Locally Ancillary Quasiscore Models for Errors-in-Covariates

Paul J. RATHOUZ and Kung-Yee LIANG

Paul J. Rathouz is Assistant Professor, Department of Health Studies, University of Chicago, Chicago, IL 60637 (email: prathouz@health.bsd.uchicago.edu); and Kung-Yee Liang is Professor, Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205 (email: kyliang@jhsph.edu). The authors thank Theodore G. Karrison, an Associate Editor and two referees for very constructive discussions, comments and suggestions; Wanli Min for computing assistance; and Rolf Loeber for permission to use the cortisol data.

Corresponding address:
Paul Rathouz
Dept. of Health Studies
5841 S. Maryland Ave., MC 2007
Chicago, IL 60637
Ph: 773-834-1970
FAX: 773-702-1979
email: prathouz@health.bsd.uchicago.edu

02 February 2001
Revised for *Journal of the American Statistical Association*

ABSTRACT

We use the notion of locally ancillary estimating functions to develop a quasiscore method for fitting regression models containing measurement error in the covariates. Suppose interest is on the model $E(Y|u, w)$ for response Y, the observed data are (y, x, w), and X is a mismeasured surrogate for u. We take a functional modelling approach, treating the u as a fixed nuisance parameter. Beginning with quasiscores for the regression parameter and the unknown u, a bias-corrected quasiscore for the regression parameter is derived that is second order locally ancillary for the nuisance u. The method used to accomplish this requires only the correct specification of the mean and variance functions for Y and X in terms of u, w and the regression parameter. When an estimator for u is plugged into the corrected quasiscore, local approximations show that the bias is small. Simulations verifying this result and an example from child psychiatry are presented, both using log-linear regression models.

KEY WORDS: ancillarity, measurement error, nuisance parameter, quasilikelihood, semiparametric model.

1 Introduction

Let (u_i, w_i) be a sequence of covariates with arbitrary joint empirical distribution function $G(\cdot)$, and let (y_i, x_i, w_i) be a sequence of observations such that the random variables $(Y_i|u_i, w_i)$ are independent conditional on the vector of (u_i, w_i)s. Assume that y_i, u_i and x_i are scalars and that w_i is a vector of dimension q. Interest is on the regression model

$$\left.\begin{array}{rcl} \mathrm{E}(Y_i \mid u_i, w_i) & = & \mu_y(\beta; u_i, w_i) = \mu_{yi} \\ \mathrm{var}(Y_i \mid u_i, w_i) & = & \phi_y \tilde{v}_y(\beta; u_i, w_i) = v_{yi}, \end{array}\right\}, \tag{1}$$

where β is a p-dimensional regression parameter and ϕ_y is a dispersion parameter. Often, $p = q + 1$. Let x_i be a mismeasured version of u_i such that

$$\left.\begin{array}{rcl} \mathrm{E}(X_i \mid u_i, w_i) & = & \mu_x(\alpha; u_i, w_i) = \mu_{xi} \\ \mathrm{var}(X_i \mid u_i, w_i) & = & \phi_x \tilde{v}_x(\alpha; u_i, w_i) = v_{xi}, \end{array}\right\}. \tag{2}$$

Again, ϕ_x is a dispersion parameter and α is a measurement error parameter. We make the common surrogacy assumption that X_i is independent of the response Y_i, conditional on the covariates (u_i, w_i).

In this paper, we propose a new method for inference in the regression model (1), subject to stochastic measurement error in the covariates following model (2). The method extends in an approximate way the functional modelling approach of Stefanski and Carroll (1987) in which the mismeasured covariate u_i is viewed as a fixed nuisance parameter. Under their generalized linear models (McCullagh and Nelder 1989) framework, the conditional score function for β (Lindsay 1982) is unbiased even when the covariate u_i is not known exactly, but rather is estimated. This elegant method is semiparametric efficient in the sense that the conditional score is optimal for β

in the absence of knowledge of u_i or of the distribution of $(u_i|w_i)$ (Lindsay 1982, 1985). However, the class of applicable models for the distributions of $(Y_i|u_i, w_i)$ and $(X_i|u_i, w_i)$ is limited to the canonical exponential family.

In separate research, Waterman and Lindsay (1996a,b) proposed a projected score method that approximates the conditional score when it exists and emulates it in terms of robustness to nuisance parameters when it does not. Robustness is operationalized in terms of local ancillarity (Small and McLeish 1994), which we define in the next section. While the Waterman-Lindsay method generates estimating functions that are locally ancillary to an arbitrary order, their work and that of Small and McLeish (1989) has also shown that second-order local ancillarity is a particularly important special case. Recently, Rathouz and Liang (1999) have extended the Waterman-Lindsay projected score method to a quasilikelihood setting, thereby obtaining a second-order locally ancillary quasiscore (SOLAQS). The measurement error method proposed here is motivated by making three observations, which synthesize these prior works: (i) the Stefanski and Carroll (1987) method for measurement error problems corresponds to the conditional score method for general nuisance parameter problems; (ii) recent work shows that second-order locally ancillary estimating functions provide very good approximations to the behavior of the conditional score; and (iii) the general method of obtaining second order locally ancillary estimating functions from quasilikelihood models, of which (1) and (2) are one example, can be exploited to develop a new method for inference in functional measurement error models. Such development is the object of this paper.

Model (1), which is the inferential target, includes linear, logistic, log-linear and polynomial regression models as special cases. Model (2) encompasses the classical measurement error model $\mathrm{E}(X_i \mid u_i, w_i) = u_i$, $\mathrm{var}(X_i \mid u_i, w_i) = \phi_x$, as well as the error calibration model $\mathrm{E}(X_i \mid u_i, w_i) = \alpha_0 + \alpha_1 u_i + \alpha_2 w_i$, $\mathrm{var}(X_i \mid u_i, w_i) = \phi_x$ (Carroll, Ruppert and Stefanski 1995) as special cases. Additionally, in model (2), the mean and variance of x_i do not have to be specified on the same scale in which u_i appears in model (1). We could for example, allow a multiplicative error model $\mathrm{E}\{\log(X_i) \mid u_i, w_i\} = \alpha_0 + \alpha_1 \log(u_i) + \alpha_2 w_i$ and $\mathrm{var}\{\log(X_i) \mid u_i, w_i\} = \phi_x$. Whatever the measurement error model, we assume throughout that the required internal or external replication or validation data (Carroll and Stefanski, 1990) are available to provide consistent estimators of the measurement error parameters (α, ϕ_x).

Besides the aforementioned conditional score method, other related methods include those of small measurement error asymptotics which provide first-order bias-corrections to the naive estimator (e.g. Stefanski 1985), and the approximate quasilikelihood-variance function (QVF) method of Carroll and Stefanski (1990), in which parametric functions are assumed only for the first two moments of $(Y_i|u_i, w_i)$ and $(X_i|u_i, w_i)$. Our approach is in the spirit of the conditional score, while making the weaker assumptions equivalent to the QVF models. We obtain a bias-corrected quasiscore function for which, using the method of Stefanski (1985), the resulting estimator would have zero first-order bias-correction. Our approach is therefore semiparametric in two senses. First, it is a functional model in that it does not require a specification of the marginal distribution of $(u_i|w_i)$, as the u_i's are treated as fixed nuisance parameters. Second, it only requires specification of the first two moments of the distributions of $(Y_i|u_i, w_i)$ and $(X_i|u_i, w_i)$. For a unified presentation of methods for errors-in-covariates, see Carroll, et al. (1995).

This paper has the following organization. In Section 2, we review second-order locally ancillary estimating functions and show that when the ancillarity applies to the mismeasured covariate u_i, the bias-correction of Stefanski (1985) obtains "automatically". Section 3 contains the main development of the SOLAQS for measurement error problems. Theoretical and practical considerations for use of the SOLAQS for inferences on β are presented in Section 4. These include variance estimation and the use of small measurement error asymptotics to examine the behavior of the resulting SOLAQS with respect to bias. In Section 5, we study the log-linear regression model in more detail in order to illustrate some advantages of the SOLAQS over other methods. We include simulation results and a small example data analysis from child psychiatry. We close with a brief discussion in Section 6.

2 Locally ancillary estimating functions in measurement error problems

Suppose that the regression parameter β is a vector of dimension p and that the $(p \times 1)$ estimating function

$$g(\beta) = \sum_{i=1}^{n} g_i(\beta; y_i, u_i, w_i) = \sum_i g_i, \tag{3}$$

such that $\mathrm{E}\{g_i(\beta; Y_i, u_i, w_i); u_i, w_i\} = 0$, is available for inferences on β. Now, treating u_i as an unknown nuisance parameter and operating only on the ith summand in (3), define the $(p \times 1)$ functional operators

$$b_k^i(g_i) = \left[\frac{\partial^k}{\partial u^{*k}} \mathrm{E}\{g_i(\beta; Y_i, u_i, w_i); u^*, w_i\} \right]_{u^*=u_i}, \tag{4}$$

for $k = 1, 2$, and the $(p \times 2)$ concatenated functional operator $b_{(2)}^i(g_i) = \{b_1^i(g_i), b_2^i(g_i)\}$. The idea in (4) is that the expectation is taken conditionally on (u^*, β), where $u^* \neq u_i$. If $b_1^i(g_i) = 0$, then g_i is said to be "first-order locally ancillary" for u_i, while if $b_{(2)}^i(g_i) = 0$, g_i is "second-order locally ancillary" for u_i (Small and McLeish 1994). One interpretation of kth-order local ancillarity is that, under regularity, it is equivalent to $\mathrm{E}\{g_i(u_i); u^*\} = o\{(u^* - u_i)^k\}$. In addition, under standard regularity conditions for estimating functions (Godambe, 1960; Godambe and Thompson 1974),

$$\mathrm{E}\left(-\frac{\partial g_i}{\partial u_i}\right) = b_1^i(g_i) \quad \text{and} \quad \mathrm{E}\left(\frac{\partial^2 g_i}{\partial u_i^2}\right) = b_2^i(g_i) - 2\frac{\partial b_1^i(g_i)}{\partial u_i}, \tag{5}$$

(Rathouz and Liang 1999). Higher orders of local ancillarity are defined similarly, the order providing a measure of the degree of robustness of g_i to u_i. The second order is most important in practice, however, because it provides a large degree of the bias correction obtained through second and successively higher orders (Waterman and Lindsay 1996a,b; Small and McLeish 1989).

