Ann Inst Stat Math (2012) 64:475–493
DOI 10.1007/s10463-010-0319-0

Instrumental variable approach to covariate measurement error in generalized linear models

Taraneh Abarin · Liqun Wang

Received: 25 November 2008 / Revised: 18 June 2010 / Published online: 17 November 2010

Abstract The paper presents the method of moments estimation for generalized linear measurement error models using the instrumental variable approach. The measurement error has a parametric distribution that is not necessarily normal, while the distributions of the unobserved covariates are nonparametric. We also propose simulation-based estimators for the situation where the closed forms of the moments are not available. The proposed estimators are strongly consistent and asymptotically normally distributed under some regularity conditions. Finite sample performances of the estimators are investigated through simulation studies.

Keywords Errors in variables · Generalized linear models · Heterogeneity · Measurement error · Instrumental variable · Method of moments · M-estimation · Nonlinear models · Simulation-based estimation

1 Introduction

Generalized linear models (GLM) are widely used in biostatistics, epidemiology, and many other areas. However, the real data analyzes using GLM often involve covariates that are not observed directly or are measured with error. See, e.g., Franks et al. (2004), Stimmer (2003), Kiechl et al. (2004), and Carroll et al. (2006). In such cases, statistical estimation and inference become very challenging. Several researchers, such as Stefanski and Carroll (1985), Aitkin (1996), and Rabe-Hesketh et al. (2003), have studied the maximum likelihood estimation of the GLM with measurement error. However, most of the proposed approaches rely on the normality assumption for the unobserved covariates and measurement error, though some other parametric distributions have

T. Abarin · L. Wang (✉)
Department of Statistics, University of Manitoba, Winnipeg, MB R3T 2N2, Canada
e-mail: liqun_wang@umanitoba.ca

been considered (Schafer 2001, Aitkin and Rocci 2002, Kukush and Schneeweiss 2005, Roy and Banerjee 2006). A computational difficulty with the likelihood approach is that the likelihood function involves multiple integrals which do not admit closed forms in general. Other approximate approaches such as corrected score functions, regression calibration and simulation extrapolation have been used (Nakamura 1990; Stefanski and Carroll 1991; Buzas and Stefanski 1996; Carroll et al. 2006). These approaches, however, give only approximately consistent estimators and are applicable when the measurement error variance is small.

General nonlinear models with classical errors-in-variables have been investigated by several authors using either replicate data (Li 1998; Thoresen and Laake 2003; Li and Hsiao 2004; Schennach 2004a) or instrumental variable methods (Wang and Hsiao 1995, 2010; Schennach 2007). In addition, non- or semi-parametric approaches have been considered (Schafer 2001; Taupin 2001; Schennach 2004b, 2008; Delaigle et al. 2006; Delaigle and Meister 2007). However, most of these papers deal with models with homoscedastic regression errors.

In this paper, we study the generalized linear models which allow very general heteroscedastic regression errors. In particular, we propose the method of moments estimation combined with instrumental variables (IV) method. This approach does not require parametric assumptions for the distributions of the unobserved covariates and of the measurement errors, which are difficult to check in practice. A similar method was used by Wang (2003, 2004), who deals with a nonlinear homoscedastic model with Berkson type measurement error. It is well-known that the Berkson and classical measurement errors lead to fundamentally different statistical structures and therefore must be treated differently (Carroll et al. 2006). In particular, a nonlinear model with Berkson error is usually identifiable without extra information if the parametric form of the model is known (Wang 2004). In contrast, in order for a classical measurement error model to be identifiable, extra information such as replicate data or instrumental data are needed.

In Sect. 2, we introduce the model and give some examples to motivate our estimation method. Then we introduce the method of moments estimators and derive their consistency and asymptotic normality in Sect. 3. In Sect. 4, we construct simulation-based estimators for the situations where the closed forms of the moments are not available. In Sect. 5, we present simulation studies of finite sample performances of the proposed estimators. Finally, conclusions and discussion are contained in Sect. 6, whereas proofs of the theorems are given in Sect. 7.

2 The model

In a generalized linear model (GLM, McCullagh and Nelder 1989), the first two conditional moments of the response variable $Y \in \mathbb{R}$ given the covariates $X \in \mathbb{R}^p$ can be written as

$$E(Y|X) = G^{-1}(\alpha + \beta' X),$$
$$V(Y|X) = \varphi K(G^{-1}(\alpha + \beta' X)),$$

where $\beta \in \mathbb{R}^p$, $\alpha \in \mathbb{R}$ and $\varphi \in \mathbb{R}$ are unknown parameters, G is the link function and K is a known function. It follows that

$$E(Y^2|X) = \varphi K(G^{-1}(\alpha + \beta' X)) + (G^{-1}(\alpha + \beta' X))^2.$$

To simplify notation, in this paper we consider the model of the form

$$E(Y|X) = f(\alpha + \beta' X), \tag{1}$$
$$E(Y^2|X) = g(\alpha + \beta' X, \varphi). \tag{2}$$

Further, suppose that X is unobservable, instead we observe

$$Z = X + \delta, \tag{3}$$

where δ is a random measurement error. Following Schennach (2007) and Wang and Hsiao (2010), we assume that an instrumental variable $W \in \mathbb{R}^q$, $q \geq p$, is available and is related to X through

$$X = HW + U, \tag{4}$$

where H is a $p \times q$ matrix of unknown parameters with rank p, U has distribution $f_U(u; \theta)$ with unknown parameters $\theta \in \Theta \subset \mathbb{R}^k$. An instrumental variable is defined as a random variable that is correlated with X but uncorrelated with measurement error and other random error terms in the model (Fuller 1987; Carroll et al. 2006). Furthermore, Fuller (1987, p. 51) showed that the conditions of an instrumental variable imply a parametric expression between X and W that can be written as (4). Note that (4) becomes a Berkson measurement error model, if H is an identity matrix and U is independent of W. However, the essential difference is that in a Berkson model W is an unbiased observation for X; while in an instrumental model, W can be a biased observation for X. Carroll et al. (2006) calls W "a weaker second observation for X" (in addition to Z). Therefore, the assumption for an instrumental variable is weaker than a replicate for X. The random variables are assumed to satisfy the following conditions.

Assumption 1 $E(Y^j|X, W) = E(Y^j|X)$, $j = 1, 2$; U is independent of W with $E(U) = 0$; and the measurement error δ is independent of X, W, Y with $E(\delta) = 0$.

