

Semiparametric quaslikelihood and variance function estimation in measurement error models*

J.H. Sepanski

University of Michigan, Ann Arbor, MI 48109, USA

R.J. Carroll

Texas A&M University, College Station, TX 77843, USA

We consider a quaslikelihood/variance function model when a predictor X is measured with error and a surrogate W is observed. When in addition to a primary data set containing (Y, W) a validation data set exists for which (X, W) is observed, we can (i) estimate the first and second moments of the response Y given W by kernel regression; (ii) use quaslikelihood and variance function techniques to estimate the regression parameters as well as variance structure parameters. The estimators are shown to be asymptotically normally distributed, with asymptotic variance depending on the size of the validation data set and not on the bandwidth used in the kernel estimates. A more refined analysis of the asymptotic covariance shows that the optimal bandwidth converges to zero at the rate $n^{-1/3}$.

1. Introduction

Let Y denote a scalar response and X denote a scalar predictor variable. Consider a heteroscedastic nonlinear regression model with

$$E(Y|X = x) = f_1(x, \beta) \quad \text{and} \quad \text{var}(Y|X = x) = \sigma^2 g(x, \beta, \theta), \quad (1.1)$$

where σ^2 is a scalar parameter and β and θ are column vectors. This model is widely applicable, including generalized linear models and many important

Correspondence to: R.J. Carroll, Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA.

*This paper is based on the first author's Ph.D. thesis at Texas A&M University. Her research and that of the second author was supported by a grant from the National Institutes of Health. The authors would like to express their sincere thanks to a referee for a careful reading of the manuscript and useful comments.

nonlinear regression models as special cases. The key feature of (1.1) is that specification of the mean and variance functions is sufficient to obtain estimates of the parameters (β, θ, σ) . There is no need to specify a likelihood function. The quasilikelihood (also known as generalized least squares) score for β is defined to be

$$\sum_{i=1}^n \{Y_i - f_1(X_i, \beta)\} f_{1\beta}(X_i, \beta) \{g(X_i, \beta, \theta)\}^{-1}, \quad (1.2)$$

where $f_{1\beta}$ is the derivative of f_1 with respect to β . Quasilikelihood or generalized least squares estimates of the unknown parameter β for given θ are obtained by setting (1.2) equal to zero. For a recent review of this problem, as well as a discussion of variance function parameter estimation, see Carroll and Ruppert (1988, chs. 2, 3).

In many circumstances, instead of observing X , a surrogate W is observed. As is well known, ignoring the measurement error and simply substituting W for X is inappropriate because doing so leads to inconsistent parameter estimates. Our interest lies in the estimation of the unknown parameters based on a sample in which we observe (Y, W) . Under the assumption that given X , W is independent of Y , the mean and variance functions of the observed data are defined by

$$\begin{aligned} E(Y | W = w) &= E\{f_1(X, \beta) | W = w\}, \\ E(Y^2 | W = w) &= E\{E(Y^2 | X) | W = w\}. \end{aligned} \quad (1.3)$$

Eq. (1.3) shows that the distribution of X given W plays an important role in the estimation of unknown parameters. If the distribution of X given W is known up to a finite set of parameters, then (1.3) gives parametric formulae for the mean and variance functions of the observed data, to which quasilikelihood and variance function estimation techniques can be applied.

The measurement error estimation problem has been addressed previously in this context. Carroll et al. (1984), Fuller (1987), and Schafer (1987) discussed full parametric estimation when a model for X given W is postulated. Stefanski (1985), Stefanski and Carroll (1985), Whittemore and Keller (1988), and Carroll and Stefanski (1990) considered various small error approximations to construct estimates adjusted for measurement error. Clark (1982), Rosner et al. (1989, 1990), Whittemore (1989), Gleser (1990), Pierce et al. (1990), and Carroll and Stefanski (1990) all consider the possibility of directly replacing X in (1.1) and (1.2) by a linear regression for $E(X | W)$.

In this paper, we consider estimation of the parameters in (1.1) in a semiparametric context, so that no assumptions need be made about the distribution of X given W . In some cases, an independent validation data set

containing (Y, X, W) or (X, W) is available; see Rosner et al. (1989) and Pepe and Fleming (1991) for examples. For this situation, it is possible to estimate the density of (X, W) and therefore the density of Y given W , an idea which was developed independently by Carroll and Wand (1991) and Pepe and Fleming (1991). For discrete W , Pepe and Fleming (1991) propose to estimate the probability density function $f(y|w)$ by its empirical distribution, which is constructed by replacing each observation X_i in the validation data with $W = w$ by a point mass $1/n_w^v$, where n_w^v is the number of observations in the validation data with $W = w$. They carry through this idea when X and W are discrete, but not otherwise. Carroll and Wand (1991) consider logistic regression and suggest using kernel regression to estimate $\Pr(Y = 1 | W = w)$ from the validation data. The use of kernel regression rather than direct estimation of the density of X given W greatly simplifies the calculations. In both the above two papers it has been shown that the semiparametric estimates are asymptotically normally distributed and that the efficiency of the semiparametric methods increases with the size of the validation data relative to the size of the primary data. As the validation sample size becomes large relative to the primary sample size, the semiparametric method becomes as efficient as in the cases where the density of X given W is known.

Neither Carroll and Wand (1991) nor Pepe and Fleming (1991) consider the general quasilikelihood/variance function model. In this paper, we will construct new estimates for this model, focusing on the continuous case. Our approach is as follows:

- Estimate $E(Y | W = w)$ and $\text{var}(Y | W = w)$ as functions of (β, θ, σ) and w by means of kernel regression applied to the validation data.
- Apply quasilikelihood and variance function estimation techniques to the primary data using the estimated mean and variance function.

In section 2, we outline the proposed methods for the case that θ is known, as well as for the case that θ is unknown and hence variance function estimation is required. In section 3, we state the main results on asymptotic normality of the estimates. We briefly discuss bandwidth selection in section 4. Simulation results are given in section 5. In the discussion, section 6, we note that our results apply almost without change to the case of likelihood estimation of parameters, and to the case that (X, W) is discrete. Proofs are contained in three appendices.

2. Description of methods

2.1. Introduction

Suppose that the structural variance parameter θ is known and the variance depends on the mean, i.e., x and β . This model is important because it includes as

special cases many GLIMs, for example, logistic regression, the Poisson model with $g(x, \beta, \theta) = f_1(x, \beta)$, and the Gamma model with $g(x, \beta, \theta) = f_1^2(x, \beta)$. When θ is unknown, overdispersion being the classical example, there are various ways to estimate θ [see Carroll and Ruppert (1988)]; this case is covered in section 2.3.

In what follows, assume that we have a validation data set containing (Y, X, W) or (X, W) of size n_v , and a primary data set containing (Y, W) of size n_p . Let $n = n_v + n_p$, $n_p/n_v = \lambda$. Derivatives are denoted by subscripts.

The algorithms studied require kernel regression estimation of functions of X regressed on W in the validation data. As discussed by Carroll and Wand (1991), the range of W in the observed validation data is usually smaller than that in the observed primary data. If kernel techniques were used blindly, this would lead to extrapolation, which is a dangerous business. To overcome this problem, as well as to avoid technical difficulties with edge effects, all kernel regression estimates used here are calculated on a compact set Ω interior to the support of W . Sums in the primary data are taken only for those $W \in \Omega$. While this truncation causes some loss in efficiency, it is counterbalanced by a gain in robustness. The truncation also leads to an easily described theory.

2.2. Variance parameters known

The simplest model to describe occurs when θ is known. Write $E(Y|X = x) = f_1(x, \beta)$ and $E(Y^2|X = x) = f_2(x, \beta)$. Let $u(w, \beta) = E(Y|W = w)$, $s(w, \beta) = E(Y^2|W = w)$, and $v(w, \beta) = \text{var}(Y|W = w)$. Using (1.3), we propose to use kernel regression [see Eubank (1988)] on the validation data to estimate $u(w, \beta)$, $u_\beta(w, \beta)$, and $s(w, \beta)$, and then to proceed with quasilikelihood estimation on the primary data using the resulting estimated quasilikelihood score.

Let K be a symmetric density function with finite support and h be a window or bandwidth. Let Ω be a compact subset of the support of W . For $w \in \Omega$, define

$$\begin{aligned} \hat{f}_w(w) &= \frac{1}{n_v h} \sum_{i=1}^{n_v} K\left(\frac{W_i - w}{h}\right), \\ k_n(w, \beta) &= \frac{1}{n_v h} \sum_{i=1}^{n_v} f_1(X_i, \beta) K\left(\frac{W_i - w}{h}\right), \\ k_{n\beta}(w, \beta) &= \frac{1}{n_v h} \sum_{i=1}^{n_v} f_{1\beta}(X_i, \beta) K\left(\frac{W_i - w}{h}\right), \\ b_n(w, \beta) &= \frac{1}{n_v h} \sum_{i=1}^{n_v} f_2(X_i, \beta) K\left(\frac{W_i - w}{h}\right). \end{aligned} \tag{2.1}$$

We estimate $u(w, \beta)$ by $u_n(w, \beta) = k_n(w, \beta) / \hat{f}_w(w)$, $u_\beta(w, \beta)$ by $u_{n\beta}(w, \beta) = k_{n\beta}(w, \beta) / \hat{f}_w(w)$, $s(w, \beta)$ by $s_n(w, \beta) = b_n(w, \beta) / \hat{f}_w(w)$, and $v(w, \beta)$ by $s_n(w, \beta) - u_n^2(w, \beta)$. The estimated quasilikelihood score for the primary data is

$$\sum_{j=n_v+1}^n \frac{Y_j - u_n(W_j, \beta)}{s_n(W_j, \beta) - u_n^2(W_j, \beta)} u_{n\beta}(W_j, \beta) = \sum_{j=n_v+1}^n \Psi_n(Y_j, W_j, \beta).$$

If Y is observed in the validation data, let $l(y, x, \beta) = \{ \{y - f_1(x, \beta)\} \times f_{1\beta}(x, \beta) \} / \{ f_2(x, \beta) - f_1^2(x, \beta) \}$, so that $\sum_{i=1}^{n_v} l(Y_i, X_i, \beta)$ is the quasilikelihood score for the validation data. Of course, one could use only the validation data to estimate the parameters, but this would ignore the information contained in the primary data. These considerations lead us to propose to estimate β by the solution $\hat{\beta}_1$ to the equation

$$0 = n^{-1/2} \sum_{i=1}^{n_v} l(Y_i, X_i, \beta) + n^{-1/2} \sum_{j=n_v+1}^n \Psi_n(Y_j, W_j, \beta). \tag{2.2}$$

When Y is not observed in the validation data set, we estimate β by the solution $\hat{\beta}_2$ to

$$0 = n_p^{-1/2} \sum_{j=n_v+1}^n \Psi_n(Y_j, W_j, \beta). \tag{2.3}$$

The asymptotic limit distribution of $\hat{\beta}_1$ will be presented in section 3. Our limit results are proved in the appendix not for the estimates which solve (2.2) and (2.3), but rather for one-step versions defined in the appendix which start from a root- n -consistent discretized estimate of β . By a discretized estimate we mean one which takes values in the set $\{0, \pm cn^{-1/2}, \pm 2cn^{-1/2}, \dots\}$, where c is an arbitrarily small constant.

2.3. Variance parameters unknown

Data in the economic, clinical, and biological sciences are often fitted by a heteroscedastic regression model (1.1) with unknown variance parameters σ^2 and θ which must be estimated. Examples of this situation are given in Carroll and Ruppert (1988). Following Davidian and Carroll (1987) and Carroll and Ruppert (1988), we propose to estimate the variance parameters σ^2 and θ by weighted least squares on squared residuals and then estimate the regression parameter β by quasilikelihood estimation. This method is essentially the same as pseudolikelihood estimation of variance parameters; see Carroll and Ruppert (1988, ch. 3).

