AISM (2006) 58: 153–169
DOI 10.1007/s10463-005-0002-z

Surupa Roy · Tathagata Banerjee

A flexible model for generalized linear regression with measurement error

Received: 1 September 2003 / Revised: 4 November 2004 / Published online: 2 February 2006

Abstract This paper focuses on the question of specification of measurement error distribution and the distribution of true predictors in generalized linear models when the predictors are subject to measurement errors. The standard measurement error model typically assumes that the measurement error distribution and the distribution of covariates unobservable in the main study are normal. To make the model flexible enough we, instead, assume that the measurement error distribution is multivariate t and the distribution of true covariates is a finite mixture of normal densities. Likelihood–based method is developed to estimate the regression parameters. However, direct maximization of the marginal likelihood is numerically difficult. Thus as an alternative to it we apply the EM algorithm. This makes the computation of likelihood estimates feasible. The performance of the proposed model is investigated by simulation study.

Key words Generalized linear model · Structural model · Surrogate · Validation data · Canonical link · Logistic regression · Mixture distribution · EM algorithm

S. Roy
Department of Statistics, St. Xavier's College,
30, Park Street, Kolkata 700 016, India

T. Banerjee (✉)
Department of Statistics, University of Calcutta,
35, Ballygunge Circular Road, Kolkata700 019, India
E-mail: btathaga@yahoo.com

Present & Corresponding Address
Production & Quantitative Methods Area, Indian Institute of Management,
Vastrapur, Ahmedabad 380015, India

1 Introduction

The conventional approach to regression modeling is to assume that the covariates are measured without error. The measurement error problems arise if instead of observing the covariate without error, at least one is measured as an error-prone surrogate. Such problems are primarily concerned with inference on regression parameters for an outcome y on covariates x where measurements on x are available only through the recording of an imperfect surrogate z. It is well known that regressing y on z and thus ignoring errors can be seriously misleading and several methods have been proposed for countering this (Stefanski 1985; Stefanski and Carroll 1985; Rosner et al. 1989). A good review of such methods can be found in Fuller (1980) for linear models and in Carroll et al. (1995) for nonlinear models.

The measurement error model based on structural specifications entail the formulation of three sub-models: an outcome model relating y to x, a measurement error model relating z to x and a probability model for the unknown covariates x. For specification of a model the functional forms of the distributions involved in the sub-models have to be chosen. We consider generalized linear model with canonical link as the outcome model. Choice of measurement error model typically uses both a priori knowledge of the measuring instrument as well as the external information from other similar studies. Thus the functional form is often well motivated and the default choices involve normal and lognormal error models corresponding to additive and multiplicative error situations. To make the measurement error model more flexible we consider a scale mixture of normal densities. This is possibly to guard against measurement error model with heavier tails and more specifically from potential outliers. In regression models, choice of t-errors with known and unknown degrees of freedom have been considered by Lange et al. (1989), Lange and Sinsheimer (1993), Liu and Rubin (1994, 1995) among others. In fact, they pointed out that normal theory based inference is significantly inefficient when kurtosis parameter is of moderately high value.

In structural set up, modeling the distribution of true predictor x deserves special attention. The reason is that there is but little information about it besides some observations on a gold standard in a small validation group or some repeated measures on the surrogate. We would naturally like to have a model for the distribution of x that would make the inference on regression parameters not very sensitive to a particular choice of the distribution of x. This has led to different lines of development. Carroll et al. (1993) proposed a pseudo-likelihood based method using validation data. Roeder et al. (1996) used a nonparametric model for the distribution of x estimated via NPML algorithms. Muller and Roeder (1997) have developed Dirichlet process priors for the joint model of x and z. Use of mixture of normal densities to enhance robustness against model misspecification of x is in recent lines of developments. Mixture with a fixed number of components have been used by Carroll et al. (1999a), Carroll et al. (1999b) in the context of linear measurement error model and also in the nonparametric context of regression splines with measurement error. In the Bayesian framework for logistic regression, Richardson et al. (2000) considered the modeling of the distribution of x by mixture of normal distributions with variable number of components. Mixture models thus provide an alternative natural framework in which to consider a flexible modeling

of the distribution of true predictors x. Here we consider mixture of normal distributions with a fixed number of components to increase robustness to model misspecification for the distribution of x.

In this paper the focus is on flexible modeling of both the measurement error distribution and the distribution of true predictors x in the generalized linear regression set up. For modeling the measurement error distribution, we use scale mixture of normal distribution and finite mixture of normal distributions is used for the latter.

The models and likelihood are introduced in Sect. 2. Section 3 discusses the implementation of the EM algorithm. A simulation study is undertaken in Sect. 4 and finally in Sect. 5 conclusions are drawn.

2 Models and likelihood

In present formulation y denote the response variable, x is the true predictor unobservable in the main study and z is the associated imperfect surrogate through which only measurements on x are available. There may be other covariates that are measured without error. To avoid notational complicacy they are not included in the present discussion. However, their inclusion does not complicate the analysis in any major way. In fact, the analysis is carried out conditional on the values of the known covariates. Further we assume here a non-differential measurement error model, i.e., z has no information regarding y other than that contained in x. Technically, it means that the conditional probability distribution of y given x, z and that of y given x are the same, i.e., $f(y\,|x, z) = f(y\,|x)$.

2.1 Outcome model

Specifically, a generalized linear model with canonical link function as an outcome model is considered i.e., the response y_i conditioned on the value of the true covariate x_i has the probability distribution

$$f(y_i\,\big|x_i; \beta) = \exp\{y_i(x_i^T\beta) - b(x_i^T\beta) + c(y_i)\ , \quad i = 1(1)n, \tag{1}$$

for some known functions $b(\cdot)$ and $c(\cdot)$. β $(p \times 1)$ denote the vector of regression parameters, which are to be estimated, and y_is are independent. No dispersion parameter has been included in the above model though its presence does not complicate the developments.

