ be a set of x-residualized z-variables such that their correlation with x is zero: in other words, $\tilde{z}$ is the residual from a regression of z on x.[6] $\tilde{z}$ will have mean zero and we specify a new latent linear model:

$$H_F^* : y^* = \beta_{yx.\tilde{z}}x + \beta_{y\tilde{z}.x}\tilde{z} + k, \quad sd(k) = \sigma_F^*\pi / \sqrt{3} \tag{8}$$

Compared to the full latent linear model in (2), which includes the control variables (z), the model in (8) includes instead the x-residualized counterparts, $\tilde{z}$. Because models (1), (2), and (8) are linear it follows, by the principles of population moments, that

$$\beta_{yx} = \beta_{yx.\tilde{z}} \tag{9}$$

That is, taking a linear regression with a predictor, x, and adding a variable uncorrelated with x will leave the coefficient for x unchanged. We want to show that the residual standard deviation of H_F equals the residual standard deviation of H_F^*:

$$\sigma_F = \sigma_F^* \tag{10}$$

and this is equivalent to demonstrating that models H_F and H_F^* return the same fitted values.[7] We show this for the general case in which $\mathbf{z}$, rather than being a scalar, is a vector of control variables. We define $\tilde{\mathbf{z}} = \mathbf{z} - \mathbf{\theta}x$ where $\mathbf{\theta}$ is the vector of parameters from the regressions of the individual elements of $\mathbf{z}$ on x.

---

[6] Here, and throughout the paper, z is a continuous variable, but the method can be applied when z is nominal or ordinal. In both cases we define one or more dummy variables (using the same omitted category for them all) and use OLS to regress each of them on x to generate a set of residualized z variables.

[7] We can write the residual standard deviation as $sd(y \mid \hat{y})$ and so if two models for the same outcome variable have the same fitted values they will have the same residual standard deviation.

We need to demonstrate that

$$\beta_{yx.z}x + \beta_{yz.x}z = \beta_{yx.\tilde{z}}x + \beta_{y\tilde{z}.x}\tilde{z}$$

Given (9) we can write the right hand side of the above as:

$$\beta_{yx}x + \beta_{y\tilde{z}.x}\tilde{z}$$

$$= (\beta_{yx.z} + \beta_{yz.x}\theta)x + \beta_{y\tilde{z}.x}\tilde{z}$$

This follows from the rules of path analysis for linear models (Goldberger 1991): we have expanded the total effect of x, β_{yx}, into its direct and indirect paths. Given the definition of a partial regression coefficient, we have $\beta_{y\tilde{z}.x} = \beta_{yz.x}$ and so

$$(\beta_{yx.z} + \beta_{yz.x}\theta)x + \beta_{yz.x}\tilde{z}$$

$$= (\beta_{yx.z} + \beta_{yz.x}\theta)x + \beta_{yz.x}(z - \theta x)$$

$$= \beta_{yx.z}x + \beta_{yz.x}z$$

Given the equality in (10), we see that H_F^* in (8) is a re-parameterization of H_F in (2), i.e., they reflect the same latent linear model.

We now rewrite the latent linear model in (8) into a corresponding logit model. We employ the same strategy defined by (3), (4), and (5) and obtain the following:[8]

$$H_F^{\text{Logit*}} : \text{logit}(\Pr(y=1)) = b_{yx \cdot \tilde{z}}x + b_{y\tilde{z} \cdot x}\tilde{z} = \frac{\beta_{yx \cdot \tilde{z}}}{\sigma_F^*}x + \frac{\beta_{y\tilde{z} \cdot x}}{\sigma_F^*}\tilde{z}. \tag{11}$$

---

[8] The probit is trivially similar: $\Pr(y=1) = \Phi\left[\dfrac{\beta_{yx \cdot \tilde{z}}}{\sigma_F^*}x + \dfrac{\beta_{y\tilde{z} \cdot x}}{\sigma_F^*}\tilde{z}\right]$ and, in equation (8), k is drawn from a normal distribution with standard deviation σ_F^*.

Similar to the linear case, this model is a re-parameterization of H_F^{Logit}, i.e., the two models have the same fit to the data. We can exploit the equalities in (9) and (10) and the specifications of the logit models to overcome the comparison issue encountered in (7). In other words, we can make an unbiased comparison of the coefficients of x without and with a confounding variable, z, in the model.

Our method is not only a solution to the coefficient comparability issue stated in (7), it also a solution to the problem mentioned earlier, that the distributions of the errors of the reduced and full model will differ, leading to cross-model differences in how closely the error term approximates the assumed logistic distribution. These differences may produce differences in coefficient estimates across models over and above those produced by confounding and rescaling. But our method is free of this problem because we do not draw comparisons between two differently specified models: instead we compare across reparameterizations of the same model, both of which have not just the same scale parameter but exactly the same error distribution.[9] We will see the advantages of this in our Monte Carlo analyses when we compare our method with alternatives.

We propose three measures of coefficient change that hold rescaling constant (i.e., that measure confounding net of rescaling). The first is a *difference measure*:

$$b_{yx\cdot\tilde{z}} - b_{yx\cdot z} = \frac{\beta_{yx\cdot\tilde{z}}}{\sigma_F^*} - \frac{\beta_{yx\cdot z}}{\sigma_F} = \frac{\beta_{yx}}{\sigma_F} - \frac{\beta_{yx\cdot z}}{\sigma_F} = \frac{\beta_{yx} - \beta_{yx\cdot z}}{\sigma_F} \ , \tag{12a}$$

---

[9] Because, as we have demonstrated, the linear predictors of the two models equal each other, $\beta_{yx\cdot z} x + \boldsymbol{\beta}_{yz\cdot x}\mathbf{z} = \beta_{yx\cdot\tilde{z}} x + \boldsymbol{\beta}_{y\tilde{z}\cdot x}\tilde{\mathbf{z}}$, we find that $y^* = \beta_{yx\cdot\tilde{z}} x + \boldsymbol{\beta}_{y\tilde{z}\cdot x}\tilde{\mathbf{z}} + k = \beta_{yx\cdot z} x + \boldsymbol{\beta}_{yz\cdot x}\mathbf{z} + k$, implying that the error terms on both sides of the final equality have the same distribution.

where the first equality follows from (6) and the second from (9) and (10). This result tells us that the difference between the two logit coefficients of x in $H_F^{\text{Logit*}}$ and H_F^{Logit} stated in (12a) measures the impact of confounding given the scale of the full model. The difference in (12a) is a logit coefficient and like usual logit coefficients it is identified only up to scale. It measures the *change* in the coefficient of x attributable to confounding due to the inclusion of z, conditional on the full model holding true. Since we usually prefer basing our inference on the full model rather than the reduced model (see Clogg, Petkova, and Haritou 1995), this is an important result.