This bias correction phenomenon arises in the measurement error literature, as follows. Let u_i be measured with error by x_i, where $\mathrm{E}(X_i|u_i, w_i) = u_i$ and $\mathrm{var}(X_i|u_i, w_i) = \phi_x \tilde{v}_x(u_i, w_i)$. Consider the plug-in estimating function $\hat{g}_i(\beta) = g_i(\beta; y_i, x_i, w_i)$, and let $\hat{\beta}$ be the solution to $\sum_i \hat{g}_i = 0$. Then for fixed $\phi_x > 0$, $\hat{\beta}$ converges in probability to $\beta + O(\phi_x)$ as $n \to \infty$. The remainder $O(\phi_x)$ refers to the measurement error bias in the limiting value of $\hat{\beta}$ as $\phi_x \to 0$. The order of operations for this argument is that $n \to \infty$ first, then $\phi_x \to 0$.

A first-order bias-corrected estimator of β using such small measurement error asymptotics would be

$$\hat{\beta}_c = \hat{\beta} + \frac{1}{2}\phi_x \left\{ \lim_{n\to\infty} \sum_i \mathrm{E}\left(\frac{\partial g_i}{\partial \beta}\right)^T \right\}^{-1} \left[\lim_{n\to\infty} \sum_i \mathrm{E}\left\{ \frac{\partial^2 g_i}{\partial u_i^2} \tilde{v}_x(u_i, w_i) \right\} \right], \qquad (6)$$

where for fixed $\phi_x > 0$, $\hat{\beta}_c$ converges to $\beta + o(\phi_x)$ as $n \to \infty$ (Stefanski 1985). If g_i is second-order locally ancillary for u_i, then by (5), $\mathrm{E}\left\{(\partial^2 g_i/\partial u_i^2)\tilde{v}_x(u_i, w_i)\right\} = 0$, and the bias correction given by (6) obtains automatically. Therefore, estimation with a second-order locally ancillary estimating function produces an estimator that is approximately consistent to order $o(\phi_x)$. In the next section, we propose a quasiscore method for obtaining second-order locally ancillary estimating functions for β in the functional measurement error modelling problem, eliminating the need for any correction term such as that in (6).

3 SOLAQS for Measurement Error Models

Rathouz and Liang (1999) recently proposed a method for constructing second-order locally ancillary estimating functions. The method differs from those previously proposed in that it does not rely on projection, and thereby avoids the need to specify a likelihood function. In this section, their idea is applied to the problem of inference in functional measurement error models. Assume that models (1) and (2) hold, that μ_{yi}, v_{yi}, μ_{xi}, v_{xi} are finite and admit continuous first and second derivatives with respect to (β, α, u_i) and that $v_{yi} > 0$ and $v_{xi} > 0$ for all (β, α, u_i). Additionally assume that $\mu_x(u_i, w_i)$ is strictly monotone (in u_i) in a neighborhood of the true u_i. These assumptions are not restrictive; for many models in practical usage, moments and their derivatives to several orders exist. The monotonicity assumption is quite reasonable, given that x_i is a mismeasured surrogate for the true u_i.

For the ith observation, we now construct estimating functions for β, ϕ_y and u_i which will act as building blocks in the development that follows. First, if u_i were known without error, one might consider the quasiscore

$$S_0 = \sum_i S_{i0} = \sum_i \left(\frac{\partial \mu_{yi}^T}{\partial \beta}\right) \frac{y_i - \mu_{yi}}{v_{yi}}$$

for consistent inferences on β (Wedderburn 1974, McCullagh 1983). In the traditional generalized linear models case (McCullagh and Nelder 1989), $h_y(\cdot)$ is a link function, η_{yi} is the linear predictor and $V(\cdot)$ is the variance function. Then $h_y(\mu_{yi}) = \eta_{yi} = \beta_0^T w_i + \beta_1 u_i$, and S_{i0} takes the well-known form

$$S_{i0} = \begin{pmatrix} w_i \\ u_i \end{pmatrix} \{h_y'(\mu_{yi})\}^{-1} \frac{y_i - \mu_{yi}}{\phi_y V(\mu_{yi})}.$$

Additionally, if ϕ_y is not known, the estimating function

$$R_0 = \sum_i R_{0i} = \sum_i (\phi_y v_{yi})^{-1}\{(y_i - \mu_{yi})^2 - \phi_y v_{yi}\}$$

would yield consistent inferences on ϕ_y. $R_0 = 0$ can be solved after β has been estimated via $S_0 = 0$, since S_0 is ancillary for ϕ_y. Note that $(S_0^T, R_0)^T$ is the quasilikelihood analogue to the score equations given in Stefanski and Carroll (1987), from which the conditional score function was constructed.

Now, similarly to S_{i0}, using the data y_i and x_i, define the u_i-quasiscore

$$T_{i1} = \left(\frac{\partial \mu_{yi}}{\partial u_i}\right) \frac{y_i - \mu_{yi}}{v_{yi}} + \left(\frac{\partial \mu_{xi}}{\partial u_i}\right) \frac{x_i - \mu_{xi}}{v_{xi}} = T_{i1y} + T_{i1x}. \qquad (7)$$

In the aforementioned case where $h_y(\mu_{yi}) = \eta_{yi}$, and where $\mu_{xi} = u_i$,

$$T_{i1} = \beta_1 \{h_y'(\mu_{yi})\}^{-1} \frac{y_i - \mu_{yi}}{\phi_y V(\mu_{yi})} + \frac{x_i - u_i}{v_{xi}}.$$

The quasiscore T_{i1} will be used to estimate the mismeasured covariate u_i. Further, it will be used as a basis for correcting the bias in S_{i0} due to measurement error. To that end, note that T_{i1} is optimal for u_i in the class of linear estimating functions and is thereby information unbiased (Crowder, 1987). In particular, $T_{i2} = \partial T_{i1}/\partial u_i + T_{i1}^2$ is a second unbiased estimating function for u_i. Indeed, T_{i2} is the quasiscore analogue of the second Bhattacharyya score for the nuisance u_i (Rathouz and Liang 1999). Letting prime ($'$) denote differentiation with respect to u_i, T_{i2} can be re-expressed as

$$T_{i2} \quad = \quad T_{i1}' + T_{i1}^2 = T_{i2y} + T_{i2x} + 2T_{i1y}T_{i1x},$$

where $T_{i2y} = T_{i1y}' + T_{i1y}^2$ and similarly for T_{i2x}. From (7), T_{i2y} takes the form

$$T_{i2y} = \frac{\partial}{\partial u_i}\left(\frac{\partial \mu_{yi}}{\partial u_i} v_{yi}^{-1}\right)(y_i - \mu_{yi}) + \left(\frac{\partial \mu_{yi}}{\partial u_i}\right)^2 v_{yi}^{-2}\left\{(y_i - \mu_{yi})^2 - v_{yi}\right\},$$

with an analogous form for T_{i2x}.

We now obtain a second order locally ancillary quasiscore, S_{2i}, as a linear combination of S_{i0}, T_{i1} and T_{i2}. Operating on the ith observation (y_i, x_i, w_i), define the matrix

$$D_i = \left[\mathrm{E}\left\{ -\frac{\partial}{\partial \beta} \begin{pmatrix} S_{i0} \\ T_{i1} \\ T_{i2} \end{pmatrix} \right\}, b_{(2)}^i \begin{pmatrix} S_{i0} \\ T_{i1} \\ T_{i2} \end{pmatrix} \right] = \begin{pmatrix} D_{i00} & D_{i01} & D_{i02} \\ D_{i10} & D_{i11} & D_{i12} \\ D_{i20} & D_{i21} & D_{i22} \end{pmatrix},$$

where the second and third columns of D_i are the maps of $(S_{i0}^T, T_{i1}, T_{i2})^T$ via $b_1^i(\cdot)$ and $b_2^i(\cdot)$ respectively. Interestingly, D_i is symmetric. Then define

$$S_{i2} = S_{i0} - a_i \begin{pmatrix} T_{i1} \\ T_{i2} \end{pmatrix},$$

where a_i is the $(p \times 2)$ matrix given by

$$a_i = b_{(2)}^i(S_0)\left\{ b_{(2)}^i \begin{pmatrix} T_{i1} \\ T_{i2} \end{pmatrix} \right\}^{-1} = (D_{i01} D_{i02}) \begin{pmatrix} D_{i11} & D_{i12} \\ D_{i21} & D_{i22} \end{pmatrix}^{-1}.$$

In order to compute S_{i2}, we must evaluate the functionals $b^i_{(2)}(S_{i0})$, $b^i_{(2)}(T_{i1})$ and $b^i_{(2)}(T_{i2})$. This is easily accomplished via the derivatives of the mean and the variance models (1) and (2) (see Appendix A).

