TECHNICAL WORKING PAPER SERIES

ECONOMETRIC METHODS FOR FRACTIONAL RESPONSE VARIABLES WITH AN APPLICATION TO 401(K) PLAN PARTICIPATION RATES

Leslie E. Papke
Jeffrey M. Wooldridge

Technical Working Paper No. 147

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 1993

The second author is grateful for financial support through a Sloan Research Fellowship. This paper is part of NBER's research program in Public Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #147
November 1993

ECONOMETRIC METHODS FOR FRACTIONAL RESPONSE VARIABLES WITH AN APPLICATION TO 401(K) PLAN PARTICIPATION RATES

ABSTRACT

We offer simple quasi-likelihood methods for estimating regression models with a fractional dependent variable and for performing asymptotically valid inference. Compared with log-odds type procedures, there is no difficulty in recovering the regression function for the fractional variable, and there is no need to use *ad hoc* transformations to handle data at the extreme values of zero and one. We also offer some new, simple specification tests by nesting the logit or probit function in a more general functional form. We apply these methods to a data set of employee participation rates in 401(k) pension plans.

Leslie E. Papke
Department of Economics, 0508
Michigan State University
Marshall Hall
East Lansing, MI 48824-1038
and NBER

Jeffrey M. Wooldridge
Department of Economics, 0508
Michigan State University
Marshall Hall
East Lansing, MI 48824-1038

1. Introduction

Suppose that a fractional variable y ($0 \leq y \leq 1$) is to be explained in terms of a 1xK vector of explanatory variables $\mathbf{x} = (x_1, x_2, \ldots, x_K)$, with the convention that $x_1 = 1$. Rarely does a model of the form

$$E(y|\mathbf{x}) = \beta_1 + \beta_2 x_2 + \ldots + \beta_K x_K = \mathbf{x}\beta, \qquad (1.1)$$

where β is a Kx1 vector, provide the best description of $E(y|\mathbf{x})$. The primary reason is that y is bounded between 0 and 1, so that the effects of any particular x_j cannot be constant throughout the range of $\mathbf{x}$ (unless the effect is zero or the range of x_j is very limited). To some extent this problem can be overcome by estimating a linear model in nonlinear functions of $\mathbf{x}$, but the predicted values from an OLS regression can never be guaranteed to lie in the unit interval. Thus, the drawbacks of linear models for fractional data are quite analogous to the drawbacks of the linear probability model for binary data.

In statistics and econometrics the most common alternative to (1.1) is to model the log-odds ratio as a linear function. If $0 < y < 1$, then a linear model for the log-odds ratio is

$$E\left(\log[y/(1-y)]\,|\,\mathbf{x}\right) = \mathbf{x}\beta. \qquad (1.2)$$

Equation (1.2) is attractive because $\log[y/(1-y)]$ can take on any real value as y varies between 0 and 1, so it is natural to model its population regression as a linear function. Nevertheless, there are two potential problems with (1.2), the first of which is well-known: Model (1.2) cannot strictly be true if y takes on the values 0 or 1 with positive probability. Consequently, given a sequence of independent observations $\{(\mathbf{x}_i, y_i): i=1,2,\ldots,N\}$, if any y_i equals 0 or 1 then an adjustment must be made to y_i before computing the log-odds ratio. When the y_i are proportions from a fixed number of groups with known group sizes, various adjustments are available in the literature. Then estimation of the log-odds model corresponds to Berkson's minimum chi-square method.

Unfortunately, the minimum chi-square method for a fixed number of categories is not applicable for certain economic problems. First, the fraction y may not be a proportion from a known group size -- for example, it could be the proportion of a geographic area -- in which case any adjustments to handle the extreme values of zero and one are suspect. Second, the number of categories -- which is essentially determined by the number of different values that the vector of explanatory variables can take on -- is often very large. In the empirical example we have in mind the fractional variable y_i is the proportion of eligible employees at firm i contributing to a 401(k) plan, and the primary explanatory variable of interest is the plan's matching rate (essentially a continuous variable). It seems more natural to treat such examples in a regression-type framework than in a minimum chi-square framework.

Another problem that arises is that (1.2) alone does not allow one to recover $E(y|\mathbf{x})$, which is our primary interest. Write (1.2) as

$$\log[y/(1-y)] = \mathbf{x}\beta + u, \quad E(u|\mathbf{x}) = 0. \tag{1.3}$$

Then

$$E(y|\mathbf{x}) = \int_{-\infty}^{\infty} \left(\frac{\exp(\mathbf{x}\beta + v)}{1 + \exp(\mathbf{x}\beta + v)}\right) f(v|\mathbf{x})dv, \tag{1.4}$$

where $f(\cdot|\mathbf{x})$ denotes the conditional density of u given $\mathbf{x}$. Even if u and $\mathbf{x}$ are assumed to be independent, so that $f(\cdot|\mathbf{x}) = f(\cdot)$,

$$E(y|\mathbf{x}) \neq \frac{\exp(\mathbf{x}\beta)}{1 + \exp(\mathbf{x}\beta)}, \tag{1.5}$$

although $E(y|\mathbf{x})$ can be estimated using, for example, Duan's (1983) smearing method. If u and $\mathbf{x}$ are not independent, (1.4) cannot be estimated without estimating $f(\cdot|\mathbf{x})$. This is either difficult or nonrobust, depending on whether a nonparametric or parametric approach is adopted. Given that $f(\cdot|\mathbf{x})$ is rarely of central importance it would be better to have methods of estimating $E(y|\mathbf{x})$ without having to estimate the density of u given $\mathbf{x}$.

Of course it is always possible to estimate $E(y|\mathbf{x})$ by directly assuming a particular distribution for y given $\mathbf{x}$ and estimating the parameters of the conditional distribution by maximum likelihood. One plausible distribution for fractional y is the beta distribution. Unfortunately, the estimates of $E(y|\mathbf{x})$ that one obtains are known not to be robust to distributional failure (this follows from Gourieroux, Monfort, and Trognon (1984); more on this below). And it is easy to see that in certain applications standard distributional assumptions can fail. For example, the beta distribution implies that each value in [0,1] is taken on with probability zero. This is difficult to justify in empirical applications where at least some portion of the sample is at the extreme values of zero and one. For example, in our application about 40 percent of the y_i take on the value one.

