Estimation in Binary Choice Models with Measurement Errors

Jochumzen, Peter

2003

Lund University, Department of Economics WP 2003:4
July 2003

Estimation in Binary Choice Models with Measurement Errors

David Edgerton and Peter Jochumzen

Department of Economics

Lund University

The authors wish to thank the Swedish Council for Research in the Humanities and Social Sciences (project F0353/97) and the Swedish Council for Social Research (project F0076/1998) for financial support. We also wish to thank those who have made helpful comments about earlier versions of this paper when it has been presented in Lund, Nottingham, Singapore and Sydney.

Communications can be sent to the authors
by *e-mail:* David.Edgerton@nek.lu.se or Peter.Jochumzen@nek.lu.se,
by *fax:* +46 - 46 222 4613, or
by *mail:* P.O. Box 7082, S - 220 07 Lund, Sweden.

ABSTRACT

In this paper we develop a simple maximum likelihood estimator for probit models where the regressors have measurement error. We first assume precise information about the reliability ratios (or, equivalently, the proxy correlations) of the regressors. We then show how reasonable bounds for the parameter estimates can be obtained when only imprecise information is available. The analysis is also extended to situations where the measurement error has non-zero mean and is correlated with the true values of the regressors. An extensive simulation study shows that the estimator works very well, even in quite small samples. Finally the method is applied to data explaining sick leave in Sweden.

Keywords: Measurement error; errors-in-variables; probit; binary choice; bounds.

JEL Classification: C25; C29

1) INTRODUCTION

Binary choice models are those where the dependent variable can only take one of two values. Such models are very common in economics. Examples include models that explain whether or not an individual is employed, whether or not a prospective buyer decides to purchase a particular house, whether or not a person registers as sick, *etc*. Models with *errors-in-variables* are those where the *explanatory* variable(s) are measured with error. When, *e.g.*, we estimate a production function it is quite possible - in fact quite probable - that the observed values of variables such as labor and capital include measurement error. A closely related problem concerns situations where some variables in a theoretical model may have no observable counterpart. An example of this is where we use unemployment as a proxy variable to measure the level of economic activity.

Both of the above areas have been investigated quite thoroughly on their own. Binary choice models are extremely common in the absence of measurement errors, while the analysis of the linear regression model with measurement errors, though not without problems, is quite standard. Introductory and advanced textbooks in econometrics usually have at least one section on each of these topics, and a number of specialized monographs in these areas are also available, the most cited of these probably being Maddala (1983) and Fuller (1987).

There are two reasons it is important to study this combination of problems: it is very common and it is very serious. If we ignore the presence of measurement errors or proxy variables in binary choice models, then our parameter estimates will be inconsistent. This means, for example, that we will obtain incorrect (poor) estimates of how the introduction of a qualification day into a sickness benefit system affects the probability of an individual reporting sick. Many political decisions depend on the correct estimation of such effects.

Very little has been done, however, with regard to models that include measurement errors *within* binary choice models. The most obvious general reference for this area is the book by Carroll *et al.* (1996), where the authors observe that there have only been a few examples of binary choice models with measurement errors in the literature, mostly within the areas of biology and medicine. The authors present three approximate methods ("regression calibration", "simulation extrapolation" and "approximated instrumental variables"), which have recently been implemented for generalized linear models in Stata 8. In a series of papers, Stefanski and Carroll (1985, 1987, 1990) have analyzed a logit binary choice model assuming either nonparametric or parametric (normally distributed) measurement errors, and assuming that the covariance matrix of the measurement errors is exactly known. These results are numerically quite complex. Kao and Schnell (1987a) have extended the results of Stefanski and Carroll (1985) to panel data models, and in Kao and Schnell (1987b) they develop similar results for the random effects probit model. Wansbeek and Meijer (2000) have suggested estimating a logit measurement error model using GMM. A number of recent articles have suggested various methods for estimating general nonlinear measurement error models (for example Hsiao and Wang (2000), Newey (2001) and Li (2002)). These methods can be extended to binary choice models, but all depend on either some kind of prior information and simulation techniques, or need replicated data.

In the next section of this paper we develop a simple maximum likelihood point estimator for probit models with measurement error, given sufficient *apriori* information to identify the model. This is done at first by imposing some quite restrictive assumptions which are then successively weakened. We then use techniques akin to the consistent bounds approach, proposed by Klepper and Leamer (1984), to extend the analysis to situations with weaker prior information. By concentrating directly on the mapping from the reliability ratios to the parameter estimates we are able, however to obtain what we call "reasonable bounds". These are narrower (often considerably narrower) than the Klepper-Leamer bounds.

In the third section we present a Monte Carlo experiment that shows how different degrees of measurement error affect the small sample properties of the traditional estimates and of the ML estimates we have developed, while the fourth section presents an empirical example concerning sick-leave in Sweden. The results of both these sections show that taking measurement error into account causes considerable changes in the inferences we draw.[1]

[1] The simulations were performed using Matlab while the empirical example was estimated using Gauss. A general Gauss program is available at http://swopec.hhs.se/lunewp/abs/lunewp2003_004.htm.

2) ESTIMATION OF THE PROBIT MODEL WITH ERRORS-IN-VARIABLES

2.1) FORMULATION OF THE MODEL

The traditional binary choice model connects an unobserved dependent variable v_i to a $k \times 1$ vector of explanatory variables x_i through the relationship

$$v_i = \alpha + \boldsymbol{\beta}' \boldsymbol{x}_i + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

where α is the constant term and $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown parameters. The error term ε_i is assumed independent of $\boldsymbol{x}_i$. The observed dependent variable (denoted y_i) takes the values one or zero depending on whether v_i is greater or less than zero, *i.e.*,

$$y_i = \begin{cases} 0 & \text{if } v_i \leq 0 \\ 1 & \text{if } v_i > 0. \end{cases} \tag{2}$$

In the context of an errors-in-variables model x_i will now denote the *true* (possibly unobserved) vector of explanatory variables, with j^{th} element x_{ij}, $j = 1,\ldots,k$. Corresponding to the true explanatory variables we observe an equal number of z-variables, which are related to x through

$$\boldsymbol{z}_i = \boldsymbol{x}_i + \boldsymbol{u}_i, \tag{3}$$

where $\boldsymbol{u}_i$ is a $k \times 1$ random vector with $E(u_i) = 0$. We assume that $\boldsymbol{u}_i$ is independent of $\boldsymbol{x}_i$ and ε_i for all i. Some of the z-variables may be miss-measured observations of the x-variables (in which case the corresponding elements of u_i are measurement errors), some may be proxy variables for the x-variables, and some may be identical to their unobserved counterparts (in which case the corresponding elements of u_i are equal to zero). This is the error model approach.[2]

[2] The other approach to modeling errors is the Berkson model where $x_i = z_i + u_i$, and u_i are the measurement errors. Here, $E(u_i \mid z_i) = 0$, which implies that the expected error conditional on the *observed* z_i is zero as opposed to the error model where the expectation conditional on the *actual* x_i is zero. In the error model, the observed value is correlated with the measurement error while in the Berkson model it is the actual value that is correlated with the measurement error. Which approach to use depends on how one believes the measurement errors are generated, or equivalently what one believes to be a fixed value for an observation: the true values of the regressors or the observed values. In our view it is natural to assume that in economics, and other non-experimental sciences, it is the true values that are fixed. We therefore follow the error model approach in this paper.

If ε_i is i.i.d. N(0,1) and $\boldsymbol{u}_i$ is equal to zero ($\boldsymbol{z}_i = \boldsymbol{x}_i$ for all i), we have a traditional probit model where the estimation of the parameters is straightforward using maximum likelihood. The log-likelihood is given by

$$\ell = \sum_{i=1}^{n} \left[y_i \ln\left\{ \Phi\left(\alpha + \boldsymbol{\beta}'\boldsymbol{z}_i\right)\right\} + \left(1 - y_i\right) \ln\left\{1 - \Phi\left(\alpha + \boldsymbol{\beta}'\boldsymbol{z}_i\right)\right\}\right] \tag{4}$$

where Φ is the standard normal c.d.f. function. Maximization of this log-likelihood will yield consistent estimates of α and $\boldsymbol{\beta}$.

However, if some or all of the observed variables deviate from their theoretical counterpart, either because of measurement errors or because they are proxy variables, the problem is more complicated. It is well known that a multiple regression model with the error model approach is not identified. The same is obviously true for the probit model defined above. To see the effect of the measurement errors, combine equations (1) and (3) to get

$$v_i = \alpha + \boldsymbol{\beta}'\boldsymbol{z}_i + \varepsilon_i - \boldsymbol{\beta}'\boldsymbol{u}_i\,, \tag{5}$$

where the error term is now composed of two parts. OLS and ML estimates that ignore measurement errors will be inconsistent since the error ($\varepsilon_i - \boldsymbol{\beta}'\boldsymbol{u}_i$) is correlated with the explanatory variable z.

To proceed further we will make the assumption that both x and u are i.i.d. normal

$$\boldsymbol{x}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_x) \text{ and } \boldsymbol{u}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_u)\,. \tag{6}$$

Since $\boldsymbol{x}_i$ and $\boldsymbol{u}_i$ are assumed independent, this specification implies that $\boldsymbol{z}_i$ is also i.i.d. normal, *i.e.*, $\boldsymbol{z}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_z)$ with

$$\boldsymbol{\Sigma}_z = \boldsymbol{\Sigma}_x + \boldsymbol{\Sigma}_u\,. \tag{7}$$

Note that we have now made a series of quite strong assumptions. Some of these seem to be intuitively reasonable; for example the normality of the errors can be justified by appealing to the central limit theorem. Others are more dubious, however. In particular there are three assumptions that we make for the sake of simplicity in development of our results, but that we will later relax in section 2.5. These assumptions are the following.

ASSUMPTION 1: The model is *non-calibrated*, *i.e.*, z is an unbiased estimate of x.

An error model is calibrated if it is specified that $\boldsymbol{z}_i = \boldsymbol{\gamma}_0 + \boldsymbol{\Gamma}_1\boldsymbol{x}_i + \boldsymbol{u}_i$, which is obviously more appropriate for proxy variables. A non-calibrated model has $\boldsymbol{\gamma}_0 = \boldsymbol{0}$ and $\boldsymbol{\Gamma}_1 = \boldsymbol{I}$.

ASSUMPTION 2: The measurement errors u are independent of the true variables x.

There are many cases where we can expect a non-zero correlation between the measurement errors and the true variables, for example in data on self-reported incomes.

ASSUMPTION 3: All the variables in x (and therefore z) are unconditionally normal.

This assumption excludes dummy variables, and demands that even those variables that are measured without error should be normal.

Even after making all of the assumptions given above it is obvious that the model defined by (1), (2), (3) and (6) is unidentifiable, since we only observe $\boldsymbol{z}_i$ while $\boldsymbol{x}_i$ and $\boldsymbol{u}_i$ are unobserved. This implies that $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_x$ cannot be identified, which in turn implies that $\boldsymbol{\beta}$ may not be consistently estimated.[3] At this point there are two ways of progressing. The first approach can be taken if one has precise additional information, either in terms of observable instruments for the z-variables or in terms of exact knowledge of $\boldsymbol{\Sigma}_u$ (or some transformation of $\boldsymbol{\Sigma}_u$). This approach leads to point estimates of the parameters. The other approach, which leads to consistent bounds for all parameters, is available even if one only has weak or no information about $\boldsymbol{\Sigma}_u$.

2.2) ESTIMATION USING INSTRUMENTAL VARIABLES

The estimation of nonlinear measurement error models using instrumental variable techniques is discussed in, for example, Bound *et al* (2001, Section 3.3). The use of such methods is much more complicated than in the linear case, since in general it is not sufficient to know that an instrument (w) is uncorrelated with the measurement error. The conditional distribu-

[3] Wansbeek and Meijer (2000, pp 327-329) proposed estimating a logit measurement error model using GMM with the same number of moment conditions as parameters to estimate. This seems to eliminate the need for precise additional information, but unfortunately one of their estimating equations (11.7a) can easily be shown to reduce to an identity, and thus gives no information about the parameters. This holds for any distribution of the errors and any symmetric distribution of the latent variable.

tion of z given w has to be known if consistent estimation is to be achieved. However, Amemiya, Y (1985) and Carroll and Stefanski (1990) have shown that using $\hat{x}$ (the predictor of x from the regression x on w) as a proxy for x often works quite well. This type of approach has been used by Iwata (2001) for estimating the measurement error probit model, under the assumption of a joint normal distribution of the variables.

In addition to the technical difficulties, another drawback with the IV method is the assumption that efficient instruments exist at all. There are a few situations where the method might be appropriate, for example poorly measured disposable income could be instrumented using more precisely measured taxable income. In most cases, however, it is as an act of faith to assume that an instrument's reduced correlation with the error term (in this context, its lower measurement error compared to the independent variable's) will more than compensate for the decreased precision (due to the non perfect correlation between the instrument and the variable).

2.3) ESTIMATION USING PRECISE INFORMATION ON $\boldsymbol{\Sigma}_U$.

If $\boldsymbol{\Sigma}_u$ is known then all the parameters are identified. $\boldsymbol{\Sigma}_x$ can be estimated from (7), which allows us to then estimate the rest of the parameters. This will obviously also apply if we have exact information on a one-to-one transformation of $\boldsymbol{\Sigma}_u$. However, knowledge concerning $\boldsymbol{\Sigma}_u$ is not very common, and we will therefore consider two transformations of $\boldsymbol{\Sigma}_u$ that are more intuitively appealing when our analysis is extended to cover weak information. These transformations are *reliability ratios* and *x-z correlations*.

2.3.1) Reliability Ratios

Reliability ratios are particularly useful if u is a measurement error. The (traditional) reliability ratio associated with variable j is defined as

$$\pi_j = \frac{\mathrm{Var}(x_{ij})}{\mathrm{Var}(z_{ij})}, \qquad j = 1, \dots, k. \tag{8}$$

If assumption 2 holds, *i.e.*, u is uncorrelated with x, then

$$\pi_j = \frac{\mathrm{Var}(x_{ij})}{\mathrm{Var}(x_{ij}) + \mathrm{Var}(u_{ij})} = \frac{1}{1 + [\mathrm{Var}(u_{ij}) / \mathrm{Var}(x_{ij})]}, \tag{9}$$

and in this case $0 \le \pi_j \le 1$. The reliability ratios will be the same for each observation i since $\boldsymbol{u}$ and $\boldsymbol{x}$ are assumed homoscedastic. Note that π_j is equal to one if and only if the j^{th} variable is measured without error, and that the reliability ratio decreases as the measurement error increases. π_j can be interpreted as the reliability we have in the j^{th} explanatory variable.

Typically, one specifies the covariance matrix of the measurement errors as diagonal (see for example Klepper and Leamer (1984)). If this is the case, the k traditional reliability ratios are all that are needed to identify the model. The variance of x_{ij} can be estimated as π_j times the sample variance of z_{ij}, while the covariance between x_{ij} and $x_{i\ell}$ can be estimated as the sample covariance between z_{ij} and $z_{i\ell}$.

However, if we wish to allow for covariances between the measurement errors, then we will have to introduce the *covariance reliability ratio*, $\pi_{j\ell}$

$$\pi_{j\ell} = \frac{\mathrm{Cov}(x_{ij}, x_{i\ell})}{\mathrm{Cov}(z_{ij}, z_{i\ell})} = \frac{\mathrm{Cov}(x_{ij}, x_{i\ell})}{\mathrm{Cov}(x_{ij}, x_{i\ell}) + \mathrm{Cov}(u_{ij}, u_{i\ell})}, \qquad j, \ell = 1, \ldots, k, \;\; \ell \ne j. \tag{10}$$

This is the ratio of the *true* covariance between variables j and ℓ to the *observed* covariance between these two variables. Note that $\pi_{j\ell}$ is equal to one if either of the j^{th} and ℓ^{th} variables are measured without error, or if the measurement errors are uncorrelated. Note also that $\pi_{j\ell}$ will only be greater than one if the measurement errors are correlated with the opposite sign to the true regressors. It follows from (8) to (10) that

$$\mathrm{Var}(u_{ij}) = (1 - \pi_j)\mathrm{Var}(z_{ij}), \qquad j = 1, \ldots, k. \tag{11}$$

$$\mathrm{Cov}(u_{ij}, u_{i\ell}) = (1 - \pi_{j\ell})\mathrm{Cov}(z_{ij}, z_{i\ell}), \qquad j, \ell = 1, \ldots, k, \;\; \ell \ne j. \tag{12}$$

It is possible to collect the reliability ratios into a single matrix $\boldsymbol{\Pi}$, where:

$$\boldsymbol{\Pi} = \begin{bmatrix} \pi_1 & \cdots & \pi_{1k} \\ \vdots & \ddots & \vdots \\ \pi_{k1} & \cdots & \pi_k \end{bmatrix}.$$

Let $\mathbf{1}$ denote a $k \times k$ matrix of ones, and "*" the Hadamard product (element by element multiplication). If we use this notation to define $\bar{\boldsymbol{\Pi}} = (\mathbf{1} - \boldsymbol{\Pi})$ we can then write $\boldsymbol{\Sigma}_x = \boldsymbol{\Pi} * \boldsymbol{\Sigma}_z$

and $\boldsymbol{\Sigma}_u = \bar{\boldsymbol{\Pi}} * \boldsymbol{\Sigma}_z$. If $\boldsymbol{\Pi}$ is known we can therefore estimate $\boldsymbol{\Sigma}_x$ using an estimate of $\boldsymbol{\Sigma}_z$, which will allow us to estimate the rest of the parameters.

It is often more convenient to specify $\boldsymbol{\Pi}$ than to specify $\boldsymbol{\Sigma}_u$, since $\boldsymbol{\Pi}$ is scale independent while $\boldsymbol{\Sigma}_u$ is not. The variances of the measurement errors are quite likely to be affected by such factors as growth in the economy, and can therefore vary over time. The reliability ratios, on the other hand, are much less likely to change from one period to another.

2.3.2) *X-Z* Correlation

If z_{ij} is a proxy variable for x_{ij}, then the concept of a reliability ratio may not be very appealing. Instead, one may be able to specify *the correlation between the proxy variable and the true variable* (the *x-z correlation*).

$$\rho_j = \frac{\operatorname{Cov}(x_{ij}, z_{ij})}{\sqrt{\operatorname{Var}(x_{ij})\operatorname{Var}(z_{ij})}} \tag{13}$$

If a proxy is perfect then it will have a correlation of one with the true variable, and the correlation will decrease the poorer the proxy becomes. Using (3) and the independence of x and u, we find

$$\rho_j = \frac{\operatorname{Cov}(x_{ij}, z_{ij})}{\sqrt{\operatorname{Var}(x_{ij})\operatorname{Var}(z_{ij})}} = \frac{\sqrt{\operatorname{Var}(x_{ij})}}{\sqrt{\operatorname{Var}(z_{ij})}} = \sqrt{\pi_j} \tag{14}$$

The correlation between z_{ij} and x_{ij} is simply the square root of the reliability ratio of the j^{th} variable. From the estimation point of view it is therefore immaterial whether one specifies the reliability ratio of a variable or its *x-z* correlation.

If correlation between the measurement errors of different variables is allowed, then we must also specify cross *x-z* correlations

$$\rho_{j\ell} = \frac{\operatorname{Cov}(x_{ij}, z_{i\ell})}{\sqrt{\operatorname{Var}(x_{ij})\operatorname{Var}(z_{i\ell})}} = \frac{\operatorname{Cov}(x_{ij}, x_{i\ell})}{\sqrt{\rho_j^2\operatorname{Var}(z_{ij})\operatorname{Var}(z_{i\ell})}} = \frac{\pi_{j\ell}}{\rho_j}\operatorname{Cor}(z_{ij}, z_{i\ell}) \tag{15}$$

If all the observed explanatory variables are perfect proxies then the ρ_j's are all equal to one, and the $\rho_{j\ell}$'s are equal to the correlation between z-variables.

Specifying the severity of the measurement error problem in terms of Σ_u, Π or the ρ's is obviously simply a matter of taste, intuitive appeal and available information. In the rest of this paper we will use the reliability ratio notation; if *x*-*z* correlations are easier to specify then these can be transformed to reliability ratios using (14) and (15).

Note that the relationship (14) is critically dependent on assumptions 1 and 2, *i.e.*, that the model is non-calibrated and that *u* is uncorrelated with *x*. We will return to this point in Section 2.5, where we will show that using the *x*-*z* correlations in these cases is often more appropriate than using reliability ratios directly.

2.4) Maximum Likelihood when Π is Known

The joint distribution of the observations on *y* and *z* will involve all the unknown parameters, $\omega' = (\mu', \mathrm{vech}(\Sigma_z)', \alpha, \beta')$. If we assume that $\boldsymbol{x}$, and therefore z, is weakly exogenous with respect to (α, β) we can write the log-likelihood of the parameters as

$$\ell(\omega) = \sum_{i=1}^{n} \ln f\left(y_i, z_i \middle| \omega\right) = \sum_{i=1}^{n} \ln f_1\left(z_i \middle| \omega_1\right) + \sum_{i=1}^{n} \ln f_2\left(y_i \middle| z_i, \omega_1, \omega_2\right),$$

where $\omega_1' = (\mu', \mathrm{vech}(\Sigma_z)')$ and $\omega_2' = (\alpha, \beta')$. Estimating these parameters by directly maximizing this likelihood function (so called full information maximum likelihood (FIML)) can be difficult in practice, mainly due to the large number of elements in Σ_z.

An alternative approach that is often used is therefore limited information maximum likelihood (LIML), also called two-step estimation, that is obtained by

- First, maximizing $\ell_1(\omega_1) = \sum_{i=1}^{n} \ln f_1\left(z_i \middle| \omega_1\right)$ over $\omega_1 \Rightarrow$ estimates $\hat{\omega}_1(z)$
- Second, maximizing $\ell_2(\hat{\omega}_1, \omega_2) = \sum_{i=1}^{n} \ln f_2\left(y_i \middle| z_i, \hat{\omega}_1(z), \omega_2\right)$ over $\omega_2 \Rightarrow$ estimates $\hat{\omega}_2(y, z)$.

LIML will yield consistent estimates, but the asymptotic covariance matrix of α and β is in general not simply the inverse of the information matrix from the second stage. Correct estimates of the covariance matrix can be obtained using the results of Murphy and Topel (1985).

In the linear measurement error model LIML is equivalent to substituting the first two sample moments of z for μ and Σ_z in the full likelihood, which can then be maximized over α

and β. Fuller (1987, section 2.2.1) shows that this two step procedure is equivalent to maximizing the full likelihood function directly.

We will prove that these results hold even for the probit errors-in-variables model. In the rest of this section we derive the LIML estimates and show that these are the same as the FIML estimates. A consistent estimate of the covariance matrix of α and β is also given. As discussed above, there are now two approaches to estimating the parameters using maximum likelihood, LIML and FIML.