We claim that S_{i2} is second order locally ancillary for u_i. To see this, first note that $b^i_{(2)}(\cdot)$ is a linear operator in the sense that $b^i_{(2)}(a_1 g_1 + a_2 g_2) = a_1 b^i_{(2)}(g_1) + a_2 b^i_{(2)}(g_2)$, where $a_k = a_k(\beta, u_i, w_i)$ does not contain the data (y_i, x_i), $k = 1, 2$. Then write

$$\begin{aligned} b^i_{(2)}(S_{i2}) &= b^i_{(2)}(S_{i0}) - a_i\, b^i_{(2)} \begin{pmatrix} T_{i1} \\ T_{i2} \end{pmatrix} \\ &= (D_{i01}, D_{i02}) - (D_{i01}, D_{i02}) \begin{pmatrix} D_{i11} & D_{i12} \\ D_{i21} & D_{i22} \end{pmatrix}^{-1} \begin{pmatrix} D_{i11} & D_{i12} \\ D_{i21} & D_{i22} \end{pmatrix} = 0. \end{aligned}$$

We refer to S_{i2} as a *second-order locally ancillary quasiscore* (SOLAQS). It is a robust version of S_{i0}, compensating for the bias introduced by the measurement error in x_i. In contrast to previous methods (Waterman and Lindsay 1996a), S_{i2} is obtained without the use of projection, and consequently depends only upon correct specification of models (1) and (2).

Summing over all observations, inferences on β can be based on the SOLAQS

$$S_2 = \sum_i S_{i2} = \sum_i \left\{ S_{i0} - a_i \begin{pmatrix} T_{i1} \\ T_{i2} \end{pmatrix} \right\},$$

which is an unbiased estimating function. However, while S_2 is second-order locally ancillary for the vector $(u_1, \ldots, u_n)^T$, the u_i's still appear in S_{i2} and therefore must be estimated. This is accomplished for each i by solving $T_{i1}(\beta, u_i) = 0$ in u_i for $\hat{u}_{i\beta}$, giving rise to the plug-in quasiscore

$$\hat{S}_2(\beta) = \sum_i \hat{S}_{i2}(\beta) = \sum_i S_{i2}(\beta, \hat{u}_{i\beta}),$$

which is used for inferences on β. In Section 4, we examine theoretical implications of and practical considerations for using $\hat{S}_2$ for inferences on β.

In the case where ϕ_y is unknown, an analogous procedure to that for deriving S_{2i} is implemented. Substituting R_{0i} for S_{0i} and ϕ_y for β, a linear combination R_{i2} of (R_{i0}, T_{i1}, T_{i2}) that is a second-order locally u_i-ancillary estimating function for ϕ_y-inferences is obtained. Estimation of $(\beta^T, \phi_y)^T$ is accomplished via solution of $\hat{S}_2 = 0$ and $\hat{R}_2 = \sum_i \hat{R}_{i2} = 0$. Simultaneous solution of $(\hat{S}_2^T, \hat{R}_2)^T = 0$ is required, however, as S_2 is not ancillary for ϕ_y.

4 Inferences with the plug-in SOLAQS $\hat{S}_2$

4.1 Introduction

In this section, we further develop the use of $\hat{S}_2$ for inferences on the regression parameter β. In the following subsection, we consider the bias in $\hat{S}_{i2}$ as the measurement error variance $\phi_x \to 0$. Having shown that the bias is small, in Section 4.3, we consider the asymptotic distribution of $\hat{\beta}$ as $n \to \infty$ for fixed ϕ_x, where $\hat{\beta}$ is the solution to $\hat{S}_2 = 0$. Finally, we address some computational issues in solving $\hat{S}_2 = 0$. The Fisher scoring computational technique for solving $\hat{S}_2 = 0$ is sketched in Appendix C.

4.2 Small measurement error asymptotic bias in $\hat{S}_2$

Here, we study the behavior of $\hat{S}_{i2}$ under small measurement error asymptotics (Carroll and Stefanski 1990). Formally, we hold n fixed and consider a series of experiments in which the measurement error dispersion $\phi_x \to 0$. That $\phi_x \to 0$ need not reflect a true limiting operation in practice. Rather, since exact bias analysis can be quite difficult, it serves as an analytic tool, yielding order-of-magnitude approximations that provide some insight to the performance of the method with respect to bias-correction. We establish that the asymptotic bias in $\hat{S}_{i2}$ is of smaller order than that of the naive plug-in score $\hat{S}_{i0}$ wherein x_i replaces u_i. We suppress the subscript i, and operate on one observation at a time. A proof is in Appendix B.

Letting prime denote differentiation with respect to u, we have

$$\hat{S}_2 - S_2 = (\hat{u}_\beta - u)S_2' + \frac{1}{2}(\hat{u}_\beta - u)^2 S_2'' + \frac{1}{6}(\hat{u}_\beta - u)^3 S_2'''(u_\beta^*) \tag{8}$$

where $|u_\beta^* - u| \leq |\hat{u}_\beta - u|$. We show in Appendix B that $(\hat{u}_\beta - u) = D_{11}^{-1}T_1 + O_p(\phi_x)$, so that the first term of (8) can be written

$$(\hat{u}_\beta - u)S_2' = S_2' T_1 D_{11}^{-1} + S_2' \left\{ (\hat{u}_\beta - u) - D_{11}^{-1} T_1 \right\}. \tag{9}$$

Regarding (8) and (9), we note here two interesting facts on which proof of the following result is founded, and which are direct results of the construction of S_2. First, $\mathrm{E}(S_2') = \mathrm{E}(S_2'') = 0$ due to second-order local ancillarity. Second, due to the joint optimality of the quasiscores (S_0, T_1) for (β, u), $\mathrm{E}(S_2' T_1) \approx 0$, i.e., S_2' and T_1 are approximately orthogonal.

Theorem 1. Let $\hat{u}_\beta$ be the solution of $T_1 = 0$ for fixed β. Then for the true β, $\hat{S}_2 = S_2(\hat{u}_\beta) = S_2 + Z + O_p(\phi_x)$, where $Z = O_p(\phi_x^{1/2})$ and is unbiased. Furthermore, under uniform integrability, $E(\hat{S}_2 - S_2) = O(\phi_x^{3/2})$.

We note as a point of comparison that using x to estimate u, $(\hat{S}_0 - S_0)$ is also $O_p(\phi_x^{1/2})$, but with bias of order $O(\phi_x)$.

4.3 Asymptotic distribution of $\hat{\beta}$

We now examine the asymptotic behavior of $\hat{\beta}$ for fixed measurement error variance ϕ_x. The following discussion applies equally to the parameter $(\beta^T, \phi_y)^T$ when ϕ_y is being estimated simultaneously with β; simply replace $\hat{S}_{i2}$ throughout with the vector $(\hat{S}_{i2}^T, \hat{R}_{i2})^T$.

By the standard theory of estimating functions (e.g. Carroll, et al. 1995, Appendix A.3), there exists $\hat{\beta}$, a sequence of solutions to $\hat{S}_2 = 0$ such that $\hat{\beta} \rightarrow \beta^*$ in probability, where β^* is the solution to the limiting equation $\lim_{n\rightarrow\infty}(1/n)\hat{S}_2 = 0$ that is closest to the true β. By Theorem 1, β^* is close to the true β; this result is similar to those of other methods (e.g. Stefanski 1985; Carroll and Stefanski 1990) in that the bias is not completely eliminated, except in very specialized cases. Of course, there is no guarantee that there is a unique solution to $\hat{S}_2 = 0$, even as $n \rightarrow \infty$, so $\hat{\beta}$ must be carefully defined in practice. Our approach to this problem is given in the next section, but a general solution may not exist without further assumptions.

By estimating function theory, $\sqrt{n}(\hat{\beta} - \beta^*) \xrightarrow{d} N\{0, A^{-1}B(A^{-1})^T\}$, where

$$A = n^{-1} \lim_{n\rightarrow\infty} \sum_{i=1}^{n} \mathrm{E}\left(-\frac{\partial \hat{S}_2}{\partial \beta}\right) \quad \text{and} \quad B = n^{-1} \lim_{n\rightarrow\infty} \sum_{i=1}^{n} \mathrm{E}\left(\hat{S}_{i2}\hat{S}_{i2}^T\right),$$

and all quantities are evaluated at β^*. The variance factor B can be consistently estimated by replacing β^* with $\hat{\beta}$ and using the empirical expected value of $\hat{S}_{i2}\hat{S}_{i2}^T$. To estimate A, we employ a numerical derivative matrix, as follows. At the estimated $\hat{\beta}$, the kth column of $(\partial\hat{S}_2/\partial\beta)$ is estimated by

$$\frac{\hat{S}_2(\hat{\beta} + D_k) - \hat{S}_2(\hat{\beta} - D_k)}{2d_k},$$

where d_k is a perturbation, and $D_k = (0, \ldots, 0, d_k, 0, \ldots, 0)^T$ is a p-vector of zeros, with d_k in the kth position. Wald-type confidence intervals for β^* can then be constructed from $\hat{A}^{-1}\hat{B}(\hat{A}^{-1})^T$ in the standard fashion.