The second measure is a *ratio measure,* which is a scale free measure of confounding net of coefficient rescaling:

$$\frac{b_{yx.z}}{b_{yx.\tilde{z}}} = \frac{\dfrac{\beta_{yx.z}}{\sigma_F^*}}{\dfrac{\beta_{yx.\tilde{z}}}{\sigma_F^*}} = \frac{\dfrac{\beta_{yx.z}}{\sigma_F}}{\dfrac{\beta_{yx.\tilde{z}}}{\sigma_F}} = \frac{\beta_{yx.z}}{\beta_{yx.\tilde{z}}} = \frac{\beta_{yx.z}}{\beta_{yx}} \tag{12b}$$

In other words, the ratio between the two logit coefficients of x in $H_F^{\text{Logit*}}$ and H_F^{Logit} measures the impact of confounding net of the rescaling. In (12b) the scale parameter disappears, making it a scale free measure of coefficient change. A third measure, which we believe has considerable practical relevance, is the *percentage change* in the coefficients attributable to confounding, net of scaling:

$$\frac{(b_{yx.\tilde{z}} - b_{yx.z})}{b_{yx.\tilde{z}}} \times 100\% = \frac{\dfrac{\beta_{yx.\tilde{z}} - \beta_{yx.z}}{\sigma_F}}{\dfrac{\beta_{yx.\tilde{z}}}{\sigma_F^*}} \times 100\% = \frac{\beta_{yx.\tilde{z}} - \beta_{yx.z}}{\beta_{x|\tilde{z}}} \times 100\% = \frac{\beta_{yx} - \beta_{yx.z}}{\beta_{yx}} \times 100\%$$
$$\tag{12c}$$

The choice of the scale dependent difference measure stated in (12a) or the scale free ratio and percentage change measures given in (12b) and (12c) should

depend on the objective of the research because they have different interpretations. The difference measure in (12a) has the same properties as a normal logit coefficient and can be treated as such: researchers interested in the coefficients of the logit model might therefore prefer this measure. The ratio and percentage change measures have a different interpretation, because they are concerned with the regression coefficients in the latent model. The ratio in (12b) and the percentage change in (12c) measure change in the underlying partial effects on the latent propensity rather than in the logit coefficients. Thus, if researchers are interested in the underlying partial effects they should probably prefer to use these scale free measures.

If, in addition, we want to know the magnitude of rescaling net of the impact of confounding, we need to know the relation $\frac{\sigma_R}{\sigma_F}$, which is equal to the ratio between the error standard deviations in the reduced and full model. Given (9) and (10), we find that

$$\frac{b_{yx\cdot\tilde{z}}}{b_{yx}} = \frac{\dfrac{\beta_{yx\cdot\tilde{z}}}{\sigma_F^*}}{\dfrac{\beta_{yx}}{\sigma_R}} = \frac{\dfrac{\beta_{yx}}{\sigma_F^*}}{\dfrac{\beta_{yx}}{\sigma_R}} = \frac{\sigma_R}{\sigma_F}. \tag{13}$$

In other words, the ratio between the two observed coefficients of x in H_R^{Logit} and H_F^{Logit*} measures the impact of rescaling, net of confounding. Because $\sigma_R \geq \sigma_F$, we know that $\frac{b_{yx\cdot\tilde{z}}}{b_{yx}} \geq 1$. From (12b) and (13) we have the ratio decomposition of the observed change in the coefficient for x across nested models:

$$\frac{b_{yx.z}}{b_{yx}} = \frac{b_{yx.z}}{b_{yx.\tilde{z}}} \times \frac{b_{yx.\tilde{z}}}{b_{yx}} \tag{14a}$$

where the first term on the right hand side captures confounding and the second captures rescaling itself. Similarly we derive the additive decomposition:

$$b_{yx} - b_{yx.z} = \frac{\beta_{yx}}{\sigma_R} - \frac{\beta_{yx.z}}{\sigma_F} = \frac{\beta_{yx}}{\sigma_R} - \frac{\beta_{yx.\tilde{z}}}{\sigma_F^*} + \frac{\beta_{yx.\tilde{z}}}{\sigma_F^*} - \frac{\beta_{yx.z}}{\sigma_F} = [b_{yx} - b_{yx.\tilde{z}}] + [b_{yx.\tilde{z}} - b_{yx.z}], \quad (14b)$$

where the first term on the right hand side equals rescaling and the second captures confounding. Notice that the component induced by rescaling is equal to

$$b_{yx} - b_{yx.\tilde{z}} = \frac{\beta_{yx}}{\sigma_R} - \frac{\beta_{yx}}{\sigma_F}, \quad (15)$$

and thus captures the effect, on the logit coefficients, of the change in the residual standard deviation, holding constant the underlying coefficient.

Significance tests

The null hypothesis we want to test is that the effect of confounding, net of rescaling, is zero. To derive a formal statistical test of confounding we use the delta method, drawing on Sobel (1982, 1987). The null is:

$$b_{yx.\tilde{z}} = b_{yx.z} \quad (16)$$

and so we need to find the standard error of the difference between the coefficients in (16).

But since, as we noted earlier, $\beta_{yx.\tilde{z}} - \beta_{yx.z} = \beta_{yz.x}\theta_{zx}$ (where θ_{zx} is the coefficient from the linear regression of z on x), testing $b_{yx.\tilde{z}} - b_{yx.z} = 0$ amounts to testing $b_{yx.z}\theta_{zx} = 0$, because

$$b_{yx.\tilde{z}} - b_{yx.z} = \frac{\beta_{yx.\tilde{z}} - \beta_{yx.z}}{\sigma_F} = \frac{\beta_{yz.x}\theta_{zx}}{\sigma_F} = b_{yz.x}\theta_{zx} \qquad (17)$$

Since the hypothesis in (16) is equal to a hypothesis that the indirect effect of x on y is zero, either $b_{yz.x} = 0$ or $\theta_{zx} = 0$ is a sufficient condition for (16) to hold but neither is necessary.

We deal with the general case in which, instead of just one confounder, we have J confounders, $z_1, z_2, ..., z_J$. This means that we have J regressions, one for each element of $\mathbf{z}$ on x, with coefficient estimates in the vector $\boldsymbol{\theta}$. We also have J coefficients for the coefficient estimates of each of the z variables on y in the vector $\mathbf{b}$. For notational convenience we define $f = \sum_j b_{yz(j).x}\theta_{z(j)x} = \mathbf{b'\theta}$ so f is equal to the difference between the uncontrolled and controlled coefficient of x measured on the scale of the full model. We want to obtain the standard error of this difference. $\Sigma_{\mathbf{b\theta}}$ is defined as the variance-covariance matrix of the estimates of the coefficients $b_{yz.x}$ and θ_{zx}. This has the form $\begin{bmatrix} \Sigma_{\mathbf{b}} & 0 \\ 0 & \Sigma_{\boldsymbol{\theta}} \end{bmatrix}$ with $\Sigma_{\mathbf{b}}$ being the variance-covariance matrix of the estimates of $\mathbf{b}$ and $\Sigma_{\boldsymbol{\theta}}$ the variance-covariance matrix of the estimates of $\boldsymbol{\theta}$. The off-diagonal blocks of $\Sigma_{\mathbf{b\theta}}$ contain zeros because the estimates of $\mathbf{b}$ and $\boldsymbol{\theta}$ are asymptotically independent (Sobel 1982: 294-5). The covariances among the $\mathbf{b}$ coefficients are easily obtained since they come from the same regression (equation 5, but now with more than one z variable), but the $\boldsymbol{\theta}$ coefficients are estimated from J separate regressions. To compute the covariances among their estimates we use the following formula from Turner and Rockel (1988). Let z_k and z_l be two confounders,

with associated regression coefficients θ_k and θ_l. Then $\text{cov}(\hat{\theta}_k \hat{\theta}_l) = \dfrac{\sum_i \tilde{z}_{ik} \tilde{z}_{il} x^2}{\text{var}(x)^2}$, where

i indexes observations. We insert these quantities into the off-diagonal positions in Σ_θ while the diagonal elements of this sub-matrix are simply the variances of the estimates of the elements of θ.[10]

Finally, we define a to be the column vector of partial derivatives of the indirect effect with respect to the parameters, $\mathbf{b}$ and θ. That is, $\mathbf{a}$ is the $2J \times 1$ column vector whose entries are

$$\left[\frac{\partial f}{\partial b_1}, \frac{\partial f}{\partial b_2}, ..., \frac{\partial f}{\partial b_J}, \frac{\partial f}{\partial \theta_1}, \frac{\partial f}{\partial \theta_2}, ..., \frac{\partial f}{\partial \theta_J} \right]$$

The asymptotic standard error of the indirect effect is given by

$$se(b_{yx.\tilde{z}} - b_{yx.z}) = \sqrt{\mathbf{a}' \Sigma_{\mathbf{b}\theta} a} \tag{18}$$

