INSTRUMENTAL VARIABLE ESTIMATION IN BINARY MEASUREMENT ERROR MODELS

by

L. A. Stefanski and J. S. Buzas

Institute of Statistics Mimeograph Series No. 2239

November 1992

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

MIMEO SERIES #2239

L.A. Stefanski and J.S. Buzas

INSTRUMENTAL VARIABLE ESTIMATION IN BINARY MEASUREMENT ERROR MODELS

NAME	DATE

INSTRUMENTAL VARIABLE ESTIMATION IN BINARY MEASUREMENT ERROR MODELS

L. A. Stefanski
Department of Statistics
North Carolina State University
Raleigh, NC 27695

J. S. Buzas
Department of Statistics
North Carolina State University
Raleigh, NC 27695

ABSTRACT

We describe an approach to instrumental variable estimation in binary regression measurement error models. The method entails constructing an approximate mean model for the binary response as a function of the measured predictor, the instrument and any covariates in the model. Estimates are obtained by exploiting relationships between various regression parameters, just as in linear instrumental variable estimation. In the course of deriving the approximate mean model we present an alternative characterization of instrumental variable estimation in linear measurement error models.

1. INTRODUCTION

Instrumental variable estimation is widely recognized as an important method of analysis of linear measurement error models, Fuller (1987, Chs. 1.4, 2.4). In this paper we describe some approaches to extending linear instrumental variable estimation to binary regression models.

Linear instrumental variable estimation exploits relationships between the parameters in the regressions of the response variable and measured predictor on the instrumental variable. This is also the rational behind our approach. It entails the construction of an approximate mean model for the observed response as a function of the instrument. The approximations are similar to those currently used in the analysis of generalized linear measurement error models, *e.g.*, Carroll and Stefanski (1990). Having fit the approximate mean model, the instrumental variable is exploited in much the same fashion as in linear instrumental variable estimation. This method applies quite generally, although it will usually only result in approximately consistent estimators. Amemiya (1985a–c) also has studied approximate instrumental variable estimation in nonlinear measurement error models. Our method differs substantially from his. We exploit the structure of generalized linear models whereas Amemiya deals with a much broader class of models.

Section 2 describes an application motivating the research. The main ideas are set forth in Section 3 in the context of simple binary regression models. An alternative representation of the linear measurement error model instrumental variable estimator is presented and used to motivate the generalization to binary regression models. Results of a simulation study are discussed. General versions of the estimators are given in Section 4 for the case of multiple regression models with covariates and multiple instruments. Section 4 concludes with an example application to data from the Framingham heart study. Summary remarks and generalizations are given in Section 5.

2. MOTIVATION

Recently MacMahon *et al.* (1990) presented results from a meta-analysis of the effect of blood pressure on stroke and coronary heart disease. Corrections for the attenuating effect of measurement error in blood pressure measurements figured prominently in their statistical analyses. Baseline measurements of diastolic blood pressure were made in all of the studies included in the meta-analysis. In one study, Framingham, post-baseline measurements were also made. The strategy of MacMahon *et al.* (1990) was to use the multiple measurements from the Framingham study to estimate a correction for attenuation and then apply the correction to all of the studies.

Definitional issues often arise in the application of measurement error models. Let X denote the measured predictor and U the so-called true value of the predictor. In some cases U is well defined, but this is not always the case. One recourse is to define U operationally, that is, $U = E(X)$. This makes sense whenever the possibility of replication exists and it can be argued that the regression of Y on $\bar{X}$ is a more meaningful parameter than the regression of Y on a single measurement.

For the case of measuring blood pressure, U is usually defined as a long-term average, although the averaging period is often left unspecified. In the meta-analytic framework of MacMahon *et al.* (1990) , it seems appropriate to define U as the expected value of the *baseline* blood pressure measurement, due to the commonality of this measurement across studies. This leaves open the interpretation of the post-baseline measurements. Are they replicates? The post-baseline measurements were taken four years post baseline.

Fuller (1987, p. 52) notes that one possible source of an instrumental variable, W, is a second measurement of U "... obtained by an independent method." The essential conditions that W must satisfy are: (i) nonzero correlation with U; and (ii) zero correlation with $X - U$ and $Y - E(Y \mid U)$. In order for W to be a replicate measurement it is necessary to strengthen (i) to: $E(W \mid U) = U$, and $Var(W \mid U) = Var(X \mid U)$. Thus the assumption that a second measurement is a replicate is stronger than the assumption that it is an instrument.

Carroll and Stefanski (1992) argue that the correction for attenuation employed by MacMahon *et al.* (1990) is generally correct only when the post-baseline measurements satisfy the assumption of replicate measurements. Using the post-baseline measurements as instruments permits estimation of a correction for attenuation and is justified more generally. The possibility that instrumental variable estimation might be the appropriate way to use the post-baseline measurements in this situation prompted the research described in this paper.

Carroll and Stefanski (1992) provide further discussion on the relative merits of instrumental variable estimation versus replicate measurement estimation in the context of meta analysis.

3. APPROXIMATE INSTRUMENTAL VARIABLE ESTIMATION

We start this section by establishing notation and describing the assumptions under which the theory is developed.

Let Y, U and Z denote the response variable, predictor subject to measurement error, and an error-free covariate respectively. The measured predictor and instrument are denoted X and W respectfully. We start with the following assumptions:

$$E(Y \mid U,\ Z) = m(\beta_{Y|\underline{1}UZ} + \beta^T_{Y|1\underline{U}Z}U + \beta^T_{Y|1U\underline{Z}}Z),$$

$$m(t) \text{ monotone}, \qquad -\infty < t < \infty; \tag{3.1}$$

$$E(U \mid W,\ Z) = \beta^T_{U|\underline{1}WZ} + \beta^T_{U|1\underline{W}Z}W + \beta^T_{U|1W\underline{Z}}Z; \tag{3.2}$$

$$E(X - U \mid W,\ Z) = 0; \tag{3.3}$$

$$E(Y \mid W,\ X,\ Z) = E\{E(Y \mid U,\ Z) \mid W,\ X,\ Z\}. \tag{3.4}$$

Our work is complicated by the need to indicate numerous regression parameters. We have adopted a notation that although somewhat cumbersome, has the advantage of being descriptive. For example, $\beta_{Y|\underline{1}UZ}$ is the coefficient of 1, *i.e.*, the intercept, in the regression of Y on 1, U and Z; $\beta^T_{U|1\underline{W}Z}$ is the coefficient of W in the regression of U on 1, W and Z. Note that (3.3) implies that $\beta_{X|\underline{1}WZ} = \beta_{U|\underline{1}WZ}$, $\beta_{X|1\underline{W}Z} = \beta_{U|1\underline{W}Z}$ and $\beta_{X|1W\underline{Z}} = \beta_{U|1W\underline{Z}}$.