The basic procedure is to first obtain a preliminary $n^{1/2}$ -consistent estimate of (β, θ) , perhaps from the validation data. In this case, the preliminary estimate $\hat{\beta}_*$ might be a robust estimate or the unweighted least squares estimate based the primary data, for example, $\hat{\beta}_*$ is such that $0 = n_p^{-1/2} \times \sum_{j=n_v+1}^n \{Y_j - f_1(W_j, \beta)\} f_{1\beta}(W_j, \beta)$. The preliminary estimate $\hat{\theta}_*$ minimizes $\sum_{j=n_v+1}^n \{r_j^2 - g(W_j, \hat{\beta}_*, \theta)\}^2$ where $r_j = Y_j - f_1(W_j, \hat{\beta}_*)$ and $\hat{\sigma}_*^2 = n_p^{-1} \times \sum_{j=n_v+1}^{n_v} \{Y_j - f_1(W_j, \hat{\beta}_*)\} / g(W_j, \hat{\beta}_*, \hat{\theta}_*)$. Both $\hat{\beta}_*$ and $\hat{\theta}_*$ are $n^{1/2}$ -consistent.

Having obtained preliminary estimates, residuals are formed and an updated estimate of θ is constructed using functions of the squared residuals. Finally, the resulting estimate of (β, θ) is used to form weights, and an updated weighted or quasilikelihood estimate of β is constructed.

In model (1.1), let $q(w, \beta, \theta) = E\{g(X, \beta, \theta) | W = w\}$, $t(w, \beta) = E\{f_1^2(X, \beta) | W = w\}$, and let u and u_β be defined as in section 2.1. Note that

$$\begin{aligned} v(w, \sigma^2, \beta, \theta) &= \text{var}(Y | W = w) \\ &= \sigma^2 E\{g(X, \beta, \theta) | W = w\} + \text{var}\{f_1(X, \beta) | W = w\}. \end{aligned}$$

Let K be a symmetric density function with finite support and h be a bandwidth. For $w \in \Omega$, define

$$\begin{aligned} r_n(w, \beta) &= \frac{1}{n_v h} \sum_{i=1}^{n_v} f_1^2(X_i, \beta) K\{W_i - w\}/h\}, \\ g_n(w, \beta, \theta) &= \frac{1}{n_v h} \sum_{i=1}^{n_v} g(X_i, \beta, \theta) K\{W_i - w\}/h\}, \end{aligned}$$

and define \hat{f}_W , u_n , and $u_{n\beta}$ as in section 2.1. Let

$$\begin{aligned} q_n(w, \beta, \theta) &= g_n(w, \beta, \theta) / \hat{f}_W(w), \\ t_n(w, \beta) &= r_n(w, \beta) / \hat{f}_W(w), \\ v_n(w, \sigma^2, \beta, \theta) &= \sigma^2 q_n(w, \beta, \theta) + t_n(w, \beta) - u_n^2(w, \beta). \end{aligned} \tag{2.4}$$

We estimate $t(w, \beta)$, $q(w, \beta, \theta)$, and $\text{var}(Y | W = w)$ by $t_n(w, \beta)$, $q_n(w, \beta, \theta)$, and $v_n(w, \sigma^2, \beta, \theta)$ respectively. Let $\hat{\sigma}_*^2$, $\hat{\theta}_*$, and $\hat{\beta}_*$ be preliminary $n^{1/2}$ -consistent estimates for σ^2 , θ , and β respectively, for example, the variance function estimates based on the (Y_i, X_i) in the validation data [see Davidian and Carroll (1987) and Carroll and Ruppert (1988)].

First, we obtain $\hat{\sigma}^2$ and $\hat{\theta}$ such that

$$\begin{aligned}
 0 = & n^{-1/2} \sum_{i=1}^{n_v} \frac{M(X_i, \hat{\sigma}^2, \hat{\beta}_*, \hat{\theta})}{g^2(X_i, \hat{\beta}_*, \hat{\theta}_*)} \left[\{Y_i - f_1(X_i, \hat{\beta}_*)\}^2 - \hat{\sigma}^2 g(X_i, \beta_*, \hat{\theta}) \right] \\
 & + n^{-1/2} \sum_{j=n_v+1}^n \frac{H_n(W_j, \hat{\sigma}^2, \hat{\beta}_*, \hat{\theta})}{v_n^2(W_j, \hat{\sigma}_*^2, \hat{\beta}_*, \hat{\theta}_*)} \\
 & \times [\{Y_j - u_n(W_j, \hat{\beta}_*)\}^2 - v_n(W_j, \hat{\sigma}^2, \hat{\beta}_*, \hat{\theta})], \tag{2.5}
 \end{aligned}$$

where

$$M^T(x, \sigma^2, \beta, \theta) = (\sigma^2 g_{\theta}(x, \beta, \theta), g(x, \beta, \theta)),$$

$$H_n^T(w, \sigma^2, \beta, \theta) = (\sigma^2 q_{n\theta}(w, \beta, \theta), q_n(w, \beta, \theta)),$$

i.e., M and H are the derivatives of g and v_n with respect to $(\theta, \sigma^2)^T$ respectively. This is a generalization of pseudolikelihood estimation appropriate for the measurement error model. When X is measured exactly, i.e., $X = W$, then since $n_p = 0$, solving (2.5) for θ is equivalent to pseudolikelihood estimation of θ .

We can now form a weighted estimate of β , estimating it by the solution $\hat{\beta}_3$ to the equation

$$\begin{aligned}
 0 = & n^{-1/2} \sum_{i=1}^{n_v} \frac{Y_i - f_1(X_i, \beta)}{g(X_i, \beta, \hat{\theta})} f_{1\beta}(X_i, \beta) \\
 & + n^{-1/2} \sum_{j=n_v+1}^n \frac{Y_j - u_n(W_j, \beta)}{v_n(w_j, \hat{\sigma}^2, \beta, \hat{\theta})} u_{n\beta}(W_j, \beta). \tag{2.6}
 \end{aligned}$$

If Y is not observed in the validation data, then we estimate $(\sigma^2, \theta, \beta)$ by the estimating equations (2.5) and (2.6) without the first terms and with $n^{-1/2}$ replaced by $n_p^{-1/2}$. That is, we estimate σ^2 and θ by the solution to the equation

$$\begin{aligned}
 0 = & n_p^{-1/2} \sum_{i=1}^{n_v} \frac{H_n(W_j, \sigma^2, \theta, \hat{\beta}_*)}{v_n^2(W_j, \hat{\sigma}_*^2, \hat{\beta}_*, \hat{\theta}_*)} \\
 & \times [\{Y_j - u_n(W_j, \hat{\beta}_*)\}^2 - v_n(W_j, \sigma^2, \theta, \hat{\beta}_*)]. \tag{2.7}
 \end{aligned}$$

We denote these estimates by $\hat{\sigma}_1^2$ and $\hat{\theta}_1$. Then, estimate β by the solution $\hat{\beta}_4$ to the equation

$$0 = n_p^{-1/2} \sum_{j=n_v+1}^n \{Y_j - u_n(W_j, \beta)\} u_{n\beta}(W_j, \beta) \{v_n(W_j, \hat{\sigma}_1^2, \hat{\theta}_1, \beta)\}^{-1} \tag{2.8}$$

As stated previously, our theoretical results apply not to the solutions to (2.6)–(2.8), but rather to one-step versions of them.

3. Main results

In this section, we state the asymptotic results for the estimators defined in section 2. The proof of Theorem 1 and the calculations for Corollaries 2 and 3 are given in appendices A, B, and C respectively. The remaining results can be obtained by similar arguments.

3.1. Asymptotic results with Y observed in validation

In this subsection, we will state the main asymptotic results for the estimates defined when Y is observed in the validation data, specifically (2.2), (2.5), and (2.6).

Let A^{-T} denote $(A^{-1})^T$. Let f and f_W denote the joint and marginal densities of (X, W) and W , respectively. Let Ω be a proper compact subset of the support of W and J be an open interval containing Ω . We first obtain the limit distribution for the one-step version of (2.2), the estimate is appropriate for the case that θ is known.

Theorem 1. Assume that

$\Gamma_1(\beta)$ is nonsingular,

$$f, f_W, u, u_\beta, u_{\beta\beta}, s, s_\beta \text{ are continuous on } J \text{ and} \\ \text{are in the class } C^2[\Omega], \tag{3.1}$$

$$\inf_{w \in J} f_W(w) > 0 \quad \text{and} \quad \inf_{w \in J} v(w, \beta) > 0, \tag{3.2}$$

$$f_1, f_{1\beta}, f_{1\beta\beta}, f_2, f_{2\beta} \in L^p, \tag{3.3}$$

then

$$\sup_w E(|f_i(X, \beta)|^p | W = w) f_W(w) < \infty,$$

$$\sup_w E(|f_{i\beta}(X, \beta)|^p | W = w) f_W(w) < \infty \quad \text{for } i = 1, 2,$$

$$\sup_w E(|f_{i\beta\beta}(X, \beta)|^p | W = w) f_W(w) < \infty \quad \text{for some } p > 2.$$

Suppose $n^{2\eta-1}h \rightarrow \infty$ for some $\eta < 1 - 1/p$, $\{n^{1/2}h/\log(h)\}^{-1} \rightarrow 0$, $nh^5/\log(h^{-1}) \rightarrow 0$, and $nh^4 \rightarrow 0$. Then, with ‘ \Rightarrow ’ denoting convergence in distribution,

$$n^{1/2}(\hat{\beta}_1 - \beta) \Rightarrow N\{0, \Gamma_1^{-1}(\beta) Y_1(\beta) \Gamma_1^{-T}(\beta)\},$$

where

$$\Gamma_1(\beta) = \frac{1}{1 + \lambda} E[l_\beta(Y, X, \beta) + \lambda \Psi_\beta(Y, W, \beta)],$$

$$\Psi(y, w, \beta) = \frac{y - u(w, \beta)}{v(w, \beta)} u_\beta(w, \beta),$$

$$Y_1(\beta) = \frac{1}{1 + \lambda} E[l(Y, X, \beta)l^T(Y, X, \beta) + \lambda \Psi(Y, W, \beta) \Psi^T(Y, W, \beta) + \lambda^2 \Delta_1(\beta)], \tag{3.4}$$

$$\Delta_1(\beta) = \text{cov} \left[\{(f_1(X, \beta) - u(W, \beta))\} \frac{u_\beta(W, \beta)}{v(W, \beta)} \right]. \tag{3.5}$$

Similar to the results of Carroll and Wand (1991) and Pepe and Flemming (1991), in the case of generalized linear models the first term in (3.4) is the Fisher information from the validation sample and the second term is the Fisher information from the primary sample. The third term is the cost due to nonparametric regression estimation of u and v .

Note that the condition $nh^4 \rightarrow 0$ forces p to be greater than 8/3. Note also that the asymptotic covariance in (3.4) involves neither the kernel density K nor the

bandwidth h . As $\lambda \rightarrow 0$, the estimate is also as efficient asymptotically as in the case where the density of X given W is completely known.

In the next result, we present the limit distribution of estimates of (θ, σ) when Y is observed in the validation data.