In the case where the measurement error parameters $\theta = (\alpha^T, \phi_x)^T$ are estimated as well by internal replication and/or validation data (Carroll, et al. 1995), a modified standard error estimator applies. Assume that θ is estimated via solution to some estimating equation $U_0 = \sum_i U_{i0} = 0$. Assume further than U_{i0} does not depend on β. This is not unreasonable, since in most settings U_{i0} will be a function of (x_i, u_i, w_i), but not y_i. Define the quantities

$$C = n^{-1} \lim_{n\to\infty} \sum_{i=1}^{n} \mathrm{E}\left(-\frac{\partial \hat{S}_2}{\partial \theta}\right) \quad \text{and} \quad D = n^{-1} \lim_{n\to\infty} \sum_{i=1}^{n} \mathrm{E}\left(-\frac{\partial U_0}{\partial \theta}\right).$$

Then, we show in Appendix D that $\sqrt{n}(\hat{\beta}-\beta^*) \xrightarrow{d} N\{0, A^{-1}B^*(A^{-1})^T\}$, where

$$B^* = n^{-1} \lim_{n\to\infty} \sum_{i=1}^{n} \mathrm{E}\{(\hat{S}_{i2} - CD^{-1}U_{i0})(\hat{S}_{i2} - CD^{-1}U_{i0})^T\},$$

and all quantities are evaluated at $(\beta^{*T}, \phi_y, \theta^T)^T$. As with A, C can be estimated using a numerical derivative, D can be estimated in the usual way from U_0, and B^* can be estimated using the empirical variance of $\hat{S}_{i2} - CD^{-1}U_{i0}$.

When θ is estimated using external replication and/or validation data (Carroll, et al. 1995), an augmented data set consisting of the concatenation of the primary data and the external data is employed. By setting $U_{i0} = 0$ for the primary data and $S_{i2} = 0$ for the external data, the augmented data can be analyzed as for internal replication and/or validation data.

4.4 Computational issues

Our experience thus far with the proposed method has suggested three computational techniques that provide for numerical stabilization of the estimation of β. First, before solving $\hat{S}_2$, we transform the design matrix containing the vectors w_i to form an orthonormal basis. Also, using ϕ_x and the sample variance of x, we transform x such that the empirical distribution function of u has mean zero and variance one. These transformations have the additional advantage of permitting the perturbations d_k to be set to a fixed constant for all k and sample sizes. We use $d_k = 0.01$.

Second, for smaller sample sizes, there is some instability in the simultaneous solution of $S_2, T_{11}, \ldots, T_{n1}$. This can be largely alleviated by estimating u_i as the solution to $T^*_{i1,n} = 0$ for each i, where $T^*_{i1,n}$ is the same as T_{i1}, replacing ϕ_x by a smaller quantity, ϕ^*_{xn}, thereby weighting the estimate $\hat{u}_{i\beta}$ towards x_i. We used $\phi^*_{xn} = (1 - 10p/n)\phi_x$ in our simulation work.

Third, the uniform integrability assumption made at several points in the proof of Theorem 1 has implications for the estimation of u_i. In some settings, it may be necessary to bound the permissible values of $\hat{u}_{i\beta}$ by quantities that are scientifically reasonable for the application at hand. As such bounds can be relatively wide, we see no way in which the need to specify them would restrict the applicability of the proposed method.

Additionally, two issues arise when ϕ_x is large: lack of convergence and multiple solutions. To choose among possible multiple solutions, we take the naive estimator of $(\beta^T, \phi_y)^T$ as a starting value for the Fisher scoring procedure (Appendix C). Lack of convergence can occur if the algorithm diverges to a point with one or more singular matrices. Alternatively, a solution may not be reached after the maximum number of iterations (we use 100). Interestingly, in our experience with simulated data, these problems are more frequent when

ϕ_y is known rather than estimated simultaneously with β. However, when ϕ_y is being estimated and ϕ_x is large (e.g., greater than $\text{var}(u_i)$), multiple solutions to $\hat{S}_2 = 0$ appear to exist, and the one obtained by starting from the naive estimator may not be consistent for β^*.

5 Example: Log-linear regression

We now illustrate the applicability and performance of the SOLAQS, by comparing it to other methods in the context of a special case of models (1) and (2), the log-linear regression model (McCullagh & Nelder 1989, Ch. 6)

$$\log(\mu_{yi}) = \beta_0^T w_i + \beta_1 u_i \quad \text{and} \quad v_{yi} = \phi_y \mu_{yi} \tag{10}$$

with additive measurement error $x_i = u_i + \delta_i$, where

$$\text{E}(\delta_i | u_i) = 0 \quad \text{and} \quad \text{var}(\delta_i | u_i) = \phi_x. \tag{11}$$

We consider assumptions underlying other approaches and present simulation results comparing the SOLAQS to these competitors.

5.1 Other approaches

Were $(Y_i | w_i, u_i)$ a true Poisson random variable and $\delta_i \sim N(0, \phi_x)$, the conditional score method of Stefanski and Carroll (1987) would apply and give rise to the semiparametric efficient estimator for β in the presence of unknown distribution for $(U_i | w_i)$. However, implementation is hindered by two concerns. First, the conditional score does not take a closed form; rather, iterative computations are required to compute it. Second, it is not known to what degree the conditional score is robust to misspecifications of the distributional form of $(Y_i, X_i|u_i, w_i)$ (Carroll and Wand, 1991). We note that the first problem can be addressed approximately via projection (Waterman and Lindsay 1996a,b). Indeed, due to the equivalence of quasiscores and likelihood scores in exponential family distributions, S_2 with $\phi_y = 1$ is the second-order projected score for the Poisson-Gaussian model.

Furthermore, in many problems, $(Y_i|u_i, w_i)$ is overdispersed relative to a true Poisson random variable, rendering a likelihood difficult to specify. Alternatives not requiring a likelihood include regression calibration (RC; Carroll, et al. 1995, Ch. 3) and the SIMEX estimator (Cook and Stefanski, 1995). Treating u_i as a random variable, the RC method replaces u_i with an estimate of $\mathrm{E}(U_i|w_i, x_i)$, thereby exploiting a distributional assumption on $(U_i|w_i)$, which may or may not be valid. However, if the distribution of $(U_i|w_i, x_i)$ is Gaussian, then in the log-linear model, the RC method is particularly applicable and relatively statistically efficient for $\hat{\beta}_1$. The fully functional SIMEX estimator is easy to implement and quite general. It does however require a distributional form for $(X_i|u_i, w_i)$. In contrast, the SOLAQS method only requires the mean and variance of $(X_i|u_i, w_i)$.

5.2 Simulation study

We now compare the empirical performance of SOLAQS to that of RC in a simulation study of models (10) and (11). In each simulation, we compare the naive estimator (the solution to $\hat{S}_0$, with u_i replaced by x_i), the RC estimator with the linear approximation calibration estimator as described in Carroll, et al (1995, Section 3.4.2), and the SOLAQS estimator using $(\hat{S}_2^T, \hat{R}_2)^T$ to estimate β and ϕ_y. We include the naive estimator to provide an indication of the degree of bias correction needed. The RC method, in cases where its assumptions are satisfied, will permit an assessment of the efficiency loss in SOLAQS by not exploiting the distribution of $(U_i|w_i)$. Computational procedures in Section 4.3, 4.4 and Appendix C were used for estimation and confidence interval construction with $(\hat{S}_2^T, \hat{R}_2)^T$. The error variance ϕ_x, which in practice is easily-estimated with replication data, was assumed known for both the RC and the SOLAQS methods. The RC model parameters of the distribution of (u_i, w_i), which are not needed for SOLAQS, were estimated for each replicate. We concentrate on the coefficient β_1 of u_i.

Overdispersed count data Y_i were generated as a $(3:7)$ Bernoulli mixture of two Poisson random variables such that the mean and variance were μ_{yi} and $\phi_y \mu_{yi}$, respectively. The u_i's were standardized to have mean zero and variance one and the set of (w_i, u_i)'s was fixed over all replicates. The errors δ_i were mean-zero Gaussian with variance ϕ_x. We considered relative rate values $\exp(\beta_1) = (1.5, 3.0)$, measurement error variance $\phi_x = (0.3, 0.7)$, and overdispersion $\phi_y = (1.5, 3.0)$, Each simulation contains 500 replicates.

Model 1. Let $w_i = 1$ so that β_0 is the intercept. Set $\beta_0 = 0$. Let the u_i's be a random Gaussian sample. Set sample size $n = 200$. Results are in Table 1.

Model 2 is identical to Model 1, except that the u_i's are uniformly distributed, then standardized. Results are also in Table 1.

Model 3. Let w_{i1}^* be Bernoulli with probability 0.3. To generate the u_is, let ϵ_i be uniformly distributed on $(0, 1)$. Then let u_i be the standardized version of $(1 + cw_i^*)\epsilon_i$, $c > 0$. Let $w_i = (1, w_{i1})^T$, where w_{i1} is the standardized version of w_{i1}^*. Fix $\beta_0 = (0, \log(1.5))^T$ and $n = 500$. Setting $c = (0.5, 1.5)$ allows for different values of $\rho_{wu} = \text{corr}(w_{i1}, u_i)$. Results are in Table 2.