2.4.1) LIML Estimation

Under the assumptions we have made the vector $(v_i, \boldsymbol{z}_i', \varepsilon_i, \boldsymbol{x}_i', \boldsymbol{u}_i')'$ is multivariate normal, where the variables are connected through the relations (1) and (3). The first-step likelihood is therefore given by

$$\ell_1 = \text{constant} - \tfrac{n}{2}\ln\left|\boldsymbol{\Sigma}_z\right| - \tfrac{n}{2}\operatorname{tr}\boldsymbol{S}_z\boldsymbol{\Sigma}_z^{-1} - \tfrac{n}{2}(\overline{\boldsymbol{z}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}_z^{-1}(\overline{\boldsymbol{z}} - \boldsymbol{\mu}) \tag{16}$$

where $\boldsymbol{S}_z = \frac{1}{n}\sum_i (\boldsymbol{z}_i - \overline{\boldsymbol{z}})(\boldsymbol{z}_i - \overline{\boldsymbol{z}})'$. It is well known that the ML estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_z$ are $\overline{\boldsymbol{z}}$ and $\boldsymbol{S}_z$, which are independently distributed $N(\boldsymbol{\mu}, \frac{1}{n}\boldsymbol{\Sigma}_z)$ and $\left(\frac{n}{n-1}\right)\text{Wishart}\left(\frac{1}{n}\boldsymbol{\Sigma}_z, n-1\right)$. The sample moments are thus consistently estimating the population moments. Somewhat less well known is that the asymptotic covariances of the elements of $\boldsymbol{S}_z$ are given by[4]

$$\text{AsyCov}(s_{jk}, s_{\ell m}) = \tfrac{1}{n}(\sigma_{j\ell}\sigma_{km} + \sigma_{jm}\sigma_{k\ell}).$$

These results enable us to calculate an estimate of $\boldsymbol{V}_1$, the asymptotic covariance matrix of $\hat{\boldsymbol{\omega}}_1' = (\hat{\boldsymbol{\mu}}', \text{vech}(\hat{\boldsymbol{\Sigma}}_z)') = (\overline{\boldsymbol{z}}', \text{vech}(\boldsymbol{S}_z)')$, see the Appendix.

To calculate the second-step likelihood we need to find the conditional distribution of y given $\boldsymbol{z}$. This can be found from the binary choice relationship (2) as

$$f_2\left(y_i \middle| \boldsymbol{z}_i\right) = \begin{cases} \int_{-\infty}^{0} f_{v|z}\left(v_i \middle| \boldsymbol{z}_i\right) dv_i & \text{for } y_i = 0 \\ \int_{0}^{\infty} f_{v|z}\left(v_i \middle| \boldsymbol{z}_i\right) dv_i & \text{for } y_i = 1 \end{cases}$$

Since v and $\boldsymbol{z}$ are jointly normal, the conditional distribution of v given $\boldsymbol{z}$ must also be normal

---

[4] The results in this paragraph can be found in *e.g.* Anderson (1984), equation 3.2.9, theorems 3.2.1, 3.3.2 and 3.4.4, and corollary 7.2.3.

$$f_{v|z}\left(v_i \mid \boldsymbol{z}_i\right) = \frac{1}{\sqrt{2\pi\sigma_*^2}} \exp\left[-\frac{\left(v_i - \mu_i^*\right)^2}{2\sigma_*^2}\right],$$

where the conditional mean and variance can be found from Anderson (1984, theorem 2.5.1) and the fact that the regression parameters have been normalized using $\mathrm{Var}(\varepsilon_i) = 1$,

$$\mu_i^* = \mathrm{E}(v_i | \boldsymbol{z}_i) = \alpha + \boldsymbol{\beta}'\mathrm{E}(\boldsymbol{x}_i | \boldsymbol{z}_i) = \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \boldsymbol{\beta}'(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}), \text{ and} \tag{17}$$

$$\sigma_*^2 = \mathrm{Var}(v_i | \boldsymbol{z}_i) = \boldsymbol{\beta}'\mathrm{Var}(\boldsymbol{x}_i | \boldsymbol{z}_i)\boldsymbol{\beta} + 1 = \boldsymbol{\beta}'\{(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z) - (\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\}\boldsymbol{\beta} + 1. \tag{18}$$

The probability of observing a particular value of y is thus

$$\Pr\left(y_i = y | \boldsymbol{z}_i\right) = \begin{cases} 1 - \Phi\left[\dfrac{\mu_i^*}{\sigma_*}\right] & \text{if } y = 0 \\ \Phi\left[\dfrac{\mu_i^*}{\sigma_*}\right] & \text{if } y = 1 \end{cases} \tag{19}$$

The second-step log-likelihood function for the probit model with measurement errors and given reliability ratios is therefore given by

$$\ell_2 = \sum_{i=1}^{n}\left[y_i \ln\left\{\Phi\left(\mu_i^* / \sigma_*\right)\right\} + \left(1 - y_i\right)\ln\left\{1 - \Phi\left(\mu_i^* / \sigma_*\right)\right\}\right]. \tag{20}$$

The LIML estimates are found by substituting the first-step estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_z$ ($\bar{\boldsymbol{z}}$ and $\boldsymbol{S}_z$) into (17) and (18) to yield $\hat{\mu}_i^*$ and $\hat{\sigma}_*^2$. Using these in (20) leads to a function that is very reminiscent to the ordinary probit likelihood, and that is simple to maximize.

The asymptotic covariance matrix of the first-step LIML estimates ($\boldsymbol{V}_1$) is standard, but the matrix for the second-step estimates ($\boldsymbol{V}_2^*$) has to be calculated through

$$\boldsymbol{V}_2^* = \boldsymbol{V}_2 + \boldsymbol{V}_2(\boldsymbol{C}\boldsymbol{V}_1\boldsymbol{C}' - \boldsymbol{R}\boldsymbol{V}_1\boldsymbol{C}' - \boldsymbol{C}\boldsymbol{V}_1\boldsymbol{R}')\boldsymbol{V}_2 \tag{21}$$

where $\boldsymbol{V}_2$ is the unadjusted second-step covariance matrix, and where

$$\boldsymbol{C} = \mathrm{E}\left[\left(\frac{\partial \ell_2}{\partial \omega_2}\right)\left(\frac{\partial \ell_2}{\partial \omega_1'}\right)\right] \qquad \boldsymbol{R} = \mathrm{E}\left[\left(\frac{\partial \ell_2}{\partial \omega_2}\right)\left(\frac{\partial \ell_1}{\partial \omega_1'}\right)\right]$$

see Murphy and Topel (1985), who also establish consistency of LIML under the usual regularity conditions. Details concerning the gradient vectors necessary for the maximization of the likelihood and the calculation of the Murphy-Topel variances can be found in *Theorem 1* and *Theorem 2* in the Appendix, where it is also shown that $\boldsymbol{R} = \mathbf{0}$.[5]

2.4.2) FIML Estimation

The FIML estimate is found by maximizing $\ell(\boldsymbol{\omega}) = \ell_1(\boldsymbol{\omega}_1) + \ell_2(\boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ simultaneously over all parameters to yield estimates $\hat{\hat{\boldsymbol{\omega}}}(y,z)$, where ℓ_1 and ℓ_2 are given by (16) and (20). It can be shown, however, that this maximization will lead to exactly the same estimates as LIML, see *Theorem 3* in the Appendix. The only possible advantage of using the FIML procedure is that we can estimate the covariance matrix using either the OPG method or the inverse of the Hessian. Some of the matrices in the Murphy-Topel formula can only be estimated by the OPG method if LIML is used.

2.4.3) Behavior of the Likelihood Function

It is well known that the probit likelihood function (4) is globally concave and goes to $-\infty$ as $\beta_j \to \pm\infty$. This is not true for the measurement error probit, however. In Appendix A.5 we show that the likelihood function (20) approaches an asymptote as (some of) the parameters approach infinity. This can lead to some quite subtle convergence problems in small samples, some examples of which are reported in connection with our simulation studies in Section 3.

2.5) Relaxing the Assumptions

2.5.1) Calibration

The use of non-calibrated models seems a reasonable assumption when the variables are only being affected by "true" measurement error. However, if we are using a proxy variable it is because we are hoping for a high correlation with the true variable, not because we are expecting it to be an unbiased estimate of that variable. A good proxy should also have zero partial correlation with the other variables in the model – *i.e.*, it should be a proxy for one and only one of the true explanatory variables.

---

[5] Note that the Murphy-Topel covariance matrix will only equal the unadjusted second stage matrix if $\boldsymbol{C} = \mathbf{0}$, which is equivalent to the information matrix of the full likelihood being block diagonal. In Appendix A.4 we show that this is not the case here.

Writing the calibration model explicitly as

$$z_i = \gamma_0 + \Gamma_1 x_i + u_i \tag{22}$$

we can solve for $\boldsymbol{x}$ to obtain

$$\tilde{z}_i = x_i + \tilde{u}_i, \quad \text{where } \tilde{z}_i = \Gamma_1^{-1}(z_i - \gamma_0) \text{ and } \tilde{u}_i = \Gamma_1^{-1} u_i, \tag{23}$$

where the fact that a one-to-one relationship between the proxy variables and the true explanatory variables implies that Γ_1 should be diagonal and non-singular.[6] Letting γ_{0j} be the j^{th} element of γ_0 and γ_{1j} the j^{th} diagonal element of Γ_1 we can classify the explanatory variables as follows

- No measurement error $\gamma_{0j} = 0, \quad \gamma_{1j} = 1, \quad u_{ij} = 0$
- True measurement error $\gamma_{0j} = 0, \quad \gamma_{1j} = 1, \quad u_{ij} \neq 0$
- Proxy variable $\gamma_{0j} \neq 0, \quad \gamma_{1j} \neq 1, \quad u_{ij} \neq 0$.

Using (1) and (23) implies that the ML estimators described in the previous section will still hold if z and Π are replaced by $\tilde{z}$ and $\tilde{\Pi}$ in (17) and (18), where $\tilde{\Pi}$ is the matrix of reliability ratios between $\tilde{z}$ and $\boldsymbol{x}$. However, we know that $\Sigma_{\tilde{z}} = \Gamma_1^{-1} \Sigma_z \Gamma_1^{-1\prime}$ and, if Γ_1 is diagonal, that $(\tilde{\Pi} * \Sigma_{\tilde{z}}) = \Gamma_1^{-1}(\tilde{\Pi} * \Sigma_z)\Gamma_1^{-1\prime}$. Substituting these results yields

$$\mu_i^* = \tilde{\alpha} + \tilde{\beta}'\mu_z + \tilde{\beta}'(\tilde{\Pi} * \Sigma_z)\Sigma_z^{-1}(z_i - \mu_z), \text{ and}$$

$$\sigma_*^2 = \tilde{\beta}'\{(\tilde{\Pi} * \Sigma_z) - (\tilde{\Pi} * \Sigma_z)\Sigma_z^{-1}(\tilde{\Pi} * \Sigma_z)\}\tilde{\beta} + 1.$$

where $\tilde{\alpha} = \alpha - \sum \dfrac{\gamma_{0j}}{\gamma_{1j}}$ and $\tilde{\beta}_j = \dfrac{\beta_j}{\gamma_{1j}}$.

To make this estimator operational we need to know $\tilde{\Pi}$, which can be a problem since $\tilde{z}$ is unobservable. However, since Γ_1 is assumed diagonal, $\text{Cor}(\tilde{z}_j, x_j) = \text{Cor}(z_j, x_j)$. In other words we can use (14) to let us obtain the reliability ratio we need from the usual *x*-*z* correlation, *i.e.*, $\tilde{\pi}_j = \rho_j^2$. A similar result follows for the cross reliability ratios.

---

[6] If z is a proxy for x_1, then it can still be correlated with x_2 if the two x-variables are correlated. The partial correlation between z and x_2 given x_1 must however be zero if z is to be a good proxy.

Calibration will therefore not affect our results as regards the parameters of the variables that are not proxies, as long as we obtain the reliability ratios as the squares of the correlations between the proxies and the true variables. The constant term and the parameters of those variables that are proxies will be affected, however, but this may be acceptable if the proxies have merely been introduced to control for heterogeneity. It is also sometimes possible to obtain independent information concerning regression coefficients between *x* and *z*. In this case we can obtain the required parameter estimates using the relationships

$$\beta_j = \tilde{\beta}_j \gamma_{1j} = \tilde{\beta}_j \frac{\rho_j}{\gamma^*_{1j}}, \tag{24}$$

where γ_{1j} is the coefficient from the regression of *z* on *x*, and γ^*_{1j} is the coefficient from the regression of *x* on *z*.

2.5.2) Correlation Between the Measurement Errors and the True Variables

The assumption that the measurement errors are uncorrelated with the true variables may sometimes be inappropriate, for example respondents with low incomes may well tend to report incomes in excess of the true value whilst those with high incomes may well tend to underreport.

Assuming joint normality of the errors and the variables implies that all regressions must be linear, and in particular that $\boldsymbol{u}_i = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \boldsymbol{x}_i + \boldsymbol{w}_i$, where $\mathrm{E}(\boldsymbol{w}_i \mid \boldsymbol{x}_i) = \boldsymbol{0}$. Substituting the above into (3) yields

$$\boldsymbol{z}_i = \boldsymbol{\phi}_0 + (\boldsymbol{I} + \boldsymbol{\Phi}_1)\boldsymbol{x}_i + \boldsymbol{w}_i ,$$

which is of the same form as (22) if $\boldsymbol{\Phi}_1$ is diagonal. Once again, we need to interpret the reliability ratio in the MLE as the square of the *x-z* correlation (forgetting this can lead to quite serious errors, for example there is nothing to stop (8) being greater than one if *u* and *x* are negatively correlated).

If the measurement error is correlated with its own true variable (but has zero partial correlations with the other variables) then the parameters of the other variables will still be correctly estimated. Only the parameter estimate of that variable in question will be affected..

If additional information is available we could use (24), which will be appropriate for calibrated models, correlated measurement errors or both. If only correlated measurement errors are involved, then we could also use

$$\beta_j = \tilde{\beta}_j(1+\phi_j) = \tilde{\beta}_j(1+\rho_{u_j x_j}\,\sigma_{u_j} \big/ \sigma_{x_j}). \tag{25}$$

2.5.3) Non-Normality of the True Variables

A severe constraint on the analysis presented above is the normality assumption, (6). While it is not too restrictive to presume that the errors ε and u are normal, the assumption that the unconditional distribution of the true regressors, x, is also normal would exclude all manner of empirically interesting situations, *e.g.*, models with dummy variables. Luckily the normality of x is merely a sufficient condition for deriving the LIML estimates, not a necessary one

The normality assumption was used in three places in our LIML analysis.

- In the derivation of the second step likelihood
- In the derivation of the first step estimates
- In the derivation of the Murphy-Topel adjustment

The first step estimates ($\bar{z}$ and $\boldsymbol{S}_z$) are, however, consistent even if the regressors are non-normal. *If* the second step likelihood is correct, then under the usual regularity conditions the LIML estimate given by *Theorem 1* of the Appendix will also be consistent.[7]

In a similar manner it can be shown that the Murphy-Topel adjustment (21) will still hold in the non-normal case *if* the second step likelihood is correct, the only difference being that $\boldsymbol{V}_1$ is estimated by (A.4) instead of (A.3). This can be seen by studying Murphy and Topel (1985), where it obviously does not matter if the set of equations used to define the first step estimates ((A.2)) is considered to be a set likelihood equations or moment equations. [8]

The properties of the LIML estimates thus only depend on the validity of (20), *i.e.*, we need to show that $v|\,z$ is normally distributed with mean and variance given by (17) and (18).

---

[7] Strictly speaking these estimates should now be referred to as mixed ML/MM not LIML, but for the sake of simplicity we retain the former notation.

[8] The main thrust of Murphy and Topel (1985) is in fact mixed LS/ML estimation; pure LIML estimation is introduced almost as an afterthought,

In appendix A.6 we will show that assumption 3 can be replaced by either of the following assumptions

ASSUMPTION 3′: The conditional distribution of x given z is normal and the regression of x on z is linear and homoscedastic.

ASSUMPTION 3′′: The explanatory variables can be divided into two groups; x_1 which are measured with error and x_2 which are measured without error. The conditional distribution of x_1 given x_2 is normal and the regression of x_1 on z is linear and homoscedastic.

The first of these alternatives is quite subtle. It implies that the conditional distribution of u given z is normal, but note that it is *not* enough to assume that the conditional distribution of u given x is normal; in fact from (3) we can see that u given x and u given z can only both be normal if the joint distribution of u, x and z is normal. Note also that the assumption that the regression of x on z (or equivalently, u on z) is linear and homoscedastic is quite a strong one, although considerably weaker than assuming that the unconditional distributions of x and z are normal. Our results should also hold approximately as long as the regression is approximately linear – a much weaker assumption again.[9]

The second alternative is easier to understand, since we are not making any assumptions about the distribution of the variables that are measured without error. However, it is not enough to merely assume that the conditional distribution of x_1 given x_2 is normal. Note also that under this assumption the unconditional distribution of x_1 (or z_1) need not be normal if the unconditional distribution of x_2 is non-normal.

These assumptions are shown in Appendix A.6 to imply that the second-step likelihood is given by (20), and that the LIML estimates therefore have the same properties as before. Obviously these estimates are only FIML under the joint normality of all the variables.

[9] Assumptions concerning linear regressions are made in several other places in this and other papers. Results will only be approximately correct if these regressions are merely approximately linear.

2.6) CONSISTENT BOUNDS AND REASONABLE BOUNDS

Often we are in a situation where we have no precise information concerning the reliability ratios and no instruments are available. In other cases instruments do exist which will yield consistent estimates, but these are so poorly correlated with the explanatory variables that for all feasible sample sizes the efficiency of the estimated parameters will be unacceptably low. When no instruments are available and we have no knowledge at all of the reliability ratios, the only approach is that of "consistent bounds" suggested by Klepper and Leamer (1984). In this approach one identifies all the parameter values that are mathematically consistent with non-negative estimates of the variance of ε_i and positive semi-definite estimates of Σ_x and Σ_u.[10] Thus, one abandons the attempt to find a single statistically consistent estimate of the parameters of interest.

There are some obvious drawbacks in the Klepper-Leamer approach, however.

- First, Klepper and Leamer (1984) identify bounds in the parameter space, but there is no connection between a particular error structure and a particular estimate of β. The only exception is the estimator where the measurement errors are assumed zero (all $u_i = 0$ or all $\pi_i = 1$).
- Second, consistent bounds estimates are often unbounded, at least for some parameters.
- Third, although prior information can be incorporated into the analysis, this can only be done in a non-intuitive manner using lower bounds on the ρ's *and* upper bounds on the R^2 of the regression equation.

The procedure developed in Klepper and Leamer (1984) assumes that v_i is directly observable (as they work with a traditional multiple regression model with errors-in-variables) Although it may be theoretically possible to derive consistent bounds for the model specified in equations (1), (2), (3) and (6), the drawbacks mentioned above have prompted us to take another route.

[10] The consistent bounds are found by running $k + 1$ regressions. If all these regressions are in the same orthant, then the set of maximum likelihood estimates will be the convex hull of the regression estimates. If this is not the case then the consistent bounds estimates become unbounded.

Instead of simply deriving *general* consistent bounds, which include all parameter values that are mathematically consistent with positive semi-definite covariance matrices of the data, we will present estimates that are mathematically consistent with *reasonable* values of Π. We will denote these as "reasonable bounds" for the parameters. [11]

Maximizing (20) for given Π defines a mapping from a particular reliability matrix to a parameter vector β. If we are able to specify bounds on the reliability ratios for those variables that we suspect to be measured with errors (or bounds on the *x-z* correlations for the proxy variables) we can define a continuous mapping from the "reasonable" bounds on Π to "reasonable" bounds on β. The drawbacks mentioned above in the Klepper-Leamer approach are now removed

- There is an obvious connection between the values of β and the values of the reliability ratios.
- If the "reasonable" bounds on β tend to become unbounded for some values of Π, this will merely indicate an incompatibility between the model, the data and these values of Π.
- Prior information is introduced in a very intuitive manner.

An important question that has to be answered is whether we have any prior information at all. It is certainly true that precise prior information about a particular problem is very rare. However, information from detailed surveys in similar areas can be used to give reasonable bounds to the reliability ratios. This has become much easier since Bound *et al* (2001) published their review of over 100 studies concerning measurement errors in surveys. A third of the studies give information which can easily be transformed into reliability ratios, while others can be used to derive these ratios under further assumptions.[12] In addition, some studies even give information concerning calibration and correlations between the measurement errors and the true variables.

[11] A similar idea can be found in Wansbeek and Meijer (2000, pp 46-58), where parameter bounds obtained in a linear regression model using bounds on the covariance matrix of the measurement errors.

[12] The information, which is usually in the form of correlations, is most often found in studies concerning earnings, hours worked, benefits and education.

3) SIMULATION STUDIES

The purpose of this section is twofold. Firstly, we wish to investigate the traditional probit estimator to see how it fares when the explanatory variable(s) have measurement error, and secondly we want to evaluate the performance of the MLE estimator developed in this paper under the same circumstances. This second estimator, found by maximizing (19) will henceforth be denoted the "EIVML-estimator" as opposed to the Probit ML estimator.

We have performed four separate simulation experiments to investigate the performance of both estimators, under varying circumstances, as the severity of the measurement errors varies. *Simulation Study 1* is the simplest, where the model contains only one explanatory variable, and this is subject to measurement error. In *Simulation Study 2* we consider a model with two uncorrelated explanatory variables, where the reliability ratios of both variables can vary.

It is well known that even under the classical assumptions the probit model will not perform satisfactorily when the proportion of successes ("ones") gets close to zero or one. In *Simulation Study 3* we investigate the effect of changing this proportion on the performance of the EIVML estimator in a single-regressor model with measurement error.

Finally, in *Simulation Study 4*, we again consider the two variable model, but where the explanatory variables are correlated. In a standard linear model if some variables ($\boldsymbol{X}_1$) are measured with errors while others ($\boldsymbol{X}_2$) are not, and if these two sets of explanatory variables are uncorrelated, then OLS estimator of $\boldsymbol{\beta}_2$ is unbiased. Typically, the higher the correlation between the explanatory variables the more biased is the OLS estimator of β_2. We wish to investigate if this argument carries over to binary choice models. This can be of practical importance in common situations where the elements of ($\boldsymbol{X}_2$) are dummy variables, which typically have low correlation with continuous variables

In all the simulation studies we estimate the model given by (1) – (3), where the constant term is set as $\alpha = 0$, the slope parameter(s) are set $\beta = 1$ and the distribution of the probit error term is $\varepsilon \sim N(0,1)$ [13]. In each experimental situation 1000 replications of the data

[13] As for all probit models, we are actually estimating β/σ, where σ is the variance of ε. Thus, simulating ε with a variance of 1 is simply a normalization which allow us to estimate β.

were generated. For each replication the parameter estimate(s) and their standard errors, using both the outer product of first derivatives and the inverse of the Hessian, were calculated for both the usual probit ML and the EIVML estimators. Note that in general the Murphy-Topel adjustments to the EIVML standard errors were not used, the only exception being Study 1C which was specially designed to investigate this question.

It should be noted that the relationship $\text{Var}(x) = \pi\,\text{Var}(z)$ implies that there are basically two different ways of comparing the results of our experiments as the reliability ratios are allowed to vary. One way is to hold $\text{Var}(z)$ constant and allow $\text{Var}(x)$ to vary, which will be appropriate when investigating a given data set where the variance of the observed data (z) is constant, but where it is not known how this variance is divided between x and u. Studies 1A, 2, 3 and 4 are performed in this manner.

The other way is to hold $\text{Var}(x)$ constant and allow $\text{Var}(z)$ to vary, which is realistic from a theoretical point of view where the variance of the true unobserved explanatory variable (x) is constant and we want to investigate the effect of lowering the reliability ratio. In this scenario the explanatory power of the latent model (1), $R_m^2 = 1 - \text{Var}(\varepsilon)/\text{Var}(\nu)$, is held constant for varying degrees of measurement error. Study 1B is performed in this manner.

STUDY 1

This is our basic study where we have a model with one explanatory variable measured with errors. The proportion of successes, P, is set at 50% by letting the mean of the explanatory variable be zero. The experiments are performed for sample sizes $n = 100$, 1000 and 10000, and for ten different values of the reliability ratios, $\pi = 1.0, 0.9, \ldots, 0.1$.