2.2 Measurement error model

The measurement error model is considered as

$$z_i = x_i + \Psi(v_i)\varepsilon_i. \tag{2}$$

Here $\varepsilon_{\rm i} \sim N_p\ (0,\ \Sigma)$; where $\Sigma(p \times p)$ is an unknown variance matrix; v_i is a positive valued random variable independent of ε_i and x_i and $\Psi(\cdot)$ is a positive valued function. Model (2) implies that conditionally on v_i and x_i,

$$z_i \sim N_p\big(x_i, \{\Psi(v_i)\}^2\Sigma\big), \quad i = 1, 2, \ldots, n, \tag{3}$$

and are independent. With the strength of Eqs. (2) and (3), it is evident that in terms of the latent variable v_i the measurement error distribution can be written as

$$f(z_i\,|x_i;\,\Sigma) = \int_0^\infty f(z_i\,|v_i, x_i;\,\Sigma) f(v_i) dv_i\,.$$

In particular if v_i is a chi-square variable with ν degrees of freedom (d.f.) and $\Psi(v_i) = (v_\mathrm{i}/\nu)^{-1/2}$, then the measurement process represented by the conditional distribution of z_i given x_i is multivariate t given by the density function

$$f(z_i\,|x_i;\,\Sigma) = \frac{|\Sigma|^{-\frac{1}{2}}\sqrt{(\nu+p)/2}}{(\pi\nu)^{\frac{p}{2}}\sqrt{\nu/2}}\left\{1+\nu^{-1}(z_i-x_i)^T\Sigma^{-1}(z_i-x_i)\right\}^{-\frac{\nu+p}{2}}. \tag{4}$$

This t model described by Eq. (4) accommodates the usual multivariate normal distribution as $\upsilon \to \infty$ and the distributions with tails heavier than multivariate normal in case υ is finite.

In measurement error analysis in order to avoid the identifiability problem it is necessary to assume that either Σ is known or an estimate of it is made available from external validation data (independent of the primary data) where both x and z are observed. In the present study, we assume the latter. Suppose $A = \sum_{\mathrm{i}=1}^{N}(\tilde{z}_\mathrm{i} - \tilde{x}_\mathrm{i})(\tilde{z}_\mathrm{i} - \tilde{x}_\mathrm{i})^T$ be the sum of squares and sum of products matrix based on external validation data $(\tilde{z}_\mathrm{i}, \tilde{x}_\mathrm{i}), i = 1, 2, \ldots, N$. Following Sutradhar and Ali (1989) it can be shown that 'A' follows a generalized Wishart distribution given by the pdf

$$f(A;\,\Sigma) = C(N,\,p)|\Sigma|^{-\frac{N}{2}}|A|^{\frac{N-p-1}{2}}\left\{\nu + \mathrm{Tr}\Sigma^{-1}A\right\}^{-\frac{(\nu+Np)}{2}}, \tag{5}$$

where, $C(N,\,p) = \frac{\nu^{\frac{\nu}{2}}\sqrt{(\nu+Np)/2}}{(\pi)^{p(p-1)/4}\sqrt{\nu/2}\prod_{i=1}^{p}\sqrt{(N-i+1)/2}}$
and N denotes the number of observations in the validation data set. For this distribution it can also be shown that for $\upsilon > 2$, $E(A) = N\nu\Sigma(\nu-2)^{-1}$. Hence an unbiased estimator of Σ is given by $A(\nu-2)(N\nu)^{-1}$.

2.3 Model for the true predictor

Broadly, there are two approaches to measurement error analysis depending on how x_is are treated. Functional model attempts to estimate the unknown x_i's along with other parameters treating these as unknown constants but the structural model specifies a probability distribution for x as a means to reduce the number of nuisance parameters. In the present discussion we restrict to structural models. The distribution of x_i is taken to be a g-component finite mixture with 'g' known a priori. The density of x_i can thus be written in the form,

$$f(x_i;\,\psi) = \sum_{j=1}^{g}\pi_j f_j(x_i;\,\theta_j), \tag{6}$$

where, $f_j(x_i;\theta_j)$ are densities and π_j's are non-negative quantities that sum to one and $\psi = \left(\pi_1, \pi_2, \ldots, \pi_{g-1}, \theta_1, \ldots, \theta_g\right)^{\mathrm{T}}$. Here θ_js are known a priori to be distinct. The quantities π_j's are called mixing proportions and are treated as unknown constants. Specifically, $f_j(x_i;\theta_j)$ is considered to be a p-variate normal with unknown mean vector μ_j and dispersion matrix $\Sigma_j (p \times p)$.

An obvious way of generating a random vector x_i from the density Eq. (6) is as follows. Let u_i be a g-dimensional component label vector, where the j-th element of u_i say u_{ij} is 1 or 0, depending on whether x_i is generated from $f_j(x_i;\theta_j)$ or not $(j = 1, 2, \ldots, g)$. Thus the distribution of u_i is multinomial with cell probabilities $\pi_1, \pi_2, \ldots, \pi_g$. The probability function of u_i is given by

$$f(u_i;\pi) = \pi_1^{u_{i1}} \pi_2^{u_{i2}} \cdots \pi_g^{u_{ig}}. \tag{7}$$

Here $u_i \sim \mathrm{mult}_{\mathrm{g}}(1, \pi)$, where $\pi = \left(\pi_1, \pi_2, \ldots, \pi_g\right)^{\mathrm{T}}$. Clearly the conditional density of x_i given $u_{ij} = 1$ is $f_j(x_i;\theta_j)$ and the unconditional density (i.e. the marginal density of x_i) is given by Eq. (6) which in terms of latent variable u_{ij} can be written as

$$f(x_i;\psi) = \sum_{j=1}^{g} \pi_j f\left(x_i \,|u_{ij} = 1;\theta_j\right). \tag{8}$$

2.4 Likelihood function

The log likelihood function of the unknown parameters $\xi = (\beta,\ \Sigma,\ \psi)$ arising in the models as presented in the earlier subsections can be written as

$$L(\xi\,|y, z, A) = \sum_{i=1}^{n} \log \left\{ \sum_{j=1}^{g} \pi_j \int_{-\infty}^{\infty} f(y_i\,|x_i;\beta) \left(\int_0^{\infty} f(z_i\,|x_i, v_i;\Sigma) f(v_i) \mathrm{d}v_i \right) \right.$$
$$\left. \times f(x_i\,|u_{ij} = 1;\theta_j) \mathrm{d}x_i \right\} + \log f(A;\Sigma).$$

It is evident that the direct maximization of $L(\xi\,|y, z, A)$ is quite difficult numerically, due to the sum of terms inside the logarithm and also multidimensional integral inside the summation sign. The EM algorithm (Dempster et al. 1977) offers an alternative simpler framework for computation of maximum likelihood estimates by treating the unobserved x_i's, u_{ij}'s and v_i's as missing data.