In the case where there is a single z, the asymptotic standard error in (18) is the square root of the following matrix expression

$$\left[\frac{\partial f}{\partial b} \frac{\partial f}{\partial \theta} \right] \begin{bmatrix} \sigma_b^2 & 0 \\ 0 & \sigma_\theta^2 \end{bmatrix} \begin{bmatrix} \dfrac{\partial f}{\partial b} \\ \dfrac{\partial f}{\partial \theta} \end{bmatrix} = b^2 \sigma_\theta^2 + \theta^2 \sigma_b^2 \tag{19}$$

We use (18) to test the hypothesis of whether change in the logit coefficient attributable to confounding, net of rescaling, is statistically significant, via the test statistic, Z_C, (where the subscript C denotes confounding) which, in large samples, will be normally distributed under the null:

---

[10] Σ_θ can also be obtained directly by estimating the regressions of each z on x as a set of seemingly unrelated regressions. Σ_θ is then the variance-covariance matrix of the estimates of the parameters of the system.

$$Z_C = \frac{b_{yx.\tilde{z}} - b_{yx.z}}{\sqrt{\mathbf{a}'\Sigma_{\mathbf{b}\theta}\mathbf{a}}} \qquad\qquad (20)$$

In other words, the statistic enables a direct test of the change in the logit coefficients that is attributable to confounding, net of rescaling.

A Monte Carlo study

In Table 1 we report the results of a Monte Carlo study to assess the performance of the statistical test across a range of different scenarios, reporting the percentage of times the null hypothesis is rejected.

The model used to generate the data is

$$y^* = x + \gamma z + u,$$
$$z = \theta x + 2v$$
$$y = 1 \text{ if } y^* > median(y^*), \ 0 \text{ otherwise}$$

Here, u has a standard logistic distribution, v has a standard normal distribution and x is drawn from a standard normal distribution (panel A of Table 1) or from a lognormal distribution (panel B).

We consider three situations in which the null is true: when both θ and γ are zero and when one is zero and the other non-zero. We also consider a further eight scenarios in which different combinations of these parameters yield situations with increasing confounding (measured as the ratio of the indirect to the total effect). For a variety of sample sizes we report how often, in 1000 replications, the null hypothesis is rejected, using a two-sided test and 0.05 significance level.

Table 1 shows that, as we would expect, increasing the sample size leads to a greater frequency of rejections of the null, but, when the null is true, our method's rejection rate is never more than five percent. Except in small samples and with a low

degree of confounding, for normal x and z the test performs well in rejecting a false null. For lognormal x and z drawn from a mixture of a normal and a lognormal distribution the test's power is lower for smaller sample sizes and less confounding. For smaller sample sizes and less confounding the test appears more sensitive to the size of θ (the effect of x on z) than to γ (the effect of z on y) . Whether the test is more sensitive to one or the other will depend on the relative sizes of the error variances of the linear model for z and of the latent linear model for y^*.

-- TABLE 1 HERE --

Comparing our method with other proposed solutions

In this section of the paper we compare our method with three other solutions to the scaling problem that have been proposed (see, e.g., Mood 2010:80): these are y-standardization, average partial effects derived from logit or probit models, and the linear probability model. Our comparisons are based on the performance of the methods in Monte Carlo simulations. Before we turn to the results of the simulations, we first briefly describe each of the three alternative methods.

Winship and Mare (1984) suggested y-standardization as a solution to the problem of making within sample comparisons of logit coefficients of x across models without and with control variables, z. The idea is to estimate the standard deviation of y^* using the formula given by McKelvey and Zavoina (1975)

$$SD(y^*) = \sqrt{\text{var}(x'\hat{b}) + \text{var}(u)} = \sqrt{\text{var}(x'\hat{b}) + \frac{\pi^2}{3}} \ . \tag{21}$$

for different nested models and then, for each model, the coefficient of *x* is divided by the estimated latent standard deviation, $SD(y^*)$. The calculated coefficients are thus y-standardized, because they compensate for the rescaling of the "non-standardized" coefficients. The y-standardized logit coefficient of *x* is $b^{sdY} = b / SD(y^*)$, and comparing y-standardized coefficients from different nested models is thus a potential method for making cross-model coefficient comparisons unaffected by rescaling.

In logit and probit models, the marginal effect, ME, of *x* is the derivative of the predicted probability with respect to *x*, given by (when *x* is continuous[11] and differentiable):

$$\frac{d\hat{p}}{dx} = \hat{p}(1-\hat{p})b = \frac{\hat{p}(1-\hat{p})}{\sigma}\beta, \qquad (24)$$

where $\hat{p} = \Pr(y=1\,|\,x)$ is the predicted probability given *x* and $b = \frac{\beta}{\sigma}$ is the logit coefficient of *x*. The ME of *x* is evaluated at some fixed values of the other predictor variables in the model, typically their means. But this implies that whenever we include control variables in a model we change the set of other variables at whose mean the ME is evaluated, so introducing indeterminacy into cross-model comparisons. We therefore ignore MEs in the following, and rather focus on the more general average partial effect, APE.

The APE of *x* is the derivative of the predicted probability with respect to *x* evaluated over the whole population. Let *x* be continuous and differentiable; then, using i (=1,…,N) to denote observations, we write the APE of *x* as:

---

[11] Whenever x is discrete, the ME is the difference(s) in expected probabilities for each discrete category. In this paper, we refer to the continuous case. The discrete case follows directly from these derivations.

$$\frac{1}{N}\sum_{i=1}^{N}\frac{d\hat{p}_i}{dx_i} = \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{p}_i(1-\hat{p}_i)}{\sigma}\beta.$$

(25)

Thus the APE is a weighted average of the marginal effects over the sample. If the sample is drawn randomly from the population, the APE estimates the average marginal effect of x in the population.[12] APEs have been suggested for cross-model comparisons because simulation studies show that they are unaffected by rescaling whenever x and z are uncorrelated (Cramer 2007).

Finally, the linear probability model (LPM) is simply an OLS regression of the dichotomous y on the predictor variables. In this model, the coefficients for x are unbiased estimates of the expected change in Pr(y=1) for a unit-change in x (either conditionally, if z is included, or unconditionally). Because of its simple interpretation, economists often prefer the LPM to the nonlinear counterparts such as the probit or logit. However, we know of no studies investigating the use of linear probability models for assessing confounding: our Monte Carlo study will be the first to show its performance.

To compare these three methods against the method proposed in this paper we conducted Monte Carlo analyses, based on the following model:

$$y^* = 1 + \beta x + \gamma z + e,$$

$$\rho(x,z) = r,\ sd(x) = sd(z) = 1,$$

$$y = 1 \text{ if } y^* > \tau, y = 0 \text{ otherwise}$$

In Table A in the Appendix we report the results from 120 simulations, each using 5,000 cases and each replicated 250 times.[13] Table A has five panels, in each of which

---

[12] For discrete x's, see Bartus (2005).
[13] The results reported here are robust to other specifications of the simulation procedure (not reported but available from authors). We find virtually identical results with N = 5,000 and 500 replications and,

we report simulations with six values of *r*: 0, 0.1, 0.3, 0.5, 0.7 and 0.9 and, for each value of *r*, we varied the cut-point, τ, letting it take the values 0.5, 0.7, 0.9 and 0.95 (the last corresponding to a very skewed *y*). The correlation and the cut-point are listed in columns 1 and 2 of each panel of the tables. Column 3 of each panel shows the true ratio of the direct (controlling for *z*) effect of *x* on *y** to the total effect of *x* on *y** (not controlling for *z*). This ratio is what we ideally want to recover with our method, APEs, y-standardization, or the LPM. Column 4 reports the average (over 250 replications) of the naïve estimate of this ratio based on the coefficients from two logit or probit models. Column 5 shows the average using the ratio of OLS regression coefficients derived from models using the latent outcome variable, *y**. The figures in column 5 are necessarily quite close to those of column 3 but, obviously, in a real application, we could not estimate either of them. Column 6 shows the mean estimate derived from our method and columns 7, 8, and 9 report, respectively, the mean estimates of the ratio based on the use of APEs, y-standardization, and the linear probability model.