3.1. Instrumental Variable Estimation in Simple Binary Models

Our approach to instrumental variable estimation in binary regression models is most easily explained and understood in the context of a simple binary regression model with a single scalar instrument. We will study this model in detail in later sections, making comparisons with, and exploiting the familiarity of linear model instrumental variable estimation.

We first review linear model instrumental variable estimation in its simplest form. Consider a model without covariates for which the regression of Y on the scalar predictor U is linear,

$$E(Y \mid U) = \beta_{Y|\underline{1}U} + \beta_{Y|1\underline{U}}U,$$

with a scalar instrument W.

In this case the instrumental variable estimators of $\beta_{Y|\underline{1}U}$ and $\beta_{Y|1\underline{U}}$ are

$$\hat\beta_{Y|\underline{1}U} = \bar{Y} - \hat\beta_{Y|1\underline{U}}\bar{X}, \qquad \hat\beta_{Y|1\underline{U}} = \frac{\sum_i (Y_i - \bar{Y})(W_i - \bar{W})}{\sum_i (X_i - \bar{X})(W_i - \bar{W})},$$

(Fuller, 1987, p. 53). Equivalent expressions are

$$\hat\beta_{Y|\underline{1}U} = \hat\beta_{Y|\underline{1}W} - \frac{\hat\beta_{X|\underline{1}W}\hat\beta_{Y|1\underline{W}}}{\hat\beta_{X|1\underline{W}}} \qquad \hat\beta_{Y|1\underline{U}} = \frac{\hat\beta_{Y|1\underline{W}}}{\hat\beta_{X|1\underline{W}}}, \tag{3.5}$$

where $\hat\beta_{Y|\underline{1}W}$, $\hat\beta_{X|\underline{1}W}$, $\hat\beta_{Y|1\underline{W}}$ and $\hat\beta_{X|1\underline{W}}$ are the least squares estimates of the intercepts and slopes in the linear regressions of Y on W, and X on W respectively.

3.2. Approximate Instrumental Variable Estimation in Simple Regression Models

Equation (3.5) suggests an immediate generalization to binary regression models simply by taking $\hat\beta_{Y|\underline{1}W}$ and $\hat\beta_{Y|1\underline{W}}$ to be the estimated coefficients of 1 and W when the regression model, $m()$, is fit to $\{Y_i, W_i\}_{i=1}^n$. Using (3.4) and (3.2) the approximation

$$\begin{aligned} E(Y \mid W) &\approx m(\beta_{Y|\underline{1}U} + \beta_{Y|1\underline{U}}E(U \mid W)) \\ &= m(\beta_{Y|\underline{1}U} + \beta_{Y|1\underline{U}}\beta_{U|\underline{1}W} + \beta_{Y|1\underline{U}}\beta_{U|1\underline{W}}W) \end{aligned} \tag{3.6}$$

shows that $\hat\beta_{Y|\underline{1}W}$ and $\hat\beta_{Y|1\underline{W}}$ are approximately consistent for $\beta_{Y|\underline{1}U} + \beta_{Y|1\underline{U}}\beta_{U|\underline{1}W}$ and $\beta_{Y|1\underline{U}}\beta_{U|1\underline{W}}$ respectively. Under (3.2) and (3.3), $\hat\beta_{X|\underline{1}W}$ and $\hat\beta_{X|1\underline{W}}$ are consistent for $\beta_{U|\underline{1}W}$ and $\beta_{U|1\underline{W}}$ respectively and thus the instrumental variable estimators (3.5) are approximately consistent for $\beta_{Y|\underline{1}U}$ and $\beta_{Y|1\underline{U}}$ in simple binary regression models. The instrumental variable estimators obtained by this method are denoted with superscript '$L1$,' *e.g.*, $\hat\beta^{L1}_{Y|\underline{1}U}$.

The validity of this procedure depends on the quality of the approximation in (3.6), which in turn is a function of the nonlinearity in $m()$ and the size of the residual variance, $V_{UU|W}$. In particular, the approximation is exact when $m()$ is linear or $V_{UU|W} = 0$.

When there is curvature in $m()$ the simple approximation in (3.6) may not be adequate. The approximations

$$E(Y \mid W) \approx \tilde m_i\left(\beta_{Y|\underline{1}U} + \beta_{Y|1\underline{U}}E(U \mid W),\ \beta^2_{Y|1\underline{U}}V_{UU|W}\right) \qquad i = 1,\ 2,\ 3 \tag{3.7}$$

where

$$\begin{aligned} \tilde m_1(t,\kappa) &= m(t) + (\kappa/2)m^{(2)}(t), \\ \tilde m_2(t,\kappa) &= m\left(t + (\kappa/2)\frac{m^{(2)}(t)}{m^{(1)}(t)}\right), \\ \tilde m_3(t,\kappa) &= m\left(\frac{t}{\sqrt{1-\kappa m^{(2)}(t)/tm^{(1)}(t)}}\right) \end{aligned}$$

account for nonlinearity in $m()$ and are equivalent in the sense that the approximation errors are of the same order of magnitude for general $m()$. The first approximation is obtained by a Taylor-series expansion. The second and third are range-preserving versions of the Taylor-series approximation. The third requires that $\kappa m^{(2)}(t)/\{tm^{(1)}(t)\} < 1$. Our interest lies primarily in $\tilde{m}_2$ since it is range preserving yet does not require the additional restrictions that $\tilde{m}_3$ does. However, $\tilde{m}_3$ is of interest because it yields exact mean functions in Probit models when U, X and W are jointly normally distributed.

