TECHNICAL REPORT — DO NOT CITE

Reducing sensitivity to nuisance parameters in semiparametric models: A quasiscore method

By Paul J. Rathouz
Department of Health Studies, University of Chicago,
5841 S. Maryland Ave., MC 2007, Chicago, Illinois 60637, U.S.A.
prathouz@health.bsd.uchicago.edu

and Kung-Yee Liang
Department of Biostatistics, Johns Hopkins University,
615 N. Wolfe St., Baltimore, Maryland 21205, U.S.A.
kyliang@jhsph.edu

SUMMARY

This paper proposes a semiparametric extension of the projected score method of Waterman & Lindsay (1996) for the elimination of nuisance parameters. The procedure addresses cases where only the mean and the variance of the response variable are specified and where the mean function involves both parameters of interest and nuisance parameters. Important applications of the semiparametric model include quasilikelihood models for matched designs and for measurement error models (Carroll & Stefanski, 1990). Due to the optimality and information-unbiasedness of the quasiscore function, a second-order quasiscore basis of estimating functions for the nuisance parameter is derived. Second-order locally ancillary estimating functions (Small & McLeish, 1994) are then obtained by solving a simple linear system that corresponds to a true projection for canonical exponential family distributions. Asymptotic arguments and simulation work show that the impact of nuisance parameters is considerably reduced when adopting the proposed approach.

Some key words: Bhattacharyya basis; Bias correction; Estimating function; Measurement error; Neyman-Scott problem; Quasilikelihood.

1 Introduction

In Neyman-Scott (1948) problems, a model for observations $Y_1, \ldots, Y_n$ from independent strata $i = 1, \ldots, n$ is given in terms of a parameter of interest θ common to all strata, and stratum-specific nuisance parameters ϕ_i on each of which a limited amount of information is available. Specifically, $y_i \sim f_i(\cdot; \theta, \phi_i)$. In this paper, using the framework of quasilikelihood (Wedderburn, 1974; McCullagh, 1983), we consider inferences for such independent-strata semi-parametric models in which $E(Y_i) = \mu_i(\theta, \phi_i)$.

Semi-parametric inferences on θ in the presence of ϕ_i can often be obtained through the unbiased estimating functions of the form

$$g(\theta, \phi_1, \ldots, \phi_n) = \sum_{i=1}^{n} g_i(\theta, \phi_i; y_i). \tag{1}$$

For example, in the regression setting, g_i would take the form $g_i = g_1(\theta, \phi_i; y_i, x_i)$, where the g_i differ only through the covariates x_i. For review, see Liang & Zeger, 1995; Desmond, 1997; and discussions and references therein. Ideally g_is are available such that

$$E\{g_i(\theta, \phi_i; Y_i); \theta, \phi_i^*\} = 0, \quad \text{for all } \phi_i^* \neq \phi. \tag{2}$$

Estimating functions that are robust in the sense of (2) are called "globally ancillary" (Small & McLeish, 1994). If a class of such functions is available, attention can then be turned to increasing efficiency for θ-inferences within that class (e.g. Godambe, 1976, 1980, 1984; Lindsay, 1982, 1985).

Globally ancillary estimating functions are available for some important classes of nuisance parameter problems. Examples include the quasilikelihood score (Wedderburn, 1974) and generalized estimating equations (Liang and Zeger, 1986) for cases in which ϕ_i does not appear in the mean $\mu_i(\theta)$. Alternatively, when ϕ is a canonical parameter in an exponential family model, the conditional score function is globally ancillary for ϕ_i (Godambe, 1976; Lindsay, 1982).

Unfortunately, when μ_i depends upon ϕ_i, and the conditional score is not available, it is often difficult to obtain globally ancillary estimating functions. For more general problems of this type, Waterman & Lindsay (1996) have developed an elegant theory of projected scores which approximate the conditional score when it exists and emulate it in terms of robustness to ϕ_i when it does not. Their method, which is fully parametric in (θ, ϕ_i), achieves a type of "local ancillarity" (Small & McLeish, 1994). This is further examined in § 2.1.