Note that the first part of the above assumption is weaker than the usual assumption of nondifferential measurement error (i.e. the conditional distribution of Y given X, W does not depend on W). There is no assumption concerning the functional forms of the distributions of X and δ. In this sense model (1)–(4) is semiparametric. In this model, the observed variables are (Y, Z, W). Our interest is to estimate $\gamma = (\alpha, \beta, \varphi, \theta)$. We propose the method of moments estimators as follows. First, substitute (4) into (3) to obtain a usual linear regression equation

$$E(Z|W) = HW. \tag{5}$$

It follows that H can be consistently estimated by the least squares estimator

$$\hat{H} = \left(\sum_{i=1}^{n} Z_i W_i'\right)\left(\sum_{i=1}^{n} W_i W_i'\right)^{-1}. \tag{6}$$

Further, by model assumptions and the law of iterative expectation we have

$$\begin{aligned} E(Y|W) &= E[E(Y|X, W)|W] \\ &= E[E(Y|X)|W] \\ &= E\left[f(\alpha + \beta' X)|W\right] \\ &= E\left[f(\alpha + \beta'(HW + U))|W\right] \\ &= \int f(\alpha + \beta' HW + \beta' u) f_U(u; \theta)\mathrm{d}u, \end{aligned} \tag{7}$$

and, similarly,

$$E(Y^2|W) = \int g(\alpha + \beta' HW + \beta' u; \varphi) f_U(u; \theta)\mathrm{d}u \tag{8}$$

and

$$\begin{aligned} E(YZ|W) &= E(YX|W) \\ &= E[XE(Y|X)|W] \\ &= E\left[Xf(\alpha + \beta' X)|W\right] \\ &= \int (HW + u) f(\alpha + \beta' HW + \beta' u) f_U(u; \theta)\mathrm{d}u. \end{aligned} \tag{9}$$

Throughout the paper, all integrals are taken over the space $\mathbb{R}^p$. Given that H is consistently estimated by (6), other parameters α, β, θ and φ can be consistently estimated using (7)–(9) and nonlinear least squares method, provided that they are identifiable by these equations. Wang and Hsiao (1995) showed that the homoscedastic model with integrable $E(Y|X) = f(\cdot)$ and constant $V(Y|X)$ can be identifiable if the number of available instrumental variables is equal to or higher than the number of unobserved covariates. Schennach (2007) showed that a model with one unobserved covariate is generally identifiable if $E(Y|X) = f(\cdot)$ is bounded by a polynomial (but not necessarily integrable). The identifiability of general nonlinear heteroscedastic models appears to be a difficult problem which deserves further study. In practice, checking the model identifiability can be a tedious task and it is usually done in an *ad hoc* way. On the other hand, however, a model in practice is usually identified as long as it is carefully defined and well studied. In the following, we use some examples to demonstrate that the mentioned parameters may indeed be identified and consistently estimated using (7)–(9). To simplify notation, we consider the case where all variables are scalars and $U \sim N(0, \theta)$. For the same reason we let $H = 1$ because it is identified by (5).

Example 1 Consider a Gamma loglinear model where Y has a continuous distribution with first two conditional moments $E(Y|X) = \exp(\alpha + \beta X)$ and $V(Y \mid X) = \varphi \exp[2(\alpha + \beta X)]$. Here φ is the dispersion parameter. This type of model has wide applications in finance, radio ligand assays and kinetic reaction experiments. Under the model assumptions we have

$$
\begin{aligned}
E(Y|W) &= E\left[\exp(\alpha + \beta X)|W\right] \\
&= E\left[\exp(\alpha + \beta(W + U))|W\right] \\
&= \exp(\alpha + \beta W) E\left[\exp(\beta U)|W\right] \\
&= \exp\left(\alpha + \beta W + \frac{\beta^2 \theta}{2}\right), && (10) \\
E(Y^2|W) &= E\left[(\varphi + 1)\exp(2(\alpha + \beta X))|W\right] \\
&= (\varphi + 1)\exp(2(\alpha + \beta W)) E\left[\exp(2\beta U)|W\right] \\
&= (\varphi + 1)\exp\left[2(\alpha + \beta W + \beta^2 \theta)\right], && (11)
\end{aligned}
$$

and

$$
\begin{aligned}
E(YZ|W) &= \exp(\alpha) E[(W + U)\exp(\beta(W + U))|W] \\
&= \exp(\alpha + \beta W)[W E(\exp(\beta U)|W) + E(U \exp(\beta U)|W)] \\
&= (W + \beta\theta)\exp\left(\alpha + \beta W + \frac{\beta^2 \theta}{2}\right). && (12)
\end{aligned}
$$

Now we show that all unknown parameters are identifiable because they can be consistently estimated by (10)–(12) which are usual nonlinear regression equations in observed variables. First, β and $\exp(\alpha + \beta^2\theta/2)$ are identifiable by (10), and $(\beta\theta)\exp(\alpha + \beta^2\theta/2)$ is identifiable by (12). It follows that θ and α are identifiable. Further, given that all other parameters are identified, φ can be identified by (11). Thus, all parameters are identified.

Example 2 Consider a Poisson loglinear model where Y is a count variable with moments $E(Y|X) = V(Y|X) = \exp(\alpha + \beta X)$. This model has applications in biology, demographics and survival analysis. Then we have

$$
\begin{aligned}
E(Y|W) &= E\left[\exp(\alpha + \beta X)|W\right] \\
&= \exp(\alpha) E\left[\exp(\beta(W + U))|W\right] \\
&= \exp\left(\alpha + \beta W + \frac{\beta^2 \theta}{2}\right), && (13) \\
E(Y^2|W) &= E\left[\exp(\alpha + \beta X)|W\right] + E\left[\exp(2(\alpha + \beta X))|W\right] \\
&= \exp\left[2(\alpha + \beta W + \beta^2 \theta)\right] + \exp\left(\alpha + \beta W + \frac{\beta^2 \theta}{2}\right), && (14)
\end{aligned}
$$

and

$$\begin{aligned}E(YZ|W) &= \exp(\alpha)E[(W+U)\exp(W+U)|W] \\ &= \exp(\alpha+\beta W)[WE(\exp(\beta U)|W)+E(U\exp(\beta U)|W)] \\ &= (W+\beta\theta)\exp\left(\alpha+\beta W+\frac{\beta^2\theta}{2}\right). \end{aligned} \tag{15}$$

Since Eqs. (13) and (15) are identical to (10) and (12) respectively, similar to the Gamma model in the previous example we can show that parameters (α, β, θ) are identified using (13)–(15).

Example 3 Consider a logistic model $E(Y|X) = [1+\exp(-\alpha-\beta X)]^{-1}$. In this mode, Y is binary and its second moment is equal to the mean. Logistic regression has being used extensively in medical and social sciences as well as marketing applications. Using the assumptions for the model we have

$$E(Y|W) = E(Y^2|W) = \frac{1}{\sqrt{2\pi\theta}}\int \frac{\exp(-u^2/2\theta)}{1+\exp(-\alpha-\beta W-\beta u)}\mathrm{d}u,$$

and

$$E(YZ|W) = \frac{1}{\sqrt{2\pi\theta}}\int \frac{(W+u)\exp(-u^2/2\theta)}{1+\exp(-\alpha-\beta W-\beta u)}\mathrm{d}u.$$

Although these moments do not have closed forms, however, the unknown parameters (α, β, θ) can be identified by the fact that the measurement error-free Logistic model is identifiable, so they can be consistently estimated by the simulation-based approach in Sect. 3.