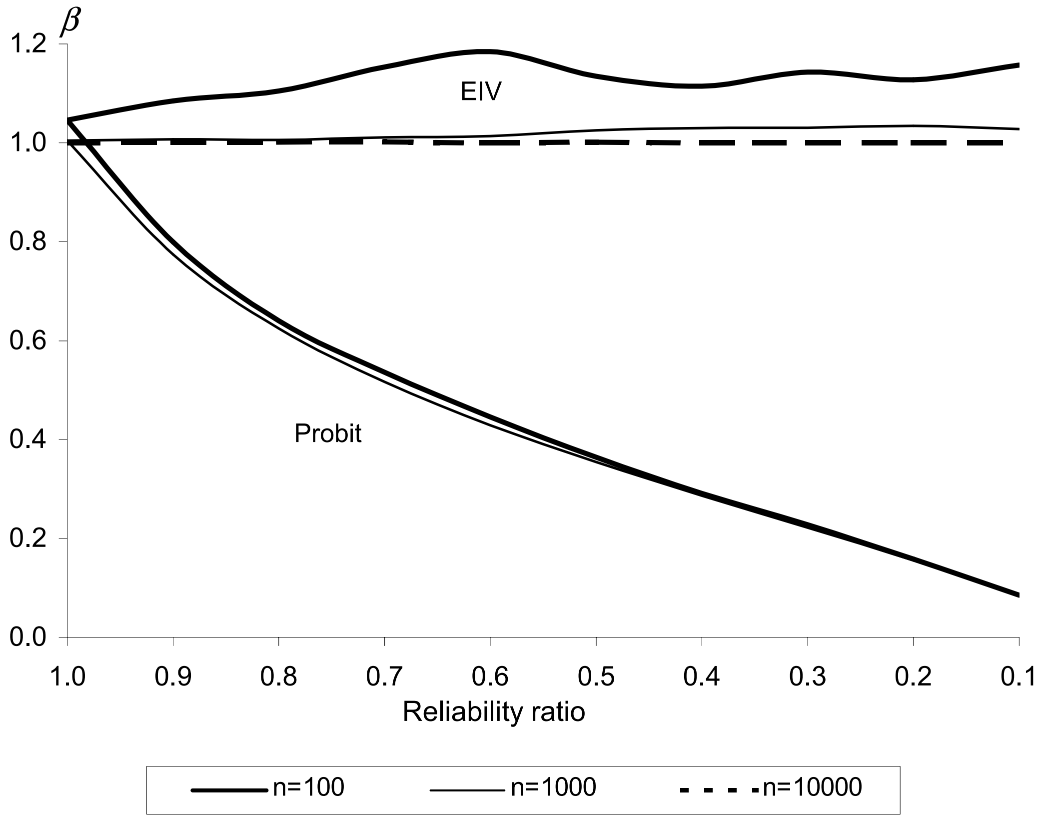
Study 1A

We simulate independently $x_i \sim N(0, 4\pi)$ and $u_i \sim N(0, 4 - 4\pi)$ for each value of π. This implies that $z_i \sim N(0, 4)$ and $R_m^2 = 4\pi/(1 + 4\pi)$ for all π. When $\pi = 1$, all the variance in z is due to the variance in x, while as π decreases, the variance in the measurement errors u increases at the expense of the variance in x. Note that the replicating for sample size 10000 is timeconsuming, and since the results are to all intents and purposes identical to those for sample size 1000 we discontinued this part of the experiment for reliability ratios less than 0.5.

Table 1. Mean Parameter Estimates and Fail Rates for Probit and EIVML in Study 1A

π	R_m^2	*Probit ML*			*EIVML*			
		n=100	*n=1000*	*n=10000*	*n=100*	*Fail Rate*	*n=1000*	*n=10000*
1.0	0.80	1.0459	1.0046	1.0009	1.0459	0.0 %	1.0046	1.0009
0.9	0.78	0.7996	0.7745	0.7723	1.0851	0.3 %	1.0073	1.0012
0.8	0.76	0.6404	0.6248	0.6250	1.1048	1.5 %	1.0060	1.0013
0.7	0.74	0.5363	0.5169	0.5164	1.1535	4.5 %	1.0114	1.0021
0.6	0.71	0.4458	0.4286	0.4284	1.1847	6.0 %	1.0135	1.0003
0.5	0.67	0.3642	0.3548	0.3535	1.1344	8.4 %	1.0254	1.0012
0.4	0.62	0.2913	0.2867		1.1144	8.9 %	1.0301	
0.3	0.55	0.2276	0.2217		1.1434	10.8 %	1.0306	
0.2	0.44	0.1587	0.1561		1.1273	11.2 %	1.0342	
0.1	0.29	0.0855	0.0848		1.1574	10.7 %	1.0282	

Figure 1. Mean Parameter Estimates for Probit and EIVML in Study 1A.



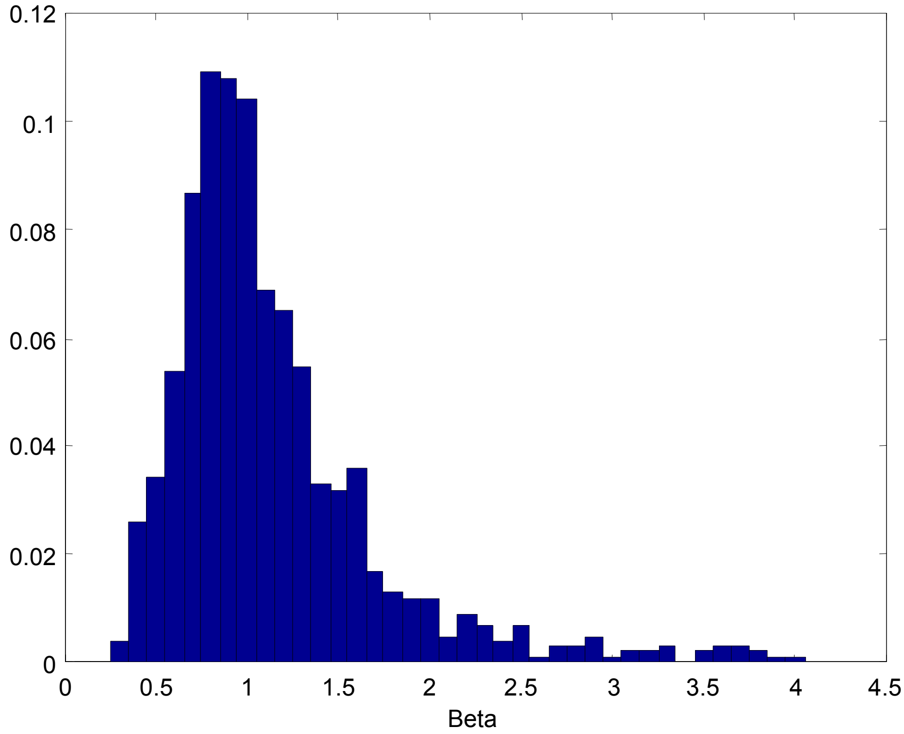
In Study 1A all the replications resulted in well-behaved likelihoods when n = 1000 and 10000. This was no longer the case, however, for n = 100 when measurement error in the data increased, for example 11% of the replications had ill-behaved likelihoods when the reliability was as low as $\pi = 0.1$. The statistics presented in the tables and graphs are based on the well-behaved likelihood functions only.

The results of this experiment are presented in Figure 1 and Table 1. The most striking observation is that the probit model will deliver severely inconsistent estimates when we have measurement errors. For $\pi = 1$ (100% reliable data, no measurement errors) the probit estimator gives results close to the true value of β. As the reliability ratio drops, the inconsistency of the probit MLE increases. Even when the reliability is as high as 80%, we have a bias of almost 40%. The reliability of the income variable, for example, is usually estimated to be between 70% to 80%. Ignoring measurement errors in economic binary choice models may thus result in a severe bias. We can also see that the standard probit method *underestimates* the true absolute value of the parameter when there is measurement error – a common feature in errors-in-variables' models that often goes under the name *attenuation*.

The estimator developed in this paper seems to works well when π is known, at least for this experiment. We do notice a small-sample upward bias in the estimate of β which disappears as n increases. Investigating the distribution of the parameter estimates for a particular case is also revealing. In Figure 2 we have created a histogram over the estimates of β from the individual replications (excluding ill-behaved likelihoods) for $n = 100$ and $\pi = 0.5$. We see clearly that the distribution of the estimates is skewed to the left, which will cause the average to be biased upwards. The *median* of the β-estimates is typically close to 1, however, for example in Figure 2 the median is 0.99 while the average is 1.14.

Tables A1 and A2 of the Appendix give additional information from Study 1A such as the standard deviation, average standard errors (OPG and inverse Hessian), and the minimum and maximum replicates. The average standard errors in general lie close to the standard deviations of the parameter, showing that the formulae we are using seem to give unbiased estimates even in small samples. This is true for both the probit and EIV variance estimators, which seems quite surprising since the probit itself is inconsistent and we are not using the Murphy-Topel correction for EIV. We shall return to the last point in Study 1C. [14]

Figure 2. Empirical Frequency Function for EIVML Estimates (Study 1A: n = 100, π = 0.5).



STUDY 1B

In Study 1B we keep the variance of x fixed at $\mathrm{Var}(x)=4$, which implies that $R_m^2=0.8$ while $\mathrm{Var}(z)=4/\pi$ and $\mathrm{Var}(u)=4(1-\pi)/\pi$. For $\pi=1$ studies 1A and 1B are identical, but when, for example, $\pi=0.1$ we have $\mathrm{Var}(x)=0.4$, $\mathrm{Var}(u)=3.6$, $\mathrm{Var}(z)=4$ and $R_m^2=0.29$ in study 1A, while we have $\mathrm{Var}(x)=4$, $\mathrm{Var}(u)=36$, $\mathrm{Var}(z)=40$ and $R_m^2=0.8$ in study 1B. In all other respects the studies are the same.

The general conclusions from Study 1B are the same as for Study 1A, *i.e.*, the probit estimator exhibits considerable inconsistencies when there is measurement error, while EIVML is consistent with only small bias when $n\geq 1000$. The results for EIVML seem at first sight, however, to be less satisfactory than in Study 1A when the reliability in the data is low. For example, from Table 2 we can see that more ill-behaved likelihoods are generated

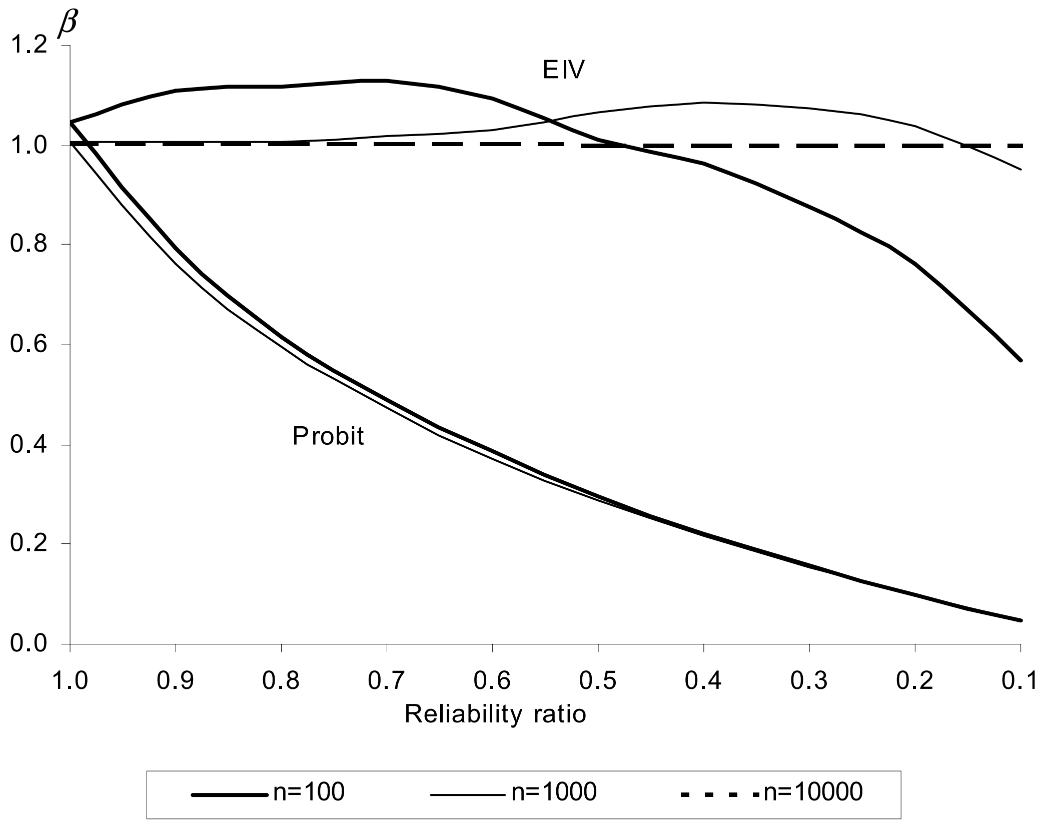
[14] The standard error estimates are even more skew than the parameter estimates, which is why we are getting bias for very low reliabilities in small samples.

when $n = 100$, and even for $n = 1000$ when $\pi \leq 0.5$, and that the small-sample biases seem to be larger. From Table A4 of the Appendix we can see a considerable bias in the standard errors for very small samples.

Table 2. Mean Parameter Estimates and Fail Rates in Study 1B, $R_m^2 = 0.8$

π	*Probit ML*			*EIVML*				
	n=100	*n=1000*	*n=10000*	*n=100*	*Fail Rate*	*n=1000*	*Fail Rate*	*n=10000*
1.0	1.0459	1.0046	1.0009	1.0459	0.0 %	1.0046	0.0 %	1.0009
0.9	0.7927	0.7635	0.7612	1.1081	0.5 %	1.0080	0.0 %	1.0012
0.8	0.6143	0.5963	0.5966	1.1168	3.3 %	1.0084	0.0 %	1.0016
0.7	0.4891	0.4731	0.4722	1.1276	9.1 %	1.0203	0.0 %	1.0027
0.6	0.3884	0.3723	0.3720	1.0944	16.1 %	1.0315	0.0 %	1.0013
0.5	0.2970	0.2900		1.0114	22.0 %	1.0651	0.2 %	
0.4	0.2225	0.2182		0.9613	26.7 %	1.0839	2.5 %	
0.3	0.1598	0.1543		0.8747	33.4 %	1.0741	5.6 %	
0.2	0.1003	0.0975		0.7599	38.3 %	1.0363	11.6 %	
0.1	0.0471	0.0466		0.5703	41.0 %	0.9495	22.0 %	

Figure 3. Mean Parameter Estimates for Probit and EIVML in Study 1B.



The reason for the difference becomes apparent when we look at the values of R_m^2 in Tables 1 and 2. Probit-type models do not work well when the goodness-of-fit of the latent regression (1) is large[15], and we can see that $R_m^2 = 0.8$ for all reliabilities in Study 1B while the goodness-of-fit is considerably less for small π in Study 1A. For a *given model* with given latent goodness-of-fit, the performance of EIVML in small-samples deteriorates quite rapidly as the reliability of the data decreases. However, for *given data*, with unknown latent goodness-of-fit, this deterioration is not at all as rapid. Sample sizes of 100 seem to be too small to obtain good estimates in all cases.

Study 1C

Calculation of the Murphy-Topel standard errors in Studies 1A and 1B gave results that were barely distinguishable from the uncorrected values (in general the differences were less than those obtained using OPG or inverse Hessian). A valid question is thus in what conditions it is necessary to use the correction.

The formula given in Theorem 2 of the Appendix simplifies considerably when there is only one explanatory variable, in particular (A.7b) reduces to $\psi = 0$. The ratio between the uncorrected and corrected standard deviations of $\hat{\beta}$ (denoted as the *rsd*) becomes a fairly simple rational function of R_m^2, P and the moments $E(\lambda^2)$, $E(\lambda^2 z)$ and $E(\lambda^2 z^2)$, where λ is given by (A.5a). These moments can be estimated very simply using simulation methods, since we do not have to perform any EIVML maximization. We therefore estimated the *rsd* using 60.000 replications, which gives a high degree of accuracy.

In Table 3 we present the results for the set-up used in Study 1A, while in Figure 4 we present the results for varying R_m^2 and P with a reliability ratio of 0.6. The results show that it is the proportion successes that is most important in deciding whether it is necessary to use the Murphy-Topel correction. For $P = 0.5$ the correction has little effect, while for $P = 0.9$ there seem to be some advantages in using the correction. The standard errors seem to be

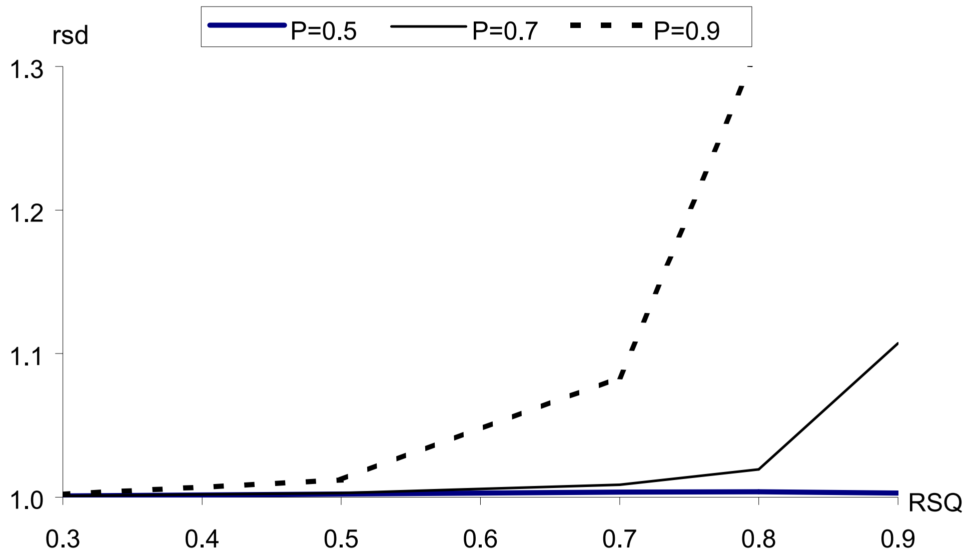
[15] This might seem to be counter intuitive, but follows from the fact that when the goodness-of-fit is high then values of x above the mean will almost always lead to one value of y, while values below the mean will lead to the other value. The likelihood will therefore be very flat, since changes in the parameters produce almost no change in the probabilities. If the fit is perfect, then we get the well-known result that the likelihood becomes horizontal and the estimation breaks down. The argument is the same for both probit and EIVML.

unbiased, at least in Study 1A, but it should be remembered that their distribution becomes very skew when R_m^2 and/or P become large.

Table 3. Corrected and Uncorrected EIVML Standard Errors ($P = 0.5$ *and* $Var(z) = 4$)

π	$\sqrt{n} \cdot \sigma(\hat{\beta})$			Study 1A (n = 10000)	
	Corrected	Uncorrected	Ratio	s.d.	Average s.e.
1.0	1.86	1.86	1.0000	0.0177	0.0183
0.9	2.48	2.48	1.0013	0.0243	0.0244
0.8	3.04	3.03	1.0028	0.0309	0.0300
0.7	3.52	3.51	1.0036	0.0354	0.0352
0.6	4.00	3.98	1.0036	0.0401	0.0402
0.5	4.56	4.55	1.0030	0.0454	0.0453
0.4	5.07	5.06	1.0022		
0.3	5.75	5.74	1.0013		
0.2	6.90	6.90	1.0005		
0.1	10.07	10.07	1.0001		

Figure 4. Ratio of Corrected to Uncorrected EIVML Standard Errors ($\pi = 0.6$)



STUDY 2

In this study we have a model with two uncorrelated explanatory variables, both measured with errors. The proportion of successes, P, is again set at 50%, while the experiments are performed for sample sizes $n = 100$, 1000 and 10000. We restrict our attention to the presentation of the experiments where $\mathrm{Var}(z)$ is held constant as π varies (case A), since the results when $\mathrm{Var}(x)$ is held constant (case B) mirror those in Study 1 (namely that EIVML performs a little worse in this case).

We simulate independently $x_{ji} \sim N(0, 4\pi_j)$ and $u_{ji} \sim N(0, 4-4\pi_j)$ for $j = 1,2$ and for each pair (π_1, π_2)[16]. This implies that $z_{ji} \sim N(0,4)$ and $R_m^2 = (4\pi_1 + 4\pi_2)/(1+4\pi_1+4\pi_2)$. We consider five values of the reliability ratios, $\pi_j = 1.0, 0.8, \ldots, 0.2$, but due to symmetry we only need to study 15 cases instead of 25.

Figure 5 shows the mean parameter estimates for EIVML and Probit ML in this study, while Tables A5 to A7 in the Appendix give some more detailed results. The general results of this study are the same as in Study 1. We can see that the Probit ML works very poorly when there is measurement error while the EIVML is consistent with quite small finite sample bias for all values of the reliability ratios. EIVML does not work well in very small samples ($n = 100$) with low reliability, however, in the sense that the failure rate increases quite dramatically.

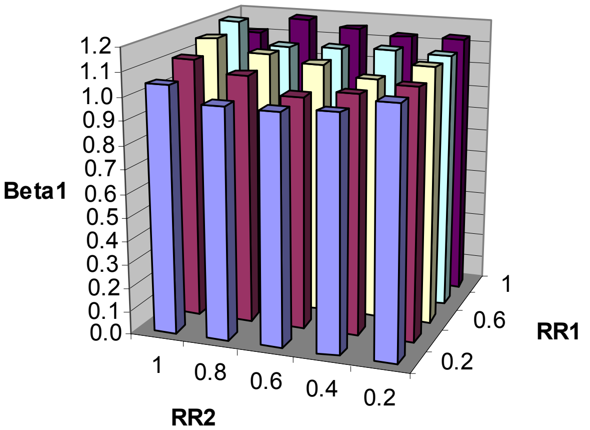
The new information given by this study concerns the interplay between measurement errors in different variables. In spite of the fact that both the true variables and the measurement errors are independent, there is obviously a spill over from one reliability ratio to the parameter estimate for the other variable. There is a large sample bias of roughly 25% in the probit estimate $\hat{\beta}_1$ when $\pi_1 = 1$ but $\pi_2 \neq 1$, thought this bias does not seem to increase as the reliability of the second variable decreases. The effect is, of course, more dramatic when the "own" reliability is low. Even the small sample bias of EIVML seems to increase when either of the reliability ratios is small.

---

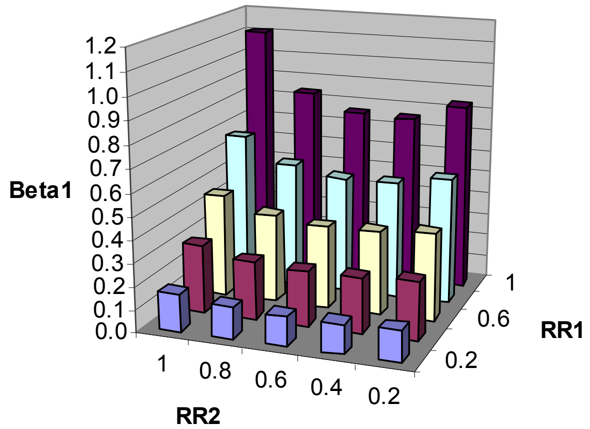
[16] Since both $\mathrm{Cov}(z_1, z_2) = 0$ and $\mathrm{Cov}(x_1, x_2) = 0$, π_{12} is not formally defined. The empirical covariances will not be exactly zero, however, and thus $\pi_{12} = 1$ is used in the formulae for the EIVML estimates.