Results. Of sixteen thousand replicates across the three models, in all but one replicate, the SOLAQS converged in less than 100 iterations (result included), while 12 others took more than 50 iterations. These instances all fell under the last case of Model 2 (Table 1). The bias in the naive estimator reflects a substantial degree of measurement error in all cases. For Model 1, the assumptions of RC are met, and therefore, by exploiting the distribution of $(U_i|w_i)$, the RC method surpassed the SOLAQS in terms of bias and efficiency. With the exception of one case, however, use of the SOLAQS resulted in less than 20 percent precision loss, as measured by the MSE.

For Models 2 and 3, where the RC assumptions are violated, SOLAQS yielded lower bias than the RC method in 19 of 24 cases. While the RC was notably biased in some cases, in no case was the bias in SOLAQS more than ten percent, and it was usually less than five percent. Compared to RC, the SOLAQS never resulted in more than 30 percent loss in precision. By contrast, the gains in precision in SOLAQS over RC were at times substantial. SOLAQS coverage probabilities of Wald-type confidence intervals using the sandwich estimator were satisfactory for $\beta_1 = \log(1.5)$, but were anti-conservative at $\beta_1 = \log(3.0)$. This suggests that, in practice, for larger values of β_1, an-other method of variance estimation, such as the BCa bootstrap (Efron and Tibshirani, 1993, Section 14.3) as suggested by Carroll, et al (1995, Sections A.6.5–A.6.6) would be more appropriate.

5.3 Cortisol data

We illustrate our method with a data set from a study examining the relationship between salivary cortisol and symptoms of conduct disorder (CD) (McBurnett, Lahey, Rathouz and Loeber 2000). One hypothesis about the psychopathology of CD is that symptomatic behaviors occur with increased frequency due to subjects' suppressed fear response to threatening stimuli, such as punishment for disruptive behaviors. Because fear response is reflected in cortisol levels, we expect symptoms to be inversely related to cortisol. This hypothesis was examined in a clinic-referred sample of $n = 38$ boys with CD. Responses Y_i are the cumulative counts over four years of reported symptoms of aggressive CD and of covert CD. We treat the two sets of symptoms separately. Symptom counts each ranged from 0 to 13, with a median of 2. The 75th percentiles were 4 aggressive symptoms and 5 covert symptoms. The covariate u_i of interest is the logarithm of salivary cortisol in Year 2 of the study. Since cortisol was measured in Years 2 and 4, we let $X_{ij} = \log(\text{measured cortisol})$, where j denotes Year. Note that $\text{corr}(x_{i2}, x_{i4}) = 0.2$, suggesting substantial within-subject variation, in this case reflecting laboratory error and temporal fluctuations. Age in the first study year was also obtained, presumably without error. Let $w_i = (1, \text{age}_i)^T$.

We model the error in the log-cortisol measurements as $\text{E}(X_{ij} \mid u_i) = u_i +$

$\alpha I(j = 4)$ and $\mathrm{var}(X_{ij} \mid u_i) = \tilde{\phi}_x$. Consequently, define $X_i = (X_{i2} + X_{i4} - \alpha)/2$, so that $\mathrm{E}(X_i \mid u_i) = u_i$ and $\mathrm{var}(X_i \mid u_i) = \tilde{\phi}_x/2 = \phi_x$. Rescaling X_i so that $\mathrm{sd}(u_i) \approx 1$, we first estimated the error model as $\hat{\alpha} = \bar{x}_4 - \bar{x}_2 = 0.84$ and $\hat{\phi}_x = 0.25$ times the sample variance of $(x_{i2} - x_{i4} + \hat{\alpha}) = 1.99$. We then estimated β and ϕ_y simultaneously using the naive and SOLAQS methods (Table 3); the correction ϕ^*_{xn} (Section 4.4) was not used. Due to the small sample size and to account for variability in estimation of α and ϕ_x, bootstrap BCa confidence intervals were generated instead of Wald-type confidence intervals. As expected, correcting for the measurement error makes moderate difference in the parameter estimates for the age coefficients, while use of $\hat{S}_2$ provides a correction for attenuation of about 70 percent in $\hat{\beta}_1$ for each outcome. Nevertheless, the effect is not significant for the Covert CD outcome. Confidence intervals are considerably wider for $\hat{S}_2$, reflecting the variability induced in correcting for the bias due to measurement error. Furthermore, the results suggest that, due to the error in u_i, the dispersion ϕ_y is substantially overestimated in the naive analysis.

6 Concluding Remarks

We have outlined a quasilikelihood-based method for obtaining second order locally ancillary estimating functions for regression problems subject to errors in covariates. Prior to Rathouz and Liang (1999), local ancillarity had been achieved using the method of $\mathcal{L}^2$ projection (Waterman and Lindsay 1996a), generally requiring a likelihood specification. By replacing the projection operator proposed by Waterman and Lindsay with the solution to a simple linear system, our method achieves local ancillarity assuming only that the first two moments of the response and the surrogate covariate are correctly specified. Furthermore, through a functional modeling approach, we avoid any assumptions on the baseline distribution of u. Simulations in the context of log-linear regression show that the SOLAQS estimator is generally less biased than the RC estimator when the RC assumptions were violated, and often results in a substantial increase in precision. Corresponding results in larger sample sizes are expected to be more dramatic due to the more important role of bias. In the following, we briefly remark on a few additional aspects of the method.

Asymptotic approximations focusing on the limiting behavior of $\hat{S}_{i2}$ as $\phi_x \to \infty$ were discussed in Section 4.2. Of more direct interest is the estimator $\hat{\beta}$ that solves $\hat{S}_2 = 0$; such approximations are considerably more difficult to study. Nevertheless, the work of Stefanski (1985) described in Section 2 (equation 6) and its relationship to second-order local ancillarity suggests the following: For fixed ϕ_x, as $n \to \infty$, $\hat{\beta}$ converges to β^* which differs from β by a quantity of order $o(\phi_x)$. Also, as with some existing estimating functions for measurement error (e.g., Stefanski and Carroll, 1987), $\hat{S}_2$ may not admit a unique root even in its limit as $n \to \infty$. This may be especially true for larger values of ϕ_x. Given that $\hat{S}_{2i} \to S_{i0}$ as $\phi_x \to 0$, this is not expected to pose practical problems for small measurement error variance.

There are several reasons to believe that $\hat{S}_2$ is reasonably efficient. First, it is based on quasiscores for β and u, which are the efficient estimating functions that are linear in the data (Crowder 1987). Second, the conditional score function gives rise to the semiparametric efficient score for β when the distribution of $(u|w)$ is unspecified. Third, the second-order locally ancillary projected score of Waterman and Lindsay (1996a) emulates the conditional score in terms of bias and efficiency when it exists. Finally, our method is the quasiscore analogue of the projected score. In depth efficiency studies are a subject for further research.

Finally, extension of the SOLAQS to multiple mismeasured covariates is straightforward in the case where the variance-covariance of $(X_i|u_i, w_i)$ is reliably estimable. If the measurement errors for the components of u_i are independent, or if the errors are additive, this will generally not pose a problem.

APPENDIX A: Components of matrix D

Straightforward calculations using derivatives, expected values and the definition of the operators $b_k(\cdot)$, $k = 1, 2$ lead to the following expressions hold for the components of D_i, where i refers to the observation. For details, see Rathouz and Liang (1999) and the technical report referred to therein.

$$
\begin{aligned}
D_{i00} &= \left(\frac{\partial \mu_{yi}^T}{\partial \beta}\right) v_{yi}^{-1} \left(\frac{\partial \mu_{yi}}{\partial \beta}\right) = D_{i00}^T \\
D_{i01} &= \left(\frac{\partial \mu_{yi}^T}{\partial \beta}\right) v_{yi}^{-1} \left(\frac{\partial \mu_{yi}}{\partial u_i}\right) = D_{i10}^T \\
D_{i02} &= \left(\frac{\partial \mu_{yi}^T}{\partial \beta}\right) v_{yi}^{-1} \left(\frac{\partial^2 \mu_{yi}}{\partial u_i^2}\right) = D_{i20}^T \\
D_{i11} &= \left(\frac{\partial \mu_{yi}}{\partial u_i}\right) v_{yi}^{-1} \left(\frac{\partial \mu_{yi}}{\partial u_i}\right) + \left(\frac{\partial \mu_{xi}}{\partial u_i}\right) v_{xi}^{-1} \left(\frac{\partial \mu_{xi}}{\partial u_i}\right) \\
D_{i12} &= \left(\frac{\partial \mu_{yi}}{\partial u_i}\right) v_{yi}^{-1} \left(\frac{\partial^2 \mu_{yi}}{\partial u_i^2}\right) + \left(\frac{\partial \mu_{xi}}{\partial u_i}\right) v_{xi}^{-1} \left(\frac{\partial^2 \mu_{xi}}{\partial u_i^2}\right) = D_{i21} \\
D_{i22} &= D_{i22y} + D_{i22x} + 2D_{i11}^2,
\end{aligned}
$$

where

$$D_{i22y} = \left(\frac{\partial \mu_{yi}}{\partial u_i}\right)^2 v_{yi}^{-2} \left(\frac{\partial^2 v_{yi}}{\partial u_i^2}\right) + \left(\frac{\partial^2 \mu_{yi}}{\partial u_i^2}\right)^2 v_{yi}^{-1} - \left(\frac{\partial \mu_{yi}}{\partial u_i}\right)\left(\frac{\partial v_{yi}}{\partial u_i}\right) v_{yi}^{-2} \left(\frac{\partial^2 \mu_{yi}}{\partial u_i^2}\right),$$

and D_{i22x} is defined analogously.