3 Method of moments estimator

Let $T = (1, Y, Z')'$ and

$$\tilde{x}' = \begin{pmatrix} 1 & 0 & x' \\ 0 & 1 & 0 \end{pmatrix},$$

then through variable substitution, (7)–(9) can be written together as

$$E(YT|W) = \int \tilde{x}h(\alpha+\beta' x;\varphi)f_U(x-HW;\theta)\mathrm{d}x, \tag{16}$$

where $h(\alpha+\beta' x;\varphi) = (f(\alpha+\beta' x), g(\alpha+\beta' x;\varphi))'$.

We define $\gamma = (\alpha, \beta', \theta', \varphi)'$ and the parameter space to be $\Gamma = A\times B\times\Theta\times\Phi$. The true parameter value of the model is denoted by $\gamma_0 \in \Gamma$. For every $v \in \mathbb{R}^p$ and

$\gamma \in \Gamma$, define

$$m(v;\gamma) = \int \tilde{x} h(\alpha + \beta' x; \varphi) f_U(x - v; \theta) \mathrm{d}x. \tag{17}$$

Then it is clear that $m(HW; \gamma_0) = E(YT|W)$.

Suppose (Y_j, Z_j, W_j), $j = 1, 2, \ldots, n$, is an $i.i.d.$ random sample, and $\hat{\rho}_j(\gamma) = Y_j T_j - m(\hat{H} W_j; \gamma)$, then the method of moments estimator (MME) for γ is defined as

$$\hat{\gamma}_n = \underset{\gamma \in \Gamma}{\operatorname{argmin}}\, Q_n(\gamma) = \underset{\gamma \in \Gamma}{\operatorname{argmin}} \sum_{j=1}^{n} \hat{\rho}_j'(\gamma) D_j \hat{\rho}_j(\gamma), \tag{18}$$

where $D_j = D(W_j)$ is a nonnegative definite matrix which may depend on W_j. Note that for the binary response Y, we have $E(Y|W) = E(Y^2|W)$, $f(x) = g(x, \varphi)$ and $\varphi = 1$. In this case the first two elements of $\rho_j(\gamma)$ are identical. This redundancy can be eliminated by setting the first row and first collum of $D(W_j)$ to be zeros. However, to simplify presentation in the following we present results in terms of general response Y. To prove the consistency of $\hat{\gamma}_n$, we assume the following regularity conditions, where μ denotes Lebesgue measure.

Assumption 2 The parameter space $\Gamma = A \times B \times \Theta \times \Phi$ is compact in $\mathbb{R}^{p+k+2}$. Furthermore, $E\,\|D(W)\|\,(\|YZ\|^2 + Y^4) < \infty$, and EWW' is nonsingular.

Assumption 3 $f(\alpha + \beta' x)$ and $g(\alpha + \beta' x; \varphi)$ are measurable functions of x for each $\alpha \in A, \beta \in B$ and $\varphi \in \Phi$. Furthermore, $f(\alpha + \beta' x) f_U(x - HW; \theta)$ and $g(\alpha + \beta' x; \varphi) f_U(x - HW; \theta)$ are uniformly bounded by a function $\eta(x, w)$, which satisfies $E\,\|D(W)\| \left(\|W\| \int \eta(x, W)(\|x\| + 1)\mathrm{d}x\right)^2 < \infty$.

Assumption 4 $E[\rho(\gamma) - \rho(\gamma_0)]' D(W)[\rho(\gamma) - \rho(\gamma_0)] = 0$ if and only if $\gamma = \gamma_0$, where $\rho(\gamma) = YT - m(HW; \gamma)$.

The above assumptions are common in nonlinear inference literature. In particular, Assumption 4 is a high-level assumption for parameter identifiability. For example, if the weight $D(W)$ is an identity or positive definite matrix then this assumption becomes that moment Eqs. (7)–(9) holds only for $\gamma = \gamma_0$. In practice this can be checked on case by case bases as demonstrated in examples 2.1–2.3.

Theorem 1 *Under Assumptions 1– 4,* $\hat{\gamma}_n \xrightarrow{a.s.} \gamma_0$*, as* $n \to \infty$.

To derive the asymptotic normality for $\hat{\gamma}_n$, we assume additional regularity conditions as follows.

Assumption 5 There exists an open subset $\gamma_0 \in \Gamma_0 \subset \Gamma$, in which all partial derivatives of orders 1 and 2 of $f(\alpha + \beta' x) f_U(x - HW; \theta)$ and $g(\alpha + \beta' x; \varphi) f_U(x - HW; \theta)$ w.r.t. γ are uniformly bounded by a function $K(x, w)$, which satisfies $E\,\|D(W)\| \left(\|W\| \int K(x, W)(\|x\| + 1)\mathrm{d}x\right)^2 < \infty$.

Assumption 6 The matrix

$$\kappa = E\left[\frac{\partial\rho'(\gamma_0)}{\partial\gamma}D(W)\frac{\partial\rho(\gamma_0)}{\partial\gamma'}\right]$$

is nonsingular.

These assumptions are equivalent to the ones that are needed for consistency and asymptotic normality of an M-estimator (see e.g., van der Vaart 2000, Sects. 5.2 and 5.3).

Theorem 2 *Under Assumptions 1–6,* $\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{L} N(0, \kappa^{-1}C\tau C'\kappa^{-1})$ *as* $n \to \infty$*, where*

$$C = \left[I_{p+k+2}, E\left(\frac{\partial\rho'(\gamma_0)}{\partial\gamma}D(W)\frac{\partial\rho(\gamma_0)}{\partial\psi'}\right)\left(EWW' \otimes I_p\right)^{-1}\right],$$

and

$$\begin{aligned}\tau &= \begin{pmatrix}\tau_{11} & \tau_{12}\\ \tau_{12}' & \tau_{22}\end{pmatrix},\\ \tau_{11} &= E\left[\frac{\partial\rho'(\gamma_0)}{\partial\gamma}D(W)\rho(\gamma_0)\rho'(\gamma_0)D(W)\frac{\partial\rho(\gamma_0)}{\partial\gamma'}\right],\\ \tau_{12} &= E\left[\frac{\partial\rho'(\gamma_0)}{\partial\gamma}D(W)\rho(\gamma_0)\left((Z-HW)'\otimes W'\right)\right]\end{aligned}$$

and $\tau_{22} = E\left[WW' \otimes (Z-HW)(Z-HW)'\right]$.