3 Implementation of the EM algorithm

It is to mention here that the complete data is denoted by $d_{\mathrm{c}} = (y, x, z, u, v, A)$ while the observed data by $d_{\mathrm{O}} = (y, z, A)$. The complete data log-likelihood is given by

$$
\begin{aligned}
L_{\mathrm{c}}(\xi) &= \sum_{i=1}^{n} \log f(y_i \,|x_i; \beta) \\
&\quad + \left\{ \sum_{i=1}^{n} \log f(z_i | x_i, v_i; \Sigma) \right. \\
&\quad \left. + \sum_{i=1}^{n} \log f(v_i; \nu) + \log f(A; \Sigma) \right\} \\
&\quad + \sum_{i=1}^{n} \sum_{j=1}^{g} u_{ij} \big\{ \log \pi_j + \log f_j(x_i; \theta_j) \big\} \\
&= L_{\mathrm{c}}(\beta) + L_{\mathrm{c}}(\Sigma) + L_{\mathrm{c}}(\psi). \qquad (9)
\end{aligned}
$$

The EM algorithm then proceeds iteratively in two steps. The basic algorithm is summarized below.

1. Start with initial guesses for the parameters ξ say $\xi^{(0)}$.
2. E step involves evaluation of the expectations of complete data log likelihood given the observed data and the parameters as obtained from the previous iteration. At the $(t+1)$-th step, compute

$$
\begin{aligned}
Q(\xi; \xi^{(t)}) &= E[L_c(\xi) \,\big|d_{\mathrm{O}}; \xi^{(t)}] = E\big[L_c(\beta) \,\big|d_{\mathrm{O}}; \xi^{(t)}\big] \\
&\quad + E\big[L_c(\Sigma) \,\big|d_{\mathrm{O}}; \xi^{(t)}\big] + E\big[L_c(\psi) \,\big|d_{\mathrm{O}}; \xi^{(t)}\big] \\
&= Q_1(\beta; \xi^{(t)}) + Q_2(\Sigma; \xi^{(t)}) + Q_3(\psi; \xi^{(t)}). \qquad (10)
\end{aligned}
$$

3. M step involves the maximization of the above conditional expectation $Q(\xi; \xi^{(t)})$ with respect to ξ over the parameter space to obtain the updated estimate $\xi^{(t+1)}$.
4. Iterate steps 2 and 3 until convergence.

It is to be noted that the augmentation of the observed data by the missing data to find the maximum likelihood estimate has two advantages. First, on knowing x_i's, y_i's are independent and on knowing v_i's and x_i's, z_i's are independent. Secondly at the M step of the EM algorithm β enters only in the first term. Hence, to find β that maximizes $Q(\xi; \xi^{(t)})$ it is enough to consider $Q_1(\beta; \xi^{(t)})$. This maximization is similar to that of a standard generalized linear model. Also, it is feasible to find Σ and ψ that maximizes $Q_2(\Sigma; \xi^{(t)})$ and $Q_3(\psi; \xi^{(t)})$, respectively.

Let us first consider $Q_3(\psi; \xi^{(t)})$. i.e.

$$
\begin{aligned}
\mathrm{Q}_3(\psi; \xi^{(t)}) &= \sum_{i=1}^{n} \sum_{j=1}^{g} E\big\{ u_{ij} \big(\log \pi_j + \log f_j(x_i; \theta_j) \big) \;\big|\; d_{\mathrm{O}};\ \xi^{(t)} \big\} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{g} \big\{ \log \pi_j + E\big(\log f_j(x_i; \theta_j) \;\big|\; d_{\mathrm{O}} ;\ \xi^{(t)} \big) \big\} \tau_j\big(y_i, z_i; \xi^{(t)}\big) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{g} \left\{ \log \pi_j - \frac{1}{2} \log \big|\Sigma_j\big| \right. \\
&\quad \left. - \frac{1}{2} \mathrm{Tr} \left\{ \Big(B_{ij}^{(t)} + (a_{ij}^{(t)} - \mu_j)(a_{ij}^{(t)} - \mu_j)^{\mathrm{T}} \Big) \Sigma_j^{-1} \right\} \right\} \tau_j\big(y_i, z_i; \xi^{(t)}\big)
\end{aligned}
$$

where,

$$\tau_j\big(y_i, z_i; \xi^{(t)}\big) = E\big(u_{ij} \,\big|\, d_{\mathrm{O}}; \xi^{(t)}\big); \;\; a_{ij}^{(t)} = E\big(x_i \,\big|d_{\mathrm{O}}, u_{ij} = 1; \xi^{(t)}\big)$$
$$B_{ij}^{(t)} = E\Big((x_i - a_{ij}^{(t)})(x_i - a_{ij}^{(t)})^{\mathrm{T}} \,\big|d_{\mathrm{O}}, u_{ij} = 1; \xi^{(t)}\Big) \tag{11}$$

Details of Eqs. (11) in are provided in the Appendix. Now $Q_3(\psi; \xi^{(t)})$ is maximized with respect to the respective parameters to obtain the updated estimates of the parameters as below,

$$\pi_j^{(t+1)} = \frac{\sum_{i=1}^{n} \tau_j(y_i, z_i; \xi^{(t)})}{n}, \tag{12}$$

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^{n} \tau_j(y_i, z_i; \xi^{(t)}) a_{ij}^{(t)}}{\sum_{i=1}^{n} \tau_j(y_i, z_i; \xi^{(t)})}, \tag{13}$$

$$\Sigma_j^{(t+1)} = \frac{\sum_{i=1}^{n} \tau_j(y_i, z_i; \xi^{(t)}) \Big\{ B_{ij}^{(t)} + (a_{ij}^{(t)} - \mu_j^{(t+1)})(a_{ij}^{(t)} - \mu_j^{(t+1)})^{\mathrm{T}} \Big\}}{\sum_{i=1}^{n} \tau_j(y_i, z_i; \xi^{(t)})}. \tag{14}$$

Next let us consider $Q_2(\Sigma; \xi^{(t)})$, i.e.,

$$Q_2(\Sigma; \xi^{(t)}) = \sum_{i=1}^{n} E\big(\log f(z_i \,|\, x_i, \upsilon_i; \Sigma) \;\; \big| d_{\mathrm{O}}; \xi^{(t)}\big)$$
$$+ \sum_{i=1}^{n} E\big(\log f(\upsilon_i) \;\; \big| d_{\mathrm{O}}; \xi^{(t)}\big) + E\big(\log f(A; \Sigma) \;\; \big| \, d_{\mathrm{O}}; \xi^{(t)}\big). \tag{15}$$