In the next section we discuss quasi-likelihood methods for estimating $E(y|\mathbf{x})$ directly. These estimators circumvent the problems raised above and are easily implemented. Section 3 covers methods that can be used when y_i is a proportion obtained from a known group of size n_i. Some new specification tests are offered in section 4, and section 5 contains an empirical application relating 401(k) plan participation rates to the generosity of the plan's matching rate and other firm characteristics.

2. Estimating the Conditional Mean Directly Using a Quasi-Likelihood

We assume the availability of an independent sequence of observations $\{(x_i,y_i):i=1,2,\ldots,N\}$, where $0 \le y_i \le 1$ and N is the sample size. (We allow the observations to be non-identically distributed, so that the procedures apply to stratified sampling.) The asymptotic analysis is carried out as $N \to \infty$. Our maintained assumption is that, for all i,

$$E(y_i|x_i) = G(x_i\beta), \tag{2.1}$$

where $G(\cdot)$ is a known function satisfying $0 < G(z) < 1$ for all $z \in \mathbb{R}$. This ensures that the predicted values of y lie in (0,1), and the equation is well-defined even if y_i can take on 0 or 1 with positive probability. Typically, $G(\cdot)$ is chosen to be a cumulative distribution function (cdf), the two most popular examples being $G(z) = \Lambda(z) = \exp(z)/[1 + \exp(z)]$ -- the logit function -- and $G(z) = \Phi(z)$, where $\Phi(\cdot)$ is the standard normal cdf. But $G(\cdot)$ need not even be a cdf in what follows.

In stating (2.1) we make no assumption about an underlying structure used to obtain y_i. In particular, if y_i is a computed proportion from a group known size n_i, the methods of this section ignore the information on n_i. There are some advantages to this approach. First, one does not always want to condition on n_i, in which case y_i contains all relevant information. Second, the method is computationally simple. Third, under the assumptions we impose the method suggested here need not be less efficient than methods that use information on group size. The next section discusses how information on n_i can be used when one explicitly wishes to do so.

Under (2.1), β can be consistently estimated by nonlinear least squares. The nonlinear nature of (2.1) is probably the leading reason a linear model for the log-odds ratio is much more popular in applied work than estimating (2.1) directly. Further, the errors in the tautological model

$$y_i = G(\mathbf{x}_i\beta) + u_i, \quad E(u_i|\mathbf{x}_i) = 0 \tag{2.2}$$

are generally heteroskedastic since $\mathrm{Var}(u_i|\mathbf{x}_i) = \mathrm{Var}(y_i|\mathbf{x}_i)$, and $\mathrm{Var}(y_i|\mathbf{x}_i)$ is unlikely to be constant when $0 \leq y_i \leq 1$. Obtaining the NLS estimates and heteroskedasticity-robust standard errors and test statistics requires special programming, and the NLS estimator does not usually have any efficiency properties in such contexts. Still, the motivation underlying NLS is sound because it directly estimates $E(y|\mathbf{x})$.

The alternative to NLS we propose is to use an appropriate quasi-likelihood method, as in Gourieroux, Monfort, and Trognon (1984) (hereafter GMT) and McCullagh and Nelder (1989). For fractional data, the Bernoulli log-likelihood function, given by

$$\ell_i(b) = y_i \log[G(\mathbf{x}_i b)] + (1 - y_i)\log[1 - G(\mathbf{x}_i b)], \tag{2.3}$$

is attractive for several reasons. First, maximizing the Bernoulli log-likelihood is fairly simple. Second, because (2.3) is a member of the linear exponential family (LEF), the quasi-maximum likelihood estimator (QMLE) of β obtained from the maximization problem

$$\max_{b} \sum_{i=1}^{N} \ell_i(b)$$

is consistent for β provided that (2.1) holds. (This follows from GMT (1984) and is also easily seen by computing the score $s_i(b) = \nabla_\beta \ell_i(b)'$ and showing that $E[s_i(\beta)|\mathbf{x}_i] = 0$.) In other words, the QMLE $\hat{\beta}$ is consistent and $\sqrt{N}$-asymptotically normal <u>regardless</u> of the distribution of y_i conditional on $\mathbf{x}_i$; y_i could be a continuous variable, a discrete variable, or have both continuous and discrete characteristics. For example, y_i could take on the values zero and one with positive probability and values between zero and one with probability zero. As we will see below, in some cases for fractional data the Bernoulli QMLE is efficient in a class of estimators containing QMLEs in the LEF and weighted nonlinear least squares estimators.

A special case of (2.3) -- namely, when $G(\cdot)$ is the logit function -- has been suggested by McCullagh and Nelder (1989) in a generalized linear models (GLM) framework. The GLM approach has two drawbacks for economic applications. First, for the logit QMLE it assumes that

$$\mathrm{Var}(y_i|\mathbf{x}_i) = \sigma^2 G(\mathbf{x}_i\beta)[1 - G(\mathbf{x}_i\beta)] \text{ for some } \sigma^2 > 0, \tag{2.4}$$

where $G(\cdot) = \Lambda(\cdot)$. While we prefer (2.4) as a nominal variance assumption to the nominal NLS homoskedasticity assumption $\mathrm{Var}(y_i|\mathbf{x}_i) = \sigma^2$, imposing any particular conditional variance when performing inference is too restrictive. It is easy to write down mechanisms for which (2.4) fails. For example, suppose that each y_i is a proportion computed as $y_i = n_i^{-1}\sum_{j=1}^{n_i} y_{ij}$, where y_{ij} is a binary variable with $P(y_{ij} = 1|\mathbf{x}_i, n_i) = P(y_{ij} = 1|\mathbf{x}_i) = G(\mathbf{x}_i\beta)$, and the y_{ij} are independent across j conditional on $(\mathbf{x}_i, n_i)$. If n_i is not observed then all that can be estimated are $E(y_i|\mathbf{x}_i)$ and $\mathrm{Var}(y_i|\mathbf{x}_i)$. But

$$E(y_i|\mathbf{x}_i) = E(y_i|\mathbf{x}_i, n_i) = G(\mathbf{x}_i\beta)$$

and

$$\begin{aligned}\mathrm{Var}(y_i|\mathbf{x}_i) &= E[\mathrm{Var}(y_i|\mathbf{x}_i, n_i)|\mathbf{x}_i] + \mathrm{Var}[E(y_i|\mathbf{x}_i, n_i)|\mathbf{x}_i] \\ &= E[G(\mathbf{x}_i\beta)(1 - G(\mathbf{x}_i\beta))/n_i|\mathbf{x}_i] + \mathrm{Var}[G(\mathbf{x}_i\beta)|\mathbf{x}_i] \\ &= E(n_i^{-1}|\mathbf{x}_i)G(\mathbf{x}_i\beta)(1 - G(\mathbf{x}_i\beta)).\end{aligned}$$

Unless n_i and $\mathbf{x}_i$ are independent (2.4) generally fails. If, say, y_{ij} is a binary indicator for whether worker j at firm i contributes to a 401(k) plan, n_i is the number of workers at firm i, and $\mathbf{x}_i$ contains firm characteristics such as annual firm sales, n_i and $\mathbf{x}_i$ are unlikely to be independent in the population. In addition, (2.4) can fail if y_{ij} and y_{ih} are correlated conditional on (x_i, n_i) for $j \neq h$, as would be the case if there are unobserved group effects and y_{ij} and y_{ih} are from the same group i. Notice, however, that neither of these situations necessarily invalidates (2.1), which is all that is needed to consistently estimate β using the Bernolli QMLE.