The above asymptotic covariance matrix depends on the weight $D(W)$. It is of interest to choose an appropriate $D(W)$ to obtain the most efficient estimator. It can be shown (Abarin 2008, Abarin and Wang 2006) that the most efficient choice of weight is $D = F^{-1}$, where $F = E(\rho(\gamma_0)\rho'(\gamma_0) \mid W)$ which leads to the asymptotic covariance matrix

$$E\left[\frac{\partial\rho'(\gamma_0)}{\partial\gamma}F^{-1}\frac{\partial\rho(\gamma_0)}{\partial\gamma'}\right]^{-1}. \tag{19}$$

In practice, F is a function of unknown parameters and therefore needs to be estimated. This can be done using the following two-stage procedure. First, minimize $Q_n(\gamma)$ with identity matrix $D = I_{p+2}$ to obtain the first-stage estimator $\hat{\gamma}_n$. Secondly, estimate F by $\hat{F} = \frac{1}{n}\sum_{j=1}^n \rho_j(\hat{\gamma}_n)\rho_j'(\hat{\gamma}_n)$ or alternatively by a nonparametric estimator, and then minimize $Q_n(\gamma)$ again with $D = \hat{F}^{-1}$ to obtain the second-stage estimator $\hat{\hat{\gamma}}_n$. Since $\hat{F}$ is consistent for F, the asymptotic covariance of $\hat{\hat{\gamma}}_n$ is given by (19). Consequently $\hat{\hat{\gamma}}_n$ is asymptotically more efficient than the first-stage estimator $\hat{\gamma}_n$.

4 Simulation-based estimator

When the explicit form of $m(v;\gamma)$ exists, the numerical computation of MME $\hat{\gamma}_n$ can be done using the usual optimization methods. However, sometimes the integrals in (17) do not have explicit forms. In this section, we use a simulation-based approach to overcome this problem. The simulation-based approach is used to approximate the multiple integrals in which they are simulated by Monte Carlo methods such as importance sampling.

We start with choosing a known density $l(x)$ and generate independent random points $\{x_{js}, s=1,2,\ldots,2S, j=1,2,\ldots,n\}$ from $l(x)$. Then we can approximate $m(HW_j;\gamma)$ by Monte Carlo simulators

$$m_S(HW_j;\gamma)=\frac{1}{S}\sum_{s=1}^{S}\frac{\tilde{x}_{js}h(\alpha+\beta' x_{js};\varphi)f_U(x_{js}-HW_j;\theta)}{l(x_{js})} \tag{20}$$

and

$$m_{2S}(HW_j;\gamma)=\frac{1}{S}\sum_{s=S+1}^{2S}\frac{\tilde{x}_{js}h(\alpha+\beta' x_{js};\varphi)f_U(x_{js}-HW_j;\theta)}{l(x_{js})}, \tag{21}$$

where

$$\tilde{x}'_{js}=\begin{pmatrix}1 & 0 & x'_{js}\\ 0 & 1 & 0\end{pmatrix}.$$

Finally, the simulation-based estimator (SBE) for γ is defined by

$$\hat{\gamma}_{n,S}=\underset{\gamma\in\Gamma}{\operatorname{argmin}}\, Q_{n,S}(\gamma)=\underset{\gamma\in\Gamma}{\operatorname{argmin}}\sum_{j=1}^{n}\hat{\rho}'_{j,S}(\gamma)D_j\hat{\rho}_{j,2S}(\gamma), \tag{22}$$

where $\hat{\rho}_{j,S}(\gamma)=Y_jT_j-m_S(\hat{H}W_j;\gamma)$ and $\hat{\rho}_{j,2S}(\gamma)=Y_jT_j-m_{2S}(\hat{H}W_j;\gamma)$.

We notice that by construction, $E[m_S(HW_j;\gamma)|W_j]=E[m_{2S}(HW_j;\gamma)|W_j]=m(HW_j;\gamma)$, and therefore $m_S(HW_j;\gamma)$ and $m_{2S}(HW_j;\gamma)$ are unbiased simulators for $m(HW_j;\gamma)$. Moreover, $Q_{n,S}(\gamma)$ is an unbiased simulator for $Q_n(\gamma)$, because $Q_{n,S}(\gamma)$ and $Q_n(\gamma)$ have the same conditional expectation given the sample (Y_j,Z_j,W_j).

Alternatives to 20 and 21 which generally yield more stable estimates are

$$m'_S(HW_j;\gamma)=\frac{\sum_{s=1}^{S}\tilde{x}_{js}h(\alpha+\beta' x_{js};\varphi)f_U(x_{js}-HW_j;\theta)/l(x_{js})}{\sum_{s=1}^{S}f_U(x_{js}-HW_j;\theta)/l(x_{js})} \tag{23}$$

and

$$m'_{2S}(HW_j;\gamma)=\frac{\sum_{s=S+1}^{2S}\tilde{x}_{js}h(\alpha+\beta' x_{js};\varphi)f_U(x_{js}-HW_j;\theta)/l(x_{js})}{\sum_{s=S+1}^{2S}f_U(x_{js}-HW_j;\theta)/l(x_{js})}, \quad (24)$$

where we have replaced S with the sum of the weights. Since $(1/S)\sum_{s=1}^{S}f_U(x_{js}-HW_j;\theta)/l(x_{js})\approx 1$, when S is large enough, $m'_S(HW_j;\gamma)\approx m_S(HW_j;\gamma)$ and $m'_{2S}(HW_j;\gamma)\approx m_{2S}(HW_j;\gamma)$. Although these simulators are biased, the biases are small, and the improvement in variance makes them preferred alternatives to 20 and 21. (See Lemma 4.3 of Robert and Cassella 2004.) We will compare the estimators in Sect. 4.

Theorem 3 *Suppose that the support of $l(x)$ covers the support of $h(\alpha+\beta' x;\varphi)f_U(x-v;\theta)$ for all $v\in\mathbb{R}^p$ and $\gamma\in\Gamma$. Then the simulation estimator $\hat{\gamma}_{n,S}$ has the following properties:*

1. *Under* Assumptions 1–4, $\hat{\gamma}_{n,S}\xrightarrow{a.s.}\gamma_0$, *as* $n\to\infty$.
2. *Under* Assumptions 1–6, $\sqrt{n}(\hat{\gamma}_{n,S}-\gamma_0)\xrightarrow{L}N(0,\kappa^{-1}C\tau_S C'\kappa^{-1})$, *where* C *is defined as in Theorem 2,* $\tau_{S,12}=\tau_{12}$, $\tau_{S,22}=\tau_{22}$ *and*

$$2\tau_{S,11}=E\left[\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma}D_1\rho_{1,2S}(\gamma_0)\rho'_{1,2S}(\gamma_0)D_1\frac{\partial\rho_{1,S}(\gamma_0)}{\partial\gamma'}\right]$$
$$+E\left[\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma}D_1\rho_{1,2S}(\gamma_0)\rho'_{1,S}(\gamma_0)D_1\frac{\partial\rho_{1,2S}(\gamma_0)}{\partial\gamma'}\right].$$

Although asymptotically, the importance density $l(x)$ does not have an effect on the efficiency of $\hat{\gamma}_{n,S}$, however, the choice of $l(x)$ will affect the finite sample variances of the simulators $m_S(HW_j;\gamma)$ and $m_{2S}(HW_j;\gamma)$. In addition, similar to Wang (2004) we can show that the efficiency loss caused by simulation is of magnitude $O(1/S)$.

5 Simulation studies

In this section we present simulation studies on three generalized linear models of Examples 2.1–2.3, to demonstrate how the proposed estimators can be calculated and their performance in finite sample sizes.