Figure 5. Mean Parameter Estimates in Study 2

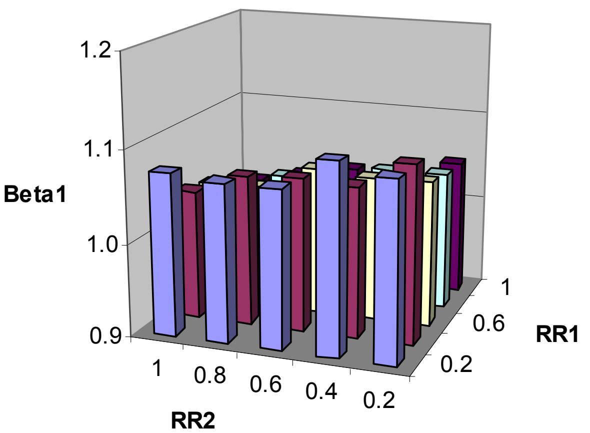
n = 100 (*a*) *EIVML*



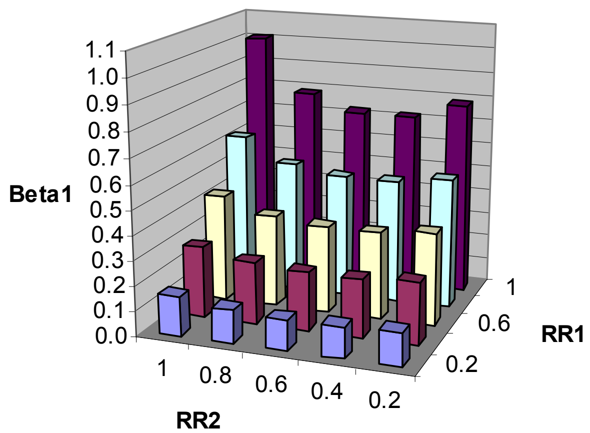
(*b*) *Probit ML*



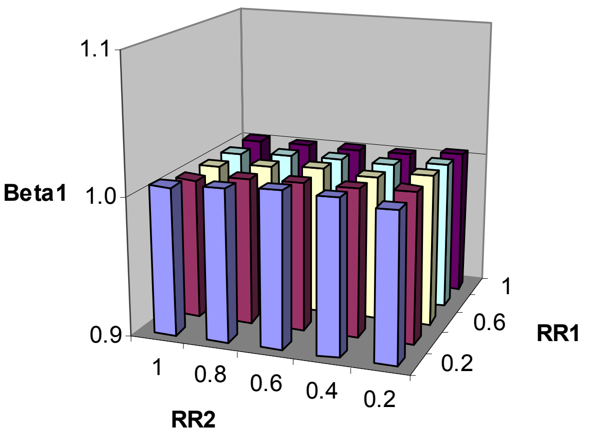
n = 1000 (*c*) *EIVML*



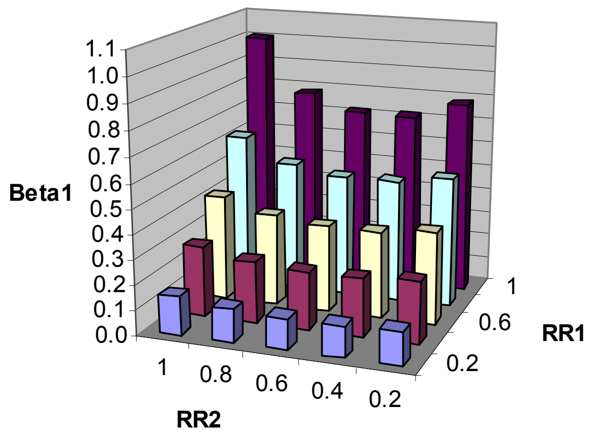
(*d*) *Probit ML*



n = 10000 (*e*) *EIVML*



(*f*) *Probit ML*



STUDY 3

In this study we extend Study 1A in such a way that the expected proportion of 1's and 0's is no longer 0.5. Simple algebra shows that the probability that y is equal to one is given by

$$P \equiv P(y=1) = \Phi\left((\alpha + \beta'\mu) \big/ \sqrt{1+\beta'\Sigma_x\beta}\right).$$

There are thus basically two ways of getting $P \neq 0.5$; we can introduce a non-zero intercept α or we can change the expected value of x and z to a non-zero value. In this study we keep $\alpha = 0$ and simulate $x_i \sim N(\mu, 4\pi)$ and $u_i \sim N(0, 4-4\pi)$. To obtain a given value of P with these parameter values we simply set $\mu = \Phi^{-1}(P) \cdot \sqrt{1+4\pi}$

In our experiments we allow the reliability ratio to take the values $\pi = 1.0, 0.8, \ldots, 0.2$ for sample sizes $n = 100$ and 1000[17]. We consider $P = 0.5, 0.75, 0.9$ for both sample sizes, while for n = 1000 we also consider P = 0.6 and 0.98.

In Figure 6 and in Table A8[18] we show some results, which are actually quite surprising. When there is no measurement error we obtain the expected result that the Probit ML is consistent, and that its standard error decreases as P gets closer to the value 0.5. The introduction of measurement error causes inconsistency for all values of P, but the surprising result is that the bias *increases* as P approaches 0.5.

The EIVML estimator is obviously consistent (note the differences in scale in the Figures 7a and 7b), but the effect of P and π on the size of the bias is quite complex. For all P the bias first increases as π decreases, but after a turning point the bias then decreases. This turning point occurs for smaller π when P gets closer to 0.5. The same pattern, though even more extreme, can be observed for the standard errors. In cases where P is large the EIVML estimator seems to perform best for small and large reliability ratios. It is important, however, to reiterate that the bias of the EIVML was very small in all cases. In addition, due to replication error, too much emphasis should not be put on the shapes of the curves in Figure 6b.

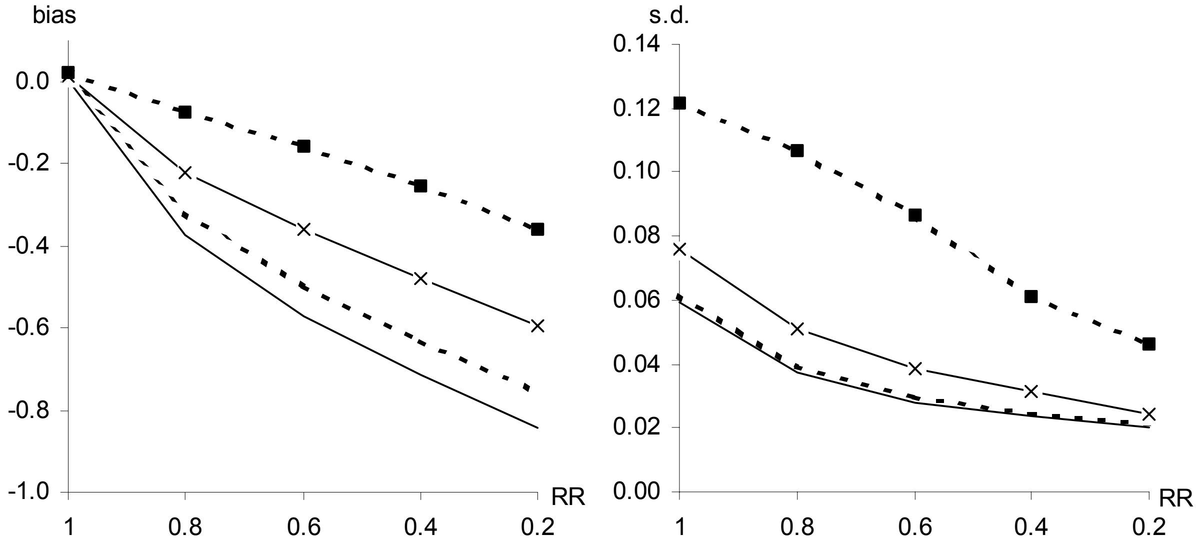
---

[17] In Studies 3 and 4 and we exclude the sample size n = 10000, since this case reveals nothing that was not already apparent in Studies 1 and 2, namely that EIVML produces estimates very close to the true parameter values while Probit ML is about as biased as we find in the case n = 1000.

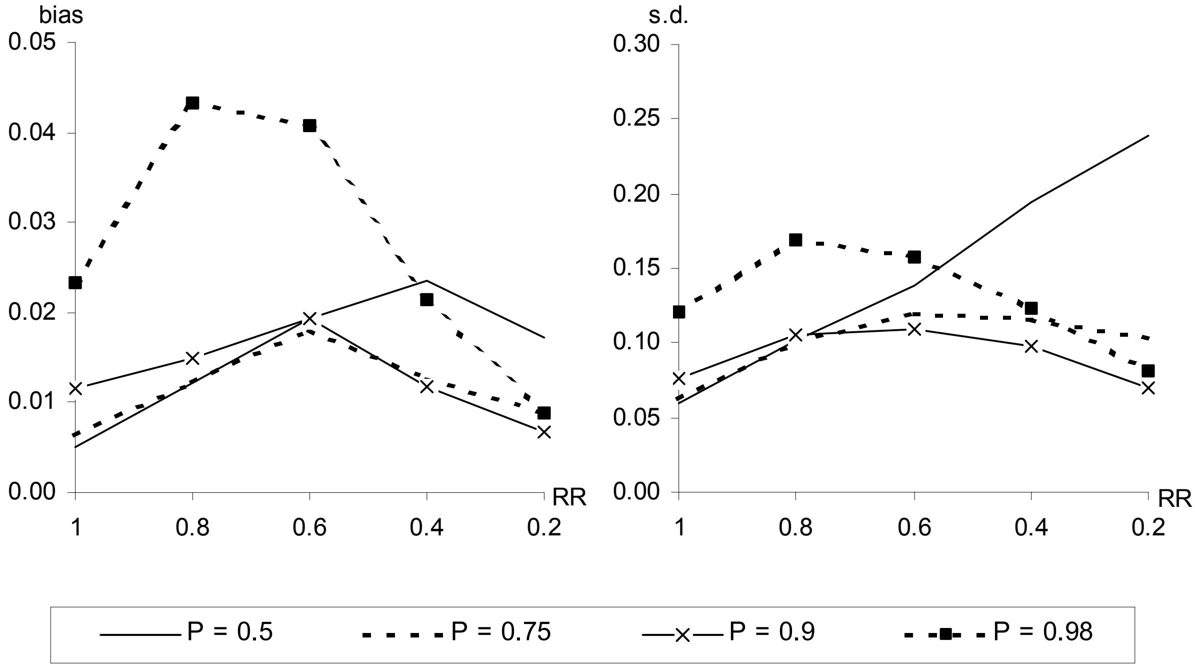
[18] Note that the results for P = 0.5 are not identical to those for Study 1A due to the use of different random numbers. They are, however, very similar. To save space the results for n = 100 are not shown in Table A8, since they do not yield any further insights in addition to those found for n = 1000.

Figure 6. Bias and Standard Deviation of the Parameter Estimates in Study 3, n = 1000

(*a*) *Probit*



(*b*) *EIVML*



STUDY 4

In the final study we again consider a two-variable model, but where the variables are now allowed to be correlated. The first variable is a correctly measured dummy variable (and therefore not normally distributed) while the second variable is continuous and has measurement errors. In this study we let the reliability ratio for the second variable take the values $\pi = 1.0, 0.8, \ldots, 0.2$, while the correlation between the variables is allowed to vary as $\rho = 0.0, 0.1, \ldots 0.6$. The sample sizes used are n = 100, 1000.

For a given reliability ratio we start by simulating (w_1, w_2) from a bivariate normal distribution with $E(w_1) = E(w_2) = 0$, $\mathrm{Var}(w_1) = 1$, $\mathrm{Var}(w_2) = 4\pi$ and $\mathrm{Cor}(w_1, w_2) = \rho_w$. The variable x_1 is defined as $x_1 = 1$ if $w_1 > 0$ and $x_1 = -1$ if $w_1 \leq 0$, while $x_2 = w_2$. It can be shown quite straightforwardly that ρ, the correlation between x_1 and x_2, is equal to ρ_w multiplied by $\sqrt{2/\pi_c} \approx 0.7979$, where π_c is the constant 3.14… (In practice we first decide on the value of ρ we wish to use in our simulations, and then calculate $\rho_w = \rho \cdot \sqrt{\pi_c/2}$).[19] Finally the observed variables are obtained as $z_1 = x_1$, indicating that the first (dummy) variable is measured without error, while $z_2 = x_2 + u_2$ where $\mathrm{Var}(u_2) = 4 - 4\pi$.

In Tables A9 and A10[20] we show the results for this study. There are two special questions that we want answered in this study.

(1) What happens to the probit estimate of the parameter of a variable measured without error when it is correlated to a variable with measurement error?

(2) What happens to the EIVML estimates when a variable measured without error is not normally distributed?

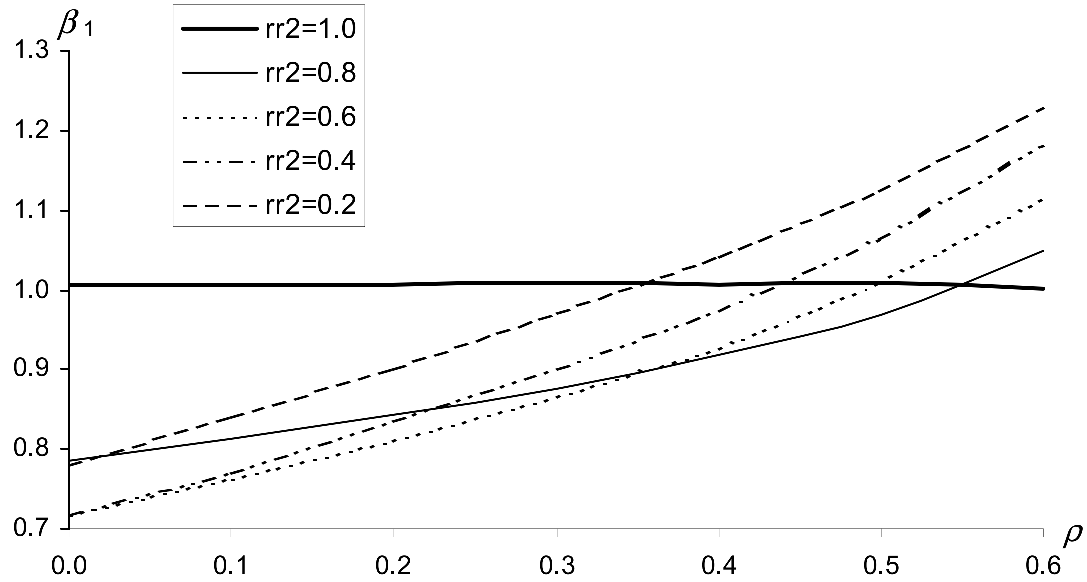
In study 2 we showed that the effect of measurement error spills over from one probit parameter estimate to another, even when the variables are uncorrelated. In this study we find that a positive correlation between the variables seems to work in the opposite direction for the variable measured without error, see Figure 7.

---

[19] Note that $\rho^2 \leq 2/\pi_c \approx 0.64$, which reaffirms the fact that severe multicollinearity is seldom a problem in models with dummy variables.

[20] To save space the results for some values of ρ are not shown in the tables

Figure 7. Probit Estimates of β_1 when π_2 Varies. Study 4, n = 1000



A general interpretation seems to be that the size of the estimate increases with the correlation between the two variables, while the slope decreases with the reliability ratio of the other variable. Multicollinearity can thus sometimes help the probit estimate against measurement error in another variable.

As to the second question, it is apparent from Tables A9 and A10 that the non-normality of the variable without measurement error does not affect the properties of the EIVML estimator in the slightest. The mean estimates of the parameter of the variable with measurement error is almost exactly the same as in Figure 3, while the estimates of the other parameter are even better. The small sample bias of the EIVML has as good as disappeared when $n = 1000$.

4) A Study Concerning Sick-leave in Sweden

4.1) Description of the Study

At the beginning of the 1990's the level of sickness benefit was 100% for most Swedes (the official insurance rate was 90%, but most employees were covered by collective labor contracts that topped up the remaining 10%). In addition no qualification time was required, benefit being paid from the first sick day. During the nineties the situation with regards to public finances has lead to more or less continuous reductions in the welfare system. Sickness benefits have been seriously affected by several of these changes; for example a qualification day (with 0% benefit) was introduced on the 1st of April 1993.

The decision of whether or not to go to work will depend on the cost in a broad sense. The price of a day of sick leave depends on, for example, the individual's wage, the benefit level and the number of qualification days. Other factors that are more difficult to quantify will also affect the decision, *e.g.,* start-up costs when returning to work after a period of unplanned absence, the risk of sanctions when illegally registering oneself as sick, the discomfort of working when sick and on the probability of thereby causing a longer period of sickness. Many of the factors that affect the cost are thus based on individual characteristics. Demographic and background variables such as sex, age, number of children, marital status, occupation, region, *etc.*, are obviously of interest.

A logit model to explain the incidence of sick leave in Sweden has been estimated by Edgerton (1997). The model used in that study involved a large number of interaction terms, but since the study is being used here mainly as an illustration, we will use a simplified version that it is still quite realistic. Probit and logit estimation of this model result in almost identical marginal effects.

The dependent variable we use in our model is the incidence of sick leave (1 = sick / 0 = not sick) under a specific week in January. The independent variables are:

- **1. Sex**: Female = 1
- **2-4. Age**: Normalized age (= [Age in years – 16]/48), its square and its cube
- **5. Nationality**: A dummy equal to 1 for individuals from the Middle East, Africa, Asia, Latin America or Stateless
- **6. Municipality type**: 1 for individuals from midsize or industrial towns
- **7. Marital status**: 1 for singles.
- **8. Children**: Dummy, 1 if individual has children less than 1 year old
- **9. Type of employment**: 1 if temporary, 0 if permanent
- **10. Degree of employment**: 1 for part time, 0 for full time

- **11. Occupation**: 1 for non industrial worker
- **12. Socioeconomic Class**: 1 for "white collar"
- **13. Blue collar trade union membership**: 1 if member
- **14. White collar trade union membership**: 1 if member
- **15. Real wage rate**: 100 Swedish Crowns / hour, in 1980 prices
- **16. Regional unemployment rates**: in percentage
- **17. Reform dummy**: 1 if there is a qualification day

4.2) THE DATA

The main source of information used in this report is the Labor Force Survey (LFS), performed on a monthly basis by Statistics Sweden. This survey consists of a number of questions, mainly concerning labor force participation but also including a number of background variables. In addition, information concerning wages and salaries is collected in the January edition of the LFS. The total number of observations containing wage information in the period 1992-1995 is 33,665 (Of these 17,809 observations are purely cross-sectional, while 15,856 consist of two observations for 7,928 individuals. We will be ignoring panel issues in this paper). Due to partial non-response only 33,498 observations were used in this study.

The quality of the LFS data is in general thought to be quite satisfactory. The wage variable, however, is collected by simply asking the interviewees what their pay-period is (monthly, weekly, hourly *etc.*) and how much they usually earn per period. A simple inspection of the data shows, for example, a considerable amount of rounding.

4.3) RESULTS

We estimate the errors-in-variables probit by maximizing the likelihood (20) for various values of Π. We believe that only the real wage is miss-measured, and we will only vary π_{15} (denoted as π). For $\pi = 1$ we have the traditional probit estimate. The results from estimating this model are given in Table 4, along with some descriptive statistics (means and, for non-dummy variables, standard deviations). We also give the correlations of all the variables with wages.

First of all we should note that the incidence of sick-leave is naturally quite small, $P = 0.058$ in our data. This causes imprecision in all binary choice modeling, with or without measurement error. Secondly, we can see from the following table that most of the variables are only slightly correlated with the miss-measured variable.

Table 4. Results for the Standard Probit Model

	Parameter	Std. error	P-value	Variable Mean	Variable Std. dev.	Corr. with Wage
Constant	−1.689	0.1014	0.0000			
Sex	0.235	0.0286	0.0000	0.5241		−0.27
Age	1.707	0.5633	0.0025	0.4957	0.2467	0.26
Age^2	−2.977	1.1847	0.0120			0.21
Age^3	2.118	0.7432	0.0044			0.17
Nationality	0.236	0.1284	0.0664	0.0062		−0.03
Municipality	−0.072	0.0310	0.0205	0.1648		−0.04
Marital	0.082	0.0265	0.0019	0.2663		−0.15
Children	−0.353	0.0766	0.0000	0.0431		−0.01
Temporary	−0.241	0.0515	0.0000	0.0839		−0.15
Part Time	−0.106	0.0285	0.0002	0.2792		−0.19
Occupation	−0.134	0.0335	0.0000	0.7814		0.11
Class	−0.100	0.0371	0.0069	0.5181		0.37
BC_Union	0.227	0.0414	0.0000	0.4536		−0.30
WC_Union	0.114	0.0436	0.0092	0.4020		0.31
Wage	−0.694	0.1273	0.0000	0.3722	0.1264	1.00
Unemp	−0.008	0.0077	0.3059	6.9365	1.8159	−0.05
Reform	−0.152	0.0284	0.0000	0.4804		0.01

Turning to the probit estimates we can see that all the parameters are significant except for the unemployment variable. The signs of the parameters also agree with the signs of the marginal effects in the more complex model given in Edgerton (1997). We can see, for example, that having less secure jobs (temporary or part-time) reduces the probability of taking sick leave, while membership of a trade union increases it. Women and industrial workers tend to be sick more often than men and non-industrial workers. It is especially interesting to note that the introduction of a qualification day does seem to reduce the tendency to take sick leave.

Table 5. Parameter Estimates for EIVML

	$\pi = 1.0$	$\pi = 0.9$	$\pi = 0.8$	$\pi = 0.7$	$\pi = 0.6$	$\pi = 0.5$
Constant	–1.689	-1.465	-1.148	-0.674	0.148	2.007
Sex	0.235	0.235	0.237	0.239	0.246	0.267
Age	1.707	1.714	1.734	1.735	1.785	1.937
Age^2	–2.977	-2.992	-3.030	-3.027	-3.113	-3.379
Age^3	2.118	2.127	2.150	2.153	2.214	2.404
Nationality	0.236	0.236	0.237	0.240	0.247	0.268
Municipality	-0.072	-0.072	-0.072	-0.073	-0.075	-0.082
Marital	0.082	0.082	0.083	0.083	0.086	0.093
Children	-0.353	-0.353	-0.355	-0.359	-0.369	-0.401
Temporary	-0.241	-0.241	-0.242	-0.245	-0.252	-0.274
Part Time	-0.106	-0.106	-0.107	-0.108	-0.111	-0.121
Occupation	-0.134	-0.134	-0.135	-0.136	-0.140	-0.152
Class	-0.100	-0.100	-0.101	-0.102	-0.105	-0.114
BC_Union	0.227	0.228	0.229	0.231	0.238	0.258
WC_Union	0.114	0.114	0.114	0.116	0.119	0.129
Wage	-0.694	-1.306	-2.183	-3.509	-5.870	-11.332
Unemp	-0.008	-0.008	-0.008	-0.008	-0.008	-0.009
Reform	-0.152	-0.152	-0.153	-0.155	-0.159	-0.173

The EIVML estimates were calculated for wage reliability ratios varying from 0.9 down to 0.5, and the results are presented in Table 5. Changing the values of the reliability ratio in the ML estimation obviously affects all the parameter estimates. The effect is quite small, however, for all parameters other than wages (plus the constant term). This can be seen as a positive empirical conclusion – measurement error in one variable does not seem to affect the parameter estimates of other variables to any significant degree. This is partly due to the low correlation between wages and the other variables.