In the present case of known υ it is sufficient to focus only on the first and last terms since the other term does not involve unknown parameter. Again the last term will be simply logarithm of $f(A;\Sigma)$. The first term may be written as (ignoring the constant term)

$$-\frac{n}{2}\log|\Sigma| + \frac{p}{2}\sum_{i=1}^{n} E\big(\log(\upsilon_i \nu^{-1}) \,\big| d_{\mathrm{O}}; \xi^{(t)}\big) - \frac{1}{2}\sum_{i=1}^{n} E\big(\upsilon_i \nu^{-1} \delta(z_i; x_i, \Sigma) \,\big| d_{\mathrm{O}}; \xi^{(t)}\big)$$

where, $\delta(z_i; x_i, \Sigma) = (z_i - x_i)^{\mathrm{T}} \Sigma^{-1} (z_i - x_i)$ denotes the Mahalanobis squared distance between z_i and x_i. Ignoring terms independent of Σ and using the result that the conditional distribution of $\upsilon_i \nu^{-1} \,|\, z_i \,; \xi^{(t)}$ is gamma with parameters

$$m_1 = \frac{\nu + p}{2}, \; m_2 = \frac{\nu + \delta(z_i; x_i, \Sigma^{(t)})}{2}$$

the expression of $Q_2(\Sigma; \xi^{(t)})$ in (Eq. 15) becomes,

$$\frac{n+N}{2}\log\left|\Sigma^{-1}\right| - \frac{\nu+p}{2}\sum_{i=1}^{n} E\left\{\delta(z_i; x_i, \Sigma)(\nu+\delta(z_i; x_i, \Sigma^{(t)})^{-1}) \;\middle|\; d_O; \xi^{(t)}\right\}$$
$$-\frac{\nu+Np}{2}\log\left(\nu+\mathrm{Tr}\left(\Sigma^{-1}A\right)\right)$$

Now maximizing $Q_2(\Sigma; \xi^{(t)})$ with respect to Σ and realizing the fact that $\frac{\partial}{\partial\theta}E\,(g(x, \theta)) = E\left(\frac{\partial}{\partial\theta}g(x, \theta)\right)$, the above yields

$$\frac{N+n}{2}(2\Sigma - \mathrm{diag}\Sigma)$$
$$-\frac{\nu+p}{2}\sum_{i=1}^{n} E\left\{(z_i - x_i)(z_i - x_i)^{\mathrm{T}}\left(\nu+\delta(z_i; x_i, \Sigma^{(t)})\right)^{-1} \middle| d_O; \xi^{(t)}\right\}$$
$$-\frac{\nu+Np}{2}\frac{(2A - \mathrm{diag}A)}{\left(\nu+\mathrm{Tr}(\Sigma^{-1}A)\right)} = 0. \tag{16}$$

The expectation in Eq. (16) above is obtained using numerical quadrature and then the equation is solved iteratively to give the updated estimate of $\Sigma^{(t+1,s+1)}$ where s is the index of iteration used to obtain Σwhich is nested within the EM iteration indexed by t.

Lastly, let us consider $Q_1(\beta; \xi^{(t)})$ which involves the parameter of interest β. Now for simplicity ignoring the constant term, $Q_1(\beta; \xi^{(t)})$ becomes

$$Q_1(\beta; \xi^{(t)}) = \sum_{i=1}^{n} y_i E\left(x_i^{\mathrm{T}} \middle| d_O; \xi^{(t)}\right)\beta - \sum_{i=1}^{n} E\left(b(x_i^{\mathrm{T}}\beta) \middle| d_O; \xi^{(t)}\right).$$

An application of Newton–Raphson method may be used iteratively to solve for β which maximises $Q_1(\beta; \xi^{(t)})$. For a given estimate $\beta^{(s)}$ an updated estimate is the solution to

$$\sum_{i=1}^{n} y_i E\left(x_i^{\mathrm{T}} \middle| d_O; \xi^{(t)}\right) - \sum_{i=1}^{n} E\left[b'(x_i^{\mathrm{T}}\beta^{(s)})x_i \middle| d_O; \xi^{(t)}\right]$$
$$+ (\beta - \beta^{(s)})\sum_{i=1}^{n} E\left[b''(x_i^{\mathrm{T}}\beta^{(s)})x_i x_i^{\mathrm{T}} \middle| d_O; \xi^{(t)}\right] = 0.$$

At this stage we approximate $E\left[b'(x_i^{\mathrm{T}}\beta^{(s)})x_i \middle| d_O; \xi^{(t)}\right]$ by $b'(a_i^{T(t)}\beta^{(s)})a_i^{(t)}$ and $E\left[b''(x_i^{\mathrm{T}}\beta^{(s)})x_i x_i^{\mathrm{T}} \middle| d_O; \xi^{(t)}\right]$ by $b''(a_i^{\mathrm{T}(t)}\beta^{(s)})a_i^{(t)}a_i^{\mathrm{T}(t)}$. The approximations are justified provided $b'(\cdot)$ and $b''(\cdot)$ are approximately linear over the support of the posterior distribution of x. This is, indeed the case for logistic regression with binary response. If $\theta_i^{(s,t)}$ is the estimate of the canonical parameter after t cycles so that $\theta_i^{(s,t)} = a_i^{(t)T}\beta^{(s,t)}$ then the weight function is $m_i^{(s,t)} = b''(\theta_i^{(s,t)})$ and the working dependent variable is

$$\tilde{y}_i^{(s,t)} = \frac{y_i - b'(\theta_i^{(s,t)})}{b''(\theta_i^{(s,t)})} + \theta_i^{(s,t)}.$$

The updated estimate of β at the $(s+1)^{\text{th}}$ iteration will then be obtained by using reweighted least squares iteratively. i.e.,

$$\beta^{(s+1,t+1)} = \left[\sum_{i=1}^{n} m_i^{(s,t)} \left\{ B_i^{(t)} + a_i^{(t)} a_i^{(t)T} \right\}\right]^{-1} \left[\sum_{i=1}^{n} m_i^{(s,t)} \tilde{y}_i^{(s,t)} a_i^{(t)T}\right]. \quad (17)$$

where s is the index of the weighted least squares iteration which is nested within the EM iteration indexed by t, and

$$a_i^{(t)} = \sum_{j=1}^{g} a_{ij}^{(t)} \pi_j^{(t)}, \quad B_i^{(t)} = \sum_{j=1}^{g} B_{ij}^{(t)} \pi_j^{(t)} + \sum_{j=1}^{g} \pi_j^{(t)} (a_{ij}^{(t)} - a_i^{(t)})(a_{ij}^{(t)} - a_i^{(t)})^{\mathrm{T}}. \quad (18)$$

At each iteration in the EM algorithm $a_{ij}^{(t)}$ and $B_{ij}^{(t)}$ are computed. The updated version of the parameter estimates $\beta^{(t+1)}$, $\Sigma^{(t+1)}$, $\mu_j^{(t+1)}$, $\Sigma_j^{(t+1)}$, $\pi_j^{(t+1)} (j = 1, 2, \ldots, g)$ are then obtained.