The semiparametric problem in which a full likelihood $f_i(\cdot; \theta, \phi)$ is not easily specified, but μ_i depends on ϕ_i, has received less attention. This report addresses this case by extending the Waterman-Lindsay (WL) method to the quasilikelihood setting, resulting in a general method for reducing sensitivity to nuisance parameters appearing in the mean function $\mu_i(\theta, \phi_i)$. These results also establish the robustness of the WL method in canonical exponential family problems in which the mean and variance models are correct, but other parts of the model might be misspecified. Two important independent-strata problems serve as motivating applications.

Example 1. Matched pair study with arbitrary link and variance functions. Let y_{i1} and y_{i2} be independent observations with means μ_{i1} and μ_{i2}, such that $h(\mu_{ij}) = \phi_i + \theta^T x_{ij}$, and variances $\nu V(\mu_{ij})$, where $h(\cdot)$ and $V(\cdot)$ are known functions and ν is the dispersion parameter (Williams, 1982). The nuisance ϕ_i is a stratum-level intercept in the regression of y_{i1} and y_{i2}, while θ provides a within-stratum contrast between y_{i1} and y_{i2} that is common to all strata.

Example 2. Errors-in-covariates models with arbitrary link and variance functions. Let Y_i be a random variable with mean $\mu_i = h_1^{-1}(\beta_0 + \beta_1 \phi_i)$ and variance $\nu V(\mu_i)$. Let the surrogate U_i be a mismeasured version of covariate ϕ_i, which is viewed as a fixed nuisance parameter (Stefanski & Carroll, 1987). Let $h_2[E\{h_3(U_i)\}] = \phi_i$ and $\text{var}\{h_3(U_i)\}$ be some known function of θ and ϕ_i (Carroll & Stefanski, 1990).

Here, $h_j(\cdot)$, $j = 1, 2, 3$ and $V(\cdot)$ are known functions. To place the problem in the independent-strata framework, define $y_{i1} = y_i$ and $y_{i2} = h_3(u_i)$ so that each observation (y_{i1}, y_{i2}) comprises its own stratum, with common regression parameter θ, and stratum-specific nuisance ϕ_i.

Example 3. Modelling within-stratum variance as a function of the mean. Suppose y^*_{i1} and y^*_{i2} are independent observations each with mean ϕ_i and variance $\sigma_i^2 = h(\theta, \phi_i)$, where $h(\cdot)$ is a known function. Interest is on the parameter θ, creating a simple extension of the classical Neyman-Scott (1948) problem of estimation of the common variance from paired observations (see also Jewell and Raab, 1981). The Neyman-Scott solution would suggest letting $y_{i1} = (y^*_{i1} - y^*_{i2})^2/2$ and $y_{i2} = (y^*_{i1} + y^*_{i2})/2$, and using y_{i1} for inference on θ. But $E(Y_{i1}) = h(\theta, \phi_i)$ contains ϕ_i, for which y_{i2} serves as an unbiased estimator that is uncorrelated with y_{i1}. Examples of $h(\cdot)$ include $h(\theta, \phi_i) = \theta V(\phi_i)$, where θ is an overdispersion parameter, and the scale regression model $h(\theta, \phi_i) = \exp(\theta_0 + \theta_1 \phi_i)$. Interestingly, viewing y_{i2} as a mismeasured surrogate for ϕ_i, the problem becomes a special case of a measurement error model.

The theory in this report applies to the stratum-specific estimating function $g_i = g_i(\theta, \phi_i; y_i)$. Presumably stratum sample sizes are fixed and asymptotic arguments would be based on summing over a large number of strata, as in (1). Focusing on finite sample robustness properties of g_i, in § 2 we review the WL projection method for obtaining locally ancillary estimating functions. We re-derive their method in a way that may be extended to settings where projection is not feasible. § 3 develops the quasiscore analogue to the WL approach. In § 4, we appeal to within-stratum asymptotic arguments to study the performance of estimating functions when an estimate, $\hat{\phi}_\theta$ is substituted for ϕ. § 5 contains some simulation results, followed by a discussion in § 6. In general, the subscript j will indicate within-stratum observations, so that y_{ij} is the jth observation in the ith stratum. Until § 5, the stratum subscript

i will be suppressed.