The second drawback to the GLM approach is related to the first: if (2.4) fails, McCullagh and Nelder (1989, p. 330) reject the logit quasi-likelihood and suggest using more complicated quasi-likelihoods. But this begs the issue of whether the conditional mean model (2.1) is appropriate. Further, if (2.4) is in fact violated, standard specification tests and inference cannot be used to analyze the conditional mean specification. Here we are primarily interested in the conditional mean. Rather than abandoning the Bernoulli QMLE because (2.4) fails, it is straightforward to conduct asymptotically robust inference by computing robust standard errors and test statistics. This is likely to be satisfactory in applications with moderately large sample sizes.

To find the asymptotic variance of the Bernoulli QMLE, define $g(z) = dG(z)/dz$, $\hat{G}_i = G(x_i\hat{\beta}) = \hat{y}_i$, and $\hat{g}_i = g(x_i\hat{\beta})$. Then the estimated information matrix is

$$\hat{A} = \sum_{i=1}^{N} \frac{\hat{g}_i^2 x_i' x_i}{[\hat{G}_i(1 - \hat{G}_i)]} . \tag{2.5}$$

Normally, the standard error of $\hat{\beta}_j$ reported from standard binary choice analysis would be obtained as the square root of the j^{th} diagonal element of $\hat{A}^{-1}$. Under (2.1) only, this is not a consistent estimator of the true asymptotic standard error. To obtain a valid estimator we also need the outer product of the score. Let $\hat{u}_i = y_i - G(x_i\hat{\beta})$ be the residuals (deviations between y_i and its estimated conditional expectation). Define

$$\hat{B} = \sum_{i=1}^{N} \frac{\hat{u}_i^2 \hat{g}_i^2 x_i' x_i}{[\hat{G}_i(1 - \hat{G}_i)]^2} . \tag{2.6}$$

Then a valid estimate of the asymptotic variance of $\hat{\beta}$ is

$$\hat{A}^{-1}\hat{B}\hat{A}^{-1}. \tag{2.7}$$

The standard errors are obtained as the square roots of the diagonal elements of (2.7); see GMT (1984) and Wooldridge (1991b).

Interestingly, the robust standard errors from (2.7) in the context of logit and probit are computed almost routinely by certain statistics and econometrics packages, such as STATA® and SST®. Unfortunately, the packages with which we are familiar automatically transform the dependent variable used in logit or probit into a binary variable before estimation, or do not allow non-binary variables at all (STATA® and SST® fall into the first category). With the very minor change of allowing for fractional y in so-called binary choice analysis, standard statistics and econometrics packages could be used to estimate the parameters in (2.1) and to perform asymptotically valid inference. Alternatively, programming the estimator in a language such as GAUSS® is fairly straightforward; this is what we do for the application in section 5.

If the GLM assumption (2.4) is maintained in addition to (2.1) then σ^2 is consistently estimated by

$$\hat{\sigma}^2 = (N - K)^{-1} \sum_{i=1}^{N} \tilde{u}_i^2, \tag{2.8}$$

where $\tilde{u}_i$ are the <u>weighted</u> residuals

$$\tilde{u}_i = \hat{u}_i/[\hat{G}_i(1 - \hat{G}_i)]^{1/2}. \tag{2.9}$$

(It is standard practice in the GLM literature to use the degrees-of-freedom adjustment in (2.8).) Then the asymptotic variance of $\hat{\beta}$ is estimated as

$\hat{\sigma}^2\hat{A}^{-1}$. In addition, because the first two moments are correctly specified under (2.1) and (2.4), the Bernoulli QMLE is efficient in the class of QMLEs in the LEF by the results of GMT (1984); this is essentially the same as the class of all weighted NLS estimators, and so it is a nontrivial efficiency result.

To summarize, we have chosen a quasi-likelihood function that leads to a relatively efficient estimator under a popular auxiliary assumption -- namely, (2.4) -- but we guard against failure of (2.4) by using (2.7) as the variance estimator. In section 4 we suggest specification tests that are valid without (2.4).

3. Methods for Proportions with Known Group Size

Suppose that each proportion y_i is obtained as

$$y_i = n_i^{-1}\sum_{j=1}^{n_i} y_{ij}, \tag{3.1}$$

where y_{ij} is a binary variable such that $P(y_{ij} = 1|\mathbf{x}_i,n_i) = G(\mathbf{x}_i\beta)$, $j=1,2,\ldots,n_i$. Then it is easily seen that

$$E(y_i|\mathbf{x}_i,n_i) = G(\mathbf{x}_i\beta). \tag{3.2}$$

Note that (3.2) differs from (2.1) in that we are conditioning on n_i in addition to $\mathbf{x}_i$. In applications where it is inappropriate to condition on n_i -- in econometrics this might occur because n_i is "endogenous" -- the methods of section 2 should be used. Here we assume that (3.2) holds. Note that n_i and functions of it can be elements of $\mathbf{x}_i$.