Theorem 2. In addition to the conditions in Theorem 1, assume that (3.1) holds for q and t and (3.3) holds for g and f_1^2 . Let $(\hat{\sigma}_^2, \hat{\theta}_*, \hat{\beta}_*)$ be $n^{1/2}$ -consistent estimates of $(\sigma^2, \theta, \beta)$. Then*

$$n^{1/2} \begin{pmatrix} \hat{\theta} - \theta \\ \hat{\sigma}^2 - \sigma^2 \end{pmatrix} = B_1 A + B_1^{-1} B_3 n^{1/2} (\hat{\beta}_* - \beta) + O_P(1), \tag{3.6}$$

where

$$\begin{aligned} A = & n^{-1/2} \sum_{i=1}^{n_v} \frac{M(X_i, \sigma^2, \beta, \theta)}{g^2(X_i, \beta, \theta)} \left[\{(Y_i - f_1(X_i, \beta))^2 - \sigma^2 g(X_i, \beta, \theta)\} \right. \\ & + n^{-1/2} \sum_{j=n_v+1}^n \frac{H(W_j, \sigma^2, \beta, \theta)}{v^2(W_j, \sigma^2, \beta, \theta)} [\{Y_j - u(W_j, \beta)\}^2 - v(W_j, \sigma^2, \beta, \theta)] \\ & - \frac{\lambda}{n^{1/2}} \sum_{i=1}^{n_v} \left[\frac{H(W_i, \sigma^2, \beta, \theta)}{v^2(W_i, \sigma^2, \beta, \theta)} \left\{ \sigma^2 \{g(X_i, \beta, \theta) - q(W_i, \beta, \theta)\} \right. \right. \\ & \left. \left. + f_1^2(X_i, \beta) - t(W_i, \beta) + 2u(W_i, \beta) \{u(W_i, \beta) - f_1(X_i, \beta)\} \right\} \right], \end{aligned}$$

$$\begin{aligned} B_1 = & \frac{1}{1 + \lambda} E \left[\frac{M(X, \sigma^2, \beta, \theta) M^T(X, \sigma^2, \beta, \theta)}{g^2(X, \beta, \theta)} \right. \\ & \left. + \lambda \frac{H(W, \sigma^2, \beta, \theta) H^T(W, \sigma^2, \beta, \theta)}{v^2(W, \sigma^2, \beta, \theta)} \right], \end{aligned}$$

$$\begin{aligned} B_3 = & \frac{1}{1 + \lambda} E \left[\frac{M(X, \sigma^2, \beta, \theta) g_\beta^T(X, \beta, \theta)}{g^2(X, \beta, \theta)} \right. \\ & \left. + \lambda \frac{H(W, \sigma^2, \beta, \theta) v_\beta^T(W, \beta, \theta)}{v^2(W, \sigma^2, \beta, \theta)} \right]. \end{aligned}$$

Asymptotic normality follows directly from (3.6). If the variance function does not depend on β , then the asymptotic distribution of $\hat{\theta}$ and $\hat{\sigma}^2$ does not depend on the method used to obtain the preliminary estimate $\hat{\beta}_*$. In fact, it is easy to

check that $B_3 = 0$ in this case. Also note that the effects of the preliminary estimates $\hat{\theta}_*$ and $\hat{\sigma}_*^2$ are asymptotically negligible. The asymptotic variance of the third term in A is the dispersion induced by the nonparametric estimation of $f_w, u, g,$ and t .

We now turn to the quasilikelihood estimate of β when θ is estimated and Y is observed in the validation data.

Theorem 3. In addition to the assumptions in Theorem 2, we assume that $\Gamma_3(\beta)$ is nonsingular. Then

$$n^{1/2}(\hat{\beta}_3 - \beta) \Rightarrow N\{0, \Gamma_3^{-1}(\beta) Y_3(\beta) \Gamma_3^{-T}(\beta)\},$$

where

$$\Gamma_3(\beta) = \frac{1}{1 + \lambda} E[l_\beta(Y, X, \beta, \theta) + \lambda \Psi_\beta(Y, W, \sigma^2, \beta, \theta)],$$

$$\Psi(y, w, \sigma^2, \beta, \theta) = \frac{y - u(w, \beta)}{v(w, \sigma^2, \beta, \theta)} u_\beta(w, \beta),$$

$$Y_3(\beta) = \frac{1}{1 + \lambda} E[l(Y, X, \beta, \theta)l(Y, X, \beta, \theta) + \lambda \Psi(Y, W, \sigma^2, \beta, \theta) \Psi^T(Y, W, \sigma^2, \beta, \theta) + \lambda^2 \Delta_3(\beta)],$$

$$\Delta_3(\beta) = \text{cov} \left[\{f_1(X, \beta) - u(W, \beta)\} \frac{u_\beta(W, \beta)}{v(W, \sigma^2, \beta, \theta)} \right].$$

3.2. Y not observed in validation

Corollaries 1–3 state the limit results for (2.3), (2.7), and (2.8), and hence are the analogues to Theorems 1–3 for the case that Y is not observed in the validation data. Since Y is not observed in the validation data, the changes in the corollaries are the absence of $l_\beta(Y, X, \beta)$ in the asymptotic variances and the constant $\lambda/(1 + \lambda)$ due to a different normalization.

Corollary 1. In addition to the conditions in Theorem 1, we assume that $\Gamma_2(\beta)$ is nonsingular. Then

$$n_p^{1/2}(\hat{\beta}_2 - \beta) \Rightarrow N\{0, \Gamma_2^{-1}(\beta) Y_2(\beta) \Gamma_2^{-T}(\beta)\},$$

where

$$\Gamma_2(\beta) = E[\Psi_\beta(Y, W, \beta)], \quad \Psi(y, w, \beta) = \frac{y - u(w, \beta)}{v(w, \beta)} u_\beta(w, \beta),$$

$$Y_2(\beta) = E[\Psi(Y, W, \beta) \Psi^T(Y, W, \beta) + \lambda \Delta_2(\beta)],$$

$$\Delta_2(\beta) = \text{cov} \left[\{(f_1(X, \beta) - u(W, \beta)) \frac{u_\beta(W, \beta)}{v(W, \beta)} \right].$$

Corollary 2. Under the conditions of Theorem 2,

$$n_p^{1/2} \begin{pmatrix} \hat{\theta}_1 - \theta \\ \hat{\sigma}_1^2 - \sigma^2 \end{pmatrix} = C_1 A_* + C_1^{-1} C_3 n_p^{1/2} (\hat{\beta}_* - \beta) + o_p(1),$$

where

$$\begin{aligned} A_* &= n_p^{-1/2} \sum_{j=n_v+1}^n \frac{H(W_j, \sigma^2, \beta, \theta)}{v^2(W_j, \sigma^2, \beta, \theta)} (\{Y_j - u(W_j, \beta)\}^2 - v(W_j, \sigma^2, \beta, \theta)) \\ &\quad - \frac{\lambda}{n_p^{1/2}} \sum_{i=1}^{n_v} \left[\frac{H(W_j, \sigma^2, \beta, \theta)}{v^2(W_j, \sigma^2, \beta, \theta)} \left\{ \sigma^2 \{g(X_i, \beta, \theta) - q(W_i, \beta, \theta)\} \right. \right. \\ &\quad \left. \left. + f_1^2(X_i, \beta) - t(W_i, \beta) + 2u(W_i, \beta) \{u(W_i, \beta) - f_1(X_i, \beta)\} \right\} \right], \\ C_1 &= E \left[\frac{H(W, \sigma^2, \beta, \theta) H^T(W, \sigma^2, \beta, \theta)}{v^2(W, \sigma^2, \beta, \theta)} \right], \\ C_3 &= E \left[\frac{H(W, \sigma^2, \beta, \theta) v_\beta(W, \beta, \theta)}{v^2(W, \sigma^2, \beta, \theta)} \right]. \end{aligned}$$

Corollary 3. In addition to the assumption of Theorem 2, we assume that $\Gamma_4(\beta)$ is nonsingular. Then

$$n_p^{1/2} (\hat{\beta}_4 - \beta) \Rightarrow N \left\{ 0, \Gamma_4^{-1}(\beta) \gamma_4(\beta) \Gamma_4^{-T}(\beta) \right\},$$

where

$$\begin{aligned} \Gamma_4(\beta) &= E[\Psi_\beta(Y, W, \sigma^2, \beta, \theta)], \\ \Psi(y, w, \sigma^2, \beta, \theta) &= \frac{y - u(w, \beta)}{v(w, \sigma^2, \beta, \theta)} u_\beta(w, \beta), \\ \gamma_4(\beta) &= E \left[\Psi_\beta(Y, W, \sigma^2, \beta, \theta) \Psi_\beta^T(Y, W, \sigma^2, \beta, \theta) + \lambda \Delta_4(\beta) \right], \\ \Delta_4(\beta) &= \text{cov} \left[\{f_1(X, \beta) - u(W, \beta)\} \frac{u_\beta(W, \beta)}{v(W, \sigma^2, \beta, \theta)} \right]. \end{aligned}$$

3.3. Covariance estimation

Since X is not observed in the primary data, we estimate the third term based on the validation data,

$$\hat{\Delta}_1(\hat{\beta}_1) = \frac{1}{n_v} \sum_{i=1}^{n_v} \left\{ \frac{f(X_i, \hat{\beta}_1) - u_n(W_i, \hat{\beta}_1)}{v_n(W_i, \hat{\beta}_1)} \right\}^2 u_{n\beta}(W_i, \hat{\beta}_1) u_{n\beta}^T(W_i, \hat{\beta}_1).$$

Define

$$S_v = \frac{1}{n} \sum_{i=1}^{n_v} \frac{f_{\hat{\beta}}(X_i, \hat{\beta}_1) f_{\hat{\beta}}^T(X_i, \hat{\beta}_1)}{g(X_i, \hat{\beta}_1)},$$

$$S_p = \frac{1}{n} \sum_{j=n_v+1}^n \frac{u_{n\beta}(X_i, \hat{\beta}_1) u_{n\beta}(X_i, \hat{\beta}_1)}{v_n(X_i, \hat{\beta}_1)}.$$

Estimate $\Gamma_1(\beta)$ and $\Upsilon_1(\beta)$ by $S_v + S_p$ and $S_v + S_p + \lambda^2 \hat{\Delta}_1(\hat{\beta}_1)/(1 + \lambda)$ respectively. Since $\hat{\beta}_1$ is consistent, S_v , S_p , and $\hat{\Delta}_1(\hat{\beta}_1)$ are consistent estimators for $\Gamma_1(\beta)$, $\gamma_1(\beta)$, and $\Delta_1(\beta)$ respectively. Similarly, we can easily obtain consistent estimators for the rest of the covariances.

4. Bandwidth

Note that the classical optimal h with rate $n^{-1/5}$ is not allowed under the condition $nh^4 \rightarrow 0$. In order to find the appropriate rate for h and perhaps later to obtain an estimate of it, we generalize the bandwidth selection method in Carroll and Wand (1991). We focus on the estimate $\hat{\beta}_1$, although the same considerations should apply to the other estimates. The method is based on a higher-order expansion of the covariance of an asymptotically equivalent form of $n^{1/2}(\hat{\beta}_1 - \beta)$. Let $\eta_n = (n^{1/2}/h)^{-1} + n^{1/2}/h^2$. By a second-order Taylor series, it can be shown that $n^{1/2}(\hat{\beta}_1 - \beta) = V_n + O_p(\eta_n)$, with

$$E(V_n V_n^T) = \Gamma_1^{-1}(\beta) \gamma_1(\beta) \Gamma_1^{-T}(\beta) + a(n, h) a^T(n, h) + O(1/nh), \tag{4.1}$$

where

$$A_1 = \frac{\lambda}{2(1 + \lambda)} \int z^2 K(z) dz E \left[\frac{u_{\beta}(W, \beta)}{v(W, \beta) f_W(W)} \right. \\ \left. \times \{2u'(W, \beta) f'_W(W) + u''(W, \beta) f_W(W)\} \right],$$

$$A_2 = \lambda \int K^2(z) dz E \left[\frac{u_{\beta}(W, \beta)}{v^2(W, \beta) f^2(W)} \right. \\ \times \{4u^3(W, \beta) - 2u(W, \beta) s(W, \beta) - 4u(W, \beta) f_1^2(X, \beta) \\ + 2f_1(X, \beta) f_2(X, \beta) + 2v(W, \beta) u(W, \beta) \\ \left. - \frac{2v(W, \beta)}{u_{\beta}(W, \beta)} f_1(X, \beta) f_{1\beta}(X, \beta)\} \right],$$

$$a(n, h) = \Gamma_1(\beta)^{-1} \{ (n^{1/2}h)^{-1} A_2 - n^{1/2}h^2 A_1 \},$$

$$u' = \partial u / \partial w, \quad f' = \partial f / \partial w, \quad u'' = \partial^2 u / \partial^2 w.$$

A technical report containing the somewhat complicated calculations for (4.1) is available upon request. In the generalized linear model, the mean $u(w, \beta) = u(w^T \beta)$, and it can be shown that $A_1 = A_2 = a(n, h) = 0$ when $\beta = 0$. It also can be shown that the asymptotic variance of $\hat{\beta}_1$ is minimized at $h = \infty$ when $\beta = 0$. In terms of rates of convergence for estimating the linear combination $\xi^T \beta$, the bandwidth h is chosen such that $\{\xi^T a(n, h)\}^2$ is minimized. Except for the case that $u(w, \beta) = u(w^T \beta)$ and $\beta = 0$ mentioned above, the optimal bandwidth is $h_0 = \gamma_0 n^{-1/3}$, for some γ_0 . Note that the rate $n^{-1/3}$ satisfies the conditions in Theorem 1.