Although the approximations $\tilde{m}_i(t, \kappa)$ improve upon (3.6), the strategy of fitting the approximate model to $\{Y_i, W_i\}_{i=1}^n$ and solving for estimates of $\beta_{Y|\underline{1}U}$ and $\beta_{Y|1\underline{U}}$ is not always possible. For example, consider Probit regression for which $m(t) = \Phi(t)$, the standard normal distribution function. Then $m^{(2)}(t)/m^{(1)}(t) = -t$, and the approximation provided by $\tilde{m}_3$ turns out to be exact for $(U,\ X,\ W)$ jointly normal. That is

$$E(Y \mid W) = \Phi\left(\frac{\beta_{Y|\underline{1}U} + \beta_{Y|1\underline{U}}\beta_{U|\underline{1}W} + \beta_{Y|1\underline{U}}\beta_{U|1\underline{W}}W}{\sqrt{1 + \beta^2_{Y|1\underline{U}}V_{UU|W}}}\right). \tag{3.8}$$

Inspection of (3.8) shows that without additional approximations it is not possible to recover estimates of either $\beta_{Y|\underline{1}U}$ or $\beta_{Y|1\underline{U}}$ from a Probit regression of Y on W and a linear regression of X on W.

Buzas and Stefanski (1992) show that $\beta_{Y|\underline{1}U}$ and $\beta_{Y|1\underline{U}}$ are identifiable in this Probit measurement error model when an instrument is available. Thus the deficiency in the procedure just described is not intrinsic to the Probit model nor is it due to the use of an approximation since $\tilde{m}_3$ is exact for this model. The deficiency is a consequence of our approach to generalizing instrumental variable estimation. The problem arises because (3.5) is not the best way to describe instrumental variable estimation for the purpose of generalization to generalized linear models.

3.3. A Second Look at Instrumental Variable Estimation in Simple Linear Regression

Consider again the linear model set forth at the start of Section 3.1. For $(Y,\ U,\ W,\ X)$ jointly normal, the regression of Y on W *and* X is

$$\begin{aligned} E(Y \mid W,\ X) &= E\big(E(Y \mid U) \mid W,\ X\big) \\ &= \beta_{Y|\underline{1}U} + \beta_{Y|1\underline{U}}E(U \mid W,\ X) \\ &= \beta_{Y|\underline{1}U} + \beta_{Y|1\underline{U}}(1-\Delta)\beta_{U|\underline{1}W} + \beta_{Y|1\underline{U}}(1-\Delta)\beta_{U|1\underline{W}}W + \beta_{Y|1\underline{U}}\Delta X \end{aligned}$$

where $\Delta = V_{UU|W}/V_{XX|W}$. In terms of the parameters in the regression of Y on W and X, we have $\beta_{Y|1\underline{U}} = \beta_{Y|1W\underline{X}} + \beta_{Y|1\underline{W}X}/\beta_{U|1\underline{W}}$ and $\beta_{Y|\underline{1}U} = \beta_{Y|\underline{1}WX} - \beta_{U|\underline{1}W}\beta_{Y|1\underline{W}X}/\beta_{U|1\underline{W}}$, and thus $\beta_{Y|\underline{1}U}$ and $\beta_{Y|1\underline{U}}$ are consistently estimated by

$$\hat{\beta}_{Y|\underline{1}U} = \hat{\beta}_{Y|\underline{1}WX} - \frac{\hat{\beta}_{X|\underline{1}W}\hat{\beta}_{Y|1\underline{W}X}}{\hat{\beta}_{X|1\underline{W}}}, \qquad \hat{\beta}_{Y|1\underline{U}} = \hat{\beta}_{Y|1W\underline{X}} + \frac{\hat{\beta}_{Y|1\underline{W}X}}{\hat{\beta}_{X|1\underline{W}}}. \tag{3.9}$$

The definitions in (3.5) and (3.9) are not inconsistent, for it can be shown that the defining expressions are equivalent.

3.4. Approximate Instrumental Variable Estimation in Simple Binary Regression Models (cont.)

With regard to nonlinear models the implication is that (3.7) should be replaced with the approximations,

$$E(Y \mid W,\ X) \approx \tilde{m}_i\left(\beta_{Y|\underline{1}U} + \beta_{Y|1\underline{U}}E(U \mid W,\ X),\ \beta^2_{Y|1\underline{U}}V_{UU|WX}\right), \tag{3.10}$$

for $i = 1,\ 2,\ 3$.

For a given $m()$ the approximation errors in (3.7) and (3.10) depend on the size of $V_{UU|W}$ and $V_{UU|WX}$ respectively. Since $V_{UU|W} \geq V_{UU|WX}$, the latter approximations are generally better. More importantly, (3.10) usually preserves identifiability of $\beta_{Y|\underline{1}U}$ and $\beta_{Y|1\underline{U}}$ with no additional assumptions or approximations beyond the initial approximation. For example, in the Probit model discussed previously, $\tilde{m}_3$ from (3.10) yields the exact regression

$$E(Y \mid W,\ X) = \Phi\left(\frac{\beta_{Y|\underline{1}U} + \beta_{Y|1\underline{U}}(1-\Delta)\beta_{U|\underline{1}W} + \beta_{Y|1\underline{U}}(1-\Delta)\beta_{U|1\underline{W}}W + \Delta\beta_{Y|1\underline{U}}X}{k}\right) \tag{3.11}$$

where $k = \sqrt{1 + \beta^2_{Y|1\underline{U}}V_{UU|WX}}$.