If we add the assumption

$$y_{ij},\ y_{ih} \text{ are independent conditional on } (\mathbf{x}_i,n_i),\text{ all } h \neq j, \tag{3.3}$$

then

$$\mathrm{Var}(y_i|\mathbf{x}_i,n_i) = G(\mathbf{x}_i\beta)[1 - G(\mathbf{x}_i\beta)]/n_i. \tag{3.4}$$

Given assumptions (3.1) and (3.3) it seems reasonable that knowledge of n_i can increase efficiency of estimation. In fact, these assumptions are enough to perform (conditional) maximum likelihood estimation. In this context the distribution is typically defined in terms of the number of successes $s_i = \sum_{j=1}^{n_i} y_{ij} = n_i y_i$ out of n_i trials, but for comparison purposes we define the likelihood in terms of the proportions y_i. Under (3.1) and (3.3), s_i given $(\mathbf{x}_i,n_i)$ has a Binomial$(n_i,G(\mathbf{x}_i\beta))$ distribution. Thus, the density of y_i given $(\mathbf{x}_i,n_i)$ is

$$f(y|\mathbf{x}_i,n_i) = \binom{n_i}{n_i y}[G(\mathbf{x}_i\beta)]^{n_i y}[G(\mathbf{x}_i\beta)[1 - G(\mathbf{x}_i\beta)]]^{n_i(1-y)}, \tag{3.5}$$

$$y = 0,\ 1/n_i,\ 2/n_i,\ \ldots,\ 1.$$

Up to an additive constant which does not depend on the parameters the conditional log-likelihood for observation i is

$$\ell_i(b) = n_i\{y_i\log[G(\mathbf{x}_i b)] + (1 - y_i)\log[1 - G(\mathbf{x}_i b)]\}, \tag{3.6}$$

which is simply n_i times the conditional log-likelihood for observation i used in section 2. Thus, we simply weight the conditional log-likelihood for observation i by the group size n_i. This will be identical to the estimator in section 2 when n_i is the same for all $i=1,2,\ldots,N$.

As in section 2 we prefer to view (3.6) as a quasi-log likelihood function because (3.4) might fail (that is, s_i conditional on $(\mathbf{x}_i,n_i)$ need not have a binomial distribution) even though (3.2) can hold. This happens if we allow for y_{ij}, y_{ih} to be dependent (conditional on $(\mathbf{x}_i,n_i)$). This kind of clustering can certainly happen in the empirical example of interest here, as a worker's decision to contribute to a 401(k) plan can be related to other workers' decisions within the same firm.

Under (3.2) only, the QMLE is consistent and $\sqrt{N}$-asymptotically normal; this again follows because the conditional log-likelihood (3.6) is a member of the linear exponential family. Only minor changes are needed in the formulas from section 2 for estimating the asymptotic variance of $\hat{\beta}$. In the sum defining $\hat{A}$ (see equation (2.5)), summand i is multiplied by n_i; summand i in the sum defining $\hat{B}$ is multiplied by n_i^2. If an estimator of σ^2 under the assumption

$$\mathrm{Var}(y_i|x_i,n_i) = \sigma^2 G(x_i\beta)(1 - G(x_i\beta))/n_i \tag{3.7}$$

is desired, then it is computed as in (2.8) except that summand i in (2.8) is multiplied by n_i (equivalently, each weighted residual $\tilde{u}_i$ in (2.9) is multiplied by $\sqrt{n_i}$). Under (3.2) and (3.7), Avar($\hat{\beta}$) is estimated as $\hat{\sigma}^2\hat{A}^{-1}$; under (3.2) only, the robust form $\hat{A}^{-1}\hat{B}\hat{A}^{-1}$ should be used.

We should emphasize that the binomial QMLE is not necessarily more efficient than the Bernoulli QMLE studied in the previous section without assumptions (3.1) and (3.3). In other words, without additional assumptions, conditioning on the group size n_i need not increase the efficiency of the QMLE.

4. Specification Testing

Specification testing in this framework can be carried out as in Wooldridge (1991a,b). We discuss two forms of the test. The first is valid under (2.1) and (2.4) or (3.2) and (3.7); these are nonrobust tests because they maintain that the nominal variance assumption is in fact true. A robust form of the test requires only (2.1) or (3.2). We explicitly outline the tests for the setup of section 2, and then discuss the simple adjustment needed for the situation in section 3.

We focus primarily on Lagrange multiplier or score tests that nest $E(y|\mathbf{x}) = G(\mathbf{x}\beta)$ within a more general model. Let $m(\mathbf{x},\mathbf{z},\beta,\gamma)$ be a model for $E(y|\mathbf{x},\mathbf{z})$, where $\mathbf{z}$ is a 1xJ vector of additional variables, which could be nonlinear functions of $\mathbf{x}$ (in which case $E(y|\mathbf{x}) = E(y|\mathbf{x},\mathbf{z})$), or variables not functionally related to $\mathbf{x}$, or both. The vector γ is a Qx1 vector of additional parameters. The null is assumed to be H_0: $\gamma = \gamma_0$ for a specified vector γ_0 (e.g., $\gamma_0 = 0$). Then by definition,

$$G(\mathbf{x}\beta) = m(\mathbf{x},\mathbf{z},\beta,\gamma_0). \tag{4.1}$$

Given the estimates under the null, $\hat{\beta}$, define the 1xK vector $\nabla_\beta \hat{m}_i = \partial m(\mathbf{x}_i,\mathbf{z}_i,\hat{\beta},\gamma_0)/\partial\beta = \hat{g}_i\mathbf{x}_i$ and the 1xQ vector $\nabla_\gamma \hat{m}_i = \partial m(\mathbf{x}_i,\mathbf{z}_i,\hat{\beta},\gamma_0)/\partial\gamma$; these are simply the gradients of the regression function with respect to β and γ, respectively, evaluated under the null hypothesis. Given the residuals $\hat{u}_i = y_i - G(\mathbf{x}_i\hat{\beta})$, define the weighted quantities

$$\tilde{u}_i = \hat{u}_i/[\hat{G}_i(1 - \hat{G}_i)]^{1/2} \tag{4.2}$$

$$\nabla_\beta \tilde{m}_i = \nabla_\beta \hat{m}_i/[\hat{G}_i(1 - \hat{G}_i)]^{1/2} = \hat{g}_i\mathbf{x}_i/[\hat{G}_i(1 - \hat{G}_i)]^{1/2} \tag{4.3}$$

$$\nabla_\gamma \tilde{m}_i = \nabla_\gamma \hat{m}_i/[\hat{G}_i(1 - \hat{G}_i)]^{1/2}. \tag{4.4}$$

Note that the weights are proportional to the inverse of the nominal standard deviation (see (2.4)). As mentioned above, a valid test of H_0: $\gamma = \gamma_0$ depends on what is maintained under the null hypothesis. Under the assumptions

$$E(y_i|\mathbf{x}_i,\mathbf{z}_i) = G(\mathbf{x}_i\beta) \tag{4.5}$$

<u>and</u>

$$\mathrm{Var}(y_i|\mathbf{x}_i,\mathbf{z}_i) = \sigma^2 G(\mathbf{x}_i\beta)[1 - G(\mathbf{x}_i\beta)], \tag{4.6}$$

a valid test statistic is obtained as NR_u^2 from the OLS regression

$$\tilde{u}_i \quad \text{on} \quad \nabla_\beta \tilde{m}_i, \ \nabla_\gamma \tilde{m}_i, \quad i=1,2,\ldots,N, \tag{4.7}$$

where R_u^2 is the constant-unadjusted r-squared. Under (4.5) and (4.6), NR_u^2 is distributed asymptotically as χ_Q^2 -- see Wooldridge (1991a).