In panel A of the table we set $\beta = 0.5$, $\gamma = 1$, *x* and *z* are draws from a normal distribution and *e* is a draw from a standard normal distribution. The reported results for the naïve approach (column 4), our method (column 6), APEs (column 7), and y-standardization (column 8) are based on the probit model. In panels B and C we set $\beta = 0.5$, $\gamma = 1$ and in both cases *e* is a draw from a standard logistic distribution and the results in columns 4 and 6 through 8 derive from logit models. Panels B and C differ, however, in the distributions of *x* and *z*. In B, *x* and *z* are both normal while in C

---

although less precise, the results from a study using N = 1,000 and 500 replications tell the same story as the results reported here.

they are both lognormal. Panels D and E replicate panels B and C except that we set $\beta = 1$, $\gamma = 0.5$, so confounding is weaker in D and E than in B and C. In all our simulations we fix the two predictor variables have a standard deviation of one.

These five sets of Monte Carlo results allow us to assess the performance of our approach both absolutely and relative to that of the other methods under different scenarios. Panel A addresses pure rescaling because the errors of both the full and reduced model have a normal distribution. The reason for this is that we fix e to be normal and, because the sum of normal variables is normal, excluding or including z in the model will not change the shape of the error distribution across models.[14] Panels B through E deal with cases in which the error distributions of the full and reduced models are different although the full model always has the true (logistic in this case) distribution. B and C differ from D and E in the extent of confounding, while B and D differ from C and E in the degree to which the error distribution of the reduced model departs from the logistic. In B and D this error is a mixture of logistic and normal while in C and E it is a mixture of logistic and lognormal. Taken together the simulations allow us to evaluate the performance of these approaches under pure rescaling and in the presence of rescaling together with differences between models in their error distributions.

We can summarize the results with the aid of Table 2, which shows the mean absolute difference, across all the simulations in each panel, between the population ratio and the estimated ratio from the method in question. The most obvious result was the one we expected: the naïve estimator performs very poorly, always overestimating the ratio (and thus underestimating the extent of true confounding) and

---

[14] This will not be true in the logit models where the error term of the reduced model will be a mixture of the distributions of z and of the error of the full model, and consequently not logistically distributed.

performing more poorly when confounding is greater (panels B and C compared with D and E). As far as our method is concerned, it always returns estimates very close to the true ratio. It is unaffected by rescaling (panel A) and it is also unaffected by differences in the error distributions of the reduced and full models (panels B through E). The reason for this is that our method compares the full model with a reparameterization of the full model, thereby holding the error distribution constant across the full and reduced models. Neither y-standardization, APEs nor the LPM are sensitive to rescaling but, as the difference between full and reduced error distributions becomes greater, so they perform less well. In particular, in simulations C and E where the error of the reduced model is the sum of a logistic distribution and a mixture of lognormals (which is the distribution of z in this case) all except our method perform poorly, and APEs and y-standardization perform more poorly the larger the effect of the confounding variable.[15]

-- TABLE 2 HERE --

Applying our method to APEs

Using APEs for assessing confounding may return noticeably different results to those produced by our method. However, by applying our method to APEs we obtain a measure which may, for many purposes, be more useful than one based on logit

---

[15] In real applications, of course, we have to make an assumption about the distribution of the error, since we cannot know what it is in either the full or reduced model. Previous work suggests that the logit and probit are fairly robust against mis-specification of the error (e.g. Cramer 2007: 554) and simulations that we have undertaken (not reported but available from authors) show that our method is very good at recovering the true confounding ratio from a logit model even when the error distribution of the full model is not logistic.

coefficients.[16] Effects measured on the probability scale are intuitive, both for researchers and policy-makers. We can apply the method developed in this paper to APEs, thereby avoiding the problems shown in the Monte Carlo study and yielding measures based on the probability scale.

Using (25) to calculate the APE for two logit models involving z and $\tilde{z}$, respectively, and taking a ratio similar to (12b), we obtain the following

$$\frac{APE(x\,|\,z)}{APE(x\,|\,\tilde{z})} = \frac{\dfrac{1}{N}\sum_{i=1}^{N}\dfrac{\hat{p}_{yx \cdot z}(1-\hat{p}_{yx \cdot z})}{\sigma_F}\beta_{yx \cdot z}}{\dfrac{1}{N}\sum_{i=1}^{N}\dfrac{\hat{p}_{yx \cdot \tilde{z}}(1-\hat{p}_{yx \cdot \tilde{z}})}{\sigma_F}\beta_{yx}} = \frac{\beta_{yx \cdot z}}{\beta_{yx}}, \qquad (26)$$

which follows because of (9) and (10) and because $\hat{p}_{yx \cdot z} = \hat{p}_{yx \cdot \tilde{z}}$. While (26) casts the change in APEs in ratios as in (12b), we can easily derive the change in differences between APEs similar to (12a):

$$APE(x\,|\,\tilde{z}) - APE(x\,|\,z) = \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{p}_{yx \cdot z}(1-\hat{p}_{yx \cdot z})}{\sigma_F}\left(\beta_{yx} - \beta_{yx \cdot z}\right). \qquad (27)$$

Note that we cannot be certain that the difference in (27) will equal the difference we would normally calculate, namely $APE(x) - APE(x\,|\,z)$, because of the sensitivity of APEs to changes in the error distributions of the models. Applying our method to APEs, however, removes this uncertainty while also providing effect measures that might be considered more interpretable than logit coefficients.

Example

To illustrate the method of decomposing the change in logit coefficients into confounding and rescaling, we turn to an example using data from the National

---

[16] We could also apply our method to y-standardization, but this would be redundant, because the predicted logit index in either of the full models (with z or residualized z) is identical.

Education Longitudinal Study of 1988 (NELS88).[17] We examine how the effect of parental income on high school graduation changes when we control for student academic ability and parental educational attainment. We use NELS88 and our final sample consists of 8,173 students. The observed outcome variable is a dichotomy indicating whether the student completed high school (= 1) or not (= 0). This variable has a skewed distribution because 95.9 percent of the sample completed high school. The predictor variable of interest is a measure of yearly family income. Although the variable is measured on an ordered scale with 15 categories, for simplicity we use it here as a continuous variable. We include three control variables: these are student academic ability and the educational attainment of the mother and of the father.[18] We derive the ability measure from test scores in four different subjects using a principal component analysis.[19] We standardize the family income variable and the three control variables to have mean zero and variance of unity. We estimate five logistic models and report the results in Table 3.

In M1 we find a positive logit coefficient of 0.800 for the effect of family income on high school completion. Controlling for student academic ability in M2 reduces the effect to 0.460. A naïve comparison would thus suggest that the effect of family income on high school graduation declines by 100*(0.800-0.460)/0.800 = 42.5 percent once academic ability is controlled for. However, such a comparison conflates

---

[17] NELS88 is a sample of students who were in 8th grade in 1988 and who were re-interviewed in 1990, 1992, 1994, and 2000. We have relevant background information and information on the educational careers of the students. For a full description of the variables used in this example, see Curtin et al (2002).

[18] Parental education is coded in seven, ordered categories. To keep the example simple, we include father's and mother's education as continuous covariates, although a dummy-specification might have given a more precise picture of the relationship with the outcome variable.

[19] These tests are in reading, mathematics, science, and history. The variables are provided in the public use version of NELS88. The eigenvalue decomposition revealed one factor accounting for 78.1 percent of the total variation in the four items.

confounding and rescaling. To remedy this deficiency, we use the estimate of family income in M3, where we have included the residualized student academic ability measure, and compare it with the estimate in M2. We obtain a 100*(1.094-0.460)/1.094 = 58.0 percent reduction due to confounding, net of rescaling. We use the test-statistic in (19) to assess whether confounding is statistically significant and obtain $Z_C = 0.634/0.042 \approx 15.00$, leading us to conclude that controlling for academic ability reduces the effect of family income on high school completion.