The plug-in method suggests estimating γ_0 by replacing the unknown quantities in γ_0 by estimates. If there exists a discretized n^a -estimator $\hat{\gamma}_0$ of the unknown γ_0 for some $0 < a \leq \frac{1}{2}$, then the limit result in Theorem 1 still holds for the estimated bandwidth $\hat{h}_0 = \hat{\gamma}_0 n^{-1/3}$. Let $\hat{\beta}(h)$ denote our parameter estimator based upon a bandwidth h . It is easy to show that

$$n^{1/2} \{ \hat{\beta}(\hat{h}_0) - \hat{\beta}(h) \} \xrightarrow{P} 0,$$

where h satisfies the conditions in Theorem 1, if we discretize $\hat{\gamma}_0$ as we discretized $\hat{\beta}_0$ in the proof of Theorem 1. From (A.12) to (A.15), it is sufficient to show that for $\gamma_n = \gamma_0 + t_n/n^a$, where $t_n \rightarrow t_0$ is a fixed deterministic sequence and $h_n = \gamma_n n^{-1/3}$, and that

$$n^{1/2} \{ A_n(\beta, h_n) - A_n(\beta, h) \} \xrightarrow{P} 0. \quad (4.2)$$

Since the conditions in Theorem 1 hold for h_n , we can apply Lemma 1 and Lemma 2 to $A_n(\beta, h_n)$ and $A_n(\beta, h)$ to show (4.2).

5. Simulation

The validation data containing (X, W) were generated such that $X = 0.5W + \delta|W|^d U$, where (W, U) has a bivariate normal distribution with mean $(4, 0)$ and covariances $\sigma_{WU} = 0$ and $\xi \sigma_W^2 = \sigma_U^2 = 1$. The primary data containing (Y, W) were generated according to the Michaelis–Menton model, i.e., such that $Y = 1/(\beta_0 + \beta_1/X) + \varepsilon$, where ε has the standard normal distribution. Two hundred simulated data sets were run for each of the three validation data sizes 50, 100, and 200 and for each of the three primary data sizes 100 and 300.

As in Carroll and Wand (1991), we used an *ad hoc* bandwidth $\hat{\sigma}_W n^{-1/3}$ which is easily programmed, where $\hat{\sigma}_W$ is an estimate of the standard deviation of W based on the validation data. A Gaussian kernel and a grid of 21 points covering the range of the validation data were used. Observations W outside the range of the validation data in the primary data should not be used in the estimating equations. To avoid boundary effects, those data points in the

Table 1

Michaelis–Menton model: $Y = (\alpha + \beta/X)^{-1} + e$, where $e \sim N(0, 1)$ and $(\alpha, \beta) = (0.2, 0.3)^a$.

$X = W + 0.25U$, where $W \sim N(4, 1)$ and $U \sim N(0, 1)$						
	$n_c = 50$		$n_c = 100$		$n_c = 200$	
	<i>Param</i>	<i>Semi</i>	<i>Param</i>	<i>Semi</i>	<i>Param</i>	<i>Semi</i>
	$n_p = 100$					
<i>Mean</i>	0.320	0.318	0.305	0.311	0.290	0.285
<i>Median</i>	0.308	0.306	0.300	0.305	0.287	0.279
<i>SD</i>	0.124	0.156	0.118	0.133	0.106	0.118
<i>RMSE</i>	0.125	0.157	0.118	0.133	0.106	0.119
<i>MAE</i>	0.070	0.097	0.073	0.089	0.068	0.079
<i>RE(%)</i>	138.3	100.0	121.7	100.0	115.4	100.0
	$n_p = 300$					
<i>Mean</i>	0.309	0.305	0.307	0.307	0.303	0.298
<i>Median</i>	0.304	0.306	0.309	0.309	0.302	0.294
<i>SD</i>	0.072	0.091	0.071	0.077	0.066	0.073
<i>RMSE</i>	0.073	0.091	0.071	0.077	0.066	0.073
<i>MAE</i>	0.048	0.053	0.045	0.052	0.045	0.048
<i>RE(%)</i>	114.4	100.0	115.1	100.0	106.5	100.0
	$X = W + 0.5U$, where $W \sim N(4, 1)$ and $U \sim N(0, 1)$					
	$n_c = 50$		$n_c = 100$		$n_c = 200$	
	<i>Param</i>	<i>Semi</i>	<i>Param</i>	<i>Semi</i>	<i>Param</i>	<i>Semi</i>
	$n_p = 100$					
<i>Mean</i>	0.330	0.312	0.316	0.310	0.301	0.282
<i>Median</i>	0.317	0.290	0.308	0.305	0.295	0.272
<i>SD</i>	0.129	0.162	0.125	0.139	0.111	0.124
<i>RMSE</i>	0.132	0.162	0.126	0.139	0.111	0.125
<i>MAE</i>	0.073	0.096	0.079	0.093	0.067	0.077
<i>RE(%)</i>	130.4	100.0	117.6	100.0	113.6	100.0
	$n_p = 300$					
<i>Mean</i>	0.323	0.299	0.317	0.301	0.314	0.297
<i>Median</i>	0.320	0.296	0.317	0.303	0.312	0.294
<i>SD</i>	0.082	0.098	0.079	0.081	0.070	0.077
<i>RMSE</i>	0.085	0.098	0.080	0.081	0.070	0.077
<i>MAE</i>	0.056	0.060	0.052	0.053	0.047	0.052
<i>RE(%)</i>	107.5	100.0	101.9	100.0	112.1	100.0

^a*Param* is the parametric correction for attenuation, while *Semi* is the semiparametric quasilikelihood.

Table 2

$X = W + 0.25 W ^{0.5}U$, where $W \sim N(4, 1)$ and $U \sim N(0, 1)$						
	$n_v = 50$		$n_v = 100$		$n_v = 200$	
	Param	Semi	Param	Semi	Param	Semi
	$n_p = 100$					
Mean	0.323	0.312	0.309	0.309	0.293	0.282
Median	0.307	0.296	0.301	0.306	0.291	0.275
SD	0.126	0.159	0.123	0.137	0.108	0.120
RMSE	0.128	0.159	0.123	0.137	0.108	0.121
MAE	0.073	0.096	0.073	0.091	0.006	0.077
RE(%)	130.6	100.0	125.7	100.0	115.9	100.0
	$n_p = 300$					
Mean	0.313	0.298	0.310	0.301	0.306	0.296
Median	0.308	0.295	0.311	0.303	0.304	0.291
SD	0.077	0.095	0.075	0.079	0.068	0.074
RMSE	0.078	0.095	0.076	0.079	0.068	0.074
MAE	0.049	0.058	0.047	0.051	0.045	0.050
RE(%)	118.0	100.0	107.8	100.0	110.8	100.0
	$X = W + 0.5 W ^{0.5}U$, where $W \sim N(4, 1)$ and $U \sim N(0, 1)$					
	$n_v = 50$		$n_v = 100$		$n_v = 200$	
	Param	Semi	Param	Semi	Param	Semi
	$n_p = 100$					
Mean	0.352	0.290	0.335	0.299	0.316	0.275
Median	0.336	0.266	0.327	0.294	0.315	0.272
SD	0.150	0.181	0.145	0.153	0.122	0.132
RMSE	0.159	0.181	0.149	0.152	0.123	0.134
MAE	0.096	0.105	0.087	0.110	0.076	0.084
RE(%)	108.7	100.0	127.0	100.0	110.0	100.0
	$n_p = 300$					
Mean	0.345	0.276	0.335	0.282	0.332	0.290
Median	0.340	0.271	0.330	0.280	0.329	0.280
SD	0.106	0.113	0.096	0.092	0.078	0.088
RMSE	0.115	0.116	0.102	0.094	0.084	0.088
MAE	0.068	0.081	0.063	0.061	0.058	0.059
RE(%)	118.5	100.0	98.1	100.0	102.6	100.0

primary data which are within one bandwidth of the minimum and maximum values of the validation data were also deleted.

The parametric regression calibration estimate (*Param*) was obtained as described in Rosner et al. (1989). First compute the estimate $\hat{E}(X|W)$ of $E(X|W)$ by using simple linear regression, and then perform the usual analysis. The proposed semiparametric estimate (*Semi*) was computed by a full iteration starting from the naive estimate, which was obtained by ignoring the measurement error. Tables 1 and 2 give the mean, median, standard deviation (*SD*), root of the mean squared error (*MSE*), median absolute error (*MAE*), and the relative efficiency (*RE*) in terms of the median absolute error with respect to the proposed estimates.

For both the linear model $X = W + \delta U$ and the linear heteroscedastic model $X = W + \delta|W|^{0.5}U$, for $\delta = 0.25, 0.5$, the semiparametric regression calibration estimates were somewhat less efficient than the regression calibration estimates, but not disastrously so.

Next, we study the model $W = XU$, where X has uniform distribution on $(0, 5)$ and U has a gamma distribution with mean 1 and variance 1. The results are given in table 3. In terms of bias and variability, the proposed estimate clearly beats the parametric regression calibration estimate. A plot of X against W for a single data set, given in fig. 1, suggests that even though the relationship

Table 3

$W = XU$, where $X \sim \text{Uniform}(0, 5)$ and $U \sim \text{Gamma}(1, 1)$.^a

	$n_c = 50$		$n_c = 100$		$n_c = 200$	
	<i>Param</i>	<i>Semi</i>	<i>Param</i>	<i>Semi</i>	<i>Param</i>	<i>Semi</i>
	$n_p = 100$					
<i>Mean</i>	0.595	0.142	0.560	0.171	0.550	0.220
<i>Median</i>	0.568	0.152	0.552	0.160	0.548	0.197
<i>SD</i>	0.212	0.176	0.164	0.179	0.155	0.164
<i>RMSE</i>	0.363	0.235	0.307	0.220	0.293	0.182
<i>MAE</i>	0.268	0.154	0.252	0.154	0.248	0.133
<i>RE</i> (%)	21.3	100.0	61.1	100.0	53.5	100.0
	$n_p = 300$					
<i>Mean</i>	0.558	0.140	0.580	0.184	0.565	0.226
<i>Median</i>	0.533	0.138	0.568	0.182	0.555	0.211
<i>SD</i>	0.168	0.101	0.138	0.100	0.096	0.101
<i>RMSE</i>	0.307	0.189	0.312	0.153	0.281	0.125
<i>MAE</i>	0.233	0.162	0.268	0.125	0.255	0.096
<i>RE</i> (%)	69.3	100.0	46.7	100.0	37.6	100.0

^a Based on 200 simulations in each case. Gaussian kernel with bandwidth $\hat{\sigma}/n^{1/3}$, where $\hat{\sigma}$ is an estimate of the standard deviation of W based on the validation data set.

between X and W seems to be linear, the parametric regression calibration estimate does not perform well.

One should not draw general conclusions from such a small study. We have observed in our example a small loss of efficiency for the semiparametric estimate when X given W is normally distributed with a linear mean, and a large gain in efficiency in a case where X given W does not have a linear mean.

6. Discussion

Due to the availability of the validation data the semiparametric quasilikelihood estimation proposed here does not require choice of a model for the measurement error. Being based on variance function and quasilikelihood techniques, it also makes no assumption about the underlying distribution of Y given X and instead depends only on the first and second moments of Y given X .