For binary choice models, Engle (1984) and Davidson and MacKinnon (1984) suggest a test based on regression (4.7) for logit and probit. Gurmu and Trivedi (1993) present results for a class of models that allows testing the logit function against a more general index function. For fractional dependent variables it is important to use the NR_u^2 form rather than the explained sum of squares form suggested in Davidson and MacKinnon (1984): the latter test requires $\sigma^2 = 1$, which is always the case for binary reponse variables but is too restrictive for fractional response variables. Alternatively, as in Gurmu and Trivedi (1993), each term in (4.7) can be divided by $\hat{\sigma}$ and then the explained sum of squares can be used.

It is often useful to have a likelihood-based statistic, especially for testing exclusion restrictions. Under the same two assumptions (4.5) and (4.6), a quasi-likelihood ratio (QLR) statistic has a limiting chi-square distribution. Let $\mathcal{L}_N(\hat{\beta},\gamma_0)$ denote the log-likelihood evaluated under the null, and let $\mathcal{L}_N(\ddot{\beta},\ddot{\gamma})$ denote the log-likelihood from the unrestricted model (that is, the Bernoulli log-likelihood with $m(\mathbf{x},\mathbf{z},\beta,\gamma)$ used in place of $G(\mathbf{x}_i\beta)$). Further, define $\ddot{m}_i = m(\mathbf{x}_i,\mathbf{z}_i,\ddot{\beta},\ddot{\gamma})$, and let the variance estimator based on the unrestricted estimates be

$$\ddot{\sigma}^2 = (N-K-Q)^{-1} \sum_{i=1}^{N} (y_i - \ddot{m}_i)^2/[\ddot{m}_i(1 - \ddot{m}_i)] \tag{4.8}$$

(note that the summation is simply the sum of weighted squared residuals from the unrestricted model). Then the QLR statistic, defined by

$$QLR = 2[\mathcal{L}_N(\ddot{\beta},\ddot{\gamma}) - \mathcal{L}_N(\hat{\beta},\gamma_0)]/\ddot{\sigma}^2, \tag{4.9}$$

is distributed asymptotically as χ^2_Q under the null hypothesis, provided (4.6) holds in addition to (4.5). The validity of this statistic follows because the usual information matrix equality holds up to the scalar σ^2 when the conditional mean and conditional variance are correctly specified.

A form of the LM statistic that is valid under (4.5) alone requires an additional regression. First regress $\nabla_\gamma \tilde{m}_i$ on $\nabla_\beta \tilde{m}_i$ and save the 1xQ residuals, $\tilde{r}_i = (\tilde{r}_{i1}, \tilde{r}_{i2}, \ldots, \tilde{r}_{iQ})$, $i=1,2,\ldots,N$. (This is the same as regressing each element of $\nabla_\gamma \tilde{m}_i$ on the entire vector $\nabla_\beta \tilde{m}_i$, and collecting the residuals.) Next, obtain the 1xQ vector $\tilde{u}_i\tilde{r}_i = (\tilde{u}_i\tilde{r}_{i1}, \tilde{u}_i\tilde{r}_{i2}, \ldots, \tilde{u}_i\tilde{r}_{iQ})$. The robust LM statistics is obtained as N - SSR, where SSR is the usual sum of squared residuals from the auxiliary regression of unity on $\tilde{u}_i\tilde{r}_i$:

$$1 \text{ on } \tilde{u}_i\tilde{r}_i, \quad i=1,\ldots,N. \tag{4.10}$$

Under H_0, which is (4.5) in this case, $N - SSR \overset{a}{\sim} \chi^2_Q$. The validity of this procedure is discussed further in Wooldridge (1991a,b). Briefly, N - SSR from (4.10) is a quadratic form in the vector $N^{-1/2}\sum_{i=1}^{N} \tilde{r}_i'\tilde{u}_i$, with a weighting matrix that is the inverse of a consistent estimator of its asymptotic variance whether or not (4.6) holds.

In testing for omitted variables, one can use the QLR statistic or LM statistic under (4.5) and (4.6) or the robust LM statistic under (4.5). (Of course, Wald statistics can also be defined for these two cases, but they are computationally more cumbersome than the QLR and LM statistics.) For omitted variables tests, $m(\mathbf{x}_i,\mathbf{z}_i,\beta,\gamma) = G(\mathbf{x}_i\beta + \mathbf{z}_i\gamma)$, $\nabla_\gamma \hat{m}_i = \hat{g}_i\mathbf{z}_i = g(\mathbf{x}_i\hat{\beta})\cdot\mathbf{z}_i$, and $\nabla_\gamma \tilde{m}_i = \hat{g}_i\mathbf{z}_i/[\hat{G}_i(1 - \hat{G}_i)]^{1/2}$. One way to test for functional form is to define $\mathbf{z}_i$ as polynomials or other functions of $\mathbf{x}_i$.

A general functional form diagnostic is obtained by extending Ramsey's (1969) RESET procedure to index models. For example, let the alternative model be

$$E(y_i|x_i) = G(x_i\beta + \gamma_1(x_i\beta)^2 + \gamma_2(x_i\beta)^3) \tag{4.11}$$

where, again, $G(\cdot)$ is typically the logit or probit function. The hypothesis that (4.5) holds (with $z_i = x_i$) is stated as

$$H_0: \gamma_1 = 0, \ \gamma_2 = 0.$$

This is easily tested using the LM procedures outlined above. (By contrast, the QLR statistic is computationally difficult and nonrobust.) First, estimate the model under the assumption $\gamma_1 = \gamma_2 = 0$, as is always done. Define $\hat{\beta}$, $\hat{G}_i$, $\hat{g}_i$, $\hat{u}_i$, $\nabla_\beta\tilde{m}_i$, and $\tilde{u}_i$ as before. The gradient with respect to $\gamma = (\gamma_1,\gamma_2)'$ is easily seen to be

$$\nabla_\gamma\hat{m}_i = (\hat{g}_i\cdot(x_i\hat{\beta})^2, \hat{g}_i\cdot(x_i\hat{\beta})^3),$$

and $\nabla_\gamma\tilde{m}_i$ is defined in (4.4). The statistic obtained from the regression (4.7) is distributed approximately as χ^2_2 under (4.5) and (4.6). The robust form is obtained from regression (4.10).