First, consider the *Gamma loglinear model* in Example 2.1 where $U\sim N(0,\theta)$. We calculated conditional moments (7)–(9) for this model in Sect. 2. Therefore, the method of moments estimators (MME) can be computed by minimizing $Q_n(\gamma)$ in (18). Specifically, the estimators are computed in two steps, using the identity and the estimated optimal weighting matrix, respectively.

To compute the simulation-based estimators (SBE), we choose the density of $N(0,2)$ to be $l(x_{js})$, and generate independent points x_{js}, $s=1,2,\ldots,2S$, $j=1,2,\ldots,n$ using $S=2{,}000$. Furthermore, the simulated moments $m_S(HW_j;\gamma)$ and $m_{2S}(HW_j;\gamma)$, and $m'_S(HW_j;\gamma)$ and $m'_{2S}(HW_j;\gamma)$ are calculated according to

Table 1 Gamma loglinear model with measurement error in covariate

	$\alpha = -0.2$	$\beta = 0.3$	$\theta = 0.8$	$\varphi = 0.1$
MME	−0.206	0.299	0.810	0.101
STD	0.001	0.001	0.006	0.001
MSE	0.002	0.002	0.033	0.001
SBE1	−0.206	0.299	0.814	0.101
STD	0.001	0.001	0.006	0.001
MSE	0.002	0.002	0.034	0.001
SBE2	−0.206	0.298	0.815	0.101
STD	0.001	0.001	0.006	0.001
MSE	0.002	0.002	0.034	0.001
z_1	0.450	0.515	0.312	0.058
z_2	0.223	0.134	1.355	0.333

(20) and (21), and (23) and (24) respectively. The two-step SBE1 $\hat{\gamma}_{n,s}$ is calculated by minimizing $Q_{n,S}(\gamma)$, using the identity and the estimated optimal weighting matrix, and $m_S(HW_j;\gamma)$ and $m_{2S}(HW_j;\gamma)$. Similarly the two-step SBE2 $\hat{\gamma}_{n,s}$ is calculated by minimizing $Q_{n,S}(\gamma)$, using the identity and the estimated optimal weighting matrix, and $m'_S(HW_j;\gamma)$ and $m'_{2S}(HW_j;\gamma)$. The data have been generated using W and δ from a standard normal distribution, and parameter values $\alpha = -0.2, \beta = 0.3, \theta = 0.8, \varphi = 0.1$. In all the simulation studies, we assumed that H is known and equal to one. We also generated the response variable from a Gamma distribution with parameters $1/\varphi$ and $\varphi\exp(\alpha+\beta X)$, respectively. $N = 1{,}000$ Monte Carlo replications have been carried out and, in each replication, $n = 400$ sample points (Y_j, Z_j, W_j) have been generated. The computation has been done using MATLAB on a workstation running LINUX operating system.

Table 1 shows the summaries of the results for the Gamma loglinear model. As we can see in the table, estimators show their asymptotic properties. However, both SBE1 and SBE2 converge slower for θ. The results for α is not satisfying and it shows that estimators have finite sample bias. We used the mean squared errors (MSE) to compare the efficiency of the estimators. There is not any significant difference between the estimators efficiencies.

To test the difference between the MME and SBE1, we calculated z statistic for the test. As we can see in the table, z_1 shows that there is no significant difference between the MME and SBE1 for each parameter. Similarly, we calculated z_2 to test the difference between SBE1 and SBE2. The results show that there is no significant difference between them.

Next, we consider the *Poisson loglinear model* in Example 2.2, where $U \sim N(0,\theta)$. We calculated conditional moments (7)–(9) for this model in Sect. 2. Parameters are $\alpha = -0.2, \beta = 0.3$, and $\theta = 1$, and we choose the density of $N(0,2)$ to be $l(x_{js})$. We generated the response variable from a Poisson distribution with parameter $\exp(\alpha + \beta X)$.

Table 2 shows the summaries of the results for the Poisson loglinear model. As we can see in the table, all the estimators show their asymptotic properties. The mean squared errors (MSE) of the estimates show that the MME and both SB estimators perform equally, for all the parameters. Furthermore, the test statistic values imply that there is no significant difference between the estimators.

Table 2 Poisson Loglinear model with measurement error in covariate

	$\alpha = -0.2$	$\beta = 0.3$	$\theta = 1$
MME	−0.201	0.295	0.998
STD	0.002	0.002	0.006
MSE	0.003	0.005	0.030
SBE1	−0.201	0.296	0.999
STD	0.002	0.002	0.006
MSE	0.003	0.005	0.031
SBE2	−0.201	0.295	0.999
STD	0.002	0.002	0.006
MSE	0.003	0.005	0.030
z_1	0.281	0.641	0.445
z_2	0.224	0.570	0.411

Table 3 Logistic model with measurement error in covariate

	$\alpha = 0.5$	$\beta = 0.3$	$\theta = 0.8$
SBE1	0.506	0.306	0.811
STD	0.004	0.004	0.008
MSE	0.018	0.014	0.037
SBE2	0.507	0.307	0.809
STD	0.004	0.004	0.007
MSE	0.017	0.014	0.035
z	0.515	0.294	0.446

Finally we consider the *logistic model* in Example 2.3, where $U \sim N(0, \theta)$. For this model, conditional moments (7)–(9) do not have closed forms. Therefore, we were only able to compute SBE1 and SBE2. Parameter values for this model are $\alpha = 0.5$, $\beta = 0.3$, and $\theta = 0.8$, and we choose the density of $N(0, 2)$ to be $l(x_{js})$. We generated the response variable from a binary distribution with parameter $(1 + \exp(-(\alpha + \beta X)))^{-1}$.

Table 3 shows the summaries of the results for the logistic model. The mean squared errors (MSE) of the estimates show that SBE1 and SBE2 perform equally well, except for slightly more efficiency in SBE2, for θ. Furthermore, the test statistic values imply that there is no significant difference between the estimators.

6 Conclusions and discussion

Estimation of generalized linear errors-in-variables models with multivariate predictor variables and possibly nonnormal random errors have being considered for decades. Since these models are quite challenging, most researchers rely on restrictive conditions to achieve consistent estimation. Moreover, most methods in the literature are designed for the case where either validation or replicate data are available, which can be restrictive in many applications.

In this work, we generalized the method of Wang (2003, 2004) to the case of classical measurement error and the heterogeneous response variable given the covariates. This case which is applicable to different fields of health science, allows researchers to apply generalized linear models to data sets with this property. We used the instrumental variable approach to study generalized linear models, where the predictor variable is multivariate and the distributions of the measurement error is parametric, but not necessarily normal. Moreover, our method does not need any parametric assumption on the distribution of the unknown covariates. This makes the method less restrictive than some other methods that need either parametric distribution of the covariates, or to estimate it using some extra information. We developed consistent and asymptotic normal estimators for the model, using the method of moments. Moreover, we addressed the practical computational issue of finding explicit form of a function that involves multiple integrals, using methods of simulated moments. Simulation studies show that the estimators perform satisfactorily in some finite sample situations.