Our results also apply to likelihood methods. Suppose that the likelihood function for Y given X is denoted by $L_X(Y, X, \beta)$. Our results apply by fitting a nonparametric regression of $L_X(Y, X, \beta)$ on W to obtain an estimate $L_W(Y, W, \beta)$, the likelihood for Y given W . That is, we estimate $L_W(Y = y, W = w, \beta)$ by

$$L_{w,n}(Y = y, W = w, \beta) = \frac{1}{n_v h} \sum_{i=1}^{n_v} L_X(y, X_i, \beta) K\left(\frac{W_i - w}{h}\right) / \hat{f}_w(w), \quad (6.1)$$

where \hat{f}_w is defined in (2.1). Then an estimate $\hat{\beta}_p$ of β is obtained by maximizing $\prod_{j=n_v+1}^n L_{w,n}(Y_j, W_j, \beta)$, this usually being done by the method of scoring.

The semiparametric likelihood method in (6.1) generalizes the logistic model in Carroll and Wand (1991) to a wider class of probability densities. If both $L_W(y, w, \beta)$ and $L_X(y, x, \beta)$ are in the class of generalized linear models, for example the logistic model of Carroll and Wand, it can be shown that the asymptotic distribution theory of $\hat{\beta}_p$ defined by (6.1) is equivalent to that of $\hat{\beta}_2$. Otherwise analogous formulae are obtained; we will forego the details.

Up until now it has been assumed that the joint and marginal densities of X and W are continuous. For a discrete surrogate W , by choosing the density kernel $K(0) = 1$, eq. (6.1) yields the empirical estimate of $P(Y | W)$ proposed by Pepe and Fleming. In this case, we need not worry about bandwidth selection. Let n_w^v denote the number of observations in the validation data with $W = w$, n_w^p denote the number of observations in the primary data with $W = w$, and $n_w^v + n_w^p = n_w$.

We assume that $\text{plim } n_p/n_v = \lambda$ in this paper, and if the validation data are a selected subset of the primary data, it follows that $\text{plim } n_w^p/n_w^v = \lambda$. Also, note that n_w/n is the empirical density estimate of $P(W = w)$. It then can be shown that, in the generalized linear model, the term in the asymptotic covariance caused by the empirical estimation is equal to $\Delta_1(\beta)$; see (3.5).

By the same token, we can define $u_n(w, \beta) = \sum_1^{n_v} f_1(X_i, \beta) K(W_i - w)/n_w^v$, $u_{n\beta}(w, \beta) = \sum_1^{n_v} f_1(X_i, \beta) K(W_i - w)/n_w^v$, and $s_n(w, \beta) = \sum_1^{n_v} f_2(X_i, \beta) K(W_i - w)/n_w^v$.

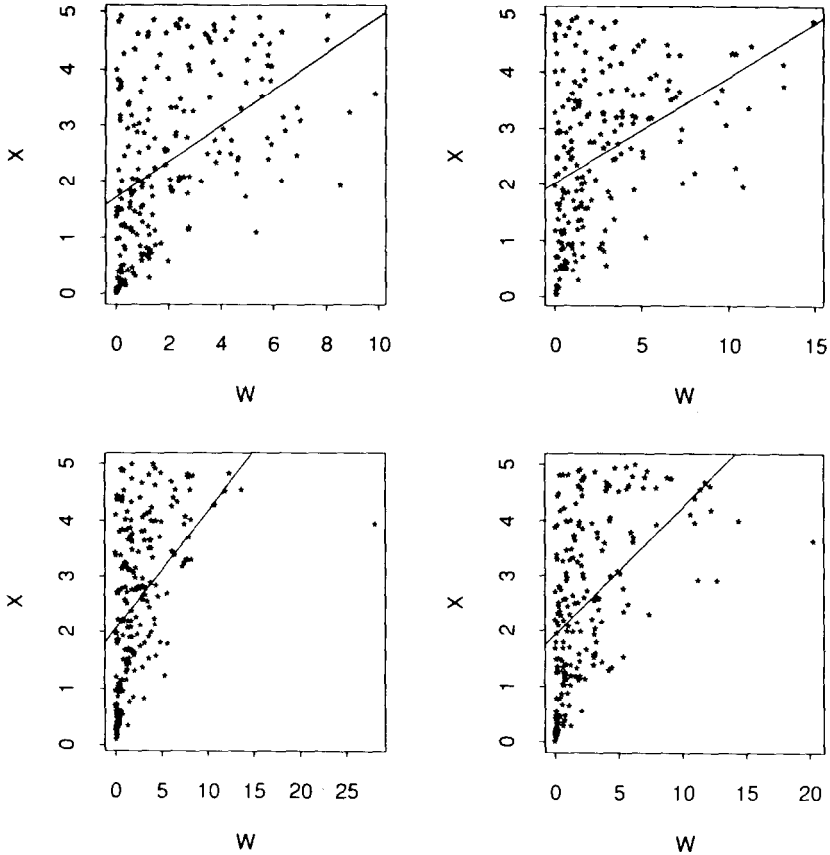


Fig. 1. Multiplicative model $W = XU$, where $X \sim \text{Uniform}(0, 5)$ and $U \sim \text{Gamma}(1, 1)$.

where $K(0) = 1$ for discrete surrogate W . The estimates u_n , $u_{n\beta}$, and s are the unbiased empirical estimates of u , u_β , and s respectively. Therefore, the method proposed here also extends to the discrete case. Only the proof, which we will not pursue here, will be different.

Appendix A

Appendix A gives a formal proof for Theorem 1; appendices B and C give the calculations for Theorem 2 and Corollary 3.

Let $K(i, j)$ denote $K(W_i - W_j)/h$, $u(i)$ denote $u(W_i, \beta)$, and similarly for the rest. Let derivatives be denoted by subscripts. Let $E(U^2)$ denote $E(UU^T)$. The following is proved under the assumptions of Theorem 1. Also note that all the

sums and expectations are taken over Ω , since the estimates in (2.1) are only defined for $w \in \Omega$. We will not indicate this restriction explicitly in the calculations.

Lemma A [Mack and Silverman (1982)]. Let $\gamma_n = ((1/nh)\log(1/h))^{1/2}$. Assume that

- f, f_w , and $u \in C^2[\Omega]$, and $\inf_{w \in \Omega} f_w(w) > 0$.
- $f_1 \in L^p$ and $\sup_w E(|f_1(X, \beta)|^p | W = w) f_w(w) < \infty$ for some $p > 2$.

Suppose $n^{2\eta-1} h \rightarrow \infty$ for some $\eta < 1 - 1/p$ and $h^2 = o(\gamma_n)$, then

$$\gamma_n^{-1} \sup_{w \in \Omega} |\hat{f}_w(w) - f_w(w)| = O_p(1),$$

$$\gamma_n^{-1} \sup_{w \in \Omega} |k_n(w) - u(w)f_w(w)| = O_p(1),$$

$$\gamma_n^{-1} \sup_{w \in \Omega} |u_n(w, \beta) - u(w, B)| = O_p(1).$$

Under mild conditions on the bandwidth, this lemma shows that the weak uniform convergence rate of the kernel estimator on Ω is $O_p(\gamma_n)$. The result can be applied to the rest of the kernel estimators we defined in section 2. Also note that $h^2 = o(\gamma_n)$ is equivalent to $nh^5/\log h^{-1} \rightarrow 0$.

We first require two technical lemmas:

Lemma 1

$$\begin{aligned} & \frac{1}{n^{1/2}} \sum_{j=n_v+1}^n \frac{Y_j - u_n(W_j, \beta)}{v_n(W_j, \beta)} u_{n\beta}(W_j, \beta) \\ &= \frac{1}{n^{1/2}} \sum_{j=n_v+1}^n \frac{Y_j - u(W_j, \beta)}{v(W_j, \beta)} u_\beta(W_j, \beta) \\ & \quad + \frac{1}{n^{1/2}} \sum_{j=n_v+1}^n \left[\frac{u(j)u_\beta(j)}{v(j)f_w(j)} \hat{f}_w(j) - \frac{u_\beta(j)}{v(j)f_w(j)} k_n(j) \right] + o_p(1), \end{aligned}$$

where $\lambda = n_p/n_v$.

Proof. From (2.1),

$$\frac{1}{n^{1/2}} \sum_{j=n_v+1}^n \frac{Y_j - u_n(j)}{v_n(j)} u_{n\beta}(j) = \frac{1}{n^{1/2}} \sum_{j=n_v+1}^n \frac{Y_j \hat{f}_W(j) - k_n(j)}{b_n(j) \hat{f}_W(j) - k_n^2(j)} k_{n\beta}(j).$$

By a Taylor series,

$$\begin{aligned} &= \frac{1}{n^{1/2}} \sum_{j=n_v+1}^n \left[\frac{Y_j - u(j)}{v(j)} u_{\beta}(j) + \frac{Y_j u_{\beta*}(j)}{v_*(j) f_{W*}(j)} \{ \hat{f}_W(j) - f_W(j) \} \right. \\ &\quad - \frac{u_{\beta*}(j)}{v_*(j) f_{W*}(j)} \{ k_n(j) - u(j) f_W(j) \} \\ &\quad + \frac{Y_j - u_*(j)}{v_*(j) f_{W*}(j)} \{ k_{n\beta}(j) - u_{\beta}(j) f_W(j) \} \\ &\quad - \frac{\{ Y_j - u_*(j) \} u_{\beta*}(j)}{v_*^2(j) f_{W*}(j)} s_*(j) \{ \hat{f}_W(j) - f_W(j) \} \\ &\quad - \frac{\{ Y_j - u_*(j) \} u_{\beta*}(j)}{v_*^2(j) f_{W*}(j)} \{ b_n(j) - s(j) f_W(j) \} \\ &\quad \left. + \frac{\{ Y_j - u_*(j) \} u_{\beta*}(j)}{v_*^2(j) f_{W*}(j)} 2u_*(j) \{ k_n(j) - u(j) f_W(j) \} \right] \\ &= U_{1n} + U_{2n}^* - U_{3n}^* + U_{4n}^* - U_{5n}^* - U_{6n}^* + U_{7n}^*, \end{aligned} \tag{A.1}$$

where $\{ f_{W*}(w), u_*(w), u_{\beta*}(w), v_*(w) \}$ lies in between $\{ \hat{f}_W(w), u_n(w, \beta), u_{n\beta}(w, \beta), v_n(w, \beta) \}$ and $\{ f_W(w), u(w, \beta), u_{\beta}(w, \beta), v(w, \beta) \}$ for each $w \in \Omega$. Note that, for example,

$$\begin{aligned} &\frac{1}{n^{1/2}} \left\| \sum_{j=n_v+1}^n \left\{ \frac{Y_j - u_*(w)}{v_*(j) f_{W*}(j)} - \frac{Y_j - u(w)}{v(j) f_W(j)} \right\} \{ k_{n\beta}(j) - u_{\beta}(j) f_W(j) \} \right\| \\ &\leq \frac{1}{n^{1/2}} \sum_{j=n_v+1}^n \| Y_j \| \sup_{w \in \Omega} \left\| \frac{1}{v_*(w) f_{W*}(w)} - \frac{1}{v(w) f_W(w)} \right\| \\ &\quad \times \sup_{w \in \Omega} \| k_{n\beta}(w) - u_{\beta}(w) f_W(w) \| \\ &\quad + n^{1/2} \frac{\lambda}{1 + \lambda} \sup_{w \in \Omega} \left\| \frac{u_*(w)}{v_*(w) f_{W*}(w)} - \frac{u(w)}{v(w) f_W(w)} \right\| \\ &\quad \times \sup_{w \in \Omega} \| k_{n\beta}(w) - u_{\beta}(w) f_W(w) \|. \end{aligned} \tag{A.2}$$