4 Simulation study

4.1 The regression estimator under different estimating conditions

In the first part of the simulation study we compare the performance of the estimator of the regression parameter β when the distribution of true predictors is mixture of normal and the error distribution is t with ν d.f. under five different situations as mentioned below.

1. $\mathbf{E_1}$: Naïve case. i.e. ignoring measurement error. In this case it is assumed that z_i's are the values of the true predictors. The maximum likelihood estimate of β is then obtained by applying iterative re-weighted least squares McCullagh and Nelder (1989). In $(s+1)$-th cycle of the iteration, the updated estimate of β is given by

$$\beta^{(s+1)} = \left\{\sum_{i=1}^{n} m_i^{(s)} (z_i z_i^T)\right\}^{-1} \times \left\{\sum_{i=1}^{n} m_i^{(s)} \tilde{y}_i^{(s)} z_i\right\}. \quad (19)$$

If $\theta_{\mathrm{i}}^{(\mathrm{s})}$ is the estimate of the canonical parameter after s cycles, so that $\theta_i^{(s)} = z_i^T \beta^{(s)}$, then the weight function is $m_i^{(s)} = b''(\theta_i^{(s)})$ and the working dependent variable is

$$\tilde{y}_i^{(s)} = \frac{\left(y_i - b'(\theta_i^{(s)})\right)}{b''(\theta_i^{(s)})} + \theta_i^{(s)},$$

where $b'(\cdot)$ and $b''(\cdot)$ are the first and second order derivatives of $b(\cdot)$, respectively.

2. $\mathbf{E_2}$: The maximum likelihood estimator obtained by the EM algorithm, assuming both the error distribution and the distribution of true predictors are normal, i.e. $z \,|\, x \tilde{} N_p(x, \Sigma)$ and $x \tilde{} N_p(\mu, \Omega)$. Following Schafer (1987), $Q_2(\Sigma; \xi^{(t)})$ is obtained as

$$Q_2(\Sigma; \xi^{(t)}) = -\frac{1}{2}\sum_{i=1}^{n} \mathrm{Tr}\left[\left\{B_i^{(t)} + (z_i - a_i^{(t)})(z_i - a_i^{(t)})^T + An^{-1}\right\}\Sigma^{-1}\right] + \frac{n+N}{2}\log\left|\Sigma^{-1}\right| \tag{20}$$

where $a_i^{(t)}$ and $B_i^{(t)}$ are as given in Eq 18. The M-step then yields

$$\Sigma^{(t+1)} = \frac{1}{n+N}\left[\sum_{i=1}^{n}\left\{B_i^{(t)} + (z_i - a_i^{(t)})(z_i - a_i^{(t)})^T\right\} + A\right]. \tag{21}$$

Now ignoring the constant term $Q_3(\psi; \xi^{(t)})$ is given by

$$Q_3(\psi\,; \xi^{(t)}) = -\frac{1}{2}\sum_{i=1}^{n} \mathrm{Tr}\left[\left\{B_i^{(t)} + (a_i^{(t)} - \mu)(a_i^{(t)} - \mu)^T\right\}\Omega^{-1}\right] - \frac{n}{2}\log|\Omega|. \tag{22}$$

Finally, $Q_3(\psi; \xi^{(t)})$ is maximized with respect to the parameters to give the updated estimates as

$$\begin{aligned} \overline{\mu}^{(t)} &= \mu^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} a_i^{(t)}. \\ \Omega^{(t+1)} &= \frac{1}{n}\sum_{i=1}^{n}\left\{B_i^{(t)} + (a_i^{(t)} - \overline{\mu}^{(t)})(a_i^{(t)} - \overline{\mu}^{(t)})^T\right\}. \end{aligned} \tag{23}$$

The estimator of the regression parameter will be as given by Eq. (17).

3. $\mathbf{E_3}$ The maximum likelihood estimator obtained by the EM algorithm, assuming that the error distribution is normal and the distribution of true predictors is mixture of normal. In this case the estimator of Σ will be given by Eq. (21). The rest of the parameters will be obtained as in Sect. 3

4. $\mathbf{E_4}$ The maximum likelihood estimator obtained by the EM algorithm, assuming that the error distribution is t with ν d.f. and the distribution of true predictors is normal. Here the estimators of μ and Ω will be given by Eq. (23). The rest of the parameters will be obtained as in Sect. 3.

5. $\mathbf{E_5}$ The maximum likelihood estimator obtained by the EM algorithm, assuming that the distribution of true predictors is mixture of normal and the error distribution is t with ν d.f. This is the proposed estimator as described in Sect. 3.

The details of the simulation study are described below:

Step 1: Generate $x_1, x_2, \ldots, x_n$ from a mixture of two independent univariate normal having parameter $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \pi)$. θ is taken as (0.45, 1.90, 0.03, 0.03, 0.60). This choice is motivated by Munich data (Carroll et al. 1995).

Step 2: Given $x_1, x_2, \ldots, x_n$ generate $y_1, y_2, \ldots, y_n$ from Bernoulli distribution with the probability of success given by

$$\text{Prob}\{y_i = 1 | x_i\} = \frac{\exp(\beta x_i)}{1 + \exp(\beta x_i)}, \quad \beta = 1.6.$$

The choice of β is taken from Munich data.

Step 3: Given $x_1, x_2, \ldots, x_n$ generate $z_1, z_2, \ldots, z_n$ from univariate t distribution with υ d.f. and origin at x_i and scale parameter Σ. Here υ=4 and Σ =0.4.

Step 4: External validation data $(\tilde{z}_i, \tilde{x}_i)$, $i = 1, 2, \ldots, N$ are generated as in Steps 1 and 3. Then the values of A and N (=10) are incorporated into the likelihood.

Step 5: Given the data y_i, z_i $(i = 1, 2, \ldots, n)$ and 'A' the five different likelihoods are fitted.

Step 6: Repeat Steps 1 to 5 a large number of times, say R times. Suppose the estimate of β at the l-th simulation be $\hat{\beta}_{(l)}$. Compute,

$$\text{bias}(\hat{\beta}_{(l)}) = \frac{1}{R}\sum_{l=1}^{R}(\hat{\beta}_{(l)} - \beta) \quad \text{and mse}(\hat{\beta}_{(l)}) = \frac{1}{R}\sum_{l=1}^{R}(\hat{\beta}_{(l)} - \beta)^2.$$

Efficiencies (reciprocal of m.s.e. expressed in percentage of the reciprocal of mse of the Naïve estimator) of each of the cases relative to the naïve case are also obtained. Efficiency of the naïve estimator is taken to be 100. The results of the simulation study are summarized in Table 1.