Looking at Table 5 also reveals what is often called *attenuation*. For all the parameters (except the constant term) we can see that

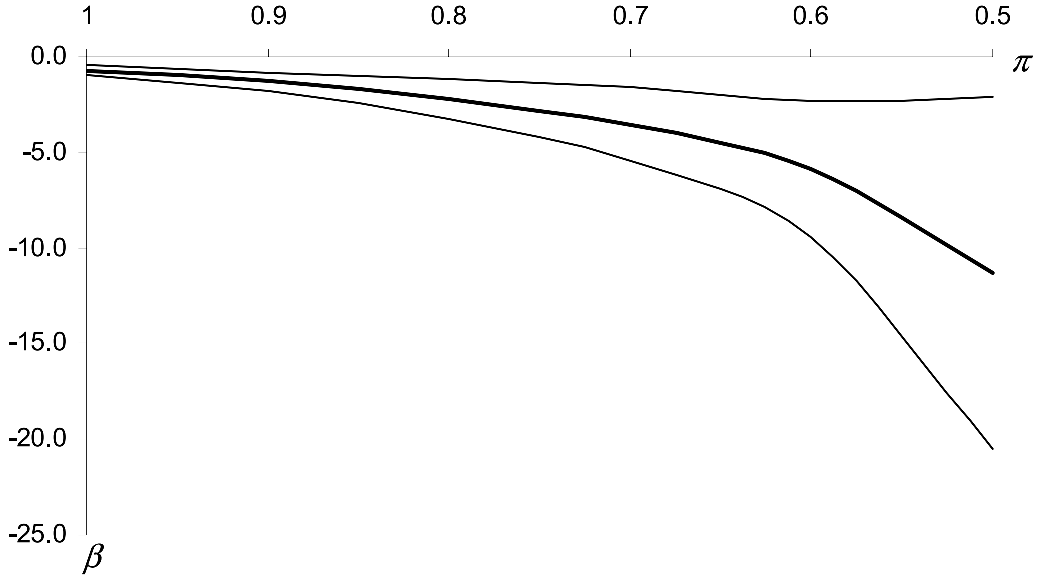
- the estimate assuming no measurement error is smallest in absolute value
- the estimates for all reliability ratios have the same sign
- the estimate assuming no measurement error is one of the bounds

Attenuation is an exact property of simple linear regression with measurement error, and a commonly observed phenomenon in other measurement error models. In other words, not taking measurement error into account will tend to make us believe that variables are less important than they really are. These conclusions reinforce what was found in the Monte Carlo study.

Another result, not shown in the table, is that the difference between the Murphy-Topel and unadjusted standard errors is quite small. For the wage parameter this difference grows from 3% to 7% as the reliability ratio decreases from 0.9 to 0.5. For the other parameters the differences are 3% at most, and often less than 1%. This is in spite of the proportion of positive y's being less than 0.07, which we saw from Study 1C tends to make the difference between the two standard error estimates increase.

In Figure 8 we show 95% confidence for the wage parameter as the reliability ratio varies. The only general conclusion that be drawn from this figure is that the parameter is significantly negative whatever the degree of measurement error. Otherwise it is apparent that the uncertainty about the parameter grows far too large for it to be useable if the reliability gets too small.

Figure 8. 95% Confidence Interval for the Wage Parameter for Different Reliability Ratios



It is therefore important to be able to try and set "reasonable bounds". Looking at the Bound *et al* (2001) survey, there are 12 studies from which we can calculate 18 different reliability ratios for self-reported earnings.[21] Four of these figures are for hourly rates, which have reliability ratios of 0.26, 0.40, 0.42 and 0.42. A closer analysis shows, however, that a large part of this unreliability is due to incorrectly reported hours worked. In the data used in our example the respondents were asked to report their earnings for the relevant period (usually monthly, sometimes hourly and occasionally weekly). The number of hours worked per period is much more standardized in Sweden, due to collective bargaining, than in the studies reported in Bound *et al.* The conversion to hourly rates is therefore much more reliable, and it is more appropriate to use the reliability ratios presented for the other intervals. In 11 of these cases the reliability ratios varied between 0.78 and 0.98[22], the three remaining "outliers" are at 0.61, 0.64 and 0.70. Reasonable bounds on π seem therefore to be between 0.7 and 1.0, which gives a wage parameter of between –0.7 and –3.5. If we are willing to

---

[21] Some of the reliability ratios are given for ln(earnings). However, the reliability ratio of ln(earnings) is approximately equal to the reliability ratio of earnings times the ratio of the of the expected values of z and x. The latter ratio is given in five studies, and is always between 0.98 and 1.035. In other words, we can ignore this. We have also ignored four results which were given for "usual" or "recent" earnings.

[22] This agrees with a further result given by Fuller (1987 p.8), who quotes an estimate of 0.85 for the reliability ratio for income. This figure was obtained from a 1970 study by the US Bureau of the Census.

accept a lower reliability bound of 0.8, then the 95% confidence interval for the parameter lies between -0.4 and –3.2.

A final piece of information that can be gleaned from Bound *et al* concerns whether the measurement error is correlated with the true values. This seems to be the case in all studies which investigate this problem – the correlation always being negative. In six studies regression coefficients are reported which can be used in (24) – strangely enough always for regressions of x on z. Two of these are for hourly rates, and the others give κ = 0.74, 0.91, 1.01 and 1.10, where $\beta = \kappa\tilde{\beta}$. The negative correlation seems only to have a small effect on the parameter estimates, and probably in the direction of reducing them in absolute value.

5) CONCLUSIONS

In this paper we have derived ML estimates and standard errors for the probit errors-in-variables model when the reliability ratio matrix is known. This has been done under quite restrictive assumptions of normality, but also when these assumptions have been relaxed somewhat. A Monte Carlo study shows that the method works well even in small samples.

We discuss how these results can be used when the reliability ratio matrix is unknown. We adapt Klepper and Leamer's consistent bounds approach to what we call "reasonable bounds" – based on a one-to-one mapping between the reliabilities and the parameter estimates. An empirical study of sick-leave in Sweden shows this approach works well, and yields insights into the model which otherwise would not be apparent.

Both the Monte Carlo and empirical studies reveal attenuation, *i.e.*, assuming no measurement error leads to an underestimation of the absolute value of the parameters. Other conclusions are that variables that are measured without error are not affected very much by measurement error in the other variables, but that if two variables both have measurement error then there can be interaction between the two reliabilities.

APPENDIX

A.1) ML and MM Estimation of μ and Σ_z

Assuming normality of $\boldsymbol{x}$ and $\boldsymbol{z}$ we can write the likelihood of the observed regressors for the i^{th} observation as (see Anderson (1984, eq 3.2.2))

$$\ell_{1i} = \text{constant} - \tfrac{1}{2}\ln\left|\boldsymbol{\Sigma}_z\right| - \tfrac{1}{2}(\boldsymbol{z}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu})$$

Let $\sigma_{j\ell}$ denote an element of $\boldsymbol{\Sigma}_z$, $\sigma^{j\ell}$ an element of $\boldsymbol{\Sigma}_z^{-1}$, $\boldsymbol{\sigma}_j$ the j^{th} column of $\boldsymbol{\Sigma}_z$ and $\boldsymbol{\sigma}^j$ the j^{th} column of $\boldsymbol{\Sigma}_z^{-1}$. We also let $\boldsymbol{E}_{ij}$ denote the matrix with $(i, j)^{\text{th}}$ element equal to one and all other elements equal to zero. Using this notation the gradients of the likelihood can be found using the matrix results of Amemiya, T (1985, pp. 461-2), Lütkepohl (1996, Chapter 10) and Harville (1997, Chapter 15). Remembering that $\boldsymbol{\Sigma}_z$ is symmetric we obtain

$$\frac{\partial \ell_{1i}}{\partial \boldsymbol{\mu}} = \boldsymbol{\Sigma}_z^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}) \tag{A.1a}$$

$$\begin{aligned}
\frac{\partial \ell_{1i}}{\partial \sigma_{jj}} &= -\frac{1}{2}\frac{\partial \ln\left|\boldsymbol{\Sigma}_z\right|}{\partial \sigma_{jj}} - \frac{1}{2}\frac{\partial (\boldsymbol{z}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu})}{\partial \sigma_{jj}} \\
&= \tfrac{1}{2}\{(\boldsymbol{z}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}_z^{-1}\boldsymbol{E}_{jj}\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}) - \sigma^{jj}\} \\
&= \tfrac{1}{2}\{(\boldsymbol{z}_i - \boldsymbol{\mu})'\boldsymbol{\sigma}^j\boldsymbol{\sigma}^{j\prime}(\boldsymbol{z}_i - \boldsymbol{\mu}) - \sigma^{jj}\} \\
&= \tfrac{1}{2}\{\boldsymbol{\sigma}^{j\prime}(\boldsymbol{z}_i - \boldsymbol{\mu})(\boldsymbol{z}_i - \boldsymbol{\mu})'\boldsymbol{\sigma}^j - \sigma^{jj}\}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \ell_{1i}}{\partial \sigma_{j\ell}} &= \tfrac{1}{2}\{(\boldsymbol{z}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{E}_{j\ell} + \boldsymbol{E}_{\ell j})\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}) - 2\sigma^{j\ell}\} \\
&= \tfrac{1}{2}\{(\boldsymbol{z}_i - \boldsymbol{\mu})'(\boldsymbol{\sigma}^j\boldsymbol{\sigma}^{\ell\prime} + \boldsymbol{\sigma}^\ell\boldsymbol{\sigma}^{j\prime})(\boldsymbol{z}_i - \boldsymbol{\mu}) - 2\sigma^{j\ell}\} \\
&= \boldsymbol{\sigma}^{j\prime}(\boldsymbol{z}_i - \boldsymbol{\mu})(\boldsymbol{z}_i - \boldsymbol{\mu})'\boldsymbol{\sigma}^\ell - \sigma^{j\ell}
\end{aligned}$$

The derivatives w.r.t. the variances and covariances can be written in vector form, which simplifies the algebra and the programming of the Murphy-Topel correction. To do this we need to define the matrix operator dg(.), which leaves the diagonal of a matrix unchanged, but puts all off-diagonal terms to zero. That is $\boldsymbol{D} = \text{dg}(\boldsymbol{C})$ implies that

$$d_{ij} = \begin{cases} c_{ii} & i = j \\ 0 & i \neq j \end{cases}$$

Using this notation we can write

$$\frac{\partial \ell_{1i}}{\partial \mathrm{vech}\boldsymbol{\Sigma}_z} = \mathrm{vech}\left(\boldsymbol{A}_i - \tfrac{1}{2}\mathrm{dg}(\boldsymbol{A}_i)\right), \text{ where} \tag{A.1b}$$

$$\boldsymbol{A}_i = \boldsymbol{\Sigma}_z^{-1}(\boldsymbol{z}_i - \bar{\boldsymbol{z}})(\boldsymbol{z}_i - \bar{\boldsymbol{z}})'\boldsymbol{\Sigma}_z^{-1} - \boldsymbol{\Sigma}_z^{-1} \tag{A.1c}$$

which can be seen directly from Harville (1997, eqs (15.6.7) and (15.8.12)).

Summing (A.1) over the observations and equating to zero yields the likelihood equations

$$\hat{\boldsymbol{\Sigma}}_z^{-1}(\bar{\boldsymbol{z}} - \hat{\boldsymbol{\mu}}) = \boldsymbol{0} \tag{A.2a}$$

$$\hat{\boldsymbol{\Sigma}}_z^{-1}\boldsymbol{S}_z'\hat{\boldsymbol{\Sigma}}_z^{-1} - \hat{\boldsymbol{\Sigma}}_z^{-1} = \boldsymbol{0} \tag{A.2b}$$

$\bar{\boldsymbol{z}}$ and $\boldsymbol{S}_z$ are obviously ML estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_z$. (A.2) can also be considered as moment equations, however, and $\bar{\boldsymbol{z}}$ and $\boldsymbol{S}_z$ are therefore also method of moment (MM) estimates. This is important, since MM estimation does not depend on the normality assumption, and in this case (A.1) can be considered the gradient of the moment equations.

The asymptotic covariance matrix of $\bar{\boldsymbol{z}}$ and $\boldsymbol{S}_z$, which is of dimension $\tfrac{1}{2}k(k+3)$, can be denoted

$$\boldsymbol{V}_1 = \begin{pmatrix} \boldsymbol{V}_{1\mu\mu} & \boldsymbol{V}_{1\sigma\mu}' \\ \boldsymbol{V}_{1\sigma\mu} & \boldsymbol{V}_{1\sigma\sigma} \end{pmatrix}.$$

Assuming normality we can use the references given in footnote 4 to show that

$$\hat{\boldsymbol{V}}_{1\mu\mu} = \tfrac{1}{n}\boldsymbol{S}_z \tag{A.3a}$$

$$\hat{\boldsymbol{V}}_{1\mu\sigma} = \boldsymbol{0} \tag{A.3b}$$

$$\hat{\boldsymbol{V}}_{1\sigma_{jk}\sigma_{\ell m}} = \tfrac{1}{n}(s_{j\ell}s_{km} + s_{jm}s_{k\ell}). \tag{A.3c}$$

If we do not assume normality, but consider $\bar{\boldsymbol{z}}$ and $\boldsymbol{S}_z$ as MM estimates, then the elements of their asymptotic covariance matrix can be estimated, using Greene (2000, eq (4.55)), as

$$\hat{\boldsymbol{V}}_{1\mu_j\sigma_{k\ell}} = \tfrac{1}{n}s_{jk\ell}$$

$$\hat{\boldsymbol{V}}_{1\sigma_{jk}\sigma_{\ell m}} = \tfrac{1}{n}(s_{jk\ell m} - s_{jk}s_{\ell m}), \text{ where}$$

$s_{jk\ell} = \tfrac{1}{n}\sum_i (z_{ji} - \bar{z}_j)(z_{ki} - \bar{z}_k)(z_{\ell i} - \bar{z}_\ell)$ and $s_{jk\ell m} = \tfrac{1}{n}\sum_i (z_{ji} - \bar{z}_j)(z_{ki} - \bar{z}_k)(z_{\ell i} - \bar{z}_\ell)(z_{mi} - \bar{z}_m)$, while (A.3a) will still hold. If we define $\boldsymbol{S}_i = (z_i - \bar{z})(z_i - \bar{z})'$ then we can write

$$\hat{\boldsymbol{V}}_{1\mu\mu} = \tfrac{1}{n}\boldsymbol{S}_z \tag{A.4a}$$

$$\hat{\boldsymbol{V}}_{1\sigma\sigma} = \tfrac{1}{n}\left\{\left(\tfrac{1}{n}\sum_i \text{vech}(\boldsymbol{S}_i)(\text{vech}(\boldsymbol{S}_i))'\right) - \text{vech}(\boldsymbol{S}_z)(\text{vech}(\boldsymbol{S}_z))'\right\} \tag{A.4b}$$

$$\hat{\boldsymbol{V}}_{1\mu\sigma} = \tfrac{1}{n^2}\left\{\sum_i (z_i - \bar{z})(\text{vech}(\boldsymbol{S}_i))'\right\} \tag{A.4c}$$

A.2) LIML Estimation of α and β

Using (20) we can see that the second stage likelihood for the i^{th} observation is

$$\ell_{2i} = y_i \ln\left\{\Phi\left(\mu_i^* / \sigma_*\right)\right\} + (1 - y_i)\ln\left\{1 - \Phi\left(\mu_i^* / \sigma_*\right)\right\},$$

where μ_i^* and σ_*^2 are given by

$$\mu_i^* = \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \boldsymbol{\beta}'(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1}(z_i - \boldsymbol{\mu})$$

$$\sigma_*^2 = \boldsymbol{\beta}'[(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z) - (\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)]\boldsymbol{\beta} + 1.$$

We need to find the derivatives of this likelihood in terms of $\boldsymbol{\omega}' = (\boldsymbol{\mu}', \text{vech}(\boldsymbol{\Sigma}_z)', \alpha, \boldsymbol{\beta}')$, Writing $\theta_i = \mu_i^*(\sigma_*^2)^{-½}$ we can adapt Amemiya, T (1985, eq (9.2.8)) and obtain

$$\frac{\partial \ell_{2i}}{\partial \boldsymbol{\omega}} = \lambda_i \frac{\partial \theta_i}{\partial \boldsymbol{\omega}}, \text{ where}$$

$$\lambda_i = \frac{y_i - \Phi_i}{\Phi_i(1 - \Phi_i)} \cdot \phi_i\,. \tag{A.5a}$$

We find for $\vartheta \in \boldsymbol{\omega}$ that

$$\frac{\partial \theta}{\partial \vartheta} = \frac{\partial \mu^*}{\partial \vartheta}\sigma_*^{-1} - \frac{1}{2}\frac{\partial \sigma_*^2}{\partial \vartheta}\mu^* \sigma_*^{-3}$$

The derivatives on the right hand side of the above equation can be found using the results in Amemiya, T (1985, pp. 461-2), Lütkepohl (1996, Chapter 10) and Harville (1997, Chapter 15). For the conditional mean we obtain

$$\frac{\partial \mu_i^*}{\partial \alpha} = 1$$

$$\frac{\partial \mu_i^*}{\partial \boldsymbol{\beta}} = \boldsymbol{\mu} + (\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1}(z_i - \boldsymbol{\mu})$$

$$\frac{\partial \mu_i^*}{\partial \boldsymbol{\mu}} = \boldsymbol{\beta} - \boldsymbol{\Sigma}_z^{-1}(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\beta}$$

$$\begin{aligned}
\frac{\partial \mu_i^*}{\partial \sigma_{jj}} &= \boldsymbol{\beta}' \frac{\partial(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)}{\partial \sigma_{jj}} \boldsymbol{\Sigma}_z^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}) + \boldsymbol{\beta}'(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\frac{\partial \boldsymbol{\Sigma}_z^{-1}}{\partial \sigma_{jj}}(\boldsymbol{z}_i - \boldsymbol{\mu}) \\
&= \boldsymbol{\beta}'\left\{\pi_{jj}\boldsymbol{E}_{jj}\boldsymbol{\Sigma}_z^{-1} - (\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1}\boldsymbol{E}_{jj}\boldsymbol{\Sigma}_z^{-1}\right\}(\boldsymbol{z}_i - \boldsymbol{\mu}) \\
&= \left\{\beta_j \pi_{jj} - \boldsymbol{\beta}'(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\sigma}^j\right\}\boldsymbol{\sigma}^{j\prime}(\boldsymbol{z}_i - \boldsymbol{\mu})
\end{aligned}$$

$$\frac{\partial \mu_i^*}{\partial \sigma_{j\ell}} = \{\pi_{j\ell}(\beta_j \boldsymbol{\sigma}^{\ell\prime} + \beta_\ell \boldsymbol{\sigma}^{j\prime}) - \boldsymbol{\beta}'(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)(\boldsymbol{\sigma}^j \boldsymbol{\sigma}^{\ell\prime} + \boldsymbol{\sigma}^\ell \boldsymbol{\sigma}^{j\prime})\}(\boldsymbol{z}_i - \boldsymbol{\mu}),$$

while for the conditional variance we obtain

$$\frac{\partial \sigma_*^2}{\partial \alpha} = 0 \text{ and } \frac{\partial \sigma_*^2}{\partial \boldsymbol{\mu}} = \boldsymbol{0}$$

$$\frac{\partial \sigma_*^2}{\partial \boldsymbol{\beta}} = 2[(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z) - (\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)]\boldsymbol{\beta}$$

$$\begin{aligned}
\frac{\partial \sigma_*^2}{\partial \sigma_{jj}} &= \boldsymbol{\beta}'\left\{\frac{\partial(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)}{\partial \sigma_{jj}} - (\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1}\frac{\partial(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)}{\partial \sigma_{jj}} - \frac{\partial(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)}{\partial \sigma_{jj}}\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z) + \right. \\
&\qquad \left. + (\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1}\frac{\partial \boldsymbol{\Sigma}_z^{-1}}{\partial \sigma_{jj}}\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\right\}\boldsymbol{\beta} \\
&= \boldsymbol{\beta}'\left\{\pi_{jj}\boldsymbol{E}_{jj} - \pi_{jj}(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1}\boldsymbol{E}_{jj} - \pi_{jj}\boldsymbol{E}_{jj}\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z) + \right. \\
&\qquad \left. + (\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1}\boldsymbol{E}_{jj}\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\right\}\boldsymbol{\beta} \\
&= \pi_{jj}\beta_j^2 - 2\pi_{jj}\beta_j\boldsymbol{\beta}'(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\sigma}^j + \boldsymbol{\beta}'(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\sigma}^j\boldsymbol{\sigma}^{j\prime}(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\beta}
\end{aligned}$$

$$\frac{\partial \sigma_*^2}{\partial \sigma_{j\ell}} = 2\pi_{j\ell}\beta_j\beta_\ell - 2\pi_{j\ell}\boldsymbol{\beta}'(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)(\beta_j\boldsymbol{\sigma}^\ell + \beta_\ell\boldsymbol{\sigma}^j) + \boldsymbol{\beta}'(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)(\boldsymbol{\sigma}^j\boldsymbol{\sigma}^{\ell\prime} + \boldsymbol{\sigma}^\ell\boldsymbol{\sigma}^{j\prime})(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\beta}$$

These results can be substituted back into the derivatives of the likelihood. If we define

$$\bar{\boldsymbol{\Pi}} = \boldsymbol{1} - \boldsymbol{\Pi} \tag{A.5b}$$

$$\boldsymbol{\Delta} = \boldsymbol{I} - (\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1} = (\bar{\boldsymbol{\Pi}} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1} \text{ and} \tag{A.5c}$$

$$\psi_{j\ell} = \beta_j\pi_{j\ell} - \boldsymbol{\beta}'(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\sigma}^j = \boldsymbol{\beta}'(\bar{\boldsymbol{\Pi}} * \boldsymbol{\Sigma}_z)\boldsymbol{\sigma}^j - \beta_j\bar{\pi}_{j\ell} \tag{A.5d}$$

then, remembering the symmetry of $\boldsymbol{\Pi}$ and $\boldsymbol{\Delta}$, we obtain the gradients

$$\frac{\partial \ell_{2i}}{\partial \alpha} = \lambda_i \sigma_*^{-1}$$

$$\frac{\partial \ell_{2i}}{\partial \boldsymbol{\beta}} = \lambda_i \sigma_*^{-1}\{\boldsymbol{\mu} + (\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu})\} - \lambda_i \mu_i^* \sigma_*^{-3}\{(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z) - (\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\}\boldsymbol{\beta}$$
$$= \lambda_i \sigma_*^{-1}\left\{\boldsymbol{\Delta\mu} + (\boldsymbol{I} - \boldsymbol{\Delta})\boldsymbol{z}_i - \mu_i^* \sigma_*^{-2}(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Delta\beta}\right\}$$

$$\frac{\partial \ell_{2i}}{\partial \boldsymbol{\mu}} = \lambda_i \sigma_*^{-1}\{\boldsymbol{\beta} - \boldsymbol{\Sigma}_z^{-1}(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\beta}\} = \lambda_i \sigma_*^{-1}\boldsymbol{\Delta\beta}$$

$$\frac{\partial \ell_{2i}}{\partial \sigma_{jj}} = \lambda_i \sigma_*^{-1}\left\{\beta_j \pi_{jj} - \boldsymbol{\beta}'(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\sigma}^j\right\}\boldsymbol{\sigma}^{j\prime}(\boldsymbol{z}_i - \boldsymbol{\mu}) - \tfrac{1}{2}\lambda_i \mu_i^* \sigma_*^{-3}\left\{\pi_{jj}\beta_j^2 - \right.$$
$$\left. -2\pi_{jj}\beta_j \boldsymbol{\beta}'(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\sigma}^j + \boldsymbol{\beta}'(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\sigma}^j \boldsymbol{\sigma}^{j\prime}(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\beta}\right\}$$
$$= \lambda_i \sigma_*^{-1}\left\{\psi_{jj}\boldsymbol{\sigma}^{j\prime}(\boldsymbol{z}_i - \boldsymbol{\mu}) - \tfrac{1}{2}\mu_i^* \sigma_*^{-2}[\pi_{jj}\beta_j^2 - 2\pi_{jj}\beta_j(\pi_{jj}\beta_j - \psi_{jj}) + (\pi_{jj}\beta_j - \psi_{jj})^2]\right\}$$
$$= \lambda_i \sigma_*^{-1}\left\{\psi_{jj}\boldsymbol{\sigma}^{j\prime}(\boldsymbol{z}_i - \boldsymbol{\mu}) - \tfrac{1}{2}\mu_i^* \sigma_*^{-2}[\psi_{jj}^2 + \beta_j^2 \pi_{jj}(1 - \pi_{jj})]\right\}$$

$$\frac{\partial \ell_{2i}}{\partial \sigma_{j\ell}} = \lambda_i \sigma_*^{-1}\left\{(\psi_{j\ell}\boldsymbol{\sigma}^{\ell\prime} + \psi_{\ell j}\boldsymbol{\sigma}^{j\prime})(\boldsymbol{z}_i - \boldsymbol{\mu}) - \mu_i^* \sigma_*^{-2}[\psi_{j\ell}\psi_{\ell j} + \beta_j \beta_\ell \pi_{j\ell}(1 - \pi_{j\ell})]\right\}$$

The gradients w.r.t. the sigmas can be simplified if we use the following lemma.