-- TABLE 3 HERE --

In Table 3 we also report estimates from two further logistic models. M4 adds father's and mother's educational attainment and M5 includes the family income residualized counterparts of all three control variables. Comparing, naively, the effect of family income in M1 (0.800) and M4 (0.311) gives a reduction of 61.1 percent. However, using our method we would compare the effect of family income in M5 (1.349) and M4 (0.311). This suggests a substantially larger reduction of 77.0 percent. Using the formula in (20) we obtain a Z_C of 14.83 and thereby conclude that the reduction is statistically significant. Moreover, our method also provides us with an estimate of how much rescaling masks the change caused by confounding. Using the decomposition expressed in ratios in (14a):

$$\frac{b_{yx.z}}{b_{yx}} = \frac{b_{yx.z}}{b_{yx.\tilde{z}}} \times \frac{b_{yx.\tilde{z}}}{b_{yx}} \Rightarrow \frac{0.311}{0.800} = \frac{0.311}{1.349} \times \frac{1.349}{0.800} \Rightarrow 0.389 = 0.231 \times 1.686$$

Interpreted naively, adding the confounders reduces the effect of x to 0.39 of its original size. But this is the product of the true reduction caused by confounding

(which would reduce the effect to 0.23 of its original size) and the offsetting impact of rescaling that increases the naïve estimate by a factor of 1.69. In this case rescaling plays an important role in masking the change due to confounding.

In Table 4 we show the APEs from the same models as reported in Table 3.[20] In M1 we observe that a standard deviation increase in family income increases, on average, the probability of completing high school by 3.0 percentage points. Controlling for student academic ability in M2 changes the average effect to 1.6 percentage points, a reduction of 46.8 percent. However, using the specification in M3 returns a different result, namely a 58.0 percent reduction. This percentage reduction exactly equals that calculated with the logit coefficients in Table 3. In light of equation (26), this finding is what we would have expected. Moreover, the bias arising from using APEs for the decomposition is smaller than the bias arising from naively comparing logit coefficients. The naïve comparison using logit coefficients returned a 42.5 percent reduction, while the naïve use of APEs returned a 46.8 percent reduction. The naïve comparison based on APEs is thus closer to the reduction we find using our method (58.0 percent).

Turning to models M4 and M5 in Table 4, naïvely comparing the APE of family income in M1 and M4 returns a 63 percent reduction, while correctly comparing the APE in M5 and M4 returns a 76 percent reduction, equal to the reduction found when applying our method to the logit coefficients in Table 4. Moreover, conditional on M1 holding true, we would estimate that a standard deviation increase in family income would, on average, increase the probability of

---

[20] We use the user-written *margeff* command in Stata to calculate the APEs (Bartus 2005). One reviewer noted that earlier versions of *margeff* suffered from errors. We have used the most recent version. Moreover, using Stata version 11 command *margins* returns identical results.

completing high school by around 3 percentage points. However, conditional on M4 (the model which, in this example, we would take as the full model) holding true, the average effect is around 4.5 percentage points (see M5). Although the results point in the same direction, the effect sizes are quite different.

Similar to the decomposition of the naïve ratio of logit coefficients into confounding and rescaling, we can report an APE counterpart:

$$\frac{APE(x\,|\,z)}{APE(x)} = \frac{APE(x\,|\,z)}{APE(x\,|\,\tilde{z})} \times \frac{APE(x\,|\,\tilde{z})}{APE(x)} \Rightarrow \frac{0.011}{0.030} = \frac{0.011}{0.046} \times \frac{0.046}{0.030}$$

$$\Rightarrow 0.367 = 0.239 \times 1.534$$

From the decomposition we see that the ratio measuring confounding equals the one found with logit coefficients (the slight discrepancy—0.239 compared with 0.231—vanishes if we go to more decimal places than shown here). However, the distortion for APEs, 1.534, is less than for logit coefficients, 1.686.

-- TABLE 4 HERE --

Conclusion

Winship and Mare (1984) noted that logit coefficients are not directly comparable across same sample nested models because the logit fixes the error variance at an arbitrary constant. While the consequences of this identification restriction for the binary logistic model may be well known in the econometric literature, many

quantitative sociologists believe that confounding works the same way for the binary logit or probit regression model as for the linear regression model.

Naïve comparisons of logit coefficients across same sample nested models should be avoided because the rescaling problem may mask or underestimate the true change due to confounding. Rescaling will always tend to increase the apparent magnitude of the coefficient of a variable[21] and this commonly counteracts the effect of the inclusion of confounding variables, which are most often expected to reduce the effect of the variable of interest. Furthermore, the error distributions will almost always differ between the reduced and the full model. Our Monte Carlo analyses indicate that although y-standardization, APEs, and the LPM are effective in dealing with pure rescaling, in certain circumstances they are sensitive to differences in the error distribution across models. But the method we have presented in this paper performs at least as well as, and in some cases better than, others when faced with differences in the error distributions, because it holds constant, across equations, scaling and error distributions.. It permits the decomposition of change in the naïve coefficient into a part attributable to confounding and a part attributable to rescaling and we have provided a test statistic to assess the significance of confounding. It can also be applied to average partial effects to yield unbiased decompositions measured on the probability scale. We applied the approach to binary non-linear probability models, but both the problem and the solution also apply to cumulative probability models for ordinal data, such as the ordered logit and probit. This follows immediately because the latent variable in such models is defined in exactly the same way as for logits and probits: the models differ only in the degree to which the latent $y*$ is

---

[21] This happens when both $y*$ and x, $y*$ and z, and x and z are all positively correlated, e.g., when $y*$ is passing an educational threshold, x is some parental background characteristic, and z is cognitive ability.

observed. A Stata program called *khb* (http://fmwww.bc.edu/RePEc/bocode/k) is available to apply our method to all these cases and others.

References

Agresti, Alan. 2002. *Categorical Data Analysis*. *Second Edition*. New Jersey: Wiley & Sons.

Allison, Paul D. 1999. "Comparing Logit and Probit Coefficients Across Groups." *Sociological Methods & Research* 28(3):186-208.

Amemiya, Takeshi. 1975. "Qualitative Response Models." *Annals of Economic and Social Measurement* 4:363-388.

Bartus, Thamás. 2005. "Estimation of marginal effects using margeff." *Stata Journal* 5(3):309-329.

Blalock, Hubert M. 1979. *Social Statistics*, 2nd ed. rev. New York: McGraw-Hill.

Clogg, Clifford C., Eva Petkova, and Adamantios Haritou. 1995. "Statistical Methods for Comparing Regression Coefficients Between Models." *The American Journal of Sociology* 100(5):1261-1293.

Cramer, J.S. 2003. *Logit Models*. *From Economics and Other Fields*. Cambridge: Cambridge University Press.

Cramer, J.S. 2007. "Robustness of Logit Analysis: Unobserved Heterogeneity and Mis-specficied Disturbances." *Oxford Bulletin of Economics and Statistics* 69(4):545-555.

Curtin, Thomas R., Steven J. Ingels, Shiying Wu, and Ruth Heuer. 2002. *User's Manual. National Education Longitudinal Study of 1988: Base-Year to Fourth Follow-up Data File User's Manual (NCES 2002-323).* Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Goldberger, Arthur S. 1999. *A Course in Econometrics*. Cambridge, Mass.: Chapman & Hall.

Hoetker, Glenn. 2004. "Confounded coefficients: Accurately comparing logit and probit coefficients across groups." Working Paper.

——— . 2007. "The use of logit and probit models in strategic management research: Critical issues." *Strategic Management Journal* 28(4):331-343.