Here R=100, the simulation is repeated for samples of size n=100, n=300. Table 1 shows the individual effects of misspecification of the measurement error distribution and the true predictor distribution as well as the joint effect. In this set-up the estimator E_5 is found to be better than the others. Comparison of its efficiency with E_2, E_3 and E_4 clearly shows that the joint misspecification of measurement error distribution and the distribution of the true predictor is nearly double the individual effect. The individual effects are substantial also.

Table 1 x ~mixnormal $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \pi)\mu_1 = 0.45, \mu_2 = 1.90, \sigma_1^2 = \sigma_2^2 = 0.03, \pi = 0.6$

n	E_1	E_2	E_3	E_4	E_5
100					
m.s.e	1.99657×10^{-1}	1.72956×10^{-1}	1.10811×10^{-1}	1.00912×10^{-1}	7.94615×10^{-2}
Bias	-2.60078×10^{-1}	2.25516×10^{-1}	2.01772×10^{-1}	2.01513×10^{-1}	1.71456×10^{-1}
Eff	100%	115%	180%	198%	251%
300					
m.s.e	1.48309×10^{-1}	1.01221×10^{-1}	4.67646×10^{-2}	4.91256×10^{-2}	3.20161×10^{-2}
Bias	-3.37272×10^{-1}	1.56156×10^{-1}	1.69125×10^{-1}	1.31251×10^{-1}	1.25123×10^{-1}
Eff	100%	146%	317%	302%	463%

4.2 Robustness study of the distribution of true predictor

The sensitivity of the performances of the estimators to the violation of the assumption of the distribution of true predictors is investigated further in the simulation study. We thus obtain the bias($\hat{\beta}$) and mse($\hat{\beta}$) for the above five cases (as described in Subsect. 4.1) when the distribution of x is:

(a) Normal with mean=1.03 and variance=0.524. The values 1.03 and 0.524 are actually the mean and variance of the mixture normal distribution.
(b) Rectangular (−0.22, 2.28) with mean=1.03 and variance=0.524.
(c) χ^2 with 1 d.f. with origin shifted and scale changed suitably to have mean 1.03 and variance 0.524.

Efficiencies of each of the cases relative to the naïve case as well as bias($\hat{\beta}$) and mse($\hat{\beta}$) are summarized in Tables 2–4. The results reveal that the proposed estimator E_5 performs uniformly better than others in all the situations considered except when the distribution of x is normal. In this particular case, it is expected that E_4 would perform better. However, the performance of E_5 is well even in this case. In particular, from Table 3 it is evident that when the distribution of the true predictor x is rectangular E_5 performs far better than others.

During the simulation study it is also observed that for small sample sizes (less than 100) the Naïve estimator E_1 is less variable but more biased than the others. But with the sample sizes becoming larger the bias increasingly dominates the variance. Hence large samples are candidates for using corrected estimators like E_2, E_3, E_4 and E_5. However, such results are not reported here.

Table 2 $x \sim \text{normal}\left(\mu, \sigma^2\right)$; $\mu = 1.03$, $\sigma^2 = 0.524$

n	E_1	E_2	E_3	E_4	E_5
100					
m.s.e	2.14981×10^{-1}	1.92965×10^{-1}	1.776702×10^{-1}	1.023719×10^{-1}	1.00912×10^{-1}
Bias	-3.70885×10^{-1}	-2.55512×10^{-1}	-3.49591×10^{-1}	-2.00061×10^{-1}	2.01513×10^{-1}
Eff	100%	111%	121%	210%	213%
300					
m.s.e	1.62577×10^{-1}	1.15126×10^{-1}	9.22727×10^{-2}	6.30143×10^{-2}	6.91256×10^{-2}
Bias	-3.37272×10^{-1}	-1.61615×10^{-1}	1.69125×10^{-1}	-1.29163×10^{-1}	1.31215×10^{-1}
Eff	100%	141%	176%	258%	235%

Table 3 $x \sim \text{Rec}(-0.22, 2.28)$

n	E_1	E_2	E_3	E_4	E_5
100					
m.s.e	1.99657×10^{-1}	1.31526×10^{-1}	9.82877×10^{-2}	9.01212×10^{-2}	6.68442×10^{-2}
Bias	-2.60078×10^{-1}	-2.15126×10^{-1}	-1.48472×10^{-1}	-1.56125×10^{-1}	-1.01262×10^{-1}
Eff	100%	152%	203%	221%	298%
300					
m.s.e	1.48309×10^{-1}	7.30586×10^{-2}	5.56336×10^{-2}	5.72189×10^{-2}	3.72819×10^{-2}
Bias	-3.37272×10^{-1}	-1.99129×10^{-1}	-1.72618×10^{-1}	-1.16125×10^{-1}	-1.00049×10^{-1}
Eff	100%	203%	267%	259%	398%

Table 4 $x \sim \chi^2(1)$

n	E_1	E_2	E_3	E_4	E_5
100					
m.s.e	3.81519×10^{-1}	3.14742×10^{-1}	2.01296×10^{-1}	2.107840×10^{-1}	1.55089×10^{-1}
Bias	-5.54706×10^{-1}	-3.80449×10^{-1}	-3.90126×10^{-1}	-2.96001×10^{-1}	-2.46001×10^{-1}
Eff	100%	121%	189%	181%	246%
300					
m.s.e	3.58892×10^{-1}	1.49538×10^{-1}	1.00121×10^{-1}	9.99821×10^{-2}	9.10893×10^{-2}
Bias	-5.82269×10^{-1}	-3.26423×10^{-1}	-3.04126×10^{-1}	-2.15126×10^{-1}	-1.99261×10^{-1}
Eff	100%	240%	358%	359%	394%

In the simulation study for the estimators obtained using EM algorithm the iteration was repeated 30 times and if the estimators failed to converge by 30 cycles the sample was rejected. This led to a rejection of 10% of the samples on an average. Here we have followed Schafer (1987); however, more iteration cycles are expected to reduce the rejection number.