7 Proofs

As we mentioned before, it is straightforward to estimate H using (5) and least squares method. In practice, one can estimate H as well as other parameters, using the main sample. In this case, the asymptotic variances can be obtained by the so-called delta-method (as we presented in Theorems 2 and 3). However, if H is estimated using an external sample or a subset of the main sample, and other parameters are estimated using the rest of the main sample, then the asymptotic covariance matrices have simpler forms and the theoretical results may be regarded as conditional on the pre-estimated H. In fact, the asymptotic covariance matrix of $\hat{\gamma}_n$ will reduce to $\kappa^{-1}\tau_{11}\kappa^{-1}$, and for the simulation-based estimator to $\kappa^{-1}\tau_{S,11}\kappa^{-1}$. To avoid complicated notations and lengthy proofs, here we present the proofs only for the case $H = 1$. The proofs for the more general case where H is unknown can be found in Abarin (2008).

7.1 Proof of Theorem 1

To prove the consistency, we use the Uniform Law of Large Numbers (ULLN, Jennrich 1969, Theorem 2). The idea is to show that $E\sup_\gamma|Q_n(\gamma)| \leq \infty$. By Assumptions 2 and 3, and the Dominated Convergence Theorem,

$$
\begin{aligned}
&E\sup_{\gamma}\left|\rho_1'(\gamma)D_1\rho_1(\gamma)\right| \leq E\,\|D_1\|\sup_{\gamma}\|\rho_1(\gamma)\|^2\\
&\leq 3E\,\|D_1\|\,(\|Y_1Z_1\|^2+Y_1^4)\\
&\quad +3E\,\|D_1\|\left(\int\sup_{\gamma}\left|f(\alpha+\beta'x)\right|f_U(x-W_1;\theta)(\|x\|+1)dx\right)^2\\
&\quad +3E\,\|D_1\|\left(\int\sup_{\gamma}\left|g(\alpha+\beta'x;\phi)\right|f_U(x-W_1;\theta)dx\right)^2\\
&< \infty.
\end{aligned}
$$

It follows from the ULLN that

$$
\sup_{\gamma}\left|\frac{1}{n}\sum_{j=1}^{n}\rho_j'(\gamma)D_j\rho_j(\gamma)-Q(\gamma)\right|\xrightarrow{a.s.}0, \tag{25}
$$

where $Q(\gamma) = E\rho_1'(\gamma)D_1\rho_1(\gamma)$. Therefore,

$$\sup_{\gamma}\left|\frac{1}{n}Q_n(\gamma) - Q(\gamma)\right| \xrightarrow{a.s.} 0. \tag{26}$$

Now for the next step, we use Lemma 3 of Amemiya (1973) to show that $\hat{\gamma}_n \xrightarrow{a.s.} \gamma_0$. Since $E(\rho_1(\gamma_0)|W_1) = 0$ and $\rho_1(\gamma) - \rho_1(\gamma_0)$ depends on W_1 only, we have

$$E[\rho_1'(\gamma_0)D_1(\rho_1(\gamma) - \rho_1(\gamma_0))] = E[E(\rho_1'(\gamma_0)|W_1)D_1(\rho_1(\gamma) - \rho_1(\gamma_0))] = 0,$$

which implies $Q(\gamma) = Q(\gamma_0) + E[(\rho_1(\gamma) - \rho_1(\gamma_0))'D_1(\rho_1(\gamma) - \rho_1(\gamma_0))]$. By Assumption 4, $Q(\gamma) \geq Q(\gamma_0)$ and equality holds if and only if $\gamma = \gamma_0$. Thus, $Q(\gamma)$ attains a unique minimum at $\gamma_0 \in \gamma$, and it follows from the Lemma that $\hat{\gamma}_n \xrightarrow{a.s.} \gamma_0$. □

7.2 Proof of Theorem 2

In the first step of the proof, using Lemma 4 of Amemiya (1973), we need to show that $E\sup_{\gamma}\left\|\frac{\partial^2 Q_n(\gamma)}{\partial\gamma\partial\gamma'}\right\| < \infty$. By Assumptions 3–5 and the Dominated Convergence Theorem, the first derivative $\partial Q_n(\gamma)/\partial\gamma$ has the first-order Taylor expansion in the open neighborhood $\Gamma_0 \subset \Gamma$ of γ_0. Since $\partial Q_n(\hat{\gamma}_n)/\partial\gamma = 0$ and $\hat{\gamma}_n \xrightarrow{a.s.} \gamma_0$, for sufficiently large n we have

$$\frac{\partial Q_n(\gamma_0)}{\partial\gamma} + \frac{\partial^2 Q_n(\tilde{\gamma})}{\partial\gamma\partial\gamma'}(\hat{\gamma}_n - \gamma_0) = 0, \tag{27}$$

where $\|\tilde{\gamma} - \gamma_0\| \leq \|\hat{\gamma}_n - \gamma_0\|$. The first and the second derivative of $Q_n(\gamma)$ in (27) is given by

$$\frac{\partial Q_n(\gamma)}{\partial\gamma} = 2\sum_{j=1}^{n}\frac{\partial\rho_j'(\gamma)}{\partial\gamma}D_j\rho_j(\gamma), \tag{28}$$

and

$$\frac{\partial^2 Q_n(\gamma)}{\partial\gamma\partial\gamma'} = 2\sum_{j=1}^{n}\left[\frac{\partial\rho_j'(\gamma)}{\partial\gamma}D_j\frac{\partial\rho_j(\gamma)}{\partial\gamma'} + (\rho_j'(\gamma)D_j \otimes I_{p+k+2})\frac{\partial\text{vec}(\partial\rho_j'(\gamma)/\partial\gamma)}{\partial\gamma'}\right],$$

respectively. Assumptions 2–5 imply that

$$E\sup_{\gamma}\left\|\frac{\partial\rho_1'(\gamma)}{\partial\gamma}D_1\frac{\partial\rho_1(\gamma)}{\partial\gamma'}\right\| < \infty.$$

and

$$\left(E \sup_{\gamma} \left\| (\rho_1'(\gamma) D_1 \otimes I_{p+k+2}) \frac{\partial \mathrm{vec}(\partial \rho_1'(\gamma)/\partial \gamma)}{\partial \gamma'} \right\| \right)^2 < \infty.$$

Since $\partial \mathrm{vec}(\partial \rho_1'(\gamma_0)/\partial \gamma)/\partial \gamma'$ depends on W_1 only and therefore

$$\begin{aligned} & E\left[\left(\rho_1'(\gamma_0) D_1 \otimes I_{p+k+2}\right) \frac{\partial \mathrm{vec}(\partial \rho_1'(\gamma_0)/\partial \gamma)}{\partial \gamma'} \right] \\ & = E\left[\left(E(\rho_1'(\gamma_0)|W_1) D_1 \otimes I_{p+k+2}\right) \frac{\partial \mathrm{vec}(\partial \rho_1'(\gamma_0)/\partial \gamma)}{\partial \gamma'} \right] \\ & = 0, \end{aligned}$$

it follows from the ULLN and Amemiya (1973, Lemma 4) that

$$\begin{aligned} \frac{1}{2n} \frac{\partial^2 Q_n(\gamma)}{\partial \gamma \partial \gamma'} & \xrightarrow{a.s.} E\left[\frac{\partial \rho_1'(\gamma_0)}{\partial \gamma} D_1 \frac{\partial \rho_1(\gamma_0)}{\partial \gamma'} + \left(\rho_1'(\gamma_0) D_1 \otimes I_{p+k+2}\right) \frac{\partial \mathrm{vec}(\partial \rho_1'(\gamma_0)/\partial \gamma)}{\partial \gamma'} \right] \\ & = \kappa. \end{aligned} \tag{29}$$