Kendall, Patricia and Paul L. Lazarsfeld. 1950. "Problems of Survey Analysis." Pp. 133-196 in *Continuities in Social Research*, edited by Merton, Robert K. and Paul L. Lazarsfeld. Glencoe, Illinois: The Free Press.

Lazarsfeld, Paul F. 1955. "The Interpretation of Statistical Relations as a Research Operation." Pp. 115-125 in *The Language of Social Research*, edited by Paul F. Lazarsfeld and Morris Rosenberg. Glencoe, Illinois: The Free Press.

Lazarsfeld, Paul F. 1958. "Evidence and Inference in Social Research." *Daedalus* 87(4):99-130.

Long, J. Scott. 1997: *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage.

Long, J.Scott. and Jeremy Freese. 2005. *Regression Models for Categorical Dependent Variables Using Stata*, 2nd ed. College Station: Stata Press.

Maddala, G. S. 1983. *Limited-Dependent Variables and Qualitative Variables in Economics*. New York: Cambridge University Press.

McKelvey, Richard D. and William Zavoina. 1975. "A Statistical Model for the Analysis of Ordinal Level Dependent Variables." *Journal of Mathematical Sociology* 4(1):103-120.

Mood, Carina. 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.

Powers, Daniel A. and Yu Xie. 2000. *Statistical Methods for Categorical Data Analysis*. San Diego: Academic Press.

Simon, Herbert A. 1954. "Spurious Correlation: a Causal Interpretation." *Journal of the American Statistical Association* 49(267):467-479.

Sobel, Michael E. 1982. "Asymptotic Confidence Intervals for Indirect Effects in Structural equation Models". *Sociological Methodology* 13: 290-312.

Sobel, Michael E. 1987. "Direct and Indirect Effects in Linear Structural Equation Models." *Sociological Methods and Research* 16(1):155-176.

Turner, Robert W. and Mark L. Rockel. 1988. "Estimating Covariances of Parameter Estimates From Different Models". *Economics Letters* 26(2): 137-40.

Yatchew, Adonis and Zvi Griliches. 1985. "Specification Error in Probit Models." *The Review of Economics and Statistics* 67(1):134-139.

White, Halbert. 1982. "Maximum likelihood estimation of misspecified models." *Econometrica* 50(1):1–25.

Williams, Richard. 2009. "Using Heterogeneous Choice Models to Compare Logit and Probit Coefficients Across Groups." *Sociological Methods & Research* 37(4):531-559.

Winship, Christopher and Robert D. Mare. 1984. "Regression Models with Ordinal Variables." *American Sociological Review* 49(4):512-525.

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

TABLES

Table 1. Monte Carlo study of power of proposed test: percentage of 1000 replications in which the test rejects H₀. Two-sided test with 5 percent significance level, $\beta = 1$ throughout

A: x normal, z normal

Scenario	θ	γ	Null true?	Percent of confounding*	% times H₀ rejected N = 200	N=500	N=1,000	N=5,000
A	0.0	0.0	Yes	0	0	0	0	0
B	1.0	0.0	Yes	0	3	5	5	4
C	0.0	1.0	Yes	0	3	5	5	5
D	0.2	0.2	No	4	6	44	85	100
E	0.2	0.6	No	11	24	59	89	100
F	0.6	0.2	No	11	55	97	100	100
G	0.5	0.5	No	20	91	100	100	100
H	0.7	0.7	No	33	100	100	100	100
I	1.0	1.0	No	50	100	100	100	100
J	1.5	3.0	No	82	100	100	100	100
K	3.0	1.5	No	82	100	100	100	100

B: x lognormal, z normal-lognormal mixture

Scenario	θ	γ	Null true?	Percent of confounding*	% times H₀ rejected N = 200	N=500	N=1,000	N=5,000
A	0.0	0.0	Yes	0	0	0	0	0
B	1.0	0.0	Yes	0	1	3	3	5
C	0.0	1.0	Yes	0	3	5	5	5
D	0.2	0.2	No	4	1	8	19	83
E	0.2	0.6	No	11	5	14	24	84
F	0.6	0.2	No	11	7	59	95	100
G	0.5	0.5	No	20	22	61	89	100
H	0.7	0.7	No	33	45	88	99	100
I	1.0	1.0	No	50	75	99	100	100
J	1.5	3.0	No	82	97	100	100	100
K	3.0	1.5	No	82	100	100	100	100

Note: * Size of indirect effect relative to the total effect.

Table 2. Mean absolute error (multiplied by 1000) in Monte Carlo simulations in Appendix Table A1.

	naïve	Our method	APE	y-stand	LPM
Panel					
A	213.1968	1.2612	1.3717	1.3264	1.3673
B	72.3232	0.9511	1.0126	20.4050	0.9534
C	75.3598	1.2180	18.5213	19.5937	49.8109
D	24.4546	1.0647	1.1610	5.6495	0.9314
E	28.4790	1.2792	5.5756	5.4452	51.8660
Mean	82.7627	1.1548	5.5284	10.4840	20.9858

Table 3. Controlling the effect of family income on high school graduation. Logit coefficients (standard errors in parenthesis)

	M1	M2	M3	M4	M5
Controls	None	z	$\tilde{z}$	z	$\tilde{z}$
Family income	0.800	0.460	1.094	0.311	1.349
	(0.047)	(0.052)	(0.058)	(0.057)	(0.073)
Academic ability		1.659	1.659	1.508	1.508
		(0.101)	(0.101)	(0.104)	(0.104)
Father's education				0.642	0.642
				(0.125)	(0.125)
Mother's education				0.340	0.340
				(0.106)	(0.106)
Intercept	3.473	4.446	4.446	4.796	4.796
	(0.070)	(0.126)	(0.126)	(0.147)	(0.147)
Pseudo-R^2	0.098	0.237	0.237	0.265	0.265
LogL	-1272.6	-1076.7	-1076.7	-1037.3	-1037.3

N = 8,173

Table 4. APE counterparts to logit coefficients in Table 3 (standard errors in parenthesis)

	M1	M2	M3	M4	M5
Controls	None	z	$\tilde{z}$	z	$\tilde{z}$
Family income	0.030 (0.002)	0.016 (0.002)	0.038 (0.002)	0.011 (0.002)	0.046 (0.003)
Academic ability		0.058 (0.004)	0.058 (0.004)	0.052 (0.004)	0.052 (0.004)
Father's education				0.022 (0.004)	0.022 (0.004)
Mother's education				0.012 (0.004)	0.012 (0.004)

N = 8,173

Appendix

Coefficient comparisons without assuming a latent outcome variable

The problem in comparing coefficients across nested logit models also applies when the logit model is derived without assuming a latent outcome variable. The problem arises because the logit indices of the reduced and full models differ. To see this, we use the rules for approximating the variance of a function. Let $g(.)$ be the logistic link function, let y be a binary outcome variable, and let x and z be centered on their respective means. For the reduced model we obtain $\text{var}(y) = \text{var}\left(g^{-1}(a_R + b_R x)\right) + \pi^2/3$, which approximates $k_R b_R^2 \text{var}(x) + \pi^2/3$, where

$$k_R = \frac{\exp(a_R + b_R E(x))}{\left(1 + \exp(a_R + b_R E(x))\right)^2} = \frac{\exp(a_R)}{\left(1 + \exp(a_R)\right)^2} = p_R(1 - p_R).$$ For the full model we

obtain the approximation $k_F\left(b_F^2 \text{var}(x) + c_F^2 \text{var}(x) - 2b_F c_F \text{cov}(x,z)\right) + \pi^2/3$, where

$$k_F = \frac{\exp(a_F + b_F E(x) + c_F E(z))}{\left(1 + \exp(a_F + b_F E(x) + c_F E(z))\right)^2} = \frac{\exp(a_F)}{\left(1 + \exp(a_F)\right)^2} = p_F(1 - p_F).$$

The ratio between the two variances is

$$m = \frac{k_R b_R^2 \text{var}(x) + \pi^2/3}{k_F\left(b_F^2 \text{var}(x) + c_F^2 \text{var}(x) - 2b_F c_F \text{cov}(x,z)\right) + \pi^2/3},$$ which approximates the

proportional reduction in the variance caused by the inclusion of the confounder z and is always less than or equal to one. It follows that $b^R \approx \left(b^F + c^F \theta\right)\sqrt{m}$, where θ is obtained from a linear regression of z on x. We can thus restate the difference in (7) in the main text as $b^R - b^F = \left(b^F + c^F \theta\right)\sqrt{m} - b^F$, from which we find that the difference in coefficients across models reflects, at least approximately, not only confounding

(captured by $c^F\theta$), but also the change in the variance of the logit index (captured by

$\sqrt{m}$).