4.3 Robustness of the choice $\upsilon = 4$

Through further simulation we investigate how robust is the choice $\upsilon = 4$. In a set up similar to Sect. 4.1 we generate the true covariate values (x) as in Step 1 and the response (y) as in Step 2. The values of the surrogate (z) given the true predictor (x) are generated from normal distribution with mean x and variance $\Sigma = 0.4$. The bias, mse and efficiencies as described in Step 6 of Sect. 4.1 are obtained for all the five cases. The results are summarized in Table 5 which show that the performance of E_3 is better as expected. It is found that with $\upsilon = 4$ and sample size n=100(300) the efficiencies of E_3 and E_5 are 259% (330%) and 220% (288%), respectively. A further study was made considering υ=6. It is to be noted that here only E_4 and E_5 will change. The details of that table are not given. It is observed that for n=100(300) the efficiencies of E_3 and E_5 are 259% (330%) and 256% (297%), respectively. Thus the study shows that the estimator E_5 assuming $\upsilon = 4$ although gives a much better performance than the naïve estimator E_1 but is around 40% less efficient than E_3. However for υ=6, the efficiencies are more or less similar. It is to be noted further from Table 1 that in the opposite case when the true measurement

Table 5 x $\sim$ mixnormal $\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \pi\right)$ $\mu_1 = 0.45, \mu_2 = 1.90, \sigma_1^2 = \sigma_2^2 = 0.03, \pi = 0.6$; $z\,|\tilde{x}N(x, \Sigma = 0.4)$

n	E_1	E_2	E_3	E_4	E_5
100					
m.s.e	1.66256×10^{-1}	7.38915×10^{-2}	6.41915×10^{-2}	8.52594×10^{-2}	7.55709×10^{-2}
Bias	-1.00162×10^{-1}	8.11562×10^{-2}	7.10264×10^{-2}	9.26182×10^{-2}	8.55162×10^{-2}
Eff	100%	225%	259%	196%	220%
300					
m.s.e	9.22568×10^{-2}	3.08551×10^{-2}	2.79566×10^{-2}	3.87636×10^{-2}	3.20336×10^{-2}
Bias	-2.12362×10^{-1}	6.30136×10^{-2}	5.11263×10^{-2}	7.11362×10^{-2}	6.33163×10^{-2}
Eff	100%	290%	330%	238%	288%

error is t with 4 d.f. the efficiency of E_3 is 70% (146%) less than that of E_5 for n=100(300). So the loss in efficiency is more in this case.

4.4 Estimation of υ from external validation data

Next we investigate the estimation of υ from an external validation data in order to choose an appropriate t measurement error model. Following Liu and Rubin (1995) we apply ECME algorithm to find the likelihood estimate of υ. Samples of sizes N = 20, 40, 60, 80 are chosen. First, x_i ($i = 1, 2, \ldots, N$) are generated as in Step 1 of Subsect. 4.1. Given x_i, $z_i's$ ($i = 1, 2, \ldots, N$) are generated from normal distribution with mean x_i and variance $\Sigma = 0.4$. Defining $t_i = z_i - x_i$ and assuming t measurement error, we apply ECME algorithm to estimate υ and Σ. Let υ_r denote the maximum likelihood estimate of υ obtained in the r-th simulation (r =1,2,. . . ,R). With R = 1,000 and N = 20, 40, 60, 80 we find that 679, 834, 908, 947 values of υ_r respectively exceed 6. The results show along with the results of the simulation study as reported in Subsect. 4.3 that even with moderate size validation data if one finds an estimate of υ and then apply the proposed methodology then one can expect to get results comparable to that obtained when the true measurement error is normal.

5 Concluding remarks

In this paper flexible parametric models have been used to model the distribution of true covariates as well as the distribution of measurement error. The advantages of using flexible parametric models are that they are easy to use, efficient and add a measure of robustness. Mixture of normal is particularly a convenient choice for a flexible parametric family. Almost all the distributions can be well approximated by mixture of normal distributions with large number of components. However, in practice there is not enough information available for the estimation of a large number of parameters. So here a mixture distribution with two components has been considered. The simulation study clearly shows that the proposed estimator performs uniformly well in all situation and is also robust to the violation of model assumption. The proposed model for measurement error although does not include distributions with tails shorter than the normal, in applications, however, departure from normality in this direction seems to be less frequent and with less serious consequences (Lange et al. 1989).

In Subsect. 4.4 a limited simulation study is made on modeling measurement error using unknown d.f. In standard regression problem with t error, Liu and Rubin (1994, 1995) have shown how the MLE of υ can be found much more efficiently by using the ECME algorithm. However, in this case, when x's are unknown the problem is much more formidable. Thus in this paper ν has been estimated from external validation data using ECME algorithm. The estimated value of υ is then treated as a tuning parameter and is used in measurement error analysis to estimate the parameter of interest β. The results are found to be extremely encouraging.

Acknowledgements The authors are thankful to the reviewers for their valuable suggestions.

Appendix

Detailed expressions of Eq. (11) of Sect. 3 are given below.

$$\begin{aligned}
&\tau_j(y_i, z_i; \xi^{(t)}) \\
&\quad = E(u_{ij} \,|d_{\mathrm{O}}\,; \xi^{(t)}) = \mathrm{Prob}\{u_{ij} = 1 \,|y_i,\, z_i; \xi^{(t)}\} \\
&\quad = \frac{\pi_j^{(t)} \int \left\{ \int f(y_i \,|x_i\,; \beta^{(t)}) f(z_i \,|x_i\,, \nu_i; \Sigma^{(t)}) f_j(x_i; \theta_j^{(t)}) \mathrm{d}x_i \right\} f(\nu_i) \mathrm{d}\nu_i}{\sum_{j=1}^{g} \pi_j^{(t)} \int \left\{ \int f(y_i \,|x_i\,; \beta^{(t)}) f(z_i \,|x_i, \nu_i\,; \Sigma^{(t)}) f_j(x_i; \theta_j^{(t)}) \mathrm{d}x_i \right\} f(\nu_i) \mathrm{d}\nu_i}.
\end{aligned} \tag{24}$$

In general the multidimensional integral within the curly bracket in Eq. (24) above is not easy to evaluate. However, in binary regression model with probit link the above integral can be explicitly expressed in terms of $\Phi(\cdot)$, the cdf of standard normal variable. For logit link the integral can be evaluated up to any desirable degree of accuracy by using a fast algorithm due to Crouch and Spiegelman (1990). If the integral within the curly bracket be denoted by $I_j(y_i, z_i, \nu_i; \xi^{(t)})$, then Eq. (24) boils down to

$$\frac{\pi_j^{(t)} \left(\int I_j(y_i, z_i, \nu_i; \xi^{(t)}) f(\nu_i) \mathrm{d}\nu_i \right)}{\sum_{j=1}^{g} \pi_j^{(t)} \left(\int I_j(y_i, z_i, \nu_i; \xi^{(t)}) f(\nu_i) \mathrm{d}\nu_i \right)}. \tag{25}$$

The integral within the parentheses in Eq. (25) is a one-dimensional integral, which can be approximated by a five point Gauss Hermite quadrature (Abramowitz and Stegun 1972) to yield $h_j(y_i, z_i; \xi^{(t)})$. Thus

$$\tau_j(y_i, z_i; \xi^{(t)}) = \frac{\pi_j^{(t)} h_j(y_i, z_i; \xi^{(t)})}{\sum_{j=1}^{g} \pi_j^{(t)} h_j(y_i, z_i; \xi^{(t)})}. \tag{26}$$