Now, we know that $\frac{\partial \rho_j'(\gamma)}{\partial \gamma} D_j \rho_j(\gamma)$, $j = 1, 2, \ldots, n$ are $i.i.d.$ with the mean vector zero and the variance covariance matrix τ, where τ is given in Theorem 2. Therefore, by Slutsky's Theorem, we have

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{L} N(0, \kappa^{-1} \tau \kappa^{-1}). \tag{30}$$

□

7.3 Proof of Theorem 3

Since $\rho_{1,S}(\gamma)$ and $\rho_{1,2S}(\gamma)$ are conditionally independent given (W_1, Y_1, Z_1), it follows from Assumptions 2 and 3 that

$$E \sup_{\gamma} \left| \rho_{1,S}'(\gamma) D_1 \rho_{1,2S}(\gamma) \right| < \infty.$$

Therefore by the ULLN,

$$\sup_{\gamma} \left| \frac{1}{n} \sum_{j=1}^{n} \rho_{j,S}'(\gamma) D_j \rho_{j,2S}(\gamma) - E \rho_{1,S}'(\gamma) D_1 \rho_{1,2S}(\gamma) \right| \xrightarrow{a.s.} 0, \tag{31}$$

where

$$\begin{aligned} E\rho'_{1,S}(\gamma)D_1\rho_{1,2S}(\gamma) &= E\left[E\left(\rho'_{1,S}(\gamma)|W_1, Y_1, Z_1\right) D_1 E\left(\rho'_{1,2S}(\gamma)|W_1, Y_1, Z_1\right)\right] \\ &= E\rho'_1(\gamma)D_1\rho_1(\gamma) \\ &= Q(\gamma). \end{aligned}$$

Therefore,

$$\sup_{\gamma}\left|\frac{1}{n}Q_{n,S}(\gamma) - Q(\gamma)\right| \xrightarrow{a.s.} 0. \tag{32}$$

We showed in the proof of Theorem 1 that $Q(\gamma)$ attains a unique minimum at $\gamma_0 \in \Gamma$. Therefore $\hat{\gamma}_{n,S} \xrightarrow{a.s.} \gamma_0$ follows from Amemiya (1973, Lemma 3). □

To prove the second part of Theorem 3, first, $\partial Q_{n,S}(\gamma)/\partial\gamma$ has the first-order Taylor expansion in an open neighborhood $\Gamma_0 \subset \Gamma$ of γ_0:

$$\frac{\partial Q_{n,S}(\gamma_0)}{\partial\gamma} + \frac{\partial^2 Q_{n,S}(\tilde{\gamma})}{\partial\gamma\partial\gamma'}(\hat{\gamma}_{n,S} - \gamma_0) = 0, \tag{33}$$

where $\|\tilde{\gamma} - \gamma_0\| \leq \left\|\hat{\gamma}_{n,S} - \gamma_0\right\|$. The first and the second derivative of $Q_{n,S}(\gamma)$ is given by

$$\frac{\partial Q_{n,S}(\gamma)}{\partial\gamma} = \sum_{j=1}^{n}\left[\frac{\partial\rho'_{j,S}(\gamma)}{\partial\gamma}D_j\hat{\rho}_{j,2S}(\gamma) + \frac{\partial\rho'_{j,2S}(\gamma)}{\partial\gamma}D_j\rho_{j,S}(\gamma)\right],$$

and

$$\begin{aligned} \frac{\partial^2 Q_{n,S}(\gamma)}{\partial\gamma\partial\gamma'} &= \sum_{j=1}^{n}\left[\frac{\partial\rho'_{j,S}(\gamma)}{\partial\gamma}D_j\frac{\partial\rho_{j,2S}(\gamma)}{\partial\gamma'} + (\rho'_{j,2S}(\gamma)D_j \otimes I_{p+k+2})\frac{\partial\mathrm{vec}(\partial\rho'_{j,S}(\gamma)/\partial\gamma)}{\partial\gamma'}\right] \\ &\quad + \sum_{j=1}^{n}\left[\frac{\partial\rho'_{j,2S}(\gamma)}{\partial\gamma}D_j\frac{\partial\rho_{j,S}(\gamma)}{\partial\gamma'} + (\rho'_{j,S}(\gamma)D_j \otimes I_{p+k+2})\frac{\partial\mathrm{vec}(\partial\rho'_{j,2S}(\gamma)/\partial\gamma)}{\partial\gamma'}\right], \end{aligned}$$

respectively. Similar to (29), we can show that $\frac{1}{n}\frac{\partial^2 Q_{n,S}(\gamma)}{\partial\gamma\partial\gamma'}$ converges $a.s.$ to

$$\begin{aligned} &E\left[\frac{\partial\rho'_{1,S}(\gamma_0)}{\partial\gamma}D_1\frac{\partial\rho_{1,2S}(\gamma_0)}{\partial\gamma'} + (\rho'_{1,2S}(\gamma_0)D_1 \otimes I_{p+k+2})\frac{\partial\mathrm{vec}(\partial\rho'_{1,S}(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right] \\ &+E\left[\frac{\partial\rho'_{1,2S}(\gamma_0)}{\partial\gamma}D_1\frac{\partial\rho_{1,S}(\gamma_0)}{\partial\gamma'} + (\rho'_{1,S}(\gamma_0)D_1 \otimes I_{p+k+2})\frac{\partial\mathrm{vec}(\partial\rho'_{1,2S}(\gamma_0)/\partial\gamma)}{\partial\gamma'}\right] \end{aligned}$$

uniformly for all $\gamma \in \Gamma$. Since

$$E\left[\frac{\partial \rho'_{1,S}(\gamma_0)}{\partial \gamma} D_1 \frac{\partial \rho_{1,2S}(\gamma_0)}{\partial \gamma'}\right] = E\left[\frac{\partial \rho'_1(\gamma_0)}{\partial \gamma} D_1 \frac{\partial \rho_1(\gamma_0)}{\partial \gamma'}\right] = \kappa$$

and

$$E\left[(\rho'_{1,2S}(\gamma_0) D_1 \otimes I_{p+k+2}) \frac{\partial \text{vec}(\partial \rho'_{1,S}(\gamma_0)/\partial \gamma)}{\partial \gamma'}\right] = 0,$$

we have

$$\frac{1}{n} \frac{\partial^2 Q_{n,S}(\gamma)}{\partial \gamma \partial \gamma'} \xrightarrow{a.s.} 2\kappa. \tag{34}$$