By Lemma A, (3.2), and $E(\|Y\|) < \infty$ following from (3.3), we can show that (A.2) is of order $O_p((1/n^{1/2}h) \log(1/h)) = o_p(1)$ as $nh^2/(\log h^{-1})^2 \rightarrow \infty$. Therefore, we can rewrite (A.1) as

$$\begin{aligned}
 &= \frac{1}{n^{1/2}} \sum_{j=n_v+1}^n \left[\frac{Y_j - u(j)}{v(j)} u_\beta(j) + \frac{Y_j u_\beta(j)}{v(j) f_W(j)} \{ \hat{f}_W(j) - f_W(j) \} \right. \\
 &\quad - \frac{u_\beta(j)}{v(j) f_W(j)} \{ k_n(j) - u(j) f_W(j) \} \\
 &\quad + \frac{Y_j - u(j)}{v(j) f_W(j)} \{ k_{n\beta}(j) - u_\beta(j) f_W(j) \} \\
 &\quad - \frac{\{ Y_j - u(j) \} u_\beta(j)}{v(j)^2 f_W(j)} s(j) \{ \hat{f}_W(j) - f_W(j) \} \\
 &\quad - \frac{\{ Y_j - u(j) \} u_\beta(j)}{v^2(j) f_W(j)} \{ b_n(j) - s(j) f_W(j) \} \\
 &\quad \left. + \frac{\{ Y_j - u(j) \} u_\beta(j)}{v^2(j) f_W(j)} 2u(j) \{ k_n(j) - u(j) f_W(j) \} \right] + o_p(1) \\
 &= U_{1n} + U_{2n} - U_{3n} + U_{4n} - U_{5n} - U_{6n} + U_{7n} + o_p(1). \tag{A.3}
 \end{aligned}$$

We first want to show that $U_{4n} = o_p(1)$, $U_{5n} = o_p(1)$, $U_{6n} = o_p(1)$, and $U_{7n} = o_p(1)$. We only show that $U_{4n} = o_p(1)$, because the rest can be done by similar arguments. Let $\mathcal{G}_p = \{(W_j)_{j=n_v+1}^n\}$ and $\mathcal{G}_v = \{(X_i, W_i)_{i=1}^{n_v}\}$. It is easy to see that $E(U_{4n} | \mathcal{G}_p, \mathcal{G}_v) = 0$; therefore

$$\begin{aligned}
 \text{cov}(U_{4n}) &= E \left[\text{cov} \left\{ \frac{1}{n^{1/2}} \sum_{j=n_v+1}^n \frac{Y_j - u(j)}{v(j) f_W(j)} \right. \right. \\
 &\quad \left. \left. \times \{ k_{n\beta}(j) - u_\beta(j) f_W(j) \} \mid \mathcal{G}_p, \mathcal{G}_v \right\} \right] \\
 &= E \left[\frac{1}{n} \sum_{j=n_v+1}^n \frac{1}{v(j) f_W^2(j)} \{ k_{n\beta}(j) - u_\beta(j) f_W(j) \} \right. \\
 &\quad \left. \times \{ k_{n\beta}(j) - u_\beta(j) f_W(j) \}^T \right].
 \end{aligned}$$

We want to show that this last term converges to zero. Let $k_{n\beta}^{(a)}$ denote the a th coordinate of the column vector $k_{n\beta}$. For $w \in \Omega$,

$$\begin{aligned} E[k_{n\beta}^{(a)}(w)] &= E\left[\frac{1}{nh} \sum_{i=1}^{n_v} f_{1\beta}^{(a)}(i) K\left(\frac{W_i - w}{h}\right)\right] \\ &= \int f_{1\beta}^{(a)}(x) K\left(\frac{u - w}{h}\right) \frac{1}{h} f_w(x, u) dx du \\ &= \int E\{f_{1\beta}^{(a)}(X) | W = u\} K\left(\frac{u - w}{h}\right) \frac{1}{h} f_w(u) du \\ &= \int u_{\beta}^{(a)}(w + zh) f_w(w + zh) K(z) dz. \end{aligned}$$

After Taylor expansions of u and f_w and by the definition of the kernel K , we obtain

$$\begin{aligned} &= u_{\beta}^{(a)}(w) f_w(w) + h^2 \int z^2 K(z) dz u_{\beta}^{(a)'}(w') f'_w(w^*) \\ &= u_{\beta}^{(a)}(w) f_w(w) + h^2 \zeta_a(w), \end{aligned} \tag{A.4}$$

where $w^*, w' \in [w - h, w + h]$ and f'_w and u' are the first derivatives of f_w and u with respect to w respectively. Note that under assumption (3.1), $\sup_{w \in \Omega} |\zeta_a(w)| < \infty$. Similarly, we have

$$\begin{aligned} &E[\{k_{n\beta}^{(a)}(w) k_{n\beta}^{(b)}(w)\}] \\ &= E\left[\frac{1}{n_v^2 h^2} \sum_{i,k=1}^{n_v} f_{1\beta}^{(a)}(X_i) f_{1\beta}^{(b)}(X_k) K\left(\frac{W_i - w}{h}\right) K\left(\frac{W_k - w}{h}\right)\right] \\ &= \frac{1}{n_v h} \int \frac{1}{h} f_{1\beta}^{(a)}(x) f_{1\beta}^{(b)}(x) K^2\left(\frac{u - w}{h}\right) f_w(x, u) dx du \\ &\quad + \frac{n_v(n_v - 1)}{n_v^2} \left[u_{\beta}^{(a)}(w) u_{\beta}^{(b)}(w) f_w^2(w) + h^2 \{u_{\beta}^{(a)}(w) f_w(w) \zeta_b(w) \right. \\ &\quad \left. + u_{\beta}^{(b)}(w) f_w(w) \zeta_a(w) + h^2 \zeta_a(w) \zeta_b(w) \right] \\ &= u_{\beta}^{(a)}(w) u_{\beta}^{(b)}(w) f_w^2(w) + \zeta(w, h), \end{aligned} \tag{A.5}$$

where ζ_b is analogous to ζ_a in (A.4) for $k_{n\beta}^{(a)}$. Note that $\zeta(w, h) \rightarrow 0$ and $\sup_{w \in \Omega} |\zeta(w, h)| < \infty$ under assumption (3.1). Let $R^{(a,b)}$ be the (a, b) entry of $\text{cov}(U_{4n})$. Then by (A.4) and (A.5),

$$R^{(a,b)} = E \left[\frac{1}{n} \sum_{j=n_v+1}^n \frac{1}{v(j)f_W^2(j)} \right. \\ \left. \times \{ \zeta(W_j) - u_\beta^{(b)} \zeta_a(W_j) f_W(j) - u_\beta^{(a)}(W_j) \zeta_b(W_j) f_W(j) \} \right].$$

Since the integrand converges to zero and is bounded, by the bounded convergence theorem $R^{(a,b)} = o(1)$. Thus, $U_{4n} = o_P(1)$.

Note that by arguments similar to the above, it can be shown that

$$\frac{1}{n^{1/2}} \sum_{j=n_v+1}^n \frac{Y_j - u(j)}{v(j)f_W(j)} u_\beta(j) \{ \hat{f}_W(j) - f_W(j) \} = o_P(1). \tag{A.6}$$

Therefore,

$$U_{2n} = \frac{1}{n^{1/2}} \sum_{j=n_v+1}^n \frac{u(j)u_\beta(j)}{v(j)f_W(j)} \{ \hat{f}_W(j) - f_W(j) \} + o_P(1) \\ = U_{2n*} + o_P(1).$$

Lemma 2

$$U_{2n*} - U_{3n} = \frac{\lambda}{n^{1/2}} \sum_{i=1}^{n_v} \left[\frac{u(W_i, \beta)}{v(W_i, \beta)} - \frac{f_1(X_i, \beta)}{v(W_i, \beta)} \right] u_\beta(W_i, \beta) + o_P(1).$$

Proof. Define $\alpha_n = E(U_{2n*} - U_{3n} | \mathcal{G}_n)$; then α_n is given by

$$= E \left[\frac{1}{n^{1/2}} \sum_{j=n_v+1}^n \left\{ \frac{u(j)u_\beta(j)}{v(j)f_W(j)} \left\{ \frac{1}{n_v h} \sum_{i=1}^{n_v} K(i, j) \right\} \right. \right. \\ \left. \left. - \frac{u_\beta(j)}{v(j)f_W(j)} \left\{ \frac{1}{n_v h} \sum_{i=1}^{n_v} f_1(i)K(i, j) \right\} \right\} \middle| \mathcal{G}_v \right]$$

$$\begin{aligned}
 &= \frac{n_p}{n_v n^{1/2}} \sum_{i=1}^{n_v} \left[\int \frac{u(w)u_\beta(w)}{v(w)} \frac{1}{h} K\left(\frac{W_i - w}{h}\right) dw \right. \\
 &\quad \left. - f_1(i) \int \frac{u_\beta(w)}{v(w)} \frac{1}{h} K\left(\frac{W_i - w}{h}\right) dw \right] \\
 &= \frac{\lambda}{n^{1/2}} \sum_{i=1}^{n_v} \left[\int \frac{u(W_i - zh)u_\beta(W_i - zh)}{v(W_i - zh)} K(z) dz \right. \\
 &\quad \left. - f_1(i) \int \frac{u_\beta(W_i - zh)}{v(W_i - zh)} K(z) dz \right]. \tag{A.7}
 \end{aligned}$$

After Taylor series expansions, (A.7) is given by

$$\begin{aligned}
 &\frac{\lambda}{n^{1/2}} \sum_{i=1}^{n_v} \left[\int \frac{\left\{ u(i) - zhu'(W_i) + \frac{z^2 h^2}{2} u''(W_i^*) \right\} \left\{ u_\beta(i) - zhu'_\beta(W_i) + \frac{z^2 h^2}{2} u''_\beta(W_i^*) \right\}}{v(i) - zhu'(W_i) + \frac{z^2 h^2}{2} v''(W_i)} \right. \\
 &\quad \left. \times K(z) dz - f_1(i) \int \frac{\left\{ u_\beta(i) - zhu'_\beta(W_i) + \frac{z^2 h^2}{2} u''_\beta(W_i^*) \right\}}{v(i) - zhv''(W_i) + \frac{z^2 h^2}{2} v''(W_i)} K(z) dz \right] \\
 &= \frac{\lambda}{n^{1/2}} \sum_{i=1}^{n_v} \left\{ \frac{u(i)}{v(i)} - \frac{f_1(i)}{v(i)} \right\} u_\beta(i) + \frac{\lambda}{n^{1/2}} \sum_{i=1}^{n_v} h^2 \gamma(X_i, W_i),
 \end{aligned}$$

where $\gamma(X_i, W_i)$ is given by

$$\begin{aligned}
 &\frac{1}{v(W_i)} \int z^2 K(z) dz \left[u'(W)u'_\beta(W_i) + \frac{1}{2} \{ u(W_i)u''_\beta(W_i^*) \right. \\
 &\quad \left. + u_\beta(W_i)u''(W_i^*) + f_1(X_i)u''_\beta(W_i^*) \} \right] + o_p(h^2),
 \end{aligned}$$

and W^\bullet and W° are in some neighborhood of W . Since $E\{\gamma(X, W)|W\}$ and $E\{\gamma(X, W)\gamma^T(X, W)|W\}$ are bounded under assumptions in Theorem 1, as $nh^4 \rightarrow 0$, $\text{cov}\{n^{-1/2}h^2\lambda\sum_{i=1}^{n_v}\gamma(X_i, W_i)\} = o(1)$. Thus we have

$$\alpha_n = \frac{1}{n^{1/2}} \sum_{i=1}^{n_v} \left[\frac{u(W_i, \beta)}{v(W_i, \beta)} - \frac{f_1(X_i, \beta)}{v(W_i, \beta)} \right] u_\beta(W_i, \beta) + o_P(1). \tag{A.8}$$

Next, we will show that $U_{2n^\bullet} - U_{3n} = \alpha_n + o_P(1)$ by a covariance calculation. We have

$$\begin{aligned} E(U_{2n^\bullet} | \mathcal{G}_v) &= \frac{1}{n^{1/2}} \sum_{i=1}^{n_v} \frac{1}{n_v} \sum_{j=n_v+1}^n \left[\int \frac{u(w)u_\beta(w)}{v(w)} \frac{1}{h} K\left(\frac{W_i - w}{h}\right) dw \right] \\ &= \frac{1}{n^{1/2}} \sum_{i=1}^{n_v} \frac{1}{n_v} \sum_{j=n_v+1}^n \left[\int \frac{u(W_i - zh)u_\beta(W_i - zh)}{v(W_i - zh)} K(z) dz \right]. \end{aligned} \tag{A.9}$$

$$\begin{aligned} E(U_{2n^\bullet} U_{2n^\bullet}^T | \mathcal{G}_v) &= \frac{1}{nn_v^2} \sum_{i,k=1}^{n_v} \sum_{p,q=n_v+1}^n E \left[\frac{u(p)u(q)u_\beta(p)u_\beta(q)^T}{v(p)v(q)f_W(p)f_W(q)} \right. \\ &\quad \left. \times \left\{ \frac{1}{h^2} K(i, p) K(k, q) \right\} \middle| \mathcal{G}_v \right]. \end{aligned}$$