Lemma

*Let **A** be any $(p \times p)$ matrix, **b** and **c** be $(p \times 1)$ vectors and **ι** be the $(p \times 1)$ vector of ones. Explicit matrix multiplication proves the following relationships*

$$\begin{pmatrix} a_{11}b_1 & \cdots & a_{1k}b_k \\ \vdots & \ddots & \vdots \\ a_{k1}b_1 & \cdots & a_{kk}b_k \end{pmatrix} = \boldsymbol{A} * (\boldsymbol{b\iota}'), \textit{ and}$$

$$\begin{pmatrix} c_1 b_1 & \cdots & c_1 b_k \\ \vdots & \ddots & \vdots \\ c_k b_1 & \cdots & c_k b_k \end{pmatrix} = \boldsymbol{cb}'$$

This leads to the following results

$$\frac{\partial \ell_{2i}}{\partial \alpha} = \lambda_i \sigma_*^{-1} \qquad \text{(A.6a)}$$

$$\frac{\partial \ell_{2i}}{\partial \boldsymbol{\beta}} = \lambda_i \sigma_*^{-1} \left\{ \boldsymbol{\Delta\mu} + (\boldsymbol{I} - \boldsymbol{\Delta})\boldsymbol{z}_i - \mu_i^* \sigma_*^{-2} (\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Delta\beta} \right\} \tag{A.6b}$$

$$\frac{\partial \ell_{2i}}{\partial \boldsymbol{\mu}} = \lambda_i \sigma_*^{-1} \boldsymbol{\Delta\beta} \tag{A.6c}$$

$$\frac{\partial \ell_{2i}}{\partial \mathrm{vech}\boldsymbol{\Sigma}_z} = \mathrm{vech}\left(\boldsymbol{B}_i - \tfrac{1}{2}\mathrm{dg}(\boldsymbol{B}_i) \right) \tag{A.6d}$$

where

$$\boldsymbol{B}_i = \lambda_i \sigma_*^{-1} \left\{ \boldsymbol{\Xi}_i + \boldsymbol{\Xi}_i' - \mu_i^* \sigma_*^{-2} \left[(\boldsymbol{\Psi} * \boldsymbol{\Psi}') + ((\boldsymbol{\beta\beta}') * \bar{\boldsymbol{\Pi}} * (\boldsymbol{1} - \bar{\boldsymbol{\Pi}})) \right] \right\} \tag{A.7a}$$

$$\boldsymbol{\Psi} = \boldsymbol{\Sigma}_z^{-1} (\bar{\boldsymbol{\Pi}} * \boldsymbol{\Sigma}_z)\boldsymbol{\beta\iota}' - [(\boldsymbol{\beta\iota}') * \bar{\boldsymbol{\Pi}}] \tag{A.7b}$$

$$\boldsymbol{\Xi}_i = \boldsymbol{\Psi} * [\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu})\boldsymbol{\iota}'] \tag{A.7c}$$

Looking at these results we can see from (A.1) and (A.6) that, conditional on $\boldsymbol{z}$, the gradients of ℓ_2 are linear functions of λ while those of ℓ_1 are not functions of λ at all. The Murphy-Topel matrix $\boldsymbol{R}$, defined after equation (21), therefore has elements that are expected values of linear functions of λ. Since we can also see from (A.5a) that $E(\lambda \mid \boldsymbol{z}) = 0$, it follows that $\boldsymbol{R}$ must also be equal to zero.

We have thus established the following two theorems

Theorem 1

The LIML estimates $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ are found by solving the equations

$$\hat{\sigma}_*^{-1}\left(\sum\nolimits_i \hat{\lambda}_i\right) = 0$$

$$\hat{\sigma}_*^{-1}\left\{ \hat{\boldsymbol{\Delta}}\bar{\boldsymbol{z}}\left(\sum\nolimits_i \hat{\lambda}_i\right) + (\boldsymbol{I} - \hat{\boldsymbol{\Delta}})\left(\sum\nolimits_i \hat{\lambda}_i \boldsymbol{z}_i\right) - \hat{\sigma}_*^{-2}(\boldsymbol{\Pi} * \boldsymbol{S}_z)\hat{\boldsymbol{\Delta}}\hat{\boldsymbol{\beta}}\left(\sum\nolimits_i \hat{\lambda}_i \hat{\mu}_i^*\right) \right\} = \boldsymbol{0}$$

where

$$\hat{\boldsymbol{\Delta}} = (\bar{\boldsymbol{\Pi}} * \boldsymbol{S}_z)\boldsymbol{S}_z^{-1}$$

$$\hat{\mu}_i^* = \hat{\alpha} + \hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\Delta}}\bar{\boldsymbol{z}} + \hat{\boldsymbol{\beta}}'(\boldsymbol{I} - \hat{\boldsymbol{\Delta}})\boldsymbol{z}_i$$

$$\hat{\sigma}_*^2 = \hat{\boldsymbol{\beta}}'(\boldsymbol{\Pi} * \boldsymbol{S}_z)\hat{\boldsymbol{\Delta}}\hat{\boldsymbol{\beta}} + 1$$

$$\hat{\lambda}_i = \frac{y_i - \Phi(\hat{\mu}_i^* / \hat{\sigma}_*)}{\Phi(\hat{\mu}_i^* / \hat{\sigma}_*)\left(1 - \Phi(\hat{\mu}_i^* / \hat{\sigma}_*)\right)} \cdot \phi(\hat{\mu}_i^* / \hat{\sigma}_*)$$

Theorem 2

Since $\boldsymbol{R} = \boldsymbol{0}$, *the Murphy-Topel adjusted covariance matrix for* $(\alpha, \boldsymbol{\beta})$ *is estimated by*

$$\hat{\boldsymbol{V}}_2^* = \hat{\boldsymbol{V}}_2 + \hat{\boldsymbol{V}}_2 \hat{\boldsymbol{C}} \hat{\boldsymbol{V}}_1 \hat{\boldsymbol{C}}' \hat{\boldsymbol{V}}_2$$

where $\hat{\boldsymbol{V}}_1$ *is given by (A.3) or (A.4), depending on whether or not the regressors are assumed to be normally distributed. Denoting* $\omega_1' = (\boldsymbol{\mu}', \text{vech}(\boldsymbol{\Sigma}_z)')$ *and* $\omega_2' = (\alpha, \boldsymbol{\beta}')$ *we have*

$$\hat{\boldsymbol{V}}_2 = \left[\sum_i \left(\frac{\partial \ell_{2i}}{\partial \hat{\omega}_2} \right) \left(\frac{\partial \ell_{2i}}{\partial \hat{\omega}_2'} \right) \right]^{-1}$$

$$\hat{\boldsymbol{C}} = \sum_i \left(\frac{\partial \ell_{2i}}{\partial \hat{\omega}_2} \right) \left(\frac{\partial \ell_{2i}}{\partial \hat{\omega}_1'} \right).$$

The elements of these matrices are found using (A.1), (A.5), (A.6) and (A.7), with $\boldsymbol{\omega}'$ *replaced by* $\hat{\boldsymbol{\omega}}' = (\bar{z}', \text{vech}(\boldsymbol{S}_z)', \hat{\alpha}, \hat{\boldsymbol{\beta}}')$.[23]

A.3) Full Information Estimation of the Parameters

The FIML method is to maximize $\ell = \ell_1 + \ell_2 = \sum_i \left(\ell_{1i} + \ell_{2i} \right)$ over all the parameters simultaneously, which leads to the likelihood equations

$$\hat{\hat{\sigma}}_*^{-1} \left(\sum_i \hat{\hat{\lambda}}_i \right) = 0 \tag{A.8a}$$

$$\hat{\hat{\sigma}}_*^{-1} \left\{ \hat{\hat{\Delta}} \hat{\hat{\mu}} \left(\sum_i \hat{\hat{\lambda}}_i \right) + (\boldsymbol{I} - \hat{\hat{\Delta}}) \left(\sum_i \hat{\hat{\lambda}}_i z_i \right) - \hat{\hat{\sigma}}_*^{-2} (\boldsymbol{\Pi} * \hat{\hat{\Sigma}}_z) \hat{\hat{\Delta}} \hat{\hat{\beta}} \left(\sum_i \hat{\lambda}_i \hat{\hat{\mu}}_i^* \right) \right\} = \boldsymbol{0} \tag{A:8b}$$

$$\hat{\hat{\sigma}}_*^{-1} \hat{\hat{\Delta}} \hat{\hat{\beta}} \left(\sum_i \hat{\hat{\lambda}}_i \right) + \hat{\hat{\Sigma}}_z^{-1} \left(\sum_i (z_i - \hat{\hat{\mu}}) \right) = \boldsymbol{0} \tag{A.8c}$$

[23] Greene (2000, p. 135) multiplies the summations in the estimates of $\boldsymbol{C}$ and $\boldsymbol{R}$ by a factor $1/n$, which is obviously a misprint. Greene (2003 p. 510) keeps these estimates of $\boldsymbol{C}$ and $\boldsymbol{R}$, but changes the definitions of $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ (though not of $\boldsymbol{V}_2^*$). Working through the algebra shows that the estimated adjusted covariance matrix in Theorem 2 is the same as that given in Greene (2003, Theorem 17.8).

$$\operatorname{vech}\left(\sum_i \hat{\hat{A}}_i - \tfrac{1}{2}\operatorname{dg}(\sum_i \hat{\hat{A}}_i)\right) + \operatorname{vech}\left(\sum_i \hat{\hat{B}}_i - \tfrac{1}{2}\operatorname{dg}(\sum_i \hat{\hat{B}}_i)\right) = \mathbf{0} \qquad \text{(A.8d)}$$

From (A.8a) we see that $\sum_i \hat{\hat{\lambda}}_i = 0$, and substitution into (A.8c) yields $\hat{\hat{\mu}} = \bar{z}$. Substituting these results in turn into (17) and (A.8b) leads to

$$\hat{\hat{\sigma}}_*^{-1}\left\{ I - \hat{\hat{\sigma}}_*^{-2}(\Pi * \hat{\hat{\Sigma}}_z)\hat{\hat{\Delta}}\hat{\hat{\beta}}\hat{\hat{\beta}}' \right\}\left(I - \hat{\hat{\Delta}} \right)\left(\sum_i \hat{\hat{\lambda}}_i z_i \right) = \mathbf{0}$$

which, assuming the first two matrices are non-singular, can only equal zero if $\sum_i \hat{\hat{\lambda}}_i z_i = 0$. This in turn implies $\sum_i \hat{\hat{\lambda}}_i \hat{\hat{\mu}}_i^* = 0$, and using these results in (A.7a) shows that $\sum_i \hat{\hat{B}}_i = 0$. Equation (A.8d) thus reduces to (A.1b), which as we have previously shown leads to $\hat{\hat{\Sigma}}_z = S_z$.

The FIML estimate of ω_1 is therefore the same as the LIML estimates. (A.8a) and (A.8b) are now identical to the likelihood equations in *Theorem 1*, which establishes the following theorem.

Theorem 3

In the measurement error probit model with known reliability ratio matrix the FIML and LIML estimates are identical.

A.4) Non-Block Diagonality of the Information Matrix

The "off-block-diagonal" part of the Hessian of the full likelihood can be obtained by summing the derivatives of (A.6c-d) w.r.t. ω_2. Deriving first of all w.r.t. α we obtain

$$\frac{\partial^2 \ell_{2i}}{\partial \mu \partial \alpha} = \frac{\partial \lambda_i}{\partial \alpha}\sigma_*^{-1}\Delta\beta + \lambda_i \frac{\partial \sigma_*^{-1}\Delta\beta}{\partial \alpha} = \frac{\partial \lambda_i}{\partial \alpha}\sigma_*^{-1}\Delta\beta,$$

since $\sigma_*^{-1}\Delta\beta$ is not a function of α. Writing $\theta_i = \mu_i^*(\sigma_*^2)^{-1/2}$, the derivatives of λ w.r.t. the parameters can be expressed as

$$\frac{\partial \lambda_i}{\partial \omega} = -\lambda_i^2 \frac{\partial \theta_i}{\partial \omega} + \frac{\lambda_i}{\phi_i}\frac{\partial \phi_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \omega} = -\lambda_i(\lambda_i + \theta_i)\frac{\partial \theta_i}{\partial \omega},$$

since $\dfrac{\partial \phi_i}{\partial \theta_i} = -\theta_i \phi_i$, see Amemiya, T (9.2.12) or Greene (19.23). To evaluate this function we need to derive θ w.r.t. ω. Letting $\vartheta \in \omega$ we have

$$\frac{\partial\theta}{\partial\vartheta}=\frac{\partial\mu^*}{\partial\vartheta}\sigma_*^{-1}-\frac{1}{2}\frac{\partial\sigma_*^2}{\partial\vartheta}\mu^*\sigma_*^{-3}.$$

In particular we have $\frac{\partial\mu^*}{\partial\alpha}=1$ and $\frac{\partial\sigma_*^2}{\partial\alpha}=0$, and thus

$$\frac{\partial\lambda_i}{\partial\alpha}=-\lambda_i^2\sigma_*^{-1}-\lambda_i\mu_i^*\sigma_*^{-1}.$$

Substituting into the expression for the second derivative and summing we obtain

$$\sum_i\frac{\partial^2\ell_{2i}}{\partial\mu\partial\alpha}=-\sigma_*^{-2}\Delta\beta\left(\Sigma\lambda_i^2+\sigma_*^{-1}\Sigma\lambda_i\mu_i^*\right)$$

Conditioning on $\boldsymbol{Z}$ and taking expectations yields the following element of the information matrix

$$\mathrm{E}\left(-\sum\frac{\partial^2\ell_{2i}}{\partial\mu\partial\alpha}\right)=\sigma_*^{-2}\Delta\beta\sum\frac{\phi_i^2}{\Phi_i(1-\Phi_i)},$$

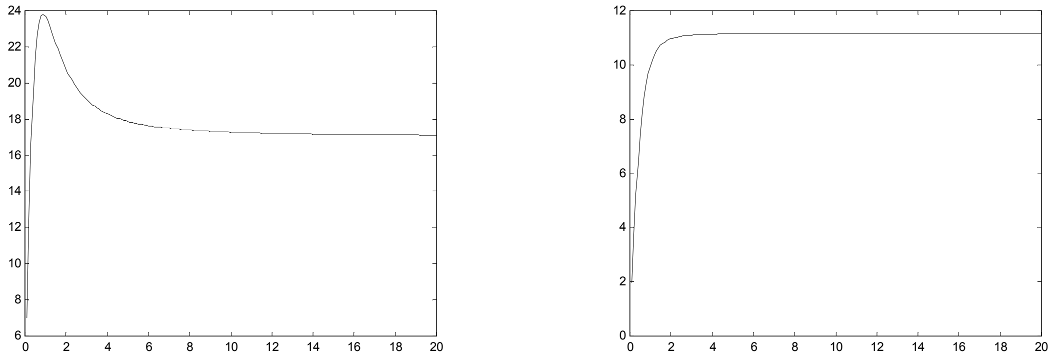
since the expected value and variance of y are Φ and $\Phi(1-\Phi)$. The above can only be zero if Δ is zero, which implies no measurement error. Similar but more complicated exercises can be performed for the other elements of the off-diagonal block.

A.5) Behavior of the Likelihood Function

Maximizing the measurement error likelihood (20) can lead to numerical problems in small samples when there is severe measurement error, in particular the iterations can appear to converge in spite of very large values of the estimated parameters being reported. This result follows, however, from the fact that the likelihood function approaches an asymptote as the absolute values of the parameters approach infinity.

A "well-behaved" likelihood will have a distinct local maximum and then approach the asymptote from above as the parameters increase, see Figure A.1a. No computational problems occur in this case. It is perfectly possible, however, for the likelihood to monotonically increase and approach the asymptote from below, see Figure A.1b. This "ill-behaved" case becomes more likely in small samples with considerable measurement error.

Figure A.1. Typical Log-Likelihood Functions for the Measurement Error
(*a*) *Well-Behaved Likelihood* (*b*) *Ill-Behaved Likelihood*



Since the likelihood is very flat when approaching the asymptote, this can lead to apparent, convergence in the ill-behaved case. Luckily, we can often distinguish between these two situations without having to plot the complete likelihood function.

In the well-behaved case we expect the maximized log-likelihood, $\ell(\hat{\boldsymbol{\beta}})$, to be larger than the asymptote, ℓ_a, whilst in the ill-behaved case they will be almost equal. To evaluate ℓ_a we first normalize the parameters as, *e.g.*, $\gamma_j = \beta_j / \max|\boldsymbol{\beta}|$ and then let $\max|\boldsymbol{\beta}| \to \infty$ in (17) and (18). This leads quite straightforwardly to

$$\frac{\mu_i^*}{\sigma_*} \to \frac{\bar{\boldsymbol{\gamma}}'\{\boldsymbol{\mu} + (\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu})\}}{\sqrt{\bar{\boldsymbol{\gamma}}'\{(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z) - (\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\}\bar{\boldsymbol{\gamma}}}}, \tag{A.9}$$

where $\bar{\boldsymbol{\gamma}} = \lim \boldsymbol{\gamma}$. Substituting the limiting value from (A.9) into (20) yields ℓ_a.

Note that $\bar{\gamma} = \pm 1$ if there is only one explanatory variable, which makes the formula for the asymptote very simple. In the general case we estimate $\bar{\boldsymbol{\gamma}}$ using $\hat{\gamma}_j = \hat{\beta}_j / \max|\hat{\boldsymbol{\beta}}|$, which implies that the limit in (A.9) can be estimated using

$$\frac{\hat{\boldsymbol{\gamma}}'\{\boldsymbol{\mu} + (\boldsymbol{\Pi} * \boldsymbol{S}_z)\boldsymbol{S}_z^{-1}(\boldsymbol{z}_i - \bar{\boldsymbol{z}})\}}{\sqrt{\hat{\boldsymbol{\gamma}}'\{(\boldsymbol{\Pi} * \boldsymbol{S}_z) - (\boldsymbol{\Pi} * \boldsymbol{S}_z)\boldsymbol{S}_z^{-1}(\boldsymbol{\Pi} * \boldsymbol{S}_z)\}\hat{\boldsymbol{\gamma}}}}.$$

A.6) The LIML Estimate with Non-Normal Regressors

In Section 2.5.3 we assert that Assumptions 3´ or 3´´, together with the assumptions we have made concerning the errors, are sufficient for the second-step likelihood to be given by (20). We will now prove this.

The distributional assumptions that we are making concerning the errors can be written

(a) ε, $\boldsymbol{u}$ and $\boldsymbol{x}$ are jointly and pairwise independent

(b) $\varepsilon \sim N(0,1)$ and $\boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_u)$

The distributional assumptions concerning the variables in Assumption 3´ are

(c´) $\boldsymbol{x} \,|\, z \sim N(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$, where $\boldsymbol{\mu}_c = \boldsymbol{\mu} + \boldsymbol{A}(z - \boldsymbol{\mu})$ and $\boldsymbol{\Sigma}_c$ is homoscedastic

The last condition implies that $\boldsymbol{u}|\boldsymbol{z}$ is normal with a linear homoscedastic regression, which together with (a), (b) and (5) means that the conditional distribution of v given z is also normal. To show that this conditional distribution has the same mean and variance as given previously we write

$$(\boldsymbol{x} - \boldsymbol{\mu}) = \boldsymbol{A}(z - \boldsymbol{\mu}) + \boldsymbol{\delta}\,, \tag{A.10}$$

where $\mathrm{E}(\boldsymbol{\delta} | \, z) = \boldsymbol{0}$. Using the usual regression results we have

$$\boldsymbol{A} = \mathrm{Cov}(\boldsymbol{x}, z)[\mathrm{Var}(z)]^{-1} = \boldsymbol{\Sigma}_x \boldsymbol{\Sigma}_z^{-1} = (\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1}\text{, and}$$

$$\mathrm{Var}(\boldsymbol{\delta} | \, z) = \mathrm{Var}(\boldsymbol{\delta}) = \boldsymbol{\Sigma}_x - \boldsymbol{A}\boldsymbol{\Sigma}_z\boldsymbol{A}' = \boldsymbol{\Pi} * \boldsymbol{\Sigma}_z - (\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\boldsymbol{\Sigma}_z^{-1}(\boldsymbol{\Pi} * \boldsymbol{\Sigma}_z)\,.$$

Substituting (A.10) into (1)gives

$$v = \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \boldsymbol{\beta}'\boldsymbol{A}(z - \boldsymbol{\mu}) + (\varepsilon - \boldsymbol{\beta}'\boldsymbol{\delta})\,,$$

and thus $\mathrm{E}(v | \, z) = \alpha + \boldsymbol{\beta}'\boldsymbol{\mu} + \boldsymbol{\beta}'\boldsymbol{A}(z - \boldsymbol{\mu})$ and $\mathrm{Var}(v | \, z) = 1 + \boldsymbol{\beta}'\mathrm{Var}(\boldsymbol{\delta})\boldsymbol{\beta}$. Substituting the results for $\boldsymbol{A}$ and $\mathrm{Var}(\boldsymbol{\delta})$ into these expressions yields (17) and (18), which proves our result.

A similar proof will apply if we use Assumption 3´´. Condition (c´) is replaced by

(c´´) $\boldsymbol{x}_1 \,|\, \boldsymbol{x}_2 \sim N$, $\mathrm{E}(\boldsymbol{x}_1 \,|\, z) = \boldsymbol{\mu}_1 + \boldsymbol{A}_1(z - \boldsymbol{\mu})$ and $\mathrm{Var}(\boldsymbol{x}_1 \,|\, z)$ is homoscedastic.

Conditions (a), (b) and (c´´) imply that both $z_1 \,|\, \boldsymbol{x}_2$ and $v \,|\, \boldsymbol{x}_2$ are normal, and thus $v, z_1 \,|\, \boldsymbol{x}_2$ and $v \,|\, z_1, \boldsymbol{x}_2$ are also normal. The last expression is simply $v \,|\, \boldsymbol{z}$, since $\boldsymbol{x}_2$ is measured without error. It also follows from (c′′) that $(\boldsymbol{x}_1 - \boldsymbol{\mu}_1) = \boldsymbol{A}_1(\boldsymbol{z} - \boldsymbol{\mu}) + \boldsymbol{\delta}_1$, which leads to (A.10) with $\boldsymbol{A} = \begin{pmatrix} & \boldsymbol{A}_1 \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix}$ and $\boldsymbol{\delta} = \begin{pmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{0} \end{pmatrix}$. The results from the proof using condition (c′) thus carry through this case.