Further, by the Central Limit Theorem we have

$$\frac{1}{2\sqrt{n}} \frac{\partial Q_{n,S}(\gamma)}{\partial \gamma} \xrightarrow{L} N(0, \tau_S), \tag{35}$$

where

$$\begin{aligned}
\tau_S &= \frac{1}{4} E\left[\left(\frac{\partial \rho'_{1,S}(\gamma_0)}{\partial \gamma} D_1 \rho_{1,2S}(\gamma_0) + \frac{\partial \rho'_{1,2S}(\gamma_0)}{\partial \gamma} D_1 \rho_{1,S}(\gamma_0)\right)\right. \\
&\quad \left. \times \left(\rho'_{1,2S}(\gamma_0) D_1 \frac{\partial \rho_{1,S}(\gamma_0)}{\partial \gamma'} + \rho'_{1,S}(\gamma_0) D_1 \frac{\partial \rho_{1,2S}(\gamma_0)}{\partial \gamma'}\right)\right] \\
&= \frac{1}{2} E\left[\frac{\partial \rho'_{1,S}(\gamma_0)}{\partial \gamma} D_1 \rho_{1,2S}(\gamma_0) \rho'_{1,2S}(\gamma_0) D_1 \frac{\partial \rho_{1,S}(\psi_0)}{\partial \gamma'}\right] \\
&\quad + \frac{1}{2} E\left[\frac{\partial \rho'_{1,S}(\gamma_0)}{\partial \gamma} D_1 \rho_{1,2S}(\gamma_0) \rho'_{1,S}(\gamma_0) D_1 \frac{\partial \rho_{1,2S}(\gamma_0)}{\partial \gamma'}\right],
\end{aligned}$$

Finally, the second part of Theorem 3 follows from (34) and (35), and Slutsky's Theorem. □

Acknowledgments The authors thank the editor and two anonymous referees for their constructive comments and suggestions that lead to improvement of this paper. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the National Institute for Complex Data Structures (NICDS).

References

Abarin, T. (2008). Second-order least squares estimation in regression models with application to measurement error problems. PhD Thesis, University of Manitoba, Winnipeg.

Abarin, T., Wang, L. (2006). Comparison of GMM with second-order least squares estimation in nonlinear models. *Far East Journal of Theoretical Statistics, 20*(2), 179–196.

Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing, 6*, 251–262.

Aitkin, M., Rocci, R. (2002). A general maximum likelihood analysis of measurement error in generalized linear models. *Statistics and Computing, 12*, 163–174.

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica, 41*, 997–1016.

Buzas, J. S., Stefanski, L. A. (1996). Instrumental variable estimation in generalized linear measurement error models. *Journal of American Statistical Association, 91*, 999–1006.

Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. (2006). *Measurement error in nonlinear models: A modern perspective* (2nd ed.). London: Chapman and Hall.

Delaigle, A., Meister, A. (2007). Nonparametric regression estimation in the heteroscedastic errors-in-variables problem. *Journal of American Statistical Association, 102*, 1416–1426.

Delaigle, A., Hall, P., Qiu, P. (2006). Nonparametric methods for solving the Berkson errors-in-variables problem. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68*, 201–220.

Franks, P. W., Ekelund, U., Brage, S., Wong, M., Wareham, N. J. (2004). Does the association of habitual physical activity with the metabolic syndrome differ by level of cardiorespiratory fitness? *Diabetes Care, 27*(5), 1187–1193.

Fuller, W. A. (1987) *Measurement error models*. New York: Wiley.

Jennrich, R. I. (1969). Asymptotic properties of nonlinear least squares estimators. *The Annals of Mathematical Statistics, 40*, 633–643.

Kiechl, S., Schett, G., Wenning, G., Redlich, K., Oberhollenzer, M., Mayr, A., Santer, P., Smolen, J., Poewe, W., Willeit, J. (2004). Osteoprotegerin is a risk factor for progressive atherosclerosis and cardiovascular disease. *Circulation, 109*, 2175–2180.

Kukush, A., Schneeweiss, H. (2005). Comparing different estimators in a nonlinear measurement error model. *Mathematical Methods of Statistics, 14*(1), 53–79.

Li, T. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis, 65*, 139–165.

Li, T., Hsiao, C. (2004). Robust estimation of generalized linear models with measurement errors. *Journal of Econometrics, 118*, 51–65.

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

Nakamura, T. (1990). Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika, 77*(1), 127–137.

Rabe-Hesketh, S., Pickles, A., Skrondal, A. (2003). Correcting for covariate measurement error in logistic regression using nonparametric maximum likelihood estimation. *Satatistical Modeling, 3*, 215–232.

Robert, C. P., Cassella, G. (2004). *Monte Carlo statistical methods* (2nd ed.). New York: Springer.

Roy, S., Banerjee, T. (2006). A flexible model for generalized linear regression with measurement error. *Annals of the Institute of Statistical Mathematics, 58*, 153–169.

Schafer, D. W. (2001). Semiparametric maximum likelihood for measurement error model regression. *Biometrics, 57*(1), 53–61.

Schennach, M. S. (2004a). Estimation of nonlinear models with measurement error. *Econometrica, 72*(1), 33–75.

Schennach, M. S. (2004b). Nonparametric Estimation in the presence of measurement error. *Econometric Theory, 20*, 1046–1093.

Schennach, M. S. (2007). Instrumental variable estimation of nonlinear errors-in-variables models. *Econometrica, 75*, 201–239.

Schennach, M. S. (2008). Quantile regression with mismeasured covariates. *Econometric Theory, 24*, 1010–1043.

Stefanski, L. A., Carroll, R. J. (1985). Covariate measurement error in logistic regression. *The Annals of Statistics, 13*(4), 1335–1351.

Stefanski, L. A., Carroll, R. J. (1991). Deconvolution-based score tests in measurement error models. *The Annals of Statistics, 19*(1), 249–259.

Stimmer, K. (2003). Modeling gene expression measurement error: A quasi-likelihood approach. *BMC Bioinformatics, 4*, 1–10.

Taupin, M. (2001). Semi-parametric estimation in the nonlinear structural errors-in-variables model. *The Annals of Statistics, 29*(1), 66–93

Thoresen, M., Laake, P. (2003). The use of replicates in logistic measurement error modeling. *Scandinavian Journal of Statistics, 30*, 625–636.
van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge: Cambridge University Press.
Wang, L. (2003). Estimation of nonlinear Berkson-type measurement error models. *Statistica Sinica, 13*, 1201–1210
Wang, L. (2004). Estimation of nonlinear models with Berkson measurement errors. *The Annals of Statistics, 32*, 2559–2579.
Wang, L., Hsiao, C. (1995). A simulated semiparametric estimation of nonlinear errors-in-variables models. *Working Paper*. Department of Economics, University of Southern California, Los Angeles.
Wang, L., Hsiao, C. (2010). Method of moments estimation and identifiability of nonlinear semiparametric errors-in-variables models. *Journal of Econometrics* (accepted).