Simulation results

Table 1: Simulation results (ratio of partial effect of *x* to gross effect) under the latent
model: $y^* = 1 + \beta x + \gamma z + e$, N=5,000, 250 reps.

Panel A: Probit model, *e*, *x* and *z* all *N(0,1)*, $\beta = 0.5$, $\gamma = 1$.

r(x,z)	Cut-point	True ratio	Naïve	Latent	Our method	APE	y-stand	LPM
0.0000	0.5000	1.0007	1.4145	1.0007	1.0008	1.0011	1.0010	1.0007
0.0017	0.7000	0.9973	1.4096	0.9973	0.9972	0.9959	0.9962	0.9972
0.0009	0.9000	0.9989	1.4144	0.9989	0.9990	0.9991	0.9989	0.9990
0.0007	0.9500	0.9994	1.4158	0.9994	0.9994	1.0008	0.9994	0.9993
0.1000	0.5000	0.8333	1.1763	0.8337	0.8343	0.8342	0.8342	0.8343
0.1000	0.7000	0.8333	1.1747	0.8330	0.8334	0.8336	0.8332	0.8335
0.1000	0.9000	0.8333	1.1748	0.8335	0.8322	0.8327	0.8325	0.8323
0.1000	0.9500	0.8333	1.1736	0.8336	0.8320	0.8305	0.8309	0.8319
0.3000	0.5000	0.6250	0.8637	0.6235	0.6236	0.6242	0.6240	0.6231
0.3000	0.7000	0.6250	0.8618	0.6255	0.6242	0.6239	0.6239	0.6243
0.3000	0.9000	0.6250	0.8649	0.6251	0.6242	0.6249	0.6246	0.6241
0.3000	0.9500	0.6250	0.8628	0.6239	0.6241	0.6236	0.6237	0.6240
0.5000	0.5000	0.5000	0.6623	0.4996	0.5003	0.5002	0.5003	0.4999
0.5000	0.7000	0.5000	0.6601	0.4995	0.4986	0.4989	0.4988	0.4988
0.5000	0.9000	0.5000	0.6611	0.4999	0.4992	0.4998	0.4994	0.4986
0.5000	0.9500	0.5000	0.6580	0.5004	0.4951	0.4958	0.4953	0.4961
0.7000	0.5000	0.4167	0.5131	0.4174	0.4174	0.4179	0.4176	0.4178
0.7000	0.7000	0.4167	0.5125	0.4162	0.4174	0.4176	0.4174	0.4170
0.7000	0.9000	0.4167	0.5115	0.4176	0.4164	0.4162	0.4163	0.4167
0.7000	0.9500	0.4167	0.5087	0.4151	0.4122	0.4128	0.4124	0.4134
0.9000	0.5000	0.3571	0.3880	0.3555	0.3557	0.3558	0.3557	0.3562
0.9000	0.7000	0.3571	0.3883	0.3564	0.3561	0.3561	0.3561	0.3558
0.9000	0.9000	0.3571	0.3868	0.3548	0.3544	0.3545	0.3545	0.3550
0.9000	0.9500	0.3571	0.3845	0.3573	0.3532	0.3528	0.3530	0.3483

Panel B: Logit model, x and z are $N(0,1)$, e is standard logistic, $\beta = 0.5$, $\gamma = 1$.

r(x,z)	Cut-point	True ratio	Naïve	Latent	Our method	APE	y-stand	LPM
0.0000	0.5000	1.0007	1.2056	1.0007	1.0007	1.0004	1.0525	1.0007
0.0012	0.7000	0.9984	1.1819	0.9984	0.9983	0.9983	1.0333	0.9983
-0.0008	0.9000	1.0022	1.1093	1.0022	1.0021	1.0021	0.9730	1.0021
-0.0006	0.9500	1.0021	1.0736	1.0022	1.0019	1.0050	0.9426	1.0019
0.1000	0.5000	0.8333	0.9993	0.8327	0.8321	0.8321	0.8727	0.8322
0.1000	0.7000	0.8333	0.9850	0.8324	0.8321	0.8322	0.8600	0.8322
0.1000	0.9000	0.8333	0.9233	0.8331	0.8318	0.8330	0.8105	0.8318
0.1000	0.9500	0.8333	0.8907	0.8330	0.8317	0.8312	0.7832	0.8320
0.3000	0.5000	0.6250	0.7387	0.6257	0.6246	0.6248	0.6498	0.6245
0.3000	0.7000	0.6250	0.7294	0.6256	0.6251	0.6249	0.6424	0.6250
0.3000	0.9000	0.6250	0.6882	0.6250	0.6235	0.6240	0.6110	0.6236
0.3000	0.9500	0.6250	0.6653	0.6246	0.6215	0.6218	0.5925	0.6216
0.5000	0.5000	0.5000	0.5762	0.5017	0.5022	0.5021	0.5168	0.5022
0.5000	0.7000	0.5000	0.5698	0.5008	0.5011	0.5010	0.5114	0.5007
0.5000	0.9000	0.5000	0.5432	0.5010	0.4994	0.4996	0.4925	0.4996
0.5000	0.9500	0.5000	0.5302	0.4991	0.4995	0.4994	0.4827	0.4998
0.7000	0.5000	0.4167	0.4584	0.4155	0.4164	0.4165	0.4238	0.4162
0.7000	0.7000	0.4167	0.4542	0.4163	0.4154	0.4154	0.4206	0.4149
0.7000	0.9000	0.4167	0.4431	0.4178	0.4168	0.4172	0.4137	0.4162
0.7000	0.9500	0.4167	0.4351	0.4156	0.4166	0.4170	0.4086	0.4172
0.9000	0.5000	0.3571	0.3714	0.3595	0.3582	0.3581	0.3602	0.3575
0.9000	0.7000	0.3571	0.3705	0.3610	0.3581	0.3581	0.3596	0.3575
0.9000	0.9000	0.3571	0.3639	0.3523	0.3554	0.3554	0.3546	0.3560
0.9000	0.9500	0.3571	0.3618	0.3554	0.3560	0.3558	0.3537	0.3540

Panel C: Logit model, *x* lognormal with sd = 1, *z* is mixture of lognormals with sd = 1 and e is standard logistic, $\beta = 0.5$, $\gamma = 1$.