In the case where ϕ_y is estimated via $\hat{R}_2$ simultaneously with β, the following additional components of D_i are required.

$$D_{i0r} = 0 \quad \text{and} \quad D_{ir0} = (\phi_y v_{yi})^{-1} \left(\frac{\partial v_{yi}^T}{\partial \beta}\right)$$

$$D_{i1r} = 0\ , \quad D_{ir1} = (\phi_y v_{yi})^{-1} \left(\frac{\partial v_{yi}}{\partial u_i}\right) \quad \text{and} \quad D_{irr} = 1$$

$$D_{i2r} = (\phi_y v_{yi})^{-1} \left(\frac{\partial \mu_{yi}}{\partial u_i}\right)^2$$

$$D_{ir2} = (\phi_y v_{yi})^{-1} \left\{ \left(\frac{\partial^2 v_{yi}}{\partial u_i^2}\right) + 2\left(\frac{\partial \mu_{yi}}{\partial u_i}\right)^2 \right\}$$

APPENDIX B: Sketch Proof of Theorem 1

To prove Theorem 1, we study (8) and (9) in four steps. In Step 1 (Lemmas 2 and 3), we examine the distribution of $\hat{u}_\beta$ via decomposition into terms of different orders. Step 2 involves deriving expressions for the first term in (9), from which the stochastic order and bias are determined. To accomplish this, we show $S_2' = \tilde{S}_2' + O_p(\phi_x)$ (Lemma 5). Then, via important information equalities (Lemma 6), Corollary 7 establishes that $\tilde{S}_2'$ is orthogonal to T_1. In Steps 3 and 4, we derive expressions for the second term in (9) (Lemma 8) and the last two terms of (8) (Lemma 9) respectively, from which stochastic order and bias are determined. Throughout, we hold β and w fixed at the true values and take the model assumptions in Section 3 as given. More detailed proofs are in a technical report available from the first author.

Lemma 2. $\hat{u}_\beta \xrightarrow{p} u$ *as* $\phi_x \to 0$, *and* $(\hat{u} - u) = O_p(\phi_x^{1/2})$.

Proof. Let u_0 be the true value of u. For $u \in \mathcal{R}$, $\phi_x T_1(u) = \phi_x T_{1y}(u) + \phi_x T_{1x}(u) = \phi_x O_p(1) + (\partial \mu_{xi}/\partial u_i)(u,w)\tilde{v}_x(u,w)^{-1}\{x - \mu_x(u,w)\}$. Since $\mathrm{var}\{X - \mu_x(u_0,w)\} = \phi_x \tilde{v}_x(u_0,w) \to 0$, $X \xrightarrow{\mathcal{L}^2} \mu_x(u_0,w)$, which implies that $X \xrightarrow{p} \mu_x(u_0,w)$. Also, $\phi_x O_p(1) \xrightarrow{p} 0$, so $\phi_x T_1(u) \xrightarrow{p} T_1^*(u)$, where
$T_1^*(u) = (\partial \mu_{xi}/\partial u_i)(u,w)\tilde{v}_x(u,w)^{-1}\{\mu_x(u_0,w) - \mu_x(u,w)\}$. By monotonicity of $\mu_x(u)$, $T_1^*(u) > 0$ if $u < u_0$, and $T_1^*(u) < 0$ if $u > u_0$. Following arguments in Serfling (1980, Section 7.2.1), $\mathrm{pr}\{|\hat{u}_\beta - u_0| < \epsilon\} \to 1$ as $\phi_x \to 0$, completing the consistency proof. The order $O_p(\phi_x^{1/2})$ is shown via Taylor series expansion.

For the remainder of the proofs, let u be the true value, let prime ($'$) denote differentiation with respect to u, and assume all functions S and T and their derivatives S', T'', etc. are evaluated at the true u. Let $\hat{u} = \hat{u}_\beta$.

Lemma 3. *Under the conditions of Lemma 2 and mild smoothness conditions on* $T_1'''(u)$,

$$(\hat{u}_\beta - u) = D_{11}^{-1}T_1 + Z_1 + Z_2 = D_{11}^{-1}T_{1x} + Z_3 = O_p(\phi_x^{1/2}),$$

where $D_{11}^{-1}T_{1x}$ *does not depend on* y, $Z_1 = Z_1(x) = O_p(\phi_x)$, $Z_2 = Z_2(y,x) = O_p(\phi_x^{3/2})$, *and* $Z_3 = Z_3(y,x) = O_p(\phi_x)$.

Proof. Recall that $D_{11} = \mathrm{E}(-T_1')$ and define $D_{11y} = \mathrm{E}(-T_{1y}')$ and $D_{11x} = \mathrm{E}(-T_{1x}')$. Straightforward inspection provides the following orders of stochastic or fixed magnitude as $\phi_x \to 0$: $T_{1y} = O_p(1)$, $T_{1x} = O_p(\phi_x^{-1/2})$, $D_{11y} =$

$O(1)$, $D_{11x} = O(\phi_x^{-1})$, $T'_{1y} = O_p(1)$, $T'_{1x} = O_p(\phi_x^{-1})$, $T'_1 = O_p(\phi_x^{-1})$, $T'_{1x} + D_{11x} = O_p(\phi_x^{-1/2})$, $T'_1 + D_{11} = O_p(\phi_x^{-1/2})$, $T''_{1x} = O_p(\phi_x^{-1})$, $T''_1 = O_p(\phi_x^{-1})$, $T''_{1x} - \mathrm{E}(T''_{1x}) = O_p(\phi_x^{-1/2})$, $T''_{1y} = O_p(1)$, $T'''_1 = O_p(\phi_x^{-1})$. Also, $T''_1(u^*)$ and $T'''_1(u^*)$ are both $O_p(\phi_x^{-1})$ for u^* in a neighborhood of u, by smoothness of T'''_1. Finally, $(\hat{u} - u) = o_p(1)$, by Lemma 2. The remainder of the proof involves third order Taylor series expansions of T_1 in u, setting $T_1 = T_{1x} + T_{1y}$.

Lemma 4. Let the matrix $a = (a_1, a_2)$, where a_k is $p \times 1$, $k = 1, 2$. Then $a_1 = \tilde{a}_1 + O(\phi_x^2) = O(\phi_x)$, where $\tilde{a}_1 = D_{01}D_{11}^{-1}$. And, $a_2 = \tilde{a}_2 + O(\phi_x^3) = O(\phi_x^2)$, where $\tilde{a}_2 = (D_{02}D_{11} - D_{01}D_{12})/(2D_{11}^3)$. Further, $a'_1 = \tilde{a}'_1 + O(\phi_x^2) = O(\phi_x)$, $a''_1 = O(\phi_x)$, $a'_2 = O(\phi_x^2)$, $a''_2 = O(\phi_x^2)$, and $\tilde{a}'_1 = (D'_{01}D_{11} - D_{01}D'_{11})/D_{11}^2$.

Proof. By the expressions in Appendix A, $D_{01} = O(1)$, $D_{02} = O(1)$, $D_{11} = O(\phi_x^{-1})$, $D_{12} = O(\phi_x^{-1})$, $D_{22} = 2D_{11}^2 + O_p(\phi_x^{-1}) = O_p(\phi_x^{-2})$. Taylor-series expansions and order-of-magnitude bookkeeping complete the proof.

Lemma 5. The u-derivative S'_2 of S_2 is $S'_2 = \tilde{S}'_2 + O_p(\phi_x)$, where

$$\tilde{S}'_2 = (S'_0 + D_{01}) - \tilde{a}_1(T'_1 + D_{11}) - (\tilde{a}'_1 - 2\tilde{a}_2 D_{11})T_1.$$

$(S'_0 + D_{01}) = O_p(1)$ and the other two terms are $O_p(\phi_x^{1/2})$.

Proof. Recall that $D_{01} = \mathrm{E}(-S'_0)$, $D_{11} = \mathrm{E}(-T'_1)$ and $D_{21} = \mathrm{E}(-T'_2)$. In addition to the orders of magnitude in the proof of Lemma 3, $S'_0 + D_{01} = O_p(1)$, $T_1 = O_p(\phi_x^{-1/2})$, $T'_1 + D_{11} = O_p(\phi_x^{-1/2})$, $T_2 = O_p(\phi_x^{-1})$, $T'_2 + D_{21} = O_p(\phi_x^{-3/2})$.