In terms of the coefficients of 1, W and X in the Probit regression of Y on 1, W and X, (3.11) induces the relationships

$$\begin{aligned} \beta_{Y|\underline{1}WX} &= k^{-1}(\beta_{Y|\underline{1}U} + \beta_{Y|1\underline{U}}(1-\Delta)\beta_{U|\underline{1}W}) \\ \beta_{Y|1\underline{W}X} &= k^{-1}\beta_{Y|1\underline{U}}(1-\Delta)\beta_{U|1\underline{W}} \\ \beta_{Y|1W\underline{X}} &= k^{-1}\Delta\beta_{Y|1\underline{U}}. \end{aligned} \tag{3.12}$$

The parameters $\beta_{Y|\underline{1}U}$ and $\beta_{Y|1\underline{U}}$ are recovered from (3.12) via the formulas

$$\beta_{Y|\underline{1}U} = \frac{1}{k^*}\left(\beta_{Y|\underline{1}WX} - \frac{\beta_{Y|1\underline{W}X}\beta_{U|\underline{1}W}}{\beta_{U|1\underline{W}}}\right), \qquad \beta_{Y|1\underline{U}} = \frac{1}{k^*}\left(\frac{\beta_{Y|1\underline{W}X}}{\beta_{U|1\underline{W}}} + \beta_{Y|1W\underline{X}}\right),$$

where

$$k^* = \sqrt{1 - \beta_{Y|1W\underline{X}} V_{XX|W} \beta_{Y|1\underline{W}X} / \beta_{U|1\underline{W}}}\,.$$

The Probit model is studied in detail elsewhere (Buzas and Stefanski, 1992). Our purpose in highlighting it here is to point out that the approximations in (3.7) entail a loss of information that is not incurred with (3.10), thus suggesting that approximate instrumental variable estimation should start with an approximation to the regression of Y on W and X and not with the regression of Y on W alone.

However, the approximation in (3.6) and the estimators it renders, (3.5), are not without application. When the curvature in $m()$ is slight and/or $\beta_{Y|1\underline{U}}$ is small the regression equation $m(\beta_{Y|\underline{1}U} + \beta_{Y|1\underline{U}}\, u)$ is approximately linear in u and therefore the approximations in (3.6), (3.7) and (3.10) are all of comparable quality. In such cases the simple estimators (3.5) should be adequate and may be superior in terms of mean squared error since estimators derived from (3.10) will generally be more variable, the price paid for having smaller bias.

3.5. Refinements

For general mean functions (3.10) may be very complicated and it is unlikely that simple analytic relationships between parameters, such as those for the linear and Probit models, can be obtained. We will investigate two approaches that sidestep this problem. The first assumes that $\beta^2_{Y|1\underline{U}} V_{UU|WX}$ is negligible. In this case all three approximations (3.10) are identical and yield the approximate mean model

$$E(Y \mid W,\ X) \approx m\left(\beta_{Y|\underline{1}U} + \beta_{Y|1\underline{U}} E(U \mid W,\ X)\right).$$

Letting $\hat{\beta}_{Y|\underline{1}WX}$, $\hat{\beta}_{Y|1\underline{W}X}$ and $\hat{\beta}_{Y|1W\underline{X}}$ denote the estimated coefficients of 1, W and X when the mean model $m()$ is fit to $\{Y_i, W_i, X_i\}_{i=1}^n$, estimates of $\beta_{Y|\underline{1}U}$ and $\beta_{Y|1\underline{U}}$ are found via (3.9). These instrumental variable estimators are denoted by superscript $L2$, *e.g.*, $\hat{\beta}^{L2}_{Y|1\underline{U}}$.

For our second approach we consider only $\tilde{m}_2$ from (3.10) which yields the approximate mean model

$$E(Y \mid W,\ X) \approx m\Bigg(\beta_{Y|\underline{1}U} + \beta_{Y|1\underline{U}}E(U \mid W,\ X) + \frac{V_{UU|WX}\beta^2_{Y|1\underline{U}}}{2}q\big(\beta_{Y|\underline{1}U} + \beta_{Y|1\underline{U}}E(U \mid W,\ X)\big)\Bigg) \tag{3.13}$$

where $q(t) = m^{(2)}(t)/m^{(1)}(t)$.

If $q(t)$ is approximated by a linear function, say $a + bt$, then (3.13) becomes

$$E(Y \mid W,\ X) \approx m\left(\beta_{Y|\underline{1}WX} + \beta_{Y|1\underline{W}X}W + \beta_{Y|1W\underline{X}}X\right) \tag{3.14}$$

where

$$\begin{aligned}
\beta_{Y|\underline{1}WX} &= k\big(\beta_{Y|\underline{1}U} + \beta_{Y|1\underline{U}}(1-\Delta)\beta_{U|\underline{1}W}\big) + (a/2)\beta^2_{Y|1\underline{U}}V_{UU|WX} \\
\beta_{Y|1\underline{W}X} &= k\beta_{Y|1\underline{U}}(1-\Delta)\beta_{U|1\underline{W}} \\
\beta_{Y|1W\underline{X}} &= k\Delta\beta_{Y|1\underline{U}}
\end{aligned}$$

and $k = 1 + (b/2)\beta^2_{Y|1\underline{U}}V_{UU|WX}$. These equations may have multiple solutions depending on the value of $\beta^2_{Y|1\underline{U}}V_{UU|WX}$. However, we're working under the assumption that $\beta^2_{Y|1\underline{U}}V_{UU|WX}$ is small and this allows us to isolate a unique solution as described below.

The parameters $\beta_{Y|\underline{1}U}$ and $\beta_{Y|1\underline{U}}$ are recovered via the formulas

$$\begin{aligned}
\beta_{Y|\underline{1}U} &= \frac{1}{k^*}\left(\beta_{Y|\underline{1}WX} - \frac{\beta_{U|\underline{1}W}\beta_{Y|1\underline{W}X}}{\beta_{U|1\underline{W}}} - \frac{a}{b}(k^*-1)\right), \\
\beta_{Y|1\underline{U}} &= \frac{1}{k^*}\left(\beta_{Y|1W\underline{X}} + \frac{\beta_{Y|1\underline{W}X}}{\beta_{U|1\underline{W}}}\right),
\end{aligned} \tag{3.15}$$

where k^* is the largest positive solution of the equation

$$2k^2(k-1) - bV^* = 0$$

and

$$V^* = \frac{\beta_{Y|1\underline{W}X}V_{XX|W}\beta_{Y|1W\underline{X}}}{\beta_{U|1\underline{W}}}. \tag{3.16}$$

The indicated root exists and is greater than 2/3 provided and $-b\beta^2_{Y|1\underline{U}}V_{UU|WX} < 8/27$. The latter inequality is always satisfied for $b > 0$. For $b < 0$, the inequality may be loosely interpreted as a condition on the size of $V_{UU|WX}$ and the curvature of $m(\beta_{Y|\underline{1}U} + \beta_{Y|1\underline{U}}\ u)$ as a function of u, for the approximation (3.14) to be acceptable.