Next,

$$\begin{aligned}
a_{ij}^{(t)} &= E\left(x_i \,|d_{\mathrm{O}}, u_{ij} = 1; \xi^{(t)}\right) = \int x_i f(x_i \,|y_i, z_i, u_{ij} = 1; \xi^{(t)}) \mathrm{d}x_i \\
&= \int \left\{ \int x_i \frac{f(y_i \,|x_i; \beta^{(t)}) f(z_i \,|x_i, \nu_i; \Sigma^{(t)}) f_j(x_i; \theta_j^{(t)}) f(\nu_i) \pi_j}{f(y_i, z_i, \nu_i, u_{ij} = 1; \xi^{(t)})} \mathrm{d}x_i \right\} \\
&\quad \times \frac{f(y_i, z_i, \nu_i, u_{ij} = 1; \xi^{(t)})}{f(y_i, z_i, u_{ij} = 1; \xi^{(t)})} \mathrm{d}\nu_i \\
&= \int \left\{ \int x_i f(x_i \,|y_i, z_i, \nu_i, u_{ij} = 1; \xi^{(t)}) \mathrm{d}x_i \right\} f(\nu_i \,|z_i; \xi^{(t)}) \mathrm{d}\nu_i.
\end{aligned} \tag{27}$$

The multidimensional integral within the curly bracket in Eq. (27) is the mean of the conditional distribution of $x_i \,|y_i, z_i, \nu_i, u_{ij} = 1$. Following Schafer (1987) it

can be replaced by the corresponding mean of a normal distribution which approximates $f(x_i \,|y_i, z_i, v_i, u_{ij}=1)$. It is then easy to approximate the mean by the mode of the distribution of $x_i \,|y_i, z_i, v_i, u_{ij}=1$. Using the fact that the conditional distribution of $v_i \,|z_i; \xi^{(t)}$ is gamma the problem then boils down to evaluation of one-dimensional integral which can be approximated by using Gauss–Hermite quadrature. Lastly,

$$\begin{aligned} B_{ij}^{(t)} &= E\left((x_i - a_{ij}^{(t)})(x_i - a_{ij}^{(t)})^T \,|d_{\mathrm{O}}, u_{ij}=1; \xi^{(t)}\right) \\ &= E\left[\left(x_i - E(x_i \,|y_i, z_i, v_i, u_{ij}=1; \xi^{(t)})\right)\right. \\ &\quad \times \left(x_i - E(x_i \,|y_i, z_i, v_i, u_{ij}=1; \xi^{(t)})\right)^T \,|y_i, z_i, u_{ij}=1; \xi^{(t)}\Big] \\ &\quad + \left(a_{ij}^{(t)} - E(x_i \,|y_i, z_i, v_i, u_{ij}=1; \xi^{(t)})\right) \\ &\quad \times \left(a_{ij}^{(t)} - E(x_i \,|y_i, z_i, v_i, u_{ij}=1; \xi^{(t)})\right)^T. \end{aligned} \tag{28}$$

Making some algebraic manipulations the first term in Eq. (28) above can be reduced to a one dimensional integral where the integrand will be the product of the observed Fisher information matrix (Schafer 1987) of the conditional distribution of $x_i \,|y_i, z_i, v_i, u_{ij}=1$ and the pdf of the gamma distribution as described while evaluating $a_{ij}^{(t)}$.

References

Abramowitz, M., Stegun, I. (eds) (1972). *Handbook of mathematical functions*. New York: Dover.

Carroll, R.J., Gail, M.H., Lubin, J.H. (1993). Case-control studies with errors in covariates. *Journal of the American Statistical Association*, *88*, 185–199.

Carroll, R.J., Ruppert, D., Stefanski, L. A. (1995). *Measurement error in non linear models*. New York: Chapman & Hall.

Carroll, R.J., Maca, J.D., Ruppert, D. (1999a). Non parametric regression in the presence of measurement error. *Biometrika*, *86* (3), 541–554.

Carroll, R.J., Roeder, K., Wassermann, L. (1999b). Flexible parametric measurement error models. *Biometrics*, *55*, 44–54.

Crouch, E.A.C., Spiegelman, D. (1990). The evaluation of integrals of the form $\int_{-\infty}^{\infty} f(t)\exp(-t^2)dt$: application to logistic-normal models. *Journal of the American Statistical Association*, *85*, 464–467.

Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood for incomplete data via EM algorithm. *Journal of the Royal Statistical Society, Series B*, *39*, 1–38.

Fuller, W.A. (1980). *Measurement error models*. New York: John Wiley.

Lange, K.L., Little, R.J.A., Taylor, J.M.G. (1989). Robust statistical modeling using the *t* distribution. *Journal of the American Statistical Association*, *84*, 881–896.

Lange, K.L., Sinsheimer, J.S. (1993). Normal independent distributions and their applications in robust regression. *Journal Computational Graphical Statistics*, *2*, 175–198.

Liu, C.H., Rubin, D.B. (1994). The ECME algorithm: a simple extension of EM and ECM with fast monotone convergence. *Biometrika*, *81*, 633–648.

Liu, C.H., Rubin, D.B. (1995). ML estimation of the *t* distribution using EM and its extensions, ECM and ECME. *Statistica Sinica*, *5*, 19–40.

McCullagh, P., Nelder, J.A. (1989). *Generalized linear models*. London: Chapman and Hall.

Muller, P., Roeder, K. (1997). A Bayesian semi parametric model for case-control studies with errors in variables. *Biometrika*, *84*(3), 523–537.

Roeder, K., Carroll, R.J., Lindsay, B.G. (1996). A semi parametric mixture approach to case control studies with errors in covariables. *Journal of the American Statistical Association*, *91*, 722–732.

Rosner, B., Willet, W.C., Spiegelman, D. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within–person measurement error. *Statistics in Medicine*, *8*, 1075–1093.

Richardson, S., Leblond, L., Jaussent, I., Green, P.J. (2000). Mixture models in measurement error problems, with references to epidemiological studies, in Press.

Schafer, D.W. (1987). Covariate measurement error in generalized linear models. *Biometrika*, *74*, 385–391.

Stefanski, L.A. (1985). The effects of measurement error on parameter estimation. *Biometrika*, *72*, 385–389.

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

Sutradhar, B.C., Ali, M.M. (1989). A generalization of the Wishart distribution for the elliptical model and its moments for the multivariate *t* model. *Journal of Multivariate Analysis*, *29*, 155–162.