Write $S'_2 = S'_0 - a_1 T'_1 - a'_1 T_1 - a_2 T'_2 - a'_2 T_2$. By second-order local ancillarity and equation (5), S'_2 is unbiased, so we may write

$$S'_2 = (S'_0 + D_{01}) - a_1(T'_1 + D_{11}) - a'_1(T_1) - a_2(T'_2 + D_{21}) - a'_2(T_2),$$

for which each term in parentheses is unbiased. Working term-by-term, the first is $O_p(1)$. Using Lemma 4, the second is $a_1(T_1 + D_{11}) = \tilde{a}_1(T_1' + D_{11}) + O_p(\phi_x^{3/2}) = O_p(\phi_x^{1/2})$. Similarly, the third term is $a_1'T_1 = \tilde{a}_1'T_1 + O_p(\phi_x^{3/2}) = O_p(\phi_x^{1/2})$. For the fourth term, use Lemma 6 and equation (5) to write

$$T_2' + D_{21} = \{T_1'' - \mathrm{E}(T_1'')\} - 2\{T_1 D_{11}\} + 2\{T_1(T_1' + D_{11}) - (D_{11}' - D_{12}/2)\}.$$

Thereby show that $a_2(T_2' + D_{21}) = -2\tilde{a}_2 T_1 D_{11} + O_p(\phi_x) = O_p(\phi_x^{1/2})$. The last term is $a_2'T_2 = O_p(\phi_x)$, completing the proof.

Lemma 6. The unbiased estimating functions $(S_0' + D_{01})$ *and* $(T_1' + D_{11})$ *are information unbiased with respect to* T_1*. That is* $E\{-(S_0' + D_{01})'\} = E\{(S_0' + D_{01})T_1\} = D_{01}' - D_{02}$ *and* $E\{-(T_1' + D_{11})'\} = E\{(T_1' + D_{11})T_1\} = D_{11}' - D_{12}$.

Proof. The results are shown through application of equation (5), manipulations of the expressions in Appendix A and the surrogacy assumption.

Corollary 7. $S_2'T_1 D_{11}^{-1} = \tilde{S}_2'T_1 D_{11}^{-1} + O_p(\phi_x^{3/2}) = O_p(\phi_x^{1/2})$*, where* $E(\tilde{S}_2'T_1) = 0$*. So, under uniform integrability,* $E(S_2'T_1 D_{11}^{-1}) = O(\phi_x^{3/2})$.

Proof. $T_1 D_{11}^{-1} = O_p(\phi_x^{1/2})$, so $S_2'T_1 D_{11}^{-1} = \tilde{S}_2'T_1 D_{11}^{-1} + O_p(\phi_x^{3/2}) = O_p(\phi_x^{1/2})$. $\mathrm{E}(\tilde{S}_2'T_1) = 0$ is shown using the expressions in Lemmas 4, 5 and 6.

Lemma 8. The quantity $\left\{(\hat{u}_\beta - u) - D_{11}^{-1}T_1\right\} S_2'$ *is of order* $O_p(\phi_x)$*, but its bias under uniform integrability is of order* $O(\phi_x^{3/2})$.

Proof. By Lemmas 3 and 5, $S_2'\left\{(\hat{u}_\beta - u) - D_{11}^{-1}T_1\right\} = O_p(\phi_x)$. By the same lemmas, we may write

$$S_2'\left\{(\hat{u}_\beta - u) - D_{11}^{-1}T_1\right\} = (S_0' + D_{01})Z_1(x) + O_p(\phi_x^{3/2}) + O_p(\phi_x^{3/2}).$$

The expectation of the first term in this expression is 0 by independence.

Lemma 9. The quantity $(\hat{u}_\beta - u)^2 S_2''$ is of order $O_p(\phi_x)$, but its bias under uniform integrability is of order $O(\phi_x^{3/2})$. The quantity $(\hat{u}_\beta - u)^3 S_2'''(u_\beta^*)$ is of order $O_p(\phi_x^{3/2})$ and its bias under uniform integrability is of order $O(\phi_x^{3/2})$.

Proof. First,

$$S_2'' = S_0'' - a_1''T_1 - 2a_1'T_1' - a_1T_1'' - a_2''T_2 - 2a_2'T_2' - a_2T_2''.$$

Then, since S_2'' is unbiased, we may replace each term with its centered version, i.e. $S_0'' - \mathrm{E}(S_0'')$, $a_1'\{T_1' - E(T_1')\}$. Then $S_0'' - \mathrm{E}(S_0'') = O_p(1)$, $T_1^{(k)} - E(T_1^{(k)}) = O_p(\phi_x^{-1/2})$, and $T_2^{(k)} - E(T_2^{(k)}) = O_p(\phi_x^{-3/2})$, $k = 1, 2$. Write $S_2'' = \{S_0'' - \mathrm{E}(S_0'')\} + \{S_2'' - S_0'' + \mathrm{E}(S_0'')\}$. Order-of-magnitude bookkeeping shows that $S_2'' - S_0'' + \mathrm{E}(S_0'') = O_p(\phi_x^{1/2})$. Therefore, $(\hat{u}_\beta - u)^2 S_2'' = O_p(\phi_x)$. Applying Lemma 3, write $(\hat{u}_\beta - u)^2 S_2'' = D_{11}^{-2} T_{1x}^2 \{S_0'' - \mathrm{E}(S_0'')\} + O_p(\phi_x^{3/2})$ The expected value of the first term in the foregoing expression is 0 by independence.

For the second result, straightforward computations show $S_2''' = O_p(1)$. Also, $(\hat{u}_\beta - u)^3 = O_p(\phi_x^{3/2})$; assuming $S_2(u)$ is sufficiently smooth in u such that $S_2'''(u^*) = O_p(1)$, then $(\hat{u}_\beta - u)^3 S_2'''(u^*) = O_p(\phi_x^{3/2})$, completing the proof.

Proof of Theorem 1. Let $Z = D_{11}^{-1}\tilde{S}_2'T_1$. Using expansions (8) and (9), and applying Corollary 7 and Lemmas 8 and 9, the proof is immediate.

APPENDIX C: Algorithm for solving $\hat{S}_2 = 0$

The equation $\hat{S}_2(\beta) = 0$ can be solved by iterating two steps: (1) For fixed $\hat{u}_i$s, take one step in the solution of $\sum_i S_{i2}(\beta, \hat{u}_i) = 0$ for β. (2) For fixed $\hat{\beta}$ and for each i, take one step in the solution of $T_{i1}^*(\hat{\beta}, u_i) = 0$ for u_i. The first step is implemented using a Fisher-scoring algorithm, exploiting the components of the matrix D_i. The precision matrix for β inferences is the first set of columns $D_{i+0} = (D_{i00}, D_{i01}, D_{i02})^T$, of D_i. Defining the matrix $L_{i2} = (I_p, -a_i)$, we have $S_{i2} = L_{i2}(S_{i0}^T, T_{i1}, T_{i2})^T$, and hence $\mathrm{E}\{-(\partial S_{i2}/\partial\beta) = L_{i2}D_{i+0} = D_{i2}$. Estimating D_{i2} by plugging in $\hat{\beta}^{(o)}$ and $\hat{u}_i^{(o)}$, $\hat{\beta}^{(o)}$ is updated to $\hat{\beta}^{(n)}$ with

$$\hat{\beta}^{(n)} = \hat{\beta}^{(o)} + \left\{\sum_i D_{i2}(\hat{\beta}^{(o)})\right\}^{-1} \hat{S}_2(\hat{\beta}^{(o)}).$$

In the second step, we use a Newton-Raphson scheme to update $\hat{u}_i^{(o)}$ to $\hat{u}_i^{(n)}$ using the observed rather than the expected derivatives of T_{i1}^* with respect to u_i. We set the maximum number of iterations between (1) and (2) at 100.

APPENDIX D: Proof of Section 4.3 result

By standard estimating function theory $(\hat{\theta} - \theta) = -(\partial U_0/\partial\theta)^{-1}U_0 + o_p(\sqrt{n})$. By Taylor-series expansion,

$$\begin{aligned}\hat{S}_2(\beta, \hat{\theta}) &= \hat{S}_2(\beta, \theta) - \left(-\frac{\partial \hat{S}_2}{\partial\theta}\right)\left(-\frac{\partial U_0}{\partial\theta}\right)^{-1} U_0 + o_p(\sqrt{n}) \\ &= \hat{S}_2(\beta, \theta) - CD^{-1}U_0 + o_p(\sqrt{n}).\end{aligned}$$

Since $\mathrm{E}(U_0) = 0$ and $-(\partial U_0/\partial\beta) = 0$, it is immediate that
$\mathrm{E}\{-(\partial\hat{S}_2(\beta, \hat{\theta})/\partial\beta)\} = \mathrm{E}\{-(\partial\hat{S}_2(\beta, \theta)/\partial\beta)\} + o(n)$, whose limit (scaled by n) is A. Also,

$$\mathrm{E}\{\hat{S}_2(\beta, \hat{\theta})\hat{S}_2(\beta, \hat{\theta})^T\} = \mathrm{E}\left[\{\hat{S}_2(\beta, \theta) - CD^{-1}U_0\}\{\hat{S}_2(\beta, \theta) - CD^{-1}U_0\}^T\right] + o(n),$$

whose limit (scaled by n) is B^*. The result that $\sqrt{n}(\hat{\beta} - \beta^*) \xrightarrow{d} N\{0, A^{-1}B^*(A^{-1})^T\}$ then follows from standard estimating function theory.

References

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

Carroll, R.J., and Stefanski, L.A. (1990), "Approximate Quasi-likelihood Estimation in Models with Surrogate Predictors," *Journal of the American Statistical Association*, 85, 652-663.

Carroll, R.J., and Wand, M.P. (1991), "Semiparametric Estimation in Logistic Measurement Error Models," *Journal of the Royal Statistical Society*, Ser. B, 53, 573-585.