There are some models for which $q(t)$ is exactly linear. This occurs when

$$m(t) = c \int_0^t \exp\left(ax + b\frac{x^2}{2}\right)\ dx + d$$

for constants a, b, c and d. Important special cases are: $\Phi(t)$, $(a = 0,\ b = -1)$; and $1 - \exp(-t)$ and $\exp(-t)$, $(a = -1,\ b = 0)$.

When $a = -1,\ b = 0$, (3.14) is exact for jointly normal $(U,\ W,\ X)$ provided the domain of $m()$ is $(-\infty,\ \infty)$, and the appropriate inversion formulas are obtained from (3.15) by passing to the limit as $b \to 0$, noting that $(k^* - 1)/b \to V^*/2$. Technically, when $a = -1$ and $b = 0$, the above stated condition on the domain of $m()$ excludes binary regression models in light of the natural range constraints on $m()$. However, in Section 5 we point out that the mean function approximations described in this paper apply equally well to non-binary regression models, and in this more general context the conditions for exactness of (3.14) are relevant.

For many models $q(t)$ is not a linear function of t throughout its domain, but may be sufficiently linear locally to justify a linear approximation. The linearizing parameters a and b then should be chosen accordingly. Thus our second proposal has the following steps:

- Obtain $\hat\beta_{Y|\underline{1}WX}$, $\hat\beta_{Y|1\underline{W}X}$ and $\hat\beta_{Y|1W\underline{X}}$, the estimated coefficients of 1, W and X when the mean model $m()$ is fit to $\{Y_i, W_i, X_i\}_{i=1}^n$;
- Obtain $\hat\beta_{X|\underline{1}W}$, $\hat\beta_{X|1\underline{W}}$ and $\hat V_{XX|W}$, from a linear regression of X on 1 and W;
- With $\hat Q_i = \hat\beta_{Y|\underline{1}WX} + \hat\beta_{Y|1\underline{W}X} W_i + \hat\beta_{Y|1W\underline{X}} X_i$, compute $\hat a$ and $\hat b$ via a least-squares regression of $q(\hat Q_i)$ on $\hat Q_i$;
- Obtain $\hat k^*$ as the largest solution, or approximate solution as described below, to the equation

$$2k^2(k-1) - \hat b \hat V^* = 0 \tag{3.17}$$

where $\hat V^*$ is a method-of-moments estimate derived from (3.16);
- Compute the instrumental variable estimators $\hat\beta^{L3}_{Y|\underline{1}U}$ and $\hat\beta^{L3}_{Y|1\underline{U}}$ by substituting estimated parameters, including $\hat a$, $\hat b$ and $\hat k^*$, into the right hand sides of (3.15), recalling that $\hat\beta_{U|\underline{1}WZ} = \hat\beta_{X|\underline{1}WZ}$, $\hat\beta_{U|1\underline{W}Z} = \hat\beta_{X|1\underline{W}Z}$, and $\hat\beta_{U|1W\underline{Z}} = \hat\beta_{X|1W\underline{Z}}$.

The third step provides the local linear approximation to $q(t)$. It could also be obtained by a Taylor-series approximation. However, the regression approximation is better suited to the task.

It remains to discuss the determination of $\hat{k}^*$. When $\hat{V}^* \leq 0$ take $\hat{k}^* = 1$. For $\hat{V}^* > 0$ the definition of $\hat{k}^*$ depends on the behavior of the estimated equation (3.17). If $\hat{b}\hat{V}^* < -8/27$, then the estimated equation has no positive real roots and the closest we can come to a positive root is to take $\hat{k}^* = 2/3$; if $-8/27 < \hat{b}\hat{V}^* \leq 0$, then the estimated equation has two roots in the unit interval, one on either side of 2/3, and $\hat{k}^*$ is the larger of the two roots; if $\hat{b}\hat{V}^* > 0$, then the estimated equation has one positive root greater than 1 and $\hat{k}^*$ is this root.

Since our mean function is only approximate, and thus so too is the determination of $\hat{k}^*$, it seems likely that the exact computation of $\hat{k}^*$ described above can be replaced by a simpler approximate computation. Three candidates are $\hat{k}^*_{A1} = 1 + \frac{1}{2}\hat{b}\hat{V}^*$, $\hat{k}^*_{A2} = \sqrt{1 + \hat{b}\hat{V}^*}$ and $\hat{k}^*_{A3} = \exp(\frac{1}{2}\hat{b}\hat{V}^*)$. The first is derived from a linearization of (3.17) for k near 1; the second is the correction factor from the exact analysis of the Probit model; the third is suggested by the work of Burr (1988). The latter two may also be viewed as solutions to an approximation to (3.17) when k is near 1. All three approximations should be set to one when when $\hat{V}^* \leq 0$.

3.6. Simulation Results

Six estimators have been defined in the previous section. The estimators derived from (3.5) are referred to as the $L1$ estimators, those from (3.9) as the $L2$ estimators, whereas those from (3.15) as the $L3$ estimators. There are four variants of the latter depending on how $\hat{k}^*$ is calculated.

We have performed several simulation studies to gain insight into the behavior of the estimators described above. The results suggest the following general conclusions. First, is that the greatest reductions in bias are found with $L3$, the least reductions with $L1$ with variance generally varying inversely to bias. Smaller sample sizes and/or mean functions with small curvature, *i.e.*, when β_1 is small, favor $L1$ whereas $L3$ is favored in large samples and when curvature is not negligible.

Second, is that $L1$ does a poor job of correcting for bias in the intercept parameter; $L2$ does much better, but only the $L3$ estimators eliminate bias more or less completely in general. This is not a problem if only the predictor regression coefficient is of interest. However, in cases where the estimated response function is the primary estimand, bias in the intercept is problematic.

A final general conclusion is that the four variants of $L3$ are nearly indistinguishable, although the variant with $\hat{k}^*$ calculated from (3.17) seems to have a slight advantage in larger samples.