Little changes if we instead use the methods of section 3. In the expectation and variance in (4.5) and (4.6) we now condition on n_i as well as x_i and z_i. Let $\hat{\beta}$ now denote the QMLE from section 3. The statistics from regressions (4.7) and (4.10) are valid if we multiply each of $\tilde{u}_i$, $\nabla_\beta\tilde{m}_i$, and $\nabla_\gamma\tilde{m}_i$ in (4.2), (4.3), and (4.4) by $\sqrt{n_i}$. For the QLR statistic, in addition to using the (3.6) as the conditional log-likelihood, we also multiply summand i in (4.8) by n_i.

5. Empirical Application: Participation in 401(k) Pension Plans

401(k) plans differ from traditional employer-sponsored pension plans in that employees are permitted to make pre-tax contributions and the employer may match part of the contribution. Since participation in these plans is voluntary, the sensitivity of participation to plan characteristics -- specifically the employer matching rate -- will play a critical role in retirement saving.

Pension plan administrators are required to file Form 5500 annually with the Internal Revenue Service, describing participation and contribution behavior for each plan offered. Papke (1993) uses the plan level data to study, among other things, the relationship between the participation rate and various plan characteristics, including the rate at which a firm matches employee contributions.

The participation rate (PRATE) is constructed as the number of active accounts divided by the number of employees eligible to participate. An active account is any existing 401(k) account -- a contribution need not have been made that plan year. The plan match rate (MRATE) is not reported directly on the Form 5500, but can be approximated by the ratio of employer to employee contributions for plans that provide some matching. This calculated match rate may exceed the plan's marginal rate because employer contributions include any flat per participant contribution or any helper contribution made to pass anti-discrimination tests. While the calculated match rate exceeds the marginal incentive facing each saver, it may be a better indicator of overall plan generosity. See Papke (1993) for additional discussion.

Papke (1993) uses a spline method to estimate models with the participation rate, PRATE, as the dependent variable. She finds a statistically significant positive relationship between PRATE and MRATE, with some evidence of a diminishing marginal effect. Here, we allow for a diminishing marginal effect of MRATE on PRATE by using a conditional mean of the form (2.1) with G(·) taken to be the logit function. We compare this directly with linear models where PRATE is the dependent variable.

Table 1 presents summary statistics for the sample of 401(k) plans from the 1987 plan year. Statistics are presented separately for the 80 percent of the plans with match rates less than or equal to 1. IRS reporting conventions mean that match rates above unity are structurally different. This discussion focuses on the subsample with MRATE ≤ 1.

Participation rates in 401(k) plans are high -- averaging about 85 percent in our sample. Over forty percent of the plans (42.73) have a participation proportion of exactly unity -- all eligible employees have an active account. This characteristic of the data would make a log-odds approach especially awkward because an adjustment would have to be made to 40 percent of the observations.

The plan match rate averages about 41 cents on the dollar. Other explanatory variables include total firm employment (EMP) which averages 4,622 across the plans. The plans average 12 years in age (AGE). Sole plan is a dummy variable which indicates that the 401(k) plan is the only pension plan offered by the employer. These sole plans constitute about 37 percent of the sample.

The first linear model we estimate is

$$E(PRATE_i|x_i) = \beta_1 + \beta_2 MRATE_i + \beta_3 \log(EMP_i) + \beta_4 \log(EMP_i)^2 \qquad (5.1)$$

$$+ \beta_5 AGE_i + \beta_6 AGE_i^2 + \beta_7 SOLE_i,$$

where the variable definitions are given above. This model is estimated by ordinary least squares (OLS), initially using the subsample for which MRATE $\leq$ 1. The results are given in the first column of Table 2. Because of the probable heteroskedasticity in this equation, the heteroskedasticity-robust standard errors are reported in brackets below the usual OLS standard errors.

All variables are highly statistically significant except for the sole plan indicator. Note that there is very little difference between the usual OLS standard errors and the heteroskedasticity-robust ones, so it really does not matter which standard errors we use. The key variable MRATE has a t-statistic well over 10. Its coefficient of .156 implies that if the match rate increases by 10 cents on the dollar, the participation rate would increase on average by almost 1.6 percentage points. This is not a small effect considering that the average participation rate is about 85 percent in the subsample. As mentioned in the introduction, the linear model implies a constant marginal effect throughout the range of MRATE, which cannot literally be true.

That the linear model does not fit as well as it should can be seen by computing Ramsey's (1969) RESET (and its heteroskedasticity-robust version). Let $\hat{u}_i$ be the OLS residuals and let $\hat{y}_i$ be the OLS fitted values. Then, the LM version of RESET is obtained as NR^2 from the regression

$$\hat{u}_i \text{ on } \mathbf{x}_i, \hat{y}_i^2, \hat{y}_i^3, \quad i=1,2,\ldots,N.$$

Under the null that (5.1) is true, $NR^2 \overset{a}{\sim} \chi^2_2$ (homoskedasticity is also maintained). The heteroskedasticity-robust version is obtained as N - SSR from regression (4.10) given the proper definitions: let $\tilde{u}_i = \hat{u}_i$ and let $\tilde{\mathbf{r}}_i$ be the 1x2 residuals from the regression of $(\hat{y}_i^2, \hat{y}_i^3)$ on $\mathbf{x}_i$. See Wooldridge (1991a) for more details. Using either nonrobust RESET or its robust form, (5.1) is strongly rejected (the one percent critical value for a χ^2_2 is 9.21). Because RESET is a test of functional form, we conclude that (5.1) misses some potentially important nonlinearities. (At this point, one should remember the potential difference between a statistical rejection of a model and the economic importance of any misspecification.)