Now, $\text{cov}(U_{2n^\bullet} | \mathcal{G}_v) = E(U_{2n^\bullet} U_{2n^\bullet}^T | \mathcal{G}_v) - E(U_{2n^\bullet} | \mathcal{G}_v)E(U_{2n^\bullet}^T | \mathcal{G}_v)$. By independence, the cases when $p \neq q$ contribute 0 to $\text{cov}(U_{2n^\bullet} | \mathcal{G}_v)$. From (A.9) and by arguments similar to (A.7), we have

$$\begin{aligned} E[\{\text{cov}(U_{2n^\bullet} | \mathcal{G}_v)\}] &= \frac{n_p}{nn_v^2} \sum_{i,k=1}^{n_v} E \left[E \left\{ \frac{u^2(p)u_\beta(p)u_\beta(p)^T}{v(p)^2 f_W(p)^2} \right. \right. \\ &\quad \left. \left. \times \left(\frac{1}{h^2} K(i, p) K(k, p) \right) \middle| W_p \right\} \right. \\ &\quad \left. - \frac{u(i)u(k)u_\beta(i)u_\beta^T(k)}{v(i)v(k)} \right] + o(1). \end{aligned} \tag{A.10}$$

The cases $i = k$, by the bounded convergence theorem, contribute $o(1)$. The cases $i \neq k$, (A.10):

$$\begin{aligned} &= \frac{n_v(n_v - 1)n_p}{nn_v^2} \left[E \left\{ \frac{u^2(p)u_\beta(p)u_\beta(p)^T}{v(p)^2 f_w(p)^2} \left(\int K(z) f_w(W_p + zh) dz \right)^2 \right\} \right. \\ &\quad \left. - E^2 \left\{ \frac{u(W)u_\beta(W)}{v(W)} \right\} \right] + o(1) \\ &= \frac{\lambda}{1 + \lambda} \text{cov} \left[\frac{u(W)u_\beta(W)}{v(W)} \right] + o(1). \end{aligned} \tag{A.11}$$

Similarly, we can easily show that $E\{\text{cov}(U_{3n}|\mathcal{G}_v)\}$ and $E\{\text{cov}(U_{2n*}, U_{3n}|\mathcal{G}_v)\}$ are equal to $\{\lambda/(1 + \lambda)\} \text{cov}\{u(W)u_\beta(W)/v(W)\}$. Therefore,

$$\text{cov}(U_{2n*} - U_{3n} - \alpha_n) = E\{\text{cov}(U_{2n*} - U_{3n}|\mathcal{G}_v)\} = o(1).$$

By (A.8), we have proved Lemma 2.

Proof of Theorem 1

Assume that we have an $n^{1/2}$ -consistent discretized estimate $\hat{\beta}_0$ of β , for example, $\hat{\beta}_0$ is a discretized version of the solution to $n^{-1/2} \sum_{i=1}^{n_v} l(Y_i, X_i, \beta) = 0$. Define $A_n(\beta) = n^{-1} \{ \sum_{i=1}^{n_v} l(Y_i, X_i, \beta) + \sum_{j=n_v+1}^{n_v} \Psi_n(Y_j, W_j, \beta) \}$. The one-step estimate $\hat{\beta}_1$ of β is

$$\hat{\beta}_1 = \hat{\beta}_0 - \left\{ \frac{\partial}{\partial \beta} A_n(\hat{\beta}_0) \right\}^{-1} A_n(\hat{\beta}_0).$$

By a Taylor series,

$$\begin{aligned} n^{1/2}\{\hat{\beta}_1 - \beta\} &= \left[\left\{ \frac{\partial}{\partial \beta} A_n(\beta_*) \right\}^{-1} - \left\{ \frac{\partial}{\partial \beta} A_n(\hat{\beta}_0) \right\}^{-1} \right] \\ &\quad \times \frac{\partial}{\partial \beta} A_n(\beta_*) n^{1/2}(\hat{\beta}_0 - \beta) - \left\{ \frac{\partial}{\partial \beta} A_n(\beta) \right\}^{-1} n^{1/2} A_n(\beta) \\ &\quad + \left[\left\{ \frac{\partial}{\partial \beta} A_n(\beta) \right\}^{-1} - \left\{ \frac{\partial}{\partial \beta} A_n(\hat{\beta}_0) \right\}^{-1} \right] n^{1/2} A_n(\beta), \end{aligned} \tag{A.12}$$

where β_* lies in between $\hat{\beta}_0$ and β . By the central limit theorem, $n^{-1/2} \times \sum_{i=1}^{n_v} l(Y_i, X_i, \beta) = O_P(1)$, and by Lemmas 1 and 2, $n^{-1/2} \sum_{j=n_v+1}^n \Psi_n(Y_j, W_j, \beta) = O_P(1)$, as $nh^2/(\log h^{-1})^2 \rightarrow \infty$ and $nh^5/\log h^{-1} \rightarrow 0$. Therefore, $n^{1/2} A_n(\beta) = O_P(1)$.

Define $A(\beta) = n^{-1} \{ \sum_{i=1}^{n_v} l(Y_i, X_i, \beta) + \sum_{j=n_v+1}^n \Psi(Y_j, W_j, \beta) \}$. First, we want to show that

$$\frac{\partial}{\partial \beta} A_n(\beta) - \frac{\partial}{\partial \beta} A(\beta) = o_P(1) \quad (\text{A.13})$$

as $nh^2/(\log h^{-1})^2 \rightarrow \infty$ and $nh^5/\log h^{-1} \rightarrow 0$, i.e., show that

$$\frac{1}{n} \sum_{j=n_v+1}^n \Psi_{n\beta}(Y_j, W_j, \beta) - \frac{1}{n} \sum_{j=n_v+1}^n \Psi_\beta(Y_j, W_j, \beta) = o_P(1).$$

The left-hand side of eq. (A.13) is

$$\begin{aligned} & \frac{1}{n} \sum_{j=n_v+1}^n \left[\frac{u_\beta(j)}{v(j)} u_\beta^T(j) + \frac{Y_j - u(j)}{v^2(j)} u_\beta(j) v_\beta^T(j) - \frac{Y_j - u(j)}{v^2(j)} u_{\beta\beta}(j) \right. \\ & \quad - \frac{k_{n\beta}(j)}{d_n(j)} k_{n\beta}^T(j) - \frac{Y_j \hat{f}_W(j) - k_n(j)}{d_n^2(j)} k_{n\beta}(j) d_{n\beta}^T(j) \\ & \quad \left. - \frac{Y_j \hat{f}_W(j) - k_n(j)}{d_n(j)} k_{n\beta\beta}(j) \right], \end{aligned}$$

where $d_n(j) = \{b_n(j) \hat{f}_W(j) - k_n^2(j)\}$, $u_{\beta\beta} = \partial u_\beta / \partial \beta$, and $k_{n\beta\beta} = \partial k_{n\beta} / \partial \beta$. Note that, for example,

$$\begin{aligned} & \frac{1}{n} \left\| \sum_{i=1}^{n_v} \left[\frac{u_\beta(j)}{v(j)} u_\beta^T(j) - \frac{k_{n\beta}(j)}{d_n(j)} k_{n\beta}^T(j) \right] \right\| \\ & \leq \frac{1}{n} \sum_{j=n_v+1}^n \sup_{w \in \Omega} \left\| \frac{u_\beta(w)}{v(w)} u_\beta^T(w) - \frac{k_{n\beta}(w)}{d_n(w)} k_{n\beta}^T(w) \right\| = O_P(\gamma_n) = o_P(1). \end{aligned}$$

Thus,

$$\begin{aligned} & \frac{\partial}{\partial \beta} A_n(\beta) - \frac{\partial}{\partial \beta} A(\beta) = o_P(1), \\ & \frac{\partial}{\partial \beta} A_n(\beta) \xrightarrow{P} E \left\{ \frac{1}{1+\lambda} l_\beta(Y, X, \beta) + \frac{\lambda}{1+\lambda} \Psi_\beta(Y, W, \beta) \right\}. \end{aligned}$$

Next, we want to show that

$$\frac{\partial}{\partial \beta} A_n(\hat{\beta}_0) - \frac{\partial}{\partial \beta} A_n(\beta) = o_P(1). \tag{A.14}$$

Because $\hat{\beta}_0$ is discretized, it suffices to show if $c_n \rightarrow c_0$ is a fixed deterministic sequence and $\beta_n = \beta + c_n/n^{1/2}$, then

$$\frac{\partial}{\partial \beta} A_n(\beta_n) - \frac{\partial}{\partial \beta} A_n(\beta) = o_P(1).$$

By a Taylor series,

$$\frac{\partial}{\partial \beta} A_n(\beta_n) - \frac{\partial}{\partial \beta} A_n(\beta) = \frac{1}{n^{1/2}} \frac{\partial^2}{\partial^2 \beta} A_n(\beta_*) c_*$$

By arguments similar to those for (A.13) and (A.6), we have

$$\frac{1}{n} \sum_{j=n_v+1}^n \frac{\partial^2}{\partial^2 \beta} A_n(\beta_*) - \frac{1}{n} \sum_{j=n_v+1}^n \frac{\partial^2}{\partial^2 \beta} A(\beta_*) = o_P(1).$$

Since $\partial^2 A(\beta)/\partial \beta^2$ is continuous, it then follows that

$$\frac{\partial^2}{\partial^2 \beta} A_n(\beta_*) - \frac{\partial^2}{\partial^2 \beta} A(\beta) = o_P(1).$$

By (A.12), (A.13), and (A.14), we have

$$n^{1/2}(\hat{\beta}_1 - \beta) = - \{A(\beta)\}^{-1} n^{1/2} A_n(\beta) + o_P(1). \tag{A.15}$$

Now apply Lemma 1 and Lemma 2 to finish the proof of Theorem 1.