Table A1. Statistics related to Study 1A – the Probit MLE ($P = 0.5$ and $Var(z) = 4$)

(a) n = 100.

π	R_m^2	*Estimates of Beta*				*Mean s.e.*	
		Mean	*Min*	*Max*	*s.d.*	*OPG*	*Hessian*
1.0	0.80	1.0459	0.6318	2.1024	0.2069	0.2152	0.1968
0.9	0.78	0.7996	0.4642	1.7229	0.1538	0.1562	0.1458
0.8	0.76	0.6404	0.2770	1.2629	0.1214	0.1243	0.1181
0.7	0.74	0.5363	0.2630	1.2377	0.1118	0.1069	0.1027
0.6	0.71	0.4458	0.2019	0.8108	0.0919	0.0943	0.0911
0.5	0.67	0.3642	0.1054	0.7416	0.0862	0.0844	0.0821
0.4	0.62	0.2913	0.0732	0.5913	0.0758	0.0767	0.0754
0.3	0.55	0.2276	0.0404	0.4759	0.0705	0.0717	0.0709
0.2	0.44	0.1587	–0.0553	0.4044	0.0695	0.0679	0.0674
0.1	0.29	0.0855	–0.1269	0.3354	0.0661	0.0650	0.0648

(b) n = 1000.

π	R_m^2	*Estimates of Beta*				*Mean s.e.*	
		Mean	*Min*	*Max*	*s.d.*	*OPG*	*Hessian*
1.0	0.80	1.0046	0.8416	1.2017	0.0574	0.0590	0.0583
0.9	0.78	0.7745	0.6440	0.9370	0.0437	0.0442	0.0439
0.8	0.76	0.6248	0.5345	0.7580	0.0368	0.0363	0.0361
0.7	0.74	0.5169	0.4290	0.6214	0.0301	0.0313	0.0311
0.6	0.71	0.4286	0.3531	0.5248	0.0284	0.0278	0.0277
0.5	0.67	0.3548	0.2846	0.4731	0.0254	0.0254	0.0253
0.4	0.62	0.2867	0.2158	0.3805	0.0240	0.0234	0.0234
0.3	0.55	0.2217	0.1515	0.2955	0.0220	0.0221	0.0220
0.2	0.44	0.1561	0.0958	0.2326	0.0213	0.0209	0.0209
0.1	0.29	0.0848	0.0207	0.1514	0.0196	0.0202	0.0201

(c) n = 10000.

π	R_m^2	*Estimates of Beta*				*Mean s.e.*	
		Mean	*Min*	*Max*	*s.d.*	*OPG*	*Hessian*
1.0	0.80	1.0009	0.9492	1.0624	0.0177	0.0184	0.0183
0.9	0.78	0.7723	0.7265	0.8155	0.0138	0.0139	0.0138
0.8	0.76	0.6250	0.5905	0.6679	0.0117	0.0114	0.0114
0.7	0.74	0.5164	0.4852	0.5604	0.0099	0.0098	0.0098
0.6	0.71	0.4284	0.3988	0.4607	0.0087	0.0088	0.0088
0.5	0.67	0.3535	0.3286	0.3766	0.0080	0.0080	0.0080

Table A2. Statistics related to Study 1A – the EIVMLE ($P = 0.5$ *and* $Var(z) = 4$)

(a) n = 100.

π	R_m^2	*Estimates of Beta*				*Mean s.e.*	
		Mean	*Min*	*Max*	*s.d.*	*OPG*	*Hessian*
1.0	0.80	1.0459	0.6318	2.1024	0.2069	0.2152	0.1968
0.9	0.78	1.0851	0.5424	3.4397	0.3373	0.3380	0.3131
0.8	0.76	1.1048	0.3603	4.5049	0.4389	0.4943	0.4623
0.7	0.74	1.1535	0.4001	4.2133	0.5481	0.7069	0.6752
0.6	0.71	1.1847	0.3564	4.1302	0.6103	0.9126	0.8735
0.5	0.67	1.1344	0.2155	4.0863	0.5982	0.9148	0.8886
0.4	0.62	1.1144	0.1860	4.3241	0.6227	1.0180	0.9846
0.3	0.55	1.1434	0.1358	4.6428	0.7140	1.2353	1.2136
0.2	0.44	1.1273	–0.2835	5.6483	0.8586	1.5510	1.5409
0.1	0.29	1.1574	–1.9554	7.1817	1.2662	2.5610	2.5612

(b) n = 1000.

π	R_m^2	*Estimates of Beta*				*Mean s.e.*	
		Mean	*Min*	*Max*	*s.d.*	*OPG*	*Hessian*
1.0	0.80	1.0046	0.8416	1.2017	0.0574	0.0590	0.0583
0.9	0.78	1.0073	0.7923	1.3333	0.0787	0.0790	0.0784
0.8	0.76	1.0060	0.7904	1.4524	0.1007	0.0975	0.0967
0.7	0.74	1.0114	0.7407	1.5267	0.1117	0.1154	0.1149
0.6	0.71	1.0135	0.7202	1.6974	0.1399	0.1336	0.1333
0.5	0.67	1.0254	0.6921	2.9242	0.1705	0.1558	0.1553
0.4	0.62	1.0301	0.6357	2.6225	0.1899	0.1765	0.1763
0.3	0.55	1.0306	0.5697	2.2908	0.2074	0.2014	0.2007
0.2	0.44	1.0342	0.5184	3.1714	0.2606	0.2436	0.2436
0.1	0.29	1.0282	0.2083	3.6136	0.3465	0.3491	0.3491

(c) n = 10000.

π	R_m^2	*Estimates of Beta*				*Mean s.e.*	
		Mean	*Min*	*Max*	*s.d.*	*OPG*	*Hessian*
1.0	0.80	1.0009	0.9492	1.0624	0.0177	0.0184	0.0183
0.9	0.78	1.0012	0.9227	1.0797	0.0243	0.0245	0.0244
0.8	0.76	1.0013	0.9146	1.1217	0.0309	0.0300	0.0300
0.7	0.74	1.0021	0.8975	1.1780	0.0354	0.0353	0.0352
0.6	0.71	1.0003	0.8759	1.1658	0.0401	0.0402	0.0402
0.5	0.67	1.0012	0.8721	1.1448	0.0454	0.0454	0.0453

Table A3. Statistics related to Study 1B – the Probit MLE ($P = 0.5$ and $R_m^2 = 0.8$)

(a) n = 100.

π	Var(z)	*Estimates of Beta*				*Mean s.e.*	
		Mean	*Min*	*Max*	*s.d.*	*OPG*	*Hessian*
1.0	4.0	1.0459	0.6318	2.1024	0.2069	0.2152	0.1968
0.9	4.4	0.7927	0.4763	1.6345	0.1557	0.1563	0.1450
0.8	5.0	0.6143	0.2631	1.1296	0.1182	0.1192	0.1125
0.7	5.7	0.4891	0.2173	1.0340	0.1010	0.0957	0.0918
0.6	6.7	0.3884	0.1865	0.7380	0.0765	0.0793	0.0760
0.5	8.0	0.2970	0.1379	0.6145	0.0654	0.0644	0.0623
0.4	10.0	0.2225	0.0808	0.4244	0.0516	0.0522	0.0511
0.3	13.3	0.1598	0.0469	0.3245	0.0417	0.0422	0.0414
0.2	20.0	0.1003	–0.0004	0.2182	0.0320	0.0321	0.0317
0.1	40.0	0.0471	–0.0276	0.1118	0.0219	0.0213	0.0211

(b) n = 1000.

π	Var(z)	*Estimates of Beta*				*Mean s.e.*	
		Mean	*Min*	*Max*	*s.d.*	*OPG*	*Hessian*
1.0	4.0	1.0046	0.8416	1.2017	0.0574	0.0590	0.0583
0.9	4.4	0.7635	0.6394	0.9500	0.0429	0.0437	0.0433
0.8	5.0	0.5963	0.5128	0.7341	0.0357	0.0344	0.0341
0.7	5.7	0.4731	0.3879	0.5722	0.0273	0.0280	0.0278
0.6	6.7	0.3723	0.3007	0.4506	0.0237	0.0231	0.0230
0.5	8.0	0.2900	0.2235	0.3562	0.0191	0.0192	0.0191
0.4	10.0	0.2182	0.1720	0.2703	0.0161	0.0158	0.0158
0.3	13.3	0.1543	0.1193	0.2005	0.0127	0.0128	0.0128
0.2	20.0	0.0975	0.0684	0.1279	0.0098	0.0098	0.0098
0.1	40.0	0.0466	0.0228	0.0682	0.0066	0.0066	0.0066

(c) n = 10000.

π	Var(z)	*Estimates of Beta*				*Mean s.e.*	
		Mean	*Min*	*Max*	*s.d.*	*OPG*	*Hessian*
1.0	4.0	1.0009	0.9492	1.0624	0.0177	0.0184	0.0183
0.9	4.4	0.7612	0.7114	0.8044	0.0136	0.0137	0.0136
0.8	5.0	0.5966	0.5639	0.6370	0.0112	0.0108	0.0108
0.7	5.7	0.4722	0.4468	0.5153	0.0090	0.0088	0.0088
0.6	6.7	0.3720	0.3470	0.4025	0.0071	0.0073	0.0073

Table A4. Statistics related to Study 1B – the EIVMLE ($P = 0.5$ *and* $R_m^2 = 0.8$)

(*a*) n = 100.

π	Var(z)	*Estimates of Beta*				*Mean s.e.*	
		Mean	*Min*	*Max*	*s.d.*	*OPG*	*Hessian*
1.0	4.0	1.0459	0.6318	2.1024	0.2069	0.2152	0.1968
0.9	4.4	1.1081	0.5615	4.9760	0.4030	0.3914	0.3605
0.8	5.0	1.1168	0.3441	3.9940	0.4637	0.5846	0.5461
0.7	5.7	1.1276	0.3301	3.5251	0.5269	0.8199	0.7844
0.6	6.7	1.0944	0.3381	3.2086	0.4755	0.9169	0.8789
0.5	8.0	1.0114	0.2994	3.0805	0.5157	1.0730	1.0437
0.4	10.0	0.9613	0.2126	2.7659	0.5113	1.2177	1.1950
0.3	13.3	0.8747	0.1620	2.6145	0.4856	1.2939	1.2661
0.2	20.0	0.7599	–0.0018	2.5147	0.4768	1.3853	1.3641
0.1	40.0	0.5703	–0.3252	2.4402	0.4358	1.3166	1.3100

(*b*) n = 1000.

π	Var(z)	*Estimates of Beta*				*Mean s.e.*	
		Mean	*Min*	*Max*	*s.d.*	*OPG*	*Hessian*
1.0	4.0	1.0046	0.8416	1.2017	0.0574	0.0590	0.0583
0.9	4.4	1.0080	0.7951	1.4179	0.0813	0.0815	0.0809
0.8	5.0	1.0084	0.7823	1.6062	0.1146	0.1072	0.1065
0.7	5.7	1.0203	0.6974	1.8361	0.1406	0.1399	0.1390
0.6	6.7	1.0315	0.6479	2.4011	0.1983	0.1842	0.1839
0.5	8.0	1.0651	0.5772	2.5826	0.2671	0.2618	0.2602
0.4	10.0	1.0839	0.5762	2.7838	0.3312	0.3643	0.3640
0.3	13.3	1.0741	0.5328	2.6898	0.3623	0.4784	0.4764
0.2	20.0	1.0363	0.4316	2.5571	0.4000	0.6278	0.6277
0.1	40.0	0.9495	0.2521	2.4415	0.4123	0.8396	0.8382

(*c*) n = 10000.

π	Var(z)	*Estimates of Beta*				*Mean s.e.*	
		Mean	*Min*	*Max*	*s.d.*	*OPG*	*Hessian*
1.0	4.0	1.0009	0.9492	1.0624	0.0177	0.0184	0.0183
0.9	4.4	1.0012	0.9128	1.0834	0.0250	0.0252	0.0252
0.8	5.0	1.0016	0.9080	1.1344	0.0341	0.0327	0.0327
0.7	5.7	1.0027	0.8929	1.2452	0.0426	0.0413	0.0413
0.6	6.7	1.0013	0.8479	1.2684	0.0499	0.0511	0.0511

Table A5. Statistics related to Study 2, n = 100. ($P = 0.5$ and $Var(z) = 4$)

π_1	π_2	R_m^2	*Fail Rate*	*Mean*		*Min*		*Max*		*Standard deviation*		*Mean s.e. (OPG)*		*Mean s.e. (Hessian)*	
Probit				$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
1.0	1.0	0.89		1.1073	1.1083	0.5562	0.5603	3.3739	3.2923	0.3065	0.3047	0.2948	0.2962	0.2513	0.2517
0.8	1.0	0.88		0.6705	0.8404	0.3070	0.4399	1.9445	2.3475	0.1659	0.2082	0.1654	0.1942	0.1481	0.1722
0.6	1.0	0.86		0.4590	0.7627	0.1847	0.4219	1.2975	1.8013	0.1251	0.1695	0.1252	0.1641	0.1145	0.1488
0.4	1.0	0.85		0.3031	0.7551	0.0374	0.3930	0.8853	1.5539	0.1025	0.1586	0.1035	0.1558	0.0968	0.1427
0.2	1.0	0.83		0.1655	0.8301	-0.1023	0.4753	0.9183	2.3035	0.0933	0.1741	0.0963	0.1696	0.0904	0.1541
0.8	0.8	0.86		0.5615	0.5579	0.2583	0.2209	1.3351	1.3867	0.1341	0.1355	0.1279	0.1295	0.1190	0.1196
0.6	0.8	0.85		0.3934	0.5189	0.1279	0.2549	0.8441	1.2543	0.1021	0.1116	0.1031	0.1160	0.0967	0.1079
0.4	0.8	0.83		0.2567	0.5197	-0.0094	0.2560	0.6418	0.9584	0.0864	0.1067	0.0881	0.1099	0.0841	0.1042
0.2	0.8	0.80		0.1403	0.5572	-0.0817	0.3051	0.4633	1.5840	0.0808	0.1141	0.0826	0.1147	0.0794	0.1074
0.6	0.6	0.83		0.3669	0.3649	0.1020	0.1293	0.7602	0.9045	0.0903	0.0934	0.0933	0.0939	0.0886	0.0893
0.4	0.6	0.80		0.2431	0.3685	-0.0598	0.1597	0.8384	0.8294	0.0880	0.0914	0.0813	0.0890	0.0785	0.0857
0.2	0.6	0.76		0.1301	0.3872	-0.1119	0.1183	0.4436	0.7212	0.0744	0.0873	0.0751	0.0895	0.0729	0.0859
0.4	0.4	0.76		0.2433	0.2391	0.0010	0.0129	0.4970	0.5953	0.0784	0.0789	0.0773	0.0773	0.0752	0.0752
0.2	0.4	0.71		0.1284	0.2559	-0.0599	0.0365	0.3821	0.5028	0.0709	0.0759	0.0711	0.0762	0.0697	0.0743
0.2	0.2	0.62		0.1342	0.1394	-0.0952	-0.0540	0.4131	0.4722	0.0649	0.0707	0.0683	0.0686	0.0675	0.0678
EIVMLE															
1.0	1.0	0.89	0 %	1.1073	1.1083	0.5562	0.5603	3.3739	3.2923	0.3065	0.3047	0.2948	0.2962	0.2513	0.2517
0.8	1.0	0.88	5.5 %	1.1948	1.1827	0.4032	0.5012	4.3379	4.3124	0.5852	0.5097	0.8085	0.7023	0.7245	0.6241
0.6	1.0	0.86	13.1 %	1.1566	1.1509	0.3229	0.4906	3.9064	4.0685	0.5989	0.4645	1.0660	0.8457	0.9810	0.7711
0.4	1.0	0.85	17.1 %	1.1050	1.1287	0.0939	0.4408	3.8464	4.3696	0.6695	0.5084	1.3692	1.0140	1.2737	0.9345
0.2	1.0	0.83	18.8 %	1.0446	1.1309	-0.5601	0.5666	4.9897	4.3902	0.8365	0.4954	1.8809	1.1133	1.7619	1.0355
0.8	0.8	0.86	15.3 %	1.0971	1.0925	0.4122	0.2996	3.5733	3.7649	0.4849	0.5051	0.8102	0.8279	0.7476	0.7579
0.6	0.8	0.85	19.0 %	1.1040	1.1051	0.2470	0.3548	3.4202	3.2154	0.5516	0.4877	1.0900	0.9747	1.0115	0.8989
0.4	0.8	0.83	19.6 %	1.0577	1.1070	-0.0471	0.3517	3.9104	3.6965	0.6240	0.5049	1.3314	1.0712	1.2593	1.0115
0.2	0.8	0.80	23.9 %	0.9774	1.1020	-0.5667	0.4399	4.1518	3.6375	0.8005	0.5009	1.8137	1.1626	1.7350	1.1010
0.6	0.6	0.83	21.4 %	1.0761	1.0706	0.2126	0.2449	3.4415	3.2230	0.5285	0.5233	1.1153	1.1059	1.0512	1.0447
0.4	0.6	0.80	26.4 %	0.9850	1.0268	-0.1644	0.2781	3.5792	3.3788	0.5937	0.4988	1.2049	1.0563	1.1555	1.0074
0.2	0.6	0.76	25.4 %	0.9768	1.0954	-2.3292	0.2016	4.8604	3.6608	0.7995	0.5345	1.8316	1.2755	1.7642	1.2187
0.4	0.4	0.76	26.0 %	1.0146	1.0054	0.0037	0.0454	3.9342	3.7763	0.5886	0.5945	1.2672	1.2715	1.2380	1.2366
0.2	0.4	0.71	26.6 %	0.9892	1.0595	-0.8803	0.1201	4.4973	4.1741	0.7959	0.6108	1.7755	1.4048	1.7505	1.3722
0.2	0.2	0.62	27.5 %	1.0432	1.0827	-0.6994	-0.3415	4.7266	5.1428	0.8163	0.8743	1.9578	2.0444	1.9354	2.0351

Table A6. Statistics related to Study 2, n = 1000. ($P = 0.5$ and $Var(z) = 4$)

π_1	π_2	R_m^2	Fail Rate	Mean		Min		Max		Standard deviation		Mean s.e. (OPG)		Mean s.e. (Hessian)	
Probit				$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
1.0	1.0	0.89		1.0068	1.0081	0.8151	0.8248	1.2629	1.2467	0.0688	0.0695	0.0690	0.0692	0.0677	0.0678
0.8	1.0	0.88		0.6296	0.7862	0.5040	0.6254	0.8033	0.9271	0.0418	0.0474	0.0433	0.0497	0.0427	0.0490
0.6	1.0	0.86		0.4313	0.7202	0.3253	0.6070	0.5402	0.9002	0.0338	0.0423	0.0339	0.0437	0.0335	0.0432
0.4	1.0	0.85		0.2865	0.7178	0.1778	0.6137	0.3903	0.8891	0.0283	0.0426	0.0292	0.0421	0.0290	0.0419
0.2	1.0	0.83		0.1572	0.7830	0.0781	0.6635	0.2502	0.9550	0.0287	0.0463	0.0273	0.0451	0.0271	0.0448
0.8	0.8	0.86		0.5330	0.5320	0.4310	0.4115	0.6587	0.6804	0.0362	0.0364	0.0357	0.0355	0.0353	0.0353
0.6	0.8	0.85		0.3722	0.4990	0.2816	0.4129	0.4678	0.6347	0.0287	0.0318	0.0291	0.0325	0.0289	0.0323
0.4	0.8	0.83		0.2497	0.4975	0.1543	0.4049	0.3348	0.6074	0.0250	0.0315	0.0257	0.0314	0.0256	0.0312
0.2	0.8	0.80		0.1325	0.5308	0.0616	0.4180	0.2146	0.6424	0.0237	0.0331	0.0241	0.0323	0.0240	0.0320
0.6	0.6	0.83		0.3531	0.3535	0.2825	0.2696	0.4388	0.4490	0.0259	0.0277	0.0271	0.0271	0.0269	0.0269
0.4	0.6	0.80		0.2350	0.3514	0.1566	0.2738	0.3147	0.4281	0.0246	0.0257	0.0240	0.0261	0.0239	0.0260
0.2	0.6	0.76		0.1236	0.3727	0.0617	0.2880	0.1890	0.4691	0.0214	0.0272	0.0225	0.0263	0.0224	0.0261
0.4	0.4	0.76		0.2334	0.2349	0.1631	0.1652	0.3157	0.3063	0.0229	0.0232	0.0231	0.0231	0.0230	0.0230
0.2	0.4	0.71		0.1249	0.2502	0.0563	0.1720	0.2030	0.3503	0.0218	0.0229	0.0215	0.0228	0.0214	0.0228
0.2	0.2	0.62		0.1321	0.1337	0.0622	0.0602	0.1948	0.2050	0.0217	0.0206	0.0209	0.0209	0.0209	0.0209
EIVMLE															
1.0	1.0	0.89	0 %	1.0068	1.0081	0.8151	0.8248	1.2629	1.2467	0.0688	0.0695	0.0690	0.0692	0.0677	0.0678
0.8	1.0	0.88	0 %	1.0206	1.0182	0.7294	0.7787	1.6859	1.5284	0.1172	0.0988	0.1188	0.1017	0.1171	0.1000
0.6	1.0	0.86	0 %	1.0325	1.0303	0.6400	0.7644	1.9106	1.8056	0.1725	0.1316	0.1698	0.1307	0.1681	0.1292
0.4	1.0	0.85	0 %	1.0396	1.0327	0.4937	0.7506	3.3253	2.6343	0.2412	0.1627	0.2311	0.1546	0.2290	0.1532
0.2	1.0	0.83	0.3 %	1.0753	1.0480	0.4112	0.7095	4.2431	2.7283	0.3791	0.1965	0.3577	0.1820	0.3561	0.1809
0.8	0.8	0.86	0 %	1.0315	1.0295	0.7294	0.6650	1.8585	2.0353	0.1596	0.1599	0.1501	0.1497	0.1486	0.1484
0.6	0.8	0.85	0 %	1.0338	1.0366	0.6261	0.6981	2.0720	2.0904	0.1982	0.1761	0.1966	0.1739	0.1952	0.1727
0.4	0.8	0.83	0.1 %	1.0616	1.0495	0.4827	0.6930	3.1767	2.4713	0.2716	0.2076	0.2677	0.2028	0.2662	0.2014
0.2	0.8	0.80	0.4 %	1.0698	1.0507	0.3804	0.6551	3.2740	2.4249	0.3802	0.2193	0.3867	0.2193	0.3842	0.2175
0.6	0.6	0.83	0.2 %	1.0596	1.0617	0.6634	0.6408	2.5985	2.5402	0.2326	0.2386	0.2314	0.2321	0.2301	0.2308
0.4	0.6	0.80	0.6 %	1.0652	1.0557	0.5007	0.6151	2.8316	2.8515	0.2985	0.2559	0.2911	0.2488	0.2898	0.2476
0.2	0.6	0.76	0.1 %	1.0701	1.0570	0.3803	0.6328	3.2705	2.5598	0.3746	0.2456	0.3964	0.2530	0.3951	0.2520
0.4	0.4	0.76	0.4 %	1.0621	1.0693	0.5487	0.5370	3.1761	3.4252	0.3104	0.3142	0.3044	0.3065	0.3037	0.3057
0.2	0.4	0.71	0.9 %	1.1032	1.0909	0.3748	0.6242	3.8762	3.5548	0.4265	0.3163	0.4332	0.3245	0.4319	0.3235
0.2	0.2	0.62	0.5 %	1.0917	1.1019	0.3737	0.3490	3.9663	4.1182	0.4092	0.4013	0.4057	0.4075	0.4046	0.4062