We next use the logit QMLE analyzed in section 2 to estimate the nonlinear model

$$E(PRATE_i|\mathbf{x}_i) = G(\beta_1 + \beta_2 MRATE_i + \beta_3 \log(EMP_i) + \beta_4 \log(EMP_i)^2 + \beta_5 AGE_i + \beta_6 AGE_i^2 + \beta_7 SOLE_i), \qquad (5.2)$$

where $G(\cdot)$ is the logit function. (The GAUSS® code used for the estimation and testing is available from the authors on request.) The partial effect of MRATE on $E(PRATE|\mathbf{x})$ is $\partial E(PRATE|\mathbf{x})/\partial MRATE$, or, for specification (5.2),

$$g(\mathbf{x}\beta)\beta_2, \qquad (5.3)$$

where $g(z) = dG(z)/dz = \exp(z)/[1 + \exp(z)]^2$. Because $g(z) \to 0$ as $z \to \infty$, the marginal effect falls to zero as MRATE gets large, holding other variables fixed. Column (2) of Table 2 contains the results of estimating (5.2). The variable MRATE is highly statistically significant and, with the exception of SOLE (which is not significant), the directions of effects of all other variables are the same as in the linear model. But unlike the linear model, the RESET statistic reveals no misspecification in (5.2); the p-value for the robust statistic is .676, and it is even larger for the nonrobust statistic. Thus, based on this particular statistic, (5.2) appears to adequately capture the nonlinear relationship between PRATE and the explanatory variables for $MRATE \leq 1$.

There is other evidence that (5.2) fits better than (5.1). Table 2 also contains an r-squared for each model, which in either case is defined as $1 - SSR/SST$, where SST is the total sum of squares. The SSRs, reported in Table 2, are based on the <u>unweighted</u> residuals, $\hat{u}_i = y_i - \hat{y}_i$ for OLS <u>and</u>

QMLE. Thus, the r-squareds are comparable across <u>any</u> model for E(PRATE|**x**). From Table 2 we see that the r-squared from the logit model is about 6 percent higher than the r-squared for the linear model. This difference is significant given the large sample size. In addition, while OLS chooses $\hat{\beta}$ to maximize the r-squared over all linear functions of **x**, the logit QMLE does <u>not</u> maximize r-squared given the logit functional form; yet the logit model has a higher r-squared than the linear model.

Before directly comparing estimates of marginal effects, there are some other comments worth making about Table 2. First, the SERs in the table, which are estimates of σ, are not directly comparable. For OLS, $\hat{\sigma}^2$ is based on the unweighted residuals, while the QMLE $\hat{\sigma}^2$ is based on the weighted residuals $\tilde{u}_i$; see (2.8) and (2.9). Because $\hat{\sigma} = .438$ for the QMLE, this implies that the usual logit standard errors obtained from the inverse of the Hessian, $\hat{\mathbf{A}}^{-1}$, are over twice as large as the GLM standard errors that are obtained as the squared roots of the diagonal elements of $\hat{\sigma}^2\hat{\mathbf{A}}^{-1}$. The latter ones are the appropriate standard errors under the GLM assumption (2.4) because they do not require $\sigma = 1$.

We now turn to a direct comparison of estimates of marginal effects for the linear and nonlinear models for various values of MRATE. As we already discussed, the partial derivative of E(PRATE|**x**) with respect to MRATE for the linear model is estimated to be .156 regardless of the values of MRATE, EMP, AGE, and SOLE. For the nonlinear model, we need to choose values for the elements of **x**. Thus, set SOLE = 0, and set EMP and AGE at roughly their sample averages: EMP = 4620 and AGE = 13. Given these values, we evaluate (5.3) (with β replaced by $\hat{\beta}$) at three different match rates: MRATE = 0, MRATE = .50, and MRATE = 1.0. The estimated derivatives are .288, .197, and .118, respectively, which illustrates the diminishing marginal effect as MRATE increases. Not surprisingly, the marginal effect estimated from the linear model is bracketed by the low and high estimates from the nonlinear model. The differences in the estimated marginal effects are not trivial; for example, the nonlinear model predicts an increase in participation of approximately 2.9 percentage points in moving from a zero match rate to MRATE = .10, rather than the 1.6 percentage point increase obtained from the linear model. Similarly, at high match rates the marginal effect from increasing the match rate is estimated to be much lower in the nonlinear model.

One way to potentially salvage the linear model is to use a more flexible functional form for the match rate. A popular functional form that allows a diminishing marginal effect is a quadratic. Column three contains estimates of the linear model that includes a quadratic in MRATE. The squared term is marginally significant (robust t-statistic ≈ -1.98), and this does give a diminishing marginal effect. But even with this additional regressor the model in (3) does not fit as well as the logit model without the quadratic term (the r-squared has only gone up to .144). Further, the rejection of the model by RESET is almost as strong as it was without the quadratic. Thus, we conclude that simply adding $MRATE^2$ to (5.1) is not sufficient. (The spline approach used by Papke (1993) is more effective in capturing a diminishing effect in this application, but the coefficients are more difficult to interpret.)

When $MRATE^2$ is added to (5.2) it turns out to be insignificant. Thus, the logit functional form, with the term linear in MRATE, appears to be enough to capture the diminishing effect, at least for MRATE $\leq$ 1. This is a useful lesson: a significant quadratic term in a linear model might be indicating that an entirely different functional form can provide a better fit. Model (5.2) is clearly the preferred specification thus far.

The basic story does not change when we estimate the models over the entire sample, except that a quadratic term is now significant in (5.2).

reflecting a faster diminishing effect at high match rates. Table 3 presents the same models as Table 2, now estimated over the full sample. First consider the models without $MRATE^2$. The discrepency in r-squareds between (5.2) and (5.1) is even greater than before, but RESET now rejects both (5.1) and (5.2), although the logit QMLE is rejected less strongly. In columns (3) and (4) we put $MRATE^2$ into each equation. Model (5.1) is still soundly rejected, whereas (5.2) with $MRATE^2$ passes the RESET test with a p-value above .50. For the full sample, it seems a quadratic in MRATE is needed to provide a reasonable fit.

The one drawback to putting $MRATE^2$ into (5.2) is the usual one for quadratics: it implies an eventual negative marginal effect. In this case, the marginal effect becomes negative at a match rate of about 2.51. This is a high value for MRATE, but there are some match rates this large in the full sample. Note that the turning point for the linear model is much lower; the estimated marginal effect becomes negative at a match rate of 1.37.

6. Conclusion

The quasi-likelihood methods studied here apply in cases where the response variable takes on fractional values. These methods offer viable alternatives to linear models using either y or the log-odds ratio of y as the dependent variable. No special data adjustments are needed for the extreme values of zero and one, and the conditional expectation of y given the explanatory values is estimated directly. The empirical application to 401(k) plan participation rates illustrates the usefulness of these methods: while a linear model to explain the fraction of participants is strongly rejected, the logit conditional mean specification is not. In addition, the estimates of the partial effects are much more plausible for the logit model.