The results from one simulation study are presented in Table 1. The model for the study was

$$\Pr(Y = 1 \mid U) = F(\beta_0 + \beta_1 U), \qquad F(t) = (1 + e^{-t})^{-1};$$

$$X \mid U \sim \mathcal{N}(U,\ 3/4);$$

$$U \mid W \sim \mathcal{N}(4W/7,\ 3/7);$$

$$W \sim \mathcal{N}(0,\ 7/4)$$

With this specification $U \sim \mathcal{N}(0,\ 1)$ and $W \mid U \sim \mathcal{N}(U,\ 3/4)$. Thus the measurement error variance is $3/4 \times$ the variance in the true predictor, and the instrument, W, is of comparable quality to the measured value X. In fact W is a replicate measurement although this information is not used. The case where the instrument and measured predictor are replicates or near replicates is of some interest as explained in the introduction.

In the simulation study $\beta_0 = -2.25$, β_1 was varied over the set $\{0.371,\ 0.742,\ 1.484\}$, and sample size was fixed at $n = 1500$. The model specification with $\beta_1 = 0.371$ is a rough match to a subset of the Framingham data. The largest value of β_1 corresponds to a relative risk of approximately 45. This is extreme by epidemiologic standards. Thus the model with $\beta_1 = 1.484$ is not representative of typical epidemiologic applications, but it does provide a good test of the quality of the the various approximations because of the greater curvature in the model. Ten thousand data sets were generated and analyzed.

Displayed in Table 1 are Monte Carlo estimates of the bias and mean absolute error of the estimators of β_0 and β_1. Also reported are two statistics that reflect the performance of the estimated response curve over a range of u values. The mean supremum absolute relative error (MSARE) of $p(u)$ is the mean of

$$\sup_{0 \le u \le 3} \frac{|\, \hat{p}(u) - p(u) \,|}{p(u)}$$

where $p(u) = F(\beta_0 + \beta_1 u)$ and $\hat{p}(u)$ is the estimated mean function. The range (0, 3) corresponds to 'high-risk' cases when U is regarded as a risk factor and thus is likely to include the range of interest in many applications.

Similarly mean supremum absolute relative error (MSARE) of $p'(u)$ is the mean of

$$\sup_{0 \le u \le 3} \frac{|\, \hat{p}'(u) - p'(u) \,|}{p'(u)}$$

where the prime " $'$ " denotes differentiation with respect to u. In the simulations neither MSARE measure was a supremum, but rather was calculated over a grid of eleven equally-spaced points between 0 and 3.

The number of Monte Carlo replications was so great (10,000) that standard errors are virtually negligible, thus we elected not to include them in the table. With one exception, all of the pairwise comparisons one might make between estimated quantities are statistically significant. The exception is highlighted in the table using the standard convention of joining the means by a common underline.

All of the general trends reported above are manifest in the table. However, we note that not all of the statistically significant differences in the table are of practical significance. For example, when $\beta_1 = 0.371$, none of the biases in β_1 are statistically significantly different from 0, even though all are statistically significantly different from one another. For the larger values of β_1 the biases of $\hat{\beta}_1^{L1}$ are detectably different from zero whereas those of $\tilde{\beta}_1^{L2}$ and $\hat{\beta}_1^{L3}$ are not.

A fair summary of the table is that for estimation of β_1 all three methods are comparable at the two smaller values of β_1, whereas $L3$ enjoys an advantage when $\beta_1 = 1.484$. For estimation of β_0 the story is different with $L2$ and especially $L3$ dominating for all three model specifications. This results in in a slight edge for $L2$ and $L3$ when estimating the response function, although here the advantages are noteworthy only at the larger value of β_1.

4. MULTIPLE REGRESSION MODELS WITH COVARIATES AND MULTIPLE INSTRUMENTS

General formulas for the various instrumental variable estimators are straightforward extensions of those for the simple regression models. The formulas are presented in Section 4.1 without derivation. In Section 4.2 an example is discussed.

4.1. Estimation Formulas

For multiple regression models with covariates and multiple instruments the estimators $(\hat{\beta}^{L1}_{Y|\underline{1}UZ}$, $\hat{\beta}^{L1}_{Y|1\underline{U}Z}$, $\hat{\beta}^{L1}_{Y|1U\underline{Z}})$ and $(\hat{\beta}^{L2}_{Y|\underline{1}UZ}$, $\hat{\beta}^{L2}_{Y|1\underline{U}Z}$, $\hat{\beta}^{L2}_{Y|1U\underline{Z}})$ are obtained from general versions of (3.5) and (3.9).

With superscript '$-$' denoting generalized inverse, *i.e.*, $M^{-} = (M^TM)^{-1}M^T$, the general version of (3.5) is

$$\begin{aligned}
\hat\beta_{Y|\underline{1}UZ} &= \hat\beta_{Y|\underline{1}WZ} - \hat\beta_{X|\underline{1}WZ}\hat\beta^{-}_{X|1\underline{W}Z}\hat\beta_{Y|1\underline{W}Z};\\
\hat\beta_{Y|1\underline{U}Z} &= \hat\beta^{-}_{X|1\underline{W}Z}\hat\beta_{Y|1\underline{W}Z};\\
\hat\beta_{Y|1U\underline{Z}} &= \hat\beta_{Y|1W\underline{Z}} - \hat\beta_{X|1W\underline{Z}}\hat\beta^{-}_{X|1\underline{W}Z}\hat\beta_{Y|1\underline{W}Z}.
\end{aligned} \tag{4.1}$$

Equation (3.9) generalizes to

$$\begin{aligned}
\hat\beta_{Y|\underline{1}UZ} &= \hat\beta_{Y|\underline{1}WXZ} - \hat\beta_{X|\underline{1}WZ}\hat\beta^{-}_{X|1\underline{W}Z}\hat\beta_{Y|1\underline{W}XZ};\\
\hat\beta_{Y|1\underline{U}Z} &= \hat\beta_{Y|1W\underline{X}Z} + \hat\beta^{-}_{X|1\underline{W}Z}\hat\beta_{Y|1\underline{W}XZ};\\
\hat\beta_{Y|1U\underline{Z}} &= \hat\beta_{Y|1WX\underline{Z}} - \hat\beta_{X|1W\underline{Z}}\hat\beta^{-}_{X|1\underline{W}Z}\hat\beta_{Y|1\underline{W}XZ}.
\end{aligned} \tag{4.2}$$