Appendix B: Proof of Theorem 2

We will prove this for the estimators in (2.7). From (2.4) and (2.7), we have

$$\begin{aligned} 0 &= \frac{1}{n_p^{1/2}} \sum_{j=n_v+1}^n \frac{Q_n(W_j, \hat{\sigma}^2, \hat{\beta}_*, \hat{\theta})}{d_n^2(W_j, \hat{\sigma}_*^2, \hat{\beta}_*, \hat{\theta}_*)} \\ &\times \left[\{ Y_j^2 \hat{J}_W - k_n(W_j, \hat{\beta}_*) \}^2 - d_n(W_j, \hat{\sigma}^2, \hat{\beta}_*, \hat{\theta}) \right], \end{aligned} \tag{B.1}$$

where

$$d_n(w, \sigma^2, \beta, \theta) = \sigma^2 g_n(w, \beta, \theta) \hat{f}_w(w) + r_n(w, \beta) \hat{f}_w(w) - k_n^2(w, \beta)$$

and

$$Q_n^T(w, \sigma^2, \beta, \theta) = \left(\frac{\partial d_n}{\partial \theta}, \frac{\partial d_n}{\partial \sigma^2} \right).$$

We will assume that the solution to (B.1) is $n^{1/2}$ -consistent. Alternatively, a one-step argument can be developed precisely as in appendix A. Drop the dependence on W , σ^2 , θ , and β . By a Taylor's expansion and by arguments similar to those of Theorem 1,

$$\begin{aligned} & \frac{1}{n_p^{1/2}} \sum_{j=n_v+1}^n \frac{Q_n(W_j, \hat{\sigma}^2, \hat{\beta}_*, \hat{\theta})}{d_n^2(W_j, \hat{\sigma}^2, \hat{\beta}_*, \hat{\theta}_*)} \\ & \times \left[\{ Y_j^2 \hat{f}_w - k_n(W_j, \hat{\beta}_*) \}^2 - d_n(W_j, \hat{\sigma}^2, \hat{\beta}_*, \hat{\theta}) \right] \\ & = \frac{1}{n_p^{1/2}} \sum_{j=n_v+1}^n \left[\frac{Q_n}{d_n^2} \{ (Y_j^2 \hat{f}_w - k_n)^2 - d_n \} \right. \\ & \quad + \left\{ \frac{\dot{Q}_n}{d_n^2} \{ (Y_j^2 \hat{f}_w - k_n)^2 - d_n \} - \frac{Q_n Q_n^T}{d_n^2} \right\} \left(\frac{\hat{\theta} - \theta}{\hat{\sigma}^2 - \sigma^2} \right) \\ & \quad - \left[\left(\frac{2Q_n d_{n\beta}}{d_n^3} - \frac{Q_{n\beta}}{d_n^2} \right) \{ (Y_j^2 \hat{f}_w - k_n)^2 - d_n \} \right. \\ & \quad \left. - \frac{Q_n}{d_n^2} \{ 2(Y_j - k_n) k_{n\beta} - d_{n\beta} \} \right] (\hat{\beta}_* - \beta) \\ & \quad - \left[\frac{2Q_n}{d_n^3} \{ (Y_j^2 \hat{f}_w - k_n)^2 - d_n \} d_{n\theta} (\hat{\theta}_* - \theta) \right. \\ & \quad \left. - \frac{2Q_n}{d_n^3} \{ (Y_j^2 \hat{f}_w - k_n)^2 \} - d_n \right] d_{n\sigma^2} (\hat{\sigma}_*^2 - \sigma^2) \Big] + o_p(1) \\ & = A_n - B_{1n} \left(\frac{\hat{\theta} - \theta}{\hat{\sigma}^2 - \sigma^2} \right) - (B_{2n} - B_{3n}) (\hat{\beta}_* - \beta) \\ & \quad - B_{4n} (\hat{\theta}_* - \theta) - B_{5n} (\hat{\sigma}_*^2 - \sigma^2) + o_p(1), \tag{B.2} \end{aligned}$$

where \hat{Q}_n is the derivative of Q_n with respect to $(\theta, \sigma^2)^T$. By a Taylor's expansion, (B.2), and arguments similar to those for (A.3),

$$n_p^{1/2} \begin{pmatrix} \hat{\theta} - \theta \\ \hat{\sigma}^2 - \sigma^2 \end{pmatrix} = B_{1n}^{-1} \{ A_n - (B_{2n} - B_{3n})(\hat{\beta}_* - \beta) - B_{4n}(\hat{\theta}_* - \theta) - B_{5n}(\hat{\sigma}_*^2 - \sigma^2) \} + o_p(1).$$

By Lemma A and arguments similar to those for (A.6), we have $B_{2n} \xrightarrow{P} 0$, $B_{4n} \xrightarrow{P} 0$, and $B_{5n} \xrightarrow{P} 0$. Also,

$$B_{1n} - \frac{1}{n_p} \sum_{j=n_v+1}^n \frac{H H^T}{v} = o_p(1), \quad B_{3n} - \frac{1}{n_p} \sum_{j=n_v+1}^n \frac{H}{v^2} v_\beta = o_p(1). \quad (B.3)$$

Since $\hat{\beta}_*$, $\hat{\theta}_*$, and $\hat{\sigma}_*^2$ are assumed to be $n^{1/2}$ -consistent, from (B.2) and (B.3), we have

$$n_p^{1/2} \begin{pmatrix} \hat{\theta} - \theta \\ \hat{\sigma}^2 - \sigma^2 \end{pmatrix} = B_1^{-1} A_n + B_1^{-1} B_3 n_p^{1/2} (\hat{\beta}_* - \beta) + o_p(1). \quad (B.4)$$

Apply a Taylor series expansion to A_n and use a similar argument to (A.3) and (A.6). We then have

$$\begin{aligned} A_n = & \frac{1}{n_p^{1/2}} \sum_{j=n_v+1}^n \left[\frac{H}{v^2} [\{ Y_j - u(j) \}^2 - v(j)] \right. \\ & + \frac{H}{v^2 f_w} \{ 2 \{ Y_j - u(j) \} Y_j - \sigma^2 q(j) - t(j) \} \{ \hat{f}_w(j) - f_w(j) \} \\ & + \frac{H}{v^2 f_w} \{ - 2 \{ Y_j - u(j) \} + 2u(j) \} \{ \hat{k}_n(j) - u(j) f_w(j) \} \\ & - \frac{H}{v^2 f_w} \sigma^2 \{ \hat{g}_n(j) - q(j) f_w(j) \} \\ & \left. - \frac{H}{v^2 f_w} \{ \hat{r}_n(j) - t(j) f_w(j) \} \right] + o_p(1) \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{n_p^{1/2}} \sum_{j=n_v+1}^n \frac{H}{v^2} \left[[\{Y_j - u(j)\}^2 - v(j)] \right. \\
 &\quad + [2\{Y_j - u(j)\}Y_j - \sigma^2 q(j) - t(j)] \hat{f}_W(j) \\
 &\quad \left. + 2u(j)\hat{k}_n(j) - \sigma^2 \hat{g}_n(j) - \hat{r}_n(j) \right] \\
 &= V_{1n} + V_{2n} + V_{3n} - V_{4n} - V_{5n} + o_P(1).
 \end{aligned}$$

Next, by calculations of covariances similar to the arguments in appendix A, we can show that

$$V_{5n} = \frac{\lambda}{n_p^{1/2}} \sum_{i=1}^{n_v} \frac{H(W_i, \sigma^2, \beta, \theta)}{v^2(W_i, \sigma^2, \beta, \theta)} f_1^2(X_i, \beta) + o_P(1). \tag{B.5}$$

Similarly, we can show that

$$\begin{aligned}
 V_{2n} &= \frac{\lambda}{n_p^{1/2}} \sum_{i=1}^{n_v} \frac{H(W_i, \sigma^2, \beta, \theta)}{v^2(W_i, \sigma^2, \beta, \theta)} \{ \sigma^2 q(W_i, \beta, \theta) + t(W_i, \beta) - 2u^2(W_i, \beta) \} \\
 &\quad + o_P(1), \\
 V_{3n} &= \frac{\lambda}{n_p^{1/2}} \sum_{i=1}^{n_v} \frac{H(W_i, \sigma^2, \beta, \theta)}{v^2(W_i, \sigma^2, \beta, \theta)} \{ 2u(W_i, \beta) f_1(X_i, \beta) \} + o_P(1), \\
 V_{4n} &= \frac{\lambda}{n_p^{1/2}} \sum_{i=1}^{n_v} \frac{H(W_i, \sigma^2, \beta, \theta)}{v^2(W_i, \sigma^2, \beta, \theta)} \{ \sigma^2 g(X_i, \beta, \theta) \} + o_P(1). \tag{B.6}
 \end{aligned}$$

By (B.4), (B.5), and (B.6), we have proved Theorem 2.

Appendix C: Proof of Corollary 3

Here we merely sketch the arguments, as they differ in no substantial way from previous calculations. From (2.8),

$$\begin{aligned}
 0 &= \frac{1}{n_p^{1/2}} \sum_{j=n_v+1}^n \frac{Y_j \hat{f}_W(W_j) - k_n(W_j, \hat{\beta}_4)}{d_n(W_j, \hat{\sigma}_1^2, \hat{\beta}_4, \hat{\theta}_1)} k_{n\beta}(W_j, \hat{\beta}_4) \\
 &= \frac{1}{n_p^{1/2}} \sum_{j=n_v+1}^n \Psi_n(Y_j, W_j, \hat{\sigma}_1^2, \hat{\beta}_4, \hat{\theta}_1). \tag{C.1}
 \end{aligned}$$

Drop the dependence on $W, \sigma^2, \theta,$ and β . After a Taylor's expansion of (C.1), we obtain

$$n_p^{1/2}(\hat{\beta}_4 - \beta_4) = D_{1n}^{-1}G_n - D_{1n}^{-1}\{D_{2n}n_p^{1/2}(\hat{\sigma}_1^2 - \sigma^2) + D_{3n}n_p^{1/2}(\hat{\theta}_1 - \theta)\} + o_p(1), \tag{C.2}$$

where

$$G_n = \frac{1}{n_p^{1/2}} \sum_{j=n_v+1}^n \Psi_n(Y_j, W_j, \sigma^2, \beta, \theta),$$

$$D_{1n} = \frac{1}{n_p} \sum_{j=n_v+1}^n \Psi_{n\beta}(Y_j, W_j, \sigma^2, \beta, \theta),$$

$$D_{2n} = \frac{1}{n_p} \sum_{j=n_v+1}^n \frac{d_{n\sigma^2}}{d_n^2} \Psi_n(Y_j, W_j, \sigma^2, \beta, \theta),$$

$$D_{3n} = \frac{1}{n_p} \sum_{j=n_v+1}^n \frac{d_{n\theta}}{d_n^2} \Psi_n(Y_j, W_j, \sigma^2, \beta, \theta).$$

By Lemma A, we can show that $D_{2n} \xrightarrow{P} 0$ and $D_{3n} \xrightarrow{P} 0$. Since $\hat{\sigma}_1^2$ and $\hat{\theta}_1$ are $n^{1/2}$ -consistent and use similar arguments to (B.5), we obtain (C.2):

$$= D_{1n}^{-1}G_n + o_p(1).$$

Now by arguments similar to those of appendix A, we finish the proof of Corollary 3.

References

Carroll, R.J. and D. Ruppert, 1988, Transformation and weighting in regression (Chapman and Hall, London).

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

Carroll, R.J. and M.P. Wand, 1991, Semiparametric estimation in logistic measurement error models, *Journal of the Royal Statistical Society B* 53, 573-585.

Carroll, R.J., C. Spiegelman, K.G. Lan, K.T. Bailey, and R.D. Abbott, 1984, On errors-in-variables for binary regression models, *Biometrika* 71, 19-26.

Clark, R., 1982, Logistic regression with measurement error in predictors, Unpublished Ph.D. dissertation (Department of Biostatistics, University of North Carolina, Chapel Hill, NC).

Davidian, M. and R.J. Carroll, 1987, Variance function estimation, *Journal of the American Statistical Association* 82, 1079-1091.

- Eubank, R.L., 1988, *Spline smoothing and nonparametric regression* (Marcel Dekker, New York, NY).
- Fuller, W.A., 1987, *Measurement error models* (Wiley, New York, NY).
- Gleser, L.J., 1990, Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models, in: P.J. Brown and W.A. Fuller, eds., *Statistical analysis of measurement error models and application* (American Mathematics Society, Providence, RI).
- Mack, Y. and B. Silverman, 1982, Weak and strong uniform consistency of kernel regression estimates, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 60, 405–415.
- Pepe, M.S. and T.R. Fleming, 1991, A general nonparametric method for dealing with errors in missing or surrogate covariate data, *Journal of the American Statistical Association* 86, 108–113.
- Pierce, D.A., D.O. Stram, M. Vaeth, and D.W. Schafer, 1990, The errors in variables problem: Considerations provided by radiation dose–response analyses of the A-bomb survivor data, Preprint.
- Rosner, B., W.C. Willett, and D. Spiegelman, 1989, Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error, *Statistics in Medicine* 8, 105–1070.
- Rosner, B., D. Spiegelman, and W.C. Willett, 1990, Correction of logistic regression relative risk estimates and confidence intervals for measurement error: The case of multiple covariates measured with error, *American Journal of Epidemiology* 132, 734–745.
- 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, 583–592.
- Stefanski, L.A. and R.J. Carroll, 1985, Covariate measurement error in logistic regression, *Annals of Statistics* 13, 1335–1351.
- Whittemore, A.S., 1989, Errors in variables regression using Stein estimates, *American Statistician* 43, 226–228.
- Whittemore, A.S. and J.B. Keller, 1988, Approximations for errors in variables regression, *Journal of the American Statistical Association* 83, 1057–1066.