Table A7. Statistics related to Study 2, n = 10000. ($P = 0.5$ and $Var(z) = 4$)

π_1	π_2	R_m^2	*Fail Rate*	*Mean*		*Min*		*Max*		*Standard deviation*		*Mean s.e. (OPG)*		*Mean s.e. (Hessian)*	
Probit				$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
1.0	1.0	0.89		1.0019	1.0014	0.9268	0.9349	1.0880	1.0824	0.0223	0.0215	0.0212	0.0212	0.0212	0.0212
0.8	1.0	0.88		0.6248	0.7814	0.5842	0.7338	0.6671	0.8370	0.0135	0.0151	0.0134	0.0154	0.0134	0.0153
0.6	1.0	0.86		0.4284	0.7145	0.3972	0.6626	0.4734	0.7637	0.0107	0.0141	0.0105	0.0135	0.0105	0.0135
0.4	1.0	0.85		0.2854	0.7142	0.2599	0.6732	0.3120	0.7606	0.0092	0.0130	0.0091	0.0132	0.0091	0.0132
0.2	1.0	0.83		0.1563	0.7819	0.1308	0.7427	0.1843	0.8330	0.0085	0.0142	0.0085	0.0141	0.0085	0.0141
0.8	0.8	0.86		0.5300	0.5308	0.4973	0.5023	0.5719	0.5626	0.0107	0.0108	0.0111	0.0111	0.0111	0.0111
0.6	0.8	0.85		0.3724	0.4963	0.3480	0.4654	0.4019	0.5307	0.0094	0.0103	0.0091	0.0102	0.0091	0.0101
0.4	0.8	0.83		0.2482	0.4959	0.2242	0.4615	0.2791	0.5313	0.0080	0.0100	0.0081	0.0098	0.0081	0.0098
0.2	0.8	0.80		0.1326	0.5304	0.1095	0.4996	0.1654	0.5654	0.0077	0.0101	0.0076	0.0101	0.0076	0.0101
0.6	0.6	0.83		0.3516	0.3514	0.3280	0.3256	0.3791	0.3775	0.0084	0.0087	0.0085	0.0085	0.0085	0.0085
0.4	0.6	0.80		0.2344	0.3508	0.2118	0.3258	0.2701	0.3767	0.0076	0.0080	0.0076	0.0082	0.0075	0.0082
0.2	0.6	0.76		0.1243	0.3726	0.1033	0.3491	0.1468	0.4020	0.0071	0.0082	0.0071	0.0082	0.0071	0.0082
0.4	0.4	0.76		0.2341	0.2343	0.2099	0.2130	0.2572	0.2564	0.0073	0.0071	0.0073	0.0073	0.0073	0.0073
0.2	0.4	0.71		0.1243	0.2482	0.1051	0.2263	0.1475	0.2762	0.0067	0.0073	0.0067	0.0072	0.0067	0.0072
0.2	0.2	0.62		0.1325	0.1325	0.1098	0.1102	0.1510	0.1524	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066
EIVMLE															
1.0	1.0	0.89	0 %	1.0019	1.0014	0.9268	0.9349	1.0880	1.0824	0.0223	0.0215	0.0212	0.0212	0.0212	0.0212
0.8	1.0	0.88	0 %	1.0009	1.0013	0.8999	0.9227	1.1192	1.1056	0.0356	0.0300	0.0353	0.0302	0.0352	0.0301
0.6	1.0	0.86	0 %	1.0011	1.0014	0.8698	0.8831	1.2434	1.2037	0.0494	0.0389	0.0483	0.0369	0.0483	0.0368
0.4	1.0	0.85	0 %	1.0008	1.0010	0.8427	0.8963	1.2100	1.1276	0.0636	0.0417	0.0630	0.0414	0.0629	0.0413
0.2	1.0	0.83	0 %	1.0060	1.0047	0.7676	0.9015	1.3643	1.1920	0.0912	0.0439	0.0909	0.0436	0.0909	0.0436
0.8	0.8	0.86	0 %	1.0033	1.0048	0.8850	0.8979	1.1660	1.1369	0.0417	0.0421	0.0428	0.0428	0.0427	0.0428
0.6	0.8	0.85	0 %	1.0044	1.0036	0.8736	0.8837	1.2072	1.1899	0.0565	0.0488	0.0550	0.0482	0.0549	0.0481
0.4	0.8	0.83	0 %	1.0047	1.0029	0.8188	0.8528	1.3070	1.1999	0.0695	0.0525	0.0693	0.0515	0.0692	0.0515
0.2	0.8	0.80	0 %	1.0090	1.0068	0.7458	0.8839	1.5823	1.2860	0.1008	0.0538	0.0985	0.0531	0.0984	0.0531
0.6	0.6	0.83	0 %	1.0065	1.0059	0.8466	0.8597	1.2353	1.2466	0.0586	0.0598	0.0593	0.0593	0.0593	0.0593
0.4	0.6	0.80	0 %	1.0064	1.0036	0.8265	0.8390	1.3879	1.2606	0.0738	0.0611	0.0725	0.0612	0.0725	0.0612
0.2	0.6	0.76	0 %	1.0114	1.0089	0.7630	0.8524	1.4548	1.3204	0.1024	0.0627	0.1007	0.0617	0.1006	0.0617
0.4	0.4	0.76	0 %	1.0059	1.0067	0.8146	0.8277	1.2925	1.3024	0.0749	0.0737	0.0730	0.0731	0.0730	0.0730
0.2	0.4	0.71	0 %	1.0106	1.0080	0.7739	0.8199	1.4487	1.3363	0.0993	0.0723	0.0981	0.0709	0.0981	0.0709
0.2	0.2	0.62	0 %	1.0067	1.0068	0.7422	0.7358	1.3437	1.3731	0.0895	0.0901	0.0908	0.0908	0.0908	0.0908

Table A8. Statistics related to Study 3, n = 1000 ($Var(z) = 4$)

	π	Probit						EIVML						
		Beta Estimates				Standard Errors		Beta Estimates				Standard Errors		Failure Rate
		Mean	Min	Max	s.d.	OPG	Hessian	Mean	Min	Max	s.d.	OPG	Hessian	
P = 0.5	1.0	1.0051	0.8258	1.2603	0.0594	0.0590	0.0583	1.0051	0.8258	1.2603	0.0594	0.0590	0.0583	0%
	0.8	0.6270	0.5302	0.7506	0.0375	0.0364	0.0362	1.0121	0.7817	1.4197	0.1020	0.0983	0.0977	0%
	0.6	0.4299	0.3453	0.5302	0.0278	0.0280	0.0278	1.0194	0.6968	1.7662	0.1385	0.1357	0.1347	0%
	0.4	0.2858	0.2059	0.3853	0.0235	0.0234	0.0234	1.0235	0.5958	2.9119	0.1941	0.1756	0.1753	0%
	0.2	0.1550	0.0855	0.2293	0.0204	0.0209	0.0209	1.0173	0.4550	2.8789	0.2396	0.2361	0.2360	0%
P = 0.6	1.0	1.0043	0.8391	1.2127	0.0575	0.0590	0.0586	1.0043	0.8391	1.2127	0.0575	0.0590	0.0586	0%
	0.8	0.6342	0.5312	0.7886	0.0369	0.0367	0.0364	1.0083	0.7690	1.5403	0.0997	0.0972	0.0964	0%
	0.6	0.4413	0.3621	0.5613	0.0270	0.0280	0.0279	1.0164	0.7291	1.9065	0.1293	0.1302	0.1295	0%
	0.4	0.3009	0.2317	0.3716	0.0235	0.0235	0.0235	1.0295	0.6666	1.7271	0.1627	0.1607	0.1605	0%
	0.2	0.1710	0.0844	0.2420	0.0220	0.0208	0.0208	1.0244	0.5939	2.0604	0.1949	0.1834	0.1833	0%
P = 0.75	1.0	1.0062	0.8505	1.2443	0.0619	0.0621	0.0613	1.0062	0.8505	1.2443	0.0619	0.0621	0.0613	0%
	0.8	0.6798	0.5748	0.8341	0.0393	0.0383	0.0383	1.0122	0.7823	1.4646	0.0999	0.0952	0.0946	0%
	0.6	0.5025	0.4236	0.6005	0.0297	0.0289	0.0290	1.0181	0.7541	1.6161	0.1193	0.1149	0.1143	0%
	0.4	0.3686	0.2972	0.4554	0.0244	0.0237	0.0239	1.0126	0.7182	1.6589	0.1159	0.1169	0.1166	0%
	0.2	0.2451	0.1664	0.3151	0.0212	0.0208	0.0209	1.0088	0.7528	1.3935	0.1028	0.1021	0.1020	0%
P = 0.9	1.0	1.0116	0.8382	1.3121	0.0758	0.0744	0.0726	1.0116	0.8382	1.3121	0.0758	0.0744	0.0726	0%
	0.8	0.7787	0.6563	0.9826	0.0511	0.0455	0.0469	1.0150	0.7807	1.5074	0.1054	0.0993	0.0985	0%
	0.6	0.6380	0.5309	0.7690	0.0383	0.0334	0.0353	1.0193	0.7645	1.6195	0.1091	0.1053	0.1047	0%
	0.4	0.5207	0.4394	0.6441	0.0314	0.0263	0.0282	1.0117	0.8039	1.5910	0.0978	0.0939	0.0937	0%
	0.2	0.4054	0.3337	0.4919	0.0245	0.0219	0.0233	1.0068	0.8425	1.3033	0.0702	0.0723	0.0723	0%
P = 0.98	1.0	1.0234	0.7768	1.8972	0.1213	0.1245	0.1155	1.0234	0.7768	1.8972	0.1213	0.1245	0.1155	0%
	0.8	0.9261	0.6927	1.6418	0.1066	0.0776	0.0832	1.0432	0.7587	2.3637	0.1693	0.1556	0.1496	0.3%
	0.6	0.8420	0.6575	1.4226	0.0868	0.0540	0.0624	1.0407	0.7603	1.8978	0.1582	0.1461	0.1427	0.1%
	0.4	0.7454	0.6110	0.9480	0.0608	0.0388	0.0466	1.0213	0.7727	2.1448	0.1238	0.1143	0.1136	0%
	0.2	0.6417	0.5152	0.7892	0.0463	0.0294	0.0356	1.0088	0.7894	1.3896	0.0817	0.0804	0.0802	0%

Table A9. Statistics related to Study 4, n = 100

	π	*Probit*						*EIVML*						
		Beta Estimates				*Standard Errors*		*Beta Estimates*				*Standard Errors*		*Failure Rate*
		Mean $\hat{\beta}_1$	*Mean* $\hat{\beta}_2$	*s.d.* $\hat{\beta}_1$	*s.d.* $\hat{\beta}_2$	se($\hat{\beta}_1$)	se($\hat{\beta}_2$)	*Mean* $\hat{\beta}_1$	*Mean* $\hat{\beta}_2$	*s.d.* $\hat{\beta}_1$	*s.d.* $\hat{\beta}_2$	se($\hat{\beta}_1$)	se($\hat{\beta}_2$)	
$\rho = 0$	1.0	1.0714	1.0852	0.2640	0.2208	0.3438	0.2942	1.0714	1.0852	0.2640	0.2208	0.3438	0.2942	0%
	0.8	0.8116	0.6597	0.1907	0.1299	0.2016	0.1320	1.0953	1.1162	0.4839	0.5205	0.4726	0.4817	3.1%
	0.6	0.7434	0.4500	0.1659	0.0990	0.1714	0.1022	0.9983	1.0005	0.6164	0.7316	0.5523	0.5850	10.8%
	0.4	0.7325	0.2936	0.1522	0.0833	0.1572	0.0837	0.9106	0.8941	0.6090	0.8335	0.5283	0.6043	14.5%
	0.2	0.7916	0.1589	0.1472	0.0762	0.1518	0.0796	0.8946	0.8503	0.6594	1.1918	0.5220	0.7296	15.1%
$\rho = 0.2$	1.0	1.0852	1.1016	0.2662	0.2370	0.3222	0.2904	1.0852	1.1016	0.2662	0.2370	0.3222	0.2904	0%
	0.8	0.8912	0.6618	0.1952	0.1389	0.2063	0.1536	1.0802	1.0932	0.4810	0.5647	0.4750	0.5272	5.2%
	0.6	0.8394	0.4482	0.1704	0.1052	0.1899	0.1106	0.9745	0.9877	0.5706	0.7558	0.5290	0.5825	11.5%
	0.4	0.8635	0.2964	0.1589	0.0893	0.1637	0.0949	0.9000	0.8829	0.5969	0.9308	0.5213	0.6383	16.6%
	0.2	0.9155	0.1551	0.1532	0.0802	0.1604	0.0812	0.8744	0.8080	0.5584	1.2090	0.4938	0.7242	16.8%
$\rho = 0.3$	1.0	1.0775	1.0929	0.2616	0.2415	0.2793	0.2699	1.0775	1.0929	0.2616	0.2415	0.2793	0.2699	0%
	0.8	0.9185	0.6512	0.1966	0.1426	0.2186	0.1585	1.0712	1.0984	0.5028	0.6187	0.4951	0.5647	5.0%
	0.6	0.8859	0.4360	0.1716	0.1081	0.1756	0.1126	0.9748	0.9853	0.5512	0.7814	0.4992	0.5838	10.7%
	0.4	0.9200	0.2859	0.1618	0.0921	0.1785	0.0960	0.8961	0.8789	0.5570	0.9448	0.5043	0.6307	15.6%
	0.2	0.9928	0.1546	0.1578	0.0833	0.1592	0.0837	0.9101	0.8542	0.6349	1.4704	0.5147	0.8036	17.7%
$\rho = 0.5$	1.0	1.0866	1.1141	0.2695	0.2771	0.3150	0.3332	1.0866	1.1141	0.2695	0.2771	0.3150	0.3332	0%
	0.8	1.0208	0.6468	0.2073	0.1634	0.2341	0.1916	1.0404	1.0835	0.4982	0.7319	0.4837	0.6256	7.8%
	0.6	1.0482	0.4082	0.1843	0.1219	0.1971	0.1318	0.9606	0.9419	0.5665	0.9400	0.5199	0.6531	13.4%
	0.4	1.1130	0.2552	0.1762	0.1022	0.1916	0.1136	0.9253	0.8286	0.5669	1.1142	0.5247	0.6872	16.8%
	0.2	1.1575	0.1329	0.1709	0.0909	0.1730	0.0969	0.9517	0.8363	0.6559	1.8672	0.5142	0.9250	16.6%
$\rho = 0.6$	1.0	1.0854	1.1320	0.2732	0.3127	0.3478	0.4142	1.0854	1.1320	0.2732	0.3127	0.3478	0.4142	0%
	0.8	1.1081	0.6073	0.2170	0.1775	0.2588	0.2054	1.0362	1.0199	0.4479	0.7137	0.4652	0.5862	7.4%
	0.6	1.1620	0.3849	0.1967	0.1330	0.2155	0.1431	0.9466	0.8952	0.5374	1.0101	0.5162	0.6733	15.0%
	0.4	1.2259	0.2383	0.1879	0.1108	0.2044	0.1219	0.9057	0.7920	0.5199	1.2503	0.5222	0.7328	19.6%
	0.2	1.2663	0.1164	0.1835	0.0975	0.2148	0.1079	0.9636	0.7342	0.6724	2.1015	0.5723	1.0032	18.8%

Table A10. Statistics related to Study 4, n = 1000

	π	Probit						EIVML						
		Beta Estimates				Standard Errors		Beta Estimates				Standard Errors		Failure Rate
		Mean $\hat{\beta}_1$	Mean $\hat{\beta}_2$	s.d. $\hat{\beta}_1$	s.d. $\hat{\beta}_2$	se($\hat{\beta}_1$)	se($\hat{\beta}_2$)	Mean $\hat{\beta}_1$	Mean $\hat{\beta}_2$	s.d. $\hat{\beta}_1$	s.d. $\hat{\beta}_2$	se($\hat{\beta}_1$)	se($\hat{\beta}_2$)	
$\rho = 0$	1.0	1.0059	1.0057	0.0757	0.0613	0.0738	0.0603	1.0059	1.0057	0.0757	0.0613	0.0738	0.0603	0%
	0.8	0.7861	0.6280	0.0576	0.0384	0.0591	0.0390	1.0158	1.0153	0.1014	0.1043	0.1066	0.1078	0%
	0.6	0.7172	0.4307	0.0506	0.0298	0.0495	0.0298	1.0207	1.0249	0.1212	0.1460	0.1209	0.1482	0%
	0.4	0.7171	0.2873	0.0471	0.0255	0.0473	0.0259	1.0307	1.0402	0.1397	0.1992	0.1548	0.2222	0%
	0.2	0.7797	0.1565	0.0456	0.0233	0.0459	0.0223	1.0213	1.0382	0.1381	0.2735	0.1410	0.2654	0.1%
$\rho = 0.2$	1.0	1.0062	1.0043	0.0750	0.0642	0.0763	0.0641	1.0062	1.0043	0.0750	0.0642	0.0763	0.0641	0%
	0.8	0.8435	0.6225	0.0580	0.0402	0.0599	0.0411	1.0120	1.0126	0.0974	0.1095	0.1026	0.1133	0%
	0.6	0.8094	0.4257	0.0515	0.0313	0.0516	0.0320	1.0196	1.0297	0.1156	0.1566	0.1224	0.1667	0%
	0.4	0.8346	0.2827	0.0485	0.0268	0.0494	0.0270	1.0354	1.0481	0.1303	0.2137	0.1452	0.2323	0%
	0.2	0.9015	0.1527	0.0473	0.0245	0.0468	0.0241	1.0297	1.0517	0.1249	0.3005	0.1271	0.3027	0%
$\rho = 0.3$	1.0	1.0091	1.0089	0.0749	0.0666	0.0801	0.0676	1.0091	1.0089	0.0749	0.0666	0.0801	0.0676	0%
	0.8	0.8763	0.6166	0.0584	0.0415	0.0612	0.0429	1.0094	1.0124	0.0953	0.1136	0.1010	0.1191	0%
	0.6	0.8655	0.4170	0.0523	0.0323	0.0521	0.0319	1.0170	1.0212	0.1096	0.1597	0.1100	0.1598	0%
	0.4	0.9011	0.2753	0.0495	0.0277	0.0511	0.0266	1.0310	1.0425	0.1203	0.2183	0.1248	0.2194	0%
	0.2	0.9704	0.1474	0.0484	0.0252	0.0479	0.0257	1.0401	1.0694	0.1399	0.3632	0.1722	0.4172	0%
$\rho = 0.5$	1.0	1.0091	1.0094	0.0744	0.0739	0.0739	0.0732	1.0091	1.0094	0.0744	0.0739	0.0739	0.0732	0%
	0.8	0.9693	0.5920	0.0600	0.0458	0.0599	0.0453	1.0072	1.0063	0.0897	0.1272	0.0916	0.1282	0%
	0.6	1.0096	0.3857	0.0550	0.0355	0.0531	0.0351	1.0143	1.0026	0.0974	0.1769	0.0983	0.1833	0%
	0.4	1.0630	0.2452	0.0526	0.0302	0.0515	0.0302	1.0185	1.0144	0.1008	0.2439	0.1102	0.2653	0%
	0.2	1.1252	0.1276	0.0517	0.0273	0.0513	0.0284	1.0302	1.0575	0.1003	0.3915	0.1099	0.4264	0.1%
$\rho = 0.6$	1.0	1.0024	1.0069	0.0741	0.0808	0.0745	0.0819	1.0024	1.0069	0.0741	0.0808	0.0745	0.0819	0%
	0.8	1.0487	0.5602	0.0621	0.0496	0.0628	0.0510	1.0140	0.9762	0.0850	0.1362	0.0851	0.1422	0%
	0.6	1.1135	0.3511	0.0576	0.0381	0.0589	0.0380	1.0109	0.9488	0.0860	0.1822	0.0897	0.1906	0%
	0.4	1.1809	0.2192	0.0557	0.0324	0.0552	0.0325	1.0255	0.9730	0.0873	0.2557	0.0901	0.2669	0%
	0.2	1.2288	0.1100	0.0545	0.0288	0.0539	0.0279	1.0342	1.0018	0.0852	0.4050	0.0792	0.4004	0%

References

Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press, Cambridge, Mass.

Amemiya, Y. (1985), "Instrumental Variable Estimator for the Nonlinear Errors-in Variables Model," *Journal of Econometrics*, **28**, 273-89.

Anderson, T. W. (1984), *An Introduction to Multivariate Statistical Analysis, Second Edition*, Wiley, New York.

Bound, J., Brown, C., and Mathiowetz, N. (2001), "Measurement Error in Survey Data," Chapter 59 in *Handbook of Econometrics, Volume 5*, Heckman, J. J., and Leamer, E. (eds), Elsevier, Amsterdam.

Carroll, R. J., and Stefanski, L. A. (1990), "Approximate Quasi-Likelihood Estimation in Models with Surrogate Predictors," *Journal of the American Statistical Association*, **85**, 652-63.

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995), *Measurement Error in Nonlinear Models*, Chapman & Hall, London.

Edgerton, D. L. (1997), "The Effects of Cutbacks in Swedish Sickness Benefits," *Working Paper 73/97*, Dept. of Economics, University of Lund. Presented at Econometric Society's European Meeting in Toulouse.

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

Greene, W. H. (2000), *Econometric Analysis*. *Fourth Edition*, Prentice Hall, Upper Saddle River, New Jersey.

Greene, W. H. (2003), *Econometric Analysis*. *Fifth Edition*, Prentice Hall, Upper Saddle River, New Jersey.

Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective*, Springer, Berlin.

Hsiao, C., and Wang, Q. K. (2000), "Estimation of Structural Nonlinear Errors-in-Variables Models by Simulated Least-Squares Method," *International Economic Review*, **41**, 523-42.

Iwata, S. (2001) "Reentered and Rescaled Instrumental Variable Estimation of Tobit and Probit Models with Errors in Variables," *Econometric Reviews*, **20**, 319-35.

Kao, C., and Schnell, J. F. (1987a), "Errors in Variables in Panel Data with a Binary Dependent Variable," *Economics Letters*, **24**, 45-49.

Kao, C., and Schnell, J. F. (1987b), "Errors in Variables in a Random-Effects Probit Model for Panel Data," *Economics Letters*, **24**, 339-342.

Klepper, S., and Leamer, E. E. (1984), "Consistent Sets of estimates for Regressions with Errors in all Variables," *Econometrica*, **52**, 163-183.

Li, T (2002), "Robust and Consistent Estimation of Nonlinear Errors-in-Variables Models,". *Journal of Econometrics*, **110**, 1-26.

Lütkepohl, H. (1996), *Handbook of Matrices*, Wiley, New York.

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

Murphy, K. V., and Topel, R. H. (1985), "Estimation and Inference in Two-Step Econometric Models," *Journal of Business and Economic Statistics*, **3**, 370-379.

Newey, W. K. (2001), "Flexible Simulated Moment Estimation of Nonlinear Errors-in-Variables Models," *Review of Economics and Statistics*, **83**, 616-27.

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.

Stefanski, L. A., and Carroll, R. J. (1990), "Structural Logistic Regression Measurement Error Models," *Contemporary Mathematics*, **112**, 115-127.

Wansbeek, T., and Meijer, E. (2000), *Measurement Error and Latent variables in Econometrics*, Elsevier, Amsterdam