Cook, J., and Stefanski, L.A. (1995), "A simulation extrapolation method for parametric measurement error models," *Journal of the American Statistical Association*, 89, 1314-1328.

Crowder, M. (1987), "On Linear and Quadratic Estimating Functions," *Biometrika*, 74, 591-597.

Efron, B., and Tibshirani, R.J. (1993), *An Introduction to the Bootstrap*, New York: Chapman and Hall.

Godambe, V.P. (1960), "An Optimum Property of Regular Maximum Likelihood Estimation," *Annals of Mathematical Statistics*, 31, 1208-1211.

Godambe, V.P., and Thompson, M.E. (1974), "Estimating Equations in the Presence of a Nuisance Parameter," *Annals of Statistics*, 2, 568-571.

Lindsay, B. (1982), "Conditional Score Functions: Some Optimality Results," *Biometrika*, 69, 503-512.

Lindsay, B.G. (1985), "Using Empirical Partially Bayes Inference for Increased Efficiency," *The Annals of Statistics*, 13, 914-931.

McBurnett, K., Lahey, B.B., Rathouz, P.J., and Loeber, R. (2000). "Low Salivary Cortisol and Persistent Aggression in Boys referred for Disruptive Behavior," *Archives of General Psychiatry*, 57, 38-43.

McCullagh, P. (1983), "Quasi-likelihood Functions," *Annals of Statistics*, 11, 59-67.

McCullagh, P., and Nelder, J.A. (1989), *Generalized Linear Models* (2nd ed.). London: Chapman and Hall.

Rathouz, P.J., and Liang, K-Y. (1999), "Reducing Sensitivity to Nuisance Parameters in Semiparametric Models: A Quasiscore Method," *Biometrika*, 86, 857-869.

Small, C.G., and McLeish, D.L. (1989), "Projection as a Method for Increasing Sensitivity and Eliminating Nuisance Parameters," *Biometrika*, 76, 693-703.

Small, C.G., and McLeish, D.L. (1994), *Hilbert Space Methods in Probability and Statistical Inference*, New York: John Wiley and Sons.

Stefanski, L.A. (1985), "The Effects of Measurement Error on Parameter Estimation," *Biometrika*, 72, 583-592.

Stefanski, L.A., and Carroll, R.J. (1987), "Conditional Scores and Optimal Scores for Generalized Linear Measurement Error Models," *Biometrika*, 74, 703-716.

Waterman, R.P., and Lindsay, B.G. (1996a), "Projected Score Methods for Approximating Conditional Scores," *Biometrika*, 83, 1-13.

Waterman, R.P., and Lindsay, B.G. (1996b), "A Simple and Accurate Method for Approximate Conditional Inference Applied to Exponential Family Models," *Journal of the Royal Statistical Society*, Ser. B, 58, 177-188.

Wedderburn, R.W.M. (1974), "Quasi Likelihood Functions, Generalized Linear Models and the Gauss-Newton Method," *Biometrika*, 61, 439-447.

Table 1. Simulation Study of Overdispersed Poisson Model with Intercept and Mismeasured Covariate. 500 Replicates.

Model			% bias in $\hat{\beta}_1$			% $\mathrm{CV}^2(\hat{\beta}_1)$		MSE	Cov.
$\exp(\beta_1)$	ϕ_y	ϕ_x	$\hat{S}_0$	RC	$\hat{S}_2$	RC	$\hat{S}_2$	Rat.	%
			$u_i \sim$ normal						
1.5	1.5	0.3	-20.6	3.4	2.5	5.5	5.5	101	94.8
1.5	1.5	0.7	-42.4	-1.2	-4.0	8.2	7.7	104	95.0
1.5	3.0	0.3	-21.0	2.9	3.2	12.5	13.6	92	92.6
1.5	3.0	0.7	-41.4	0.9	1.5	16.8	20.1	84	92.8
3.0	1.5	0.3	-25.2	-2.6	-1.4	0.9	1.2	80	91.0
3.0	1.5	0.7	-43.5	-3.2	-7.4	1.7	1.5	87	82.6
3.0	3.0	0.3	-24.4	-1.7	2.0	1.3	2.1	61	94.8
3.0	3.0	0.7	-43.4	-2.9	-4.0	2.3	2.5	89	86.2
			$u_i \sim$ uniform						
1.5	1.5	0.3	-23.5	-0.3	1.3	5.7	6.6	87	94.4
1.5	1.5	0.7	-44.1	-4.3	-4.3	8.3	8.9	94	92.0
1.5	3.0	0.3	-23.2	0.1	2.7	12.0	14.4	82	94.4
1.5	3.0	0.7	-41.7	-0.1	4.4	15.1	20.1	74	92.6
3.0	1.5	0.3	-32.9	-12.6	-2.0	0.9	1.5	161	93.6
3.0	1.5	0.7	-51.0	-16.1	-8.9	1.5	1.8	154	82.2
3.0	3.0	0.3	-33.0	-12.6	1.3	1.2	3.3	85	91.8
3.0	3.0	0.7	-51.7	-17.2	-6.2	2.0	3.5	128	82.6

NOTE: Full model is $\mu_y = \beta_0 + \beta_1 u$, with $\beta_0 = 0$. Variable u has standard deviation one; distribution is given in the text. $\mathrm{CV}^2(\hat{\beta}_1)$ is the squared coefficient of variation, relative to the true β_1. MSE Ratio is the mean squared error of the RC estimator relative to the $\hat{S}_2$ estimator. Coverage percent is for nominal 95% Wald-type confidence intervals for β_1 using $\hat{S}_2 = 0$ and variance estimator in Section 4.3.

Table 2. Simulation Study of Overdispersed Poisson Model with Intercept, Binary Covariate and Mismeasured Covariate. 500 Relicates.

Model				% bias in $\hat{\beta}_1$			% $\mathrm{CV}^2(\hat{\beta}_1)$		MSE	Cov.
$\exp(\beta_1)$	ρ_{wu}	ϕ_y	ϕ_x	$\hat{S}_0$	RC	$\hat{S}_2$	RC	$\hat{S}_2$	Rat.	%
1.5	0.34	1.5	0.3	-23.0	3.2	-0.6	2.0	2.0	103	94.6
1.5	0.34	1.5	0.7	-40.8	6.4	-2.1	2.8	2.6	122	94.4
1.5	0.34	3.0	0.3	-23.1	3.1	-0.2	3.5	3.7	99	96.8
1.5	0.34	3.0	0.7	-40.6	7.4	1.5	5.9	6.7	97	94.6
1.5	0.61	1.5	0.3	-24.3	12.1	-0.8	2.7	2.2	188	93.0
1.5	0.61	1.5	0.7	-42.1	24.3	-1.1	5.1	3.3	332	94.2
1.5	0.61	3.0	0.3	-23.6	13.2	0.8	5.2	4.5	153	94.0
1.5	0.61	3.0	0.7	-43.2	21.6	-0.3	8.2	6.8	190	96.2
3.0	0.34	1.5	0.3	-29.0	-4.8	-1.3	0.3	0.4	117	93.0
3.0	0.34	1.5	0.7	-46.9	-4.3	-8.0	0.8	0.6	78	76.4
3.0	0.34	3.0	0.3	-29.4	-5.5	0.8	0.4	0.9	79	93.4
3.0	0.34	3.0	0.7	-46.9	-4.7	-4.5	0.9	1.2	78	87.0
3.0	0.61	1.5	0.3	-28.9	5.4	-1.2	0.5	0.4	190	92.2
3.0	0.61	1.5	0.7	-46.4	14.8	-7.8	1.6	0.6	323	76.4
3.0	0.61	3.0	0.3	-28.6	5.7	2.3	0.6	1.1	84	92.6
3.0	0.61	3.0	0.7	-46.2	15.2	-2.6	1.9	1.4	289	86.0

NOTE: $\mu_y = \beta_{00} + \beta_{01} w_1 + \beta_1 u$, with $\beta_{00} = 0$, $\beta_{01} = \log(1.5)$. Variables w_1 and u have standard deviation one; distribution is given in the text. $\mathrm{CV}^2(\hat{\beta}_1)$ is the squared coefficient of variation, relative to the true β_1. MSE Ratio is the mean squared error of the RC estimator relative to the $\hat{S}_2$ estimator. Coverage percent is for nominal 95% Wald-type confidence intervals for β_1 using $\hat{S}_2 = 0$ and variance estimator in Section 4.3.

Table 3. Log-linear Models of Mean Symptom Counts in Cortisol Study

		Estimator			
		$\hat{S}_0$		$\hat{S}_2$	
Aggressive CD	age	0.057	[-0.095,0.180]	0.043	[-0.189,0.266]
	log(cortisol)	-0.45	[-0.67,-0.30]	-0.76	[-1.05,-0.49]
	$\hat{\phi}_y$	1.53		0.89	
Covert CD	age	0.145	[-0.091,0.279]	0.152	[-0.146,0.400]
	log(cortisol)	-0.23	[-0.52,0.16]	-0.40	[-1.13,0.30]
	$\hat{\phi}_y$	4.56		3.56	

NOTE: Entries are parameter estimates [95% BCa bootstrap confidence intervals] obtained by solving $\hat{S}_0 = 0$ or $\hat{S}_2 = 0$. Age is in years. Log(cortisol) is standardized to have population standard deviation one.