r(x,z)	Cut-point	True ratio	Naïve	Latent	Our method	APE	y-stand	LPM
-0.0004	0.5000	1.0015	1.0801	1.0017	1.0018	0.9803	0.9486	1.0010
0.0003	0.7000	1.0003	1.1176	1.0002	1.0003	0.9739	0.9808	0.9997
-0.0003	0.9000	1.0013	1.1789	1.0012	1.0010	0.9618	1.0298	1.0008
-0.0002	0.9500	1.0011	1.2179	1.0011	1.0011	0.9510	1.0630	1.0015
0.1000	0.5000	0.8333	0.8980	0.8330	0.8327	0.8147	0.7913	0.8582
0.1000	0.7000	0.8333	0.9285	0.8326	0.8318	0.8101	0.8154	0.8543
0.1000	0.9000	0.8333	0.9804	0.8336	0.8330	0.8013	0.8570	0.8331
0.1000	0.9500	0.8333	1.0069	0.8336	0.8324	0.7872	0.8785	0.8171
0.3000	0.5000	0.6250	0.6712	0.6241	0.6249	0.6117	0.5984	0.6521
0.3000	0.7000	0.6250	0.6912	0.6253	0.6227	0.6071	0.6127	0.6433
0.3000	0.9000	0.6250	0.7280	0.6254	0.6277	0.6044	0.6427	0.6292
0.3000	0.9500	0.6250	0.7471	0.6255	0.6264	0.5952	0.6569	0.6109
0.5000	0.5000	0.5000	0.5319	0.4985	0.4995	0.4900	0.4837	0.4727
0.5000	0.7000	0.5000	0.5447	0.5020	0.4986	0.4867	0.4932	0.4746
0.5000	0.9000	0.5000	0.5679	0.4995	0.5000	0.4846	0.5100	0.4919
0.5000	0.9500	0.5000	0.5794	0.4999	0.4991	0.4790	0.5184	0.5022
0.7000	0.5000	0.4167	0.4360	0.4171	0.4152	0.4095	0.4086	0.2767
0.7000	0.7000	0.4167	0.4444	0.4169	0.4152	0.4084	0.4140	0.3040
0.7000	0.9000	0.4167	0.4590	0.4163	0.4190	0.4114	0.4252	0.4046
0.7000	0.9500	0.4167	0.4606	0.4157	0.4141	0.4037	0.4245	0.4494
0.9000	0.5000	0.3571	0.3611	0.3542	0.3528	0.3506	0.3518	0.0022
0.9000	0.7000	0.3571	0.3673	0.3576	0.3565	0.3544	0.3570	0.1323
0.9000	0.9000	0.3571	0.3727	0.3585	0.3591	0.3578	0.3613	0.3596
0.9000	0.9500	0.3571	0.3706	0.3541	0.3562	0.3548	0.3589	0.4819

Panel D: Logit model, *x* and *z* are *N(0,1)*, e is standard logistic, $\beta = 1$, $\gamma = 0.5$.

r(x,z)	Cut-point	True ratio	Naïve	Latent	Our method	APE	y-stand	LPM
0.0006	0.5000	0.9998	1.0521	0.9997	0.9998	0.9997	1.0110	0.9998
0.0006	0.7000	0.9998	1.0481	0.9998	0.9997	0.9998	1.0078	0.9997
-0.0002	0.9000	1.0002	1.0305	1.0002	1.0002	1.0001	0.9944	1.0002
-0.0005	0.9500	1.0003	1.0225	1.0003	1.0003	1.0021	0.9884	1.0003
0.1000	0.5000	0.9524	1.0013	0.9521	0.9524	0.9523	0.9625	0.9524
0.1000	0.7000	0.9524	0.9977	0.9523	0.9521	0.9522	0.9594	0.9521
0.1000	0.9000	0.9524	0.9809	0.9524	0.9527	0.9526	0.9478	0.9527
0.1000	0.9500	0.9524	0.9721	0.9524	0.9527	0.9528	0.9412	0.9526
0.3000	0.5000	0.8696	0.9108	0.8699	0.8697	0.8699	0.8776	0.8697
0.3000	0.7000	0.8696	0.9077	0.8699	0.8711	0.8709	0.8764	0.8711
0.3000	0.9000	0.8696	0.8937	0.8696	0.8691	0.8691	0.8654	0.8694
0.3000	0.9500	0.8696	0.8860	0.8692	0.8694	0.8691	0.8603	0.8694
0.5000	0.5000	0.8000	0.8296	0.8001	0.7989	0.7986	0.8041	0.7987
0.5000	0.7000	0.8000	0.8281	0.7992	0.7999	0.7999	0.8038	0.7997
0.5000	0.9000	0.8000	0.8197	0.8007	0.8008	0.8008	0.7983	0.8007
0.5000	0.9500	0.8000	0.8131	0.8010	0.7986	0.7991	0.7927	0.7986
0.7000	0.5000	0.7407	0.7567	0.7395	0.7372	0.7373	0.7404	0.7375
0.7000	0.7000	0.7407	0.7578	0.7398	0.7402	0.7400	0.7423	0.7399
0.7000	0.9000	0.7407	0.7520	0.7405	0.7401	0.7398	0.7385	0.7393
0.7000	0.9500	0.7407	0.7506	0.7405	0.7420	0.7418	0.7384	0.7407
0.9000	0.5000	0.6897	0.6986	0.6905	0.6921	0.6922	0.6931	0.6921
0.9000	0.7000	0.6897	0.6993	0.6916	0.6933	0.6933	0.6940	0.6931
0.9000	0.9000	0.6897	0.6909	0.6911	0.6864	0.6865	0.6860	0.6874
0.9000	0.9500	0.6897	0.6964	0.6913	0.6934	0.6933	0.6923	0.6917

Panel E: Logit model, *x* lognormal with sd = 1, *z* is mixture of lognormals with sd = 1 and e is standard logistic, $\beta = 1,\ \gamma = 0.5$.

r(x,z)	Cut-point	True ratio	Naïve	Latent	Our method	APE	y-stand	LPM
0.0005	0.5000	0.9998	1.0293	0.9998	0.9998	0.9915	0.9935	0.9997
-0.0003	0.7000	1.0002	1.0403	1.0002	1.0002	0.9921	1.0023	1.0002
-0.0010	0.9000	1.0006	1.0518	1.0006	1.0006	0.9939	1.0110	1.0006
-0.0009	0.9500	1.0005	1.0566	1.0005	1.0005	0.9919	1.0139	1.0004
0.1000	0.5000	0.9524	0.9797	0.9524	0.9522	0.9444	0.9465	0.9372
0.1000	0.7000	0.9524	0.9905	0.9519	0.9517	0.9441	0.9537	0.9417
0.1000	0.9000	0.9524	1.0006	0.9524	0.9523	0.9464	0.9621	0.9521
0.1000	0.9500	0.9524	1.0039	0.9525	0.9522	0.9443	0.9642	0.9574
0.3000	0.5000	0.8696	0.8928	0.8690	0.8680	0.8617	0.8641	0.8168
0.3000	0.7000	0.8696	0.9018	0.8693	0.8692	0.8631	0.8712	0.8357
0.3000	0.9000	0.8696	0.9092	0.8699	0.8693	0.8641	0.8765	0.8674
0.3000	0.9500	0.8696	0.9140	0.8695	0.8698	0.8640	0.8799	0.8880
0.5000	0.5000	0.8000	0.8207	0.8007	0.8018	0.7969	0.7995	0.7018
0.5000	0.7000	0.8000	0.8230	0.7994	0.7980	0.7931	0.7994	0.7370
0.5000	0.9000	0.8000	0.8305	0.7995	0.8001	0.7971	0.8055	0.8018
0.5000	0.9500	0.8000	0.8331	0.8002	0.8002	0.7966	0.8069	0.8340
0.7000	0.5000	0.7407	0.7522	0.7399	0.7394	0.7361	0.7385	0.5822
0.7000	0.7000	0.7407	0.7570	0.7407	0.7406	0.7378	0.7420	0.6447
0.7000	0.9000	0.7407	0.7608	0.7417	0.7416	0.7405	0.7448	0.7484
0.7000	0.9500	0.7407	0.7632	0.7422	0.7435	0.7425	0.7470	0.8075
0.9000	0.5000	0.6897	0.6926	0.6907	0.6874	0.6861	0.6875	0.4274
0.9000	0.7000	0.6897	0.6893	0.6897	0.6829	0.6822	0.6837	0.5376
0.9000	0.9000	0.6897	0.6997	0.6917	0.6935	0.6937	0.6943	0.7313
0.9000	0.9500	0.6897	0.7004	0.6883	0.6946	0.6951	0.6952	0.8138