The estimators $(\hat\beta^{L3}_{Y|\underline{1}UZ}, \hat\beta^{L3}_{Y|1\underline{U}Z}, \hat\beta^{L3}_{Y|1U\underline{Z}})$ are obtained by substituting estimators into the right hand sides of

$$\begin{aligned}
\beta_{Y|\underline{1}UZ} =& \frac{1}{k^*}\left(\beta_{Y|\underline{1}WXZ} - \beta_{U|\underline{1}WZ}\beta^{-}_{U|1\underline{W}Z}\beta_{Y|1\underline{W}XZ} - \frac{a}{b}(k^*-1)\right);\\
\beta_{Y|1\underline{U}Z} =& \frac{1}{k^*}\left(\beta_{Y|1W\underline{X}Z} + \beta^{-}_{U|1\underline{W}Z}\beta_{Y|1\underline{W}XZ}\right);\\
\beta_{Y|1U\underline{Z}} =& \frac{1}{k^*}\left(\beta_{Y|1WX\underline{Z}} - \beta_{U|1W\underline{Z}}\beta^{-}_{U|1\underline{W}Z}\beta_{Y|1\underline{W}XZ}\right).
\end{aligned} \tag{4.3}$$

Equation (4.3) generalizes (3.15). In (4.3) estimates of a and b are the least-squares estimates of intercept and slope from the fit of $q(\hat Q_i)$ on $\hat Q_i = \hat\beta_{Y|\underline{1}WXZ} + \hat\beta^T_{Y|1\underline{W}XZ}W_i + \hat\beta^T_{Y|1W\underline{X}Z}X_i + \hat\beta^T_{Y|1WX\underline{Z}}Z_i$. Also k^* solves (3.17) with $\hat V^*$ estimated by

$$\hat\beta^T_{Y|1\underline{W}XZ}\left(\hat\beta^{-}_{X|1\underline{W}Z}\right)^T \hat V_{XX|WZ}\hat\beta_{Y|1W\underline{X}Z}.$$

4.2. An Example

We now describe an example application of the proposed methods. Data for the example are from the Framingham heart study. The subset of the data used in this example contained the variables: CHD, an indicator of coronary heart disease during an eight year period following the second exam; A, age at second exam; S, an indicator of smoking/nonsmoking; and $P_{i,j}$, systolic blood pressure measurements made by the j^{th} examiner during the i^{th} exam, $j = 1, 2$, $i = 2, 3$. Exams two and three took place two and four years post baseline respectively. Records with incomplete information were dropped from the data set, leaving a total of 1660 observations of which 134 are cases of CHD.

The data indicate that the two measurements from the second exam are not replicates — measurements by the second examiner tend to be lower. Thus it is not immediately clear how the two measurements from the second exam should be combined into a single measurement. Since our analysis is for illustrative purposes only, we avoided this issue by using only the second measurement from the first exam.

For the example we took Y = CHD, $X = \ln(P_{2,2})$, Z = (A, S), and $W = (\ln P_{3,1}, \ln P_{3,2})$. Thus U is defined operationally to be the natural logarithm of the 'true' (= long term average) systolic blood pressure at the time of the second exam. Interest lies in fitting a logistic regression model of Y on U and Z.

Our definition of W as an instrument is valid only if the errors in W as possibly biased measurements of U are independent of the measurement error in X. This seems reasonable given the two year time span between exams.

The working model for the moments of X and W given U is thus

$$E\begin{pmatrix} X \\ W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} U \\ \beta_{W_1|\underline{1}U} + \beta_{W_1|\underline{U}}U \\ \beta_{W_2|\underline{1}U} + \beta_{W_2|\underline{U}}U \end{pmatrix}, \qquad V\begin{pmatrix} X \\ W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} \sigma_{XX} & 0 & 0 \\ 0 & \sigma_{W_1W_1} & \sigma_{W_1W_2} \\ 0 & \sigma_{W_1W_2} & \sigma_{W_2W_2} \end{pmatrix}.$$

The sample mean and covariance matrix of $(X,\ W_1,\ W_2)$ are

$$\hat{\mu} = \begin{pmatrix} 4.859 \\ 4.866 \\ 4.847 \end{pmatrix}, \qquad \hat{\Sigma} = \begin{pmatrix} .021 & .015 & .015 \\ .015 & .022 & .019 \\ .015 & .019 & .021 \end{pmatrix}.$$

Under the assumptions that $\beta_{W_2|\underline{U}} = 1$ and $\sigma_{W_2W_2} = \sigma_{XX}$ (and/or $\beta_{W_1|\underline{U}} = 1$ and $\sigma_{W_1W_1} = \sigma_{XX}$), the instruments are, apart from an estimable constant, replicates. In this case the covariance of X and W_1 (and/or W_2) provides an estimate of $n^{-1}\sum(U_i - \bar{U}.)^2$, equal to .015, from which the measurement error variance is estimated to be .006 (=.021 - .015). With this estimate of the measurement error variance the parameters in the logistic regression of Y on U and Z can be estimated using the so-called sufficiency estimator described in Stefanski and Carroll (1985, 1987). Since there is some interest in the replicate-measurements model, we computed the sufficiency estimate in addition to the proposed instrumental variable estimators for purposes of comparison.

Parameter estimates are displayed in Table 2. The naive estimate is obtained from the logistic regression of Y on X and Z. Attenuation in the coefficient of ln(SBP) is apparent. The differences between the three IV estimators and the sufficiency estimator are slight relative to their differences from the naive estimator and, without further assumptions, there is no basis on which to distinguish between them. However, if the assumptions underlying the components of variance analysis are valid, *i.e.*, the instruments are essentially replicates, then asymptotic efficiency arguments suggest that the sufficiency estimator is best. Using the sufficiency estimator as a benchmark in turn suggests the conclusion that the IV2 and IV3 estimators marginally outperformed the IV1 estimator in this example.

We have not yet discussed the estimation of standard errors for the instrumental variable estimators. IV1 and IV2 are functions of logistic and linear regression coefficients and their asymptotic distributions can be derived by the Δ-method. IV3 is a somewhat more complicated estimator although its asymptotic distribution is also amenable to standard large-sample approximations. However, the calculations are tedious, not very interesting and will not be described here.