Table 1: Summary Statistics

FULL SAMPLE

Number of Observations = 4734

Variable	Mean	Standard Deviation
PRATE	.869	.167
MRATE	.746	.844
EMPLOYMENT	4621.01	16299.64
AGE	13.14	9.63
SOLE	.415	————

RESTRICTED SAMPLE (MRATE ≤ 1)

Number of Observations = 3874

Variable	Mean	Standard Deviation
PRATE	.848	.170
MRATE	.408	.228
EMPLOYMENT	4621.91	17037.11
AGE	12.24	8.91
SOLE	.373	————

Table 2: Results for Restricted Sample

Variable	(1) OLS	(2) QMLE	(3) OLS	(4) QMLE
MRATE	.156	1.390	.239	1.218
	(.012)	(0.100)	(.042)	(0.342)
	[.011]	[0.108]	[.046]	[0.378]
$MRATE^2$	———	———	-.087	.196
			(.043)	(.373)
			[.044]	[.425]
log(EMP)	-.112	-1.002	-.112	-1.002
	(.014)	(0.111)	(.014)	(0.111)
	[.013]	[0.110]	[.013]	[0.110]
$\log(EMP)^2$	.0057	.0522	.0057	.0522
	(.0009)	(.0071)	(.0009)	(.0071)
	[.0009]	[.0071]	[.0009]	[.0071]
AGE	.0060	.0501	.0059	.0503
	(.0010)	(.0087)	(.0010)	(.0087)
	[.0009]	[.0089]	[.0009]	[.0088]
AGE^2	-.00007	-.00052	-.00007	-.00052
	(.00002)	(.00021)	(.00002)	(.00021)
	[.00002]	[.00021]	[.00002]	[.00021]
SOLE	-.0001	.0080	.0008	.0061
	(.0058)	(.0468)	(.0058)	(.0470)
	[.0060]	[.0502]	[.0060]	[.0504]
ONE	1.213	5.058	1.198	5.085
	(0.051)	(0.427)	(0.052)	(0.430)
	[0.048]	[0.421]	[0.049]	[0.423]
Observations:	3784	3784	3784	3784
SSR:	93.67	92.70	93.56	92.69
SER:	.157	.438	.157	.438
R-Squared:	.143	.152	.144	.152
RESET:	39.55	0.606	35.06	0.732
	(.000)	(.738)	(.000)	(.693)
Robust RESET:	45.36	0.782	40.08	0.836
	(.000)	(.676)	(.000)	(.658)

Notes: The quantities in (·) below estimates are the OLS standard errors or, for QMLE, the GLM standard errors; the quantities in [·] are the standard errors robust to variance misspecification. SSR is the sum of squared residuals and SER is the standard error of the regression; for QMLE, the SER is defined in terms of the weighted residuals. The values in parentheses below the RESET statistics are p-values; these are obtained from a chi-square distribution with two degrees-of-freedom.

Table 3: Results for Full Sample

Variable	(1) OLS	(2) QMLE	(3) OLS	(4) QMLE
MRATE	.034 (.003) [.003]	.542 (.045) [.079]	.143 (.008) [.008]	1.665 (0.089) [0.104]
$MRATE^2$	———	———	-.029 (.002) [.002]	-.332 (.021) [.026]
log(EMP)	-.101 (.012) [.012]	-1.038 (0.121) [0.110]	-.099 (.012) [.012]	-1.030 (0.112) [0.110]
$\log(EMP)^2$	.0051 (.0008) [.0008]	.0540 (.0078) [.0071]	.0050 (.0008) [.0008]	.0536 (.0072) [.0071]
AGE	.0064 (.0008) [.0007]	.0621 (.0089) [.0078]	.0056 (.0008) [.0007]	.0548 (.0082) [.0077]
AGE^2	-.00008 (.00002) [.00002]	-.00071 (.00021) [.00018]	-.00007 (.00002) [.00001]	-.00063 (.00019) [.00018]
SOLE	.0140 (.0050) [.0052]	.1190 (.0510) [.0503]	.0066 (.0049) [.0051]	.0642 (.0471) [.0498]
ONE	1.213 (0.045) [0.044]	5.429 (0.467) [0.422]	1.170 (0.044) [0.042]	5.105 (0.431) [0.416]
Observations:	4734	4734	4734	4734
SSR:	112.70	109.51	107.76	105.73
SER:	.154	.502	.151	.461
R-Squared:	.144	.168	.182	.197
RESET:	85.22 (.000)	50.56 (.000)	83.80 (.000)	1.370 (.504)
Robust RESET:	69.15 (.000)	9.666 (.008)	98.51 (.000)	1.275 (.529)

Notes: See Table 2.

References

Davidson, R. and J.G. MacKinnon (1984), "Convenient Specification Tests for Logit and Probit Models," *Journal of Econometrics* 24, 241-262.

Duan, N. (1983), "Smearing Estimate: A Nonparametric Retransformation Method," *Journal of the American Statistical Association* 78, 605-610.

Engle, R.F. (1984), "Wald, Likelihood Ratio, and Lagrange Multiplier Statistics in Econometrics," in Z. Griliches and M.D. Intriligator, eds., *Handbook of Econometrics*, Volume II, 776-828.

Gourieroux, C., A. Monfort, and A. Trognon (1984), "Pseudo-Maximum Likelihood Methods: Theory," *Econometrica* 52, 681-700.

Gurmu, S. and P.K. Trivedi (1993), "Variable Augmentation Specification Tests in the Exponential Family," *Econometric Theory* 9, 94-113.

Maddala, G.S. (1983), *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge: Cambridge University Press.

McCullagh, P. and J.A. Nelder (1989), *Generalized Linear Models*, New York: Chapman and Hall, second edition.

Papke, L.E. (1993), "Participation in and Contributions to 401(k) Pension Plans: Evidence from Plan Data," mimeo, Michigan State University Department of Economics. Revision of NBER Working Paper No. 4199.

Ramsey, J.B. (1969), "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis," *Journal of the Royal Statistical Society* Series B 31, 350-371.

Wooldridge, J.M. (1991a), "On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances," *Journal of Econometrics* 47, 5-46.

Wooldridge, J.M. (1991b), "Specification Testing and Quasi-Maximum Likelihood Estimation," *Journal of Econometrics* 48, 29-55.