In our work we have opted to compute jackknife standard errors. Table 3 displays jackknife standard errors for the IV3 estimator, and jackknife, information and information-sandwich standard errors for the sufficiency estimator. The latter were computed as a check on the differences between the jackknife and analytically computed (approximate) standard errors. There are no surprises in Table 3. The three sets of standard errors for the sufficiency estimator are all similar suggesting that the jackknife standard errors for IV3 are not too different from those that would be obtained by the Δ-method.

5. EXTENSIONS AND CONCLUDING REMARKS

Our interest lies primarily with binary regression although it is worth noting that the methods described in this paper have application to other generalized linear models. Binary regression offers the simplicity of a variance function that is completely determined by the mean function. An approximation of the former yields an approximation of the latter. This is not the case with non-binary regression models.

The mean function approximations derived in Sections 3 generalize to non-binary models without modification. Thus the main obstacle to extending our methods to non-binary models is the simultaneous approximation of a variance function for the approximate regression of Y on X, W and Z. The latter is not difficult as approximations can be derived along the lines of Carroll and Stefanski (1990). However, the resulting mean/variance function models generally will not correspond to simple generalized linear models even when the original model does. This complicates the estimation procedure. For example, the first-order approximation to the variance function in Poisson regression will not equal the first-order approximation to the mean function, there is extra-Poisson variation introduced by the variation of U around $E(U \mid W, X)$. Thus the added complications are similar to those encountered in fitting over-dispersed models.

BIBLIOGRAPHY

Amemiya, Y. (1985a), "Instrumental variable estimator for the nonlinear errors-in-variables model," *J. Econometrics*, 28, 273–289.

Amemiya, Y. (1985b), "Two-stage instrumental variable estimator for the nonlinear errors-in-variables model," *J. Econometrics*, 44, 311–332.

Amemiya, Y. (1985c), "Instrumental variable estimation of the nonlinear measurement error model," *Statistical Analysis of Measurement Error Models and Applications*, 147–156, P. J. Brown and W. A. Fuller, eds., American Mathematical Society, Providence.

Burr, D. (1988), "On errors-in-variables in binary regression — Berkson case," *Journal of the American Statistical Association*, 83, 739–743.

Buzas, J. and Stefanski, L. A. (1992), "Instrumental variable estimation in Probit measurement error models," *in preparation.*

Carroll, R. J. and Stefanski, L. A. (1990), "Approximate quasi-likelihood estimation in models with surrogate predictors," *Journal of the American Statistical Association*, 85, 652–663.

Carroll, R. J. and Stefanski, L. A. (1992), "Measurement error and corrections for attenuation in meta-analyses," *Institute of Statistics Mimeo Series*, No. 2222, North Carolina State University.

Fuller, W. A. (1987), *Measurement Error Models*, New York: John Wiley.

MacMahon, S., Peto, R., Cutler, J., Collins, R., Sorlie, P., Neaton, J., Abbott, R., Godwin, J., Dyer, A. and Stamler, J. (1990), "Blood pressure, stroke, and coronary heart disease. Part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias," *The Lancet*, 335, 765–773.

Stefanski, L. A. and Carroll, R. J. (1985), "Covariate measurement error in logistic regression," *Annals of Statistics*, 13, 1335–1351.

Stefanski, L. A. and Carroll, R. J. (1987), "Conditional scores and optimal scores for generalized linear measurement error models," *Biometrika*, 74, 703-716.

Table 1. Results of the simulation study described in Section 3.6. The NAIVE estimator is from a logistic regression of Y on X. The other estimators are identified by IV1, IV2 and IV3 as in the text. Table entries are: BIAS=mean bias $\times$ 10; MAE=mean absolute error $\times$ 10; MSARE= mean sup absolute relative error $\times$ 10, (defined in Section 3.6).

Parameter			Estimator Naive	IV1	IV2	IV3
$\beta_1 = 0.371$	β_0	BIAS	0.18	0.18	0.06	-0.07
		MAE	0.72	0.72	0.71	0.73
	β_1	BIAS	-1.59	-0.01	-0.00	0.01
		MAE	1.60	0.93	0.94	0.94
	$p(x)$	MSARE	3.01	2.12	2.12	2.08
	$p'(x)$	MSARE	5.84	4.68	4.68	4.65
$\beta_1 = 0.742$	β_0	BIAS	0.85	0.85	0.50	-0.07
		MAE	1.02	1.03	0.86	0.81
	β_1	BIAS	-3.28	-0.17	-0.09	0.01
		MAE	3.28	0.93	0.93	0.96
	$p(x)$	MSARE	4.22	1.48	1.43	1.39
	$p'(x)$	MSARE	6.71	2.61	2.60	2.62
$\beta_1 = 1.484$	β_0	BIAS	3.01	3.00	1.98	-0.14
		MAE	3.01	3.00	2.00	1.01
	β_1	BIAS	-7.37	-1.72	-1.14	0.16
		MAE	7.37	1.81	1.39	1.17
	$p(x)$	MSARE	4.23	1.17	0.97	0.79
	$p'(x)$	MSARE	7.08	1.83	1.60	1.49

Table 2. Estimates from the Framingham data. Naive refers to the estimate obtained by fitting the logistic model to the observed data; IV1, IV2, IV3 and Sufficiency are defined in the text. Entries are estimated regression coefficients.

	Estimator				
Coefficient	Naive	IV1	IV2	IV3	Sufficiency
Intercept	-16.854	-21.252	-20.950	-20.989	-19.899
Ln(SBP)	2.291	3.020	2.959	2.963	2.946
Age	.057	.054	.054	.054	.054
Smoke	.622	.654	.653	.654	.638

Table 3. Standard errors for the IV3 instrumental variable estimator and the sufficiency estimator; nc = not computed.

	IV3 Instrument				Sufficiency			
	-20.988	2.963	.054	.654	-19.899	2.946	.054	.638
Jackknife	3.582	.840	.011	.249	3.354	.709	.011	.249
Information		nc			3.488	.734	.012	.248
Sandwich		nc			3.312	.700	.011	.244