2 Locally Ancillary Estimating Functions

2.1 Definitions and properties

Let Y be a random vector distributed according to density $f(\cdot)$, belonging to a class $\mathcal{F}$ of densities with common support $\mathcal{Y}$. A semi-parametric model for Y is a partition of $\mathcal{F}$ indexed by the parameter $(\theta, \phi) \in \Theta \times \Phi$. Throughout this paper, θ is a $(p \times 1)$ vector parameter of scientific interest, ϕ is a scalar nuisance parameter, Θ is an interval of $\mathcal{R}^p$, and Φ is an interval of $\mathcal{R}^1$. For inferences on θ, assume existence of a $(p \times 1)$ unbiased estimating function $g(y; \theta, \phi)$ (Godambe, 1960; Durbin, 1960), and that $\mathcal{F}$ and g satisfy the standard information and regularity conditions given in Small & McLeish (1994, Ch. 4). Interest is on θ-inferences that are robust to ϕ.

When no globally ancillary g is available, local ancillarity, which we now define, serves as a sensible alternative robustness criterion. Consider the functional

$$b_k(g) \equiv \left[\frac{\partial^k}{\partial {\phi^*}^k} E\{g(\phi); \phi^*\} \right]_{\phi^*=\phi}.$$

Small & McLeish (1994, Ch. 4) defined an estimating function g to be "r-th order locally ancillary" if $b_k(g) = 0$, for $k = 0, 1, \ldots, r$. Under mild regularity conditions, this is equivalent to $E\{g(\phi); \phi^*\} = o\{(\phi^* - \phi)^r\}$, so that global ancillarity is the limiting form of local ancillarity as $r \to \infty$, the index r describing the degree of dependence of g on ϕ. Note that $b_k(\cdot)$ is a linear operator in the sense that $b_k(a_1 g_1 + a_2 g_2) = a_1 b_k(g_1) + a_2 b_k(g_2)$, where the $a_l = a_l(\theta, \phi)$s do not contain y. Under some regularity conditions, an alternative representation of $b_k(g)$ is

$$b_k(g) = E\{gV_k(\phi)\}, \tag{3}$$

where $V_k(\phi)$ is the unbiased estimating function for ϕ given by

$$V_k = \frac{\partial^k f(x;\theta,\phi)/\partial\phi^k}{f(x;\theta,\phi)}. \tag{4}$$

Thus, to say that g is rth order locally ancillary is to say that it is uncorrelated with the first r ϕ-scores V_k from the model generating Y. We note that the V_ks satisfy the recursive relationship

$$V_r = \frac{\partial}{\partial\phi}V_{r-1} + V_{r-1}V_1, \tag{5}$$

for $r = 2, 3, \ldots$ (WL, 1996), and the fact that V_2 is unbiased reflects the information-unbiasedness of V_1 (Lindsay, 1982). A final representation of $b_k(g)$ for $k = 1, 2$ is given by the following lemma, which is proved in the appendix.

LEMMA 1 *Let $g(y;\theta,\phi)$ be an unbiased estimating function for θ as defined above and satisfying regularity (smoothness) conditions given in situ. Then,*

$$E\left(-\frac{\partial g}{\partial\phi}\right) = b_1(g) \quad \textit{and} \quad E\left(\frac{\partial^2 g}{\partial\phi^2}\right) = b_2(g) - 2\frac{\partial b_1(g)}{\partial\phi}.$$

2.2 Locally ancillary estimating functions via $\mathcal{L}^2$ projection

Waterman & Lindsay (1996) provide a constructive method for obtaining rth-order locally ancillary estimating functions in the fully parametric model $\mathcal{F} = \{f(\cdot;\theta,\phi) : (\theta,\phi) \in \Theta \times \Phi\}$. The main idea is to exploit the equivalence (3) to render an arbitrary estimating function, $g_0 = g_0(\theta,\phi;y)$, rth-order locally ancillary. Define the "Bhattacharyya basis of order r" to be the vector $\mathcal{V}_r = (V_1,\ldots,V_r)^T$ of the first r ϕ-scores (4). Let Λ_r^F denote the space spanned by $\mathcal{V}_2$, and let $\Pi_r^F g_0$ be the $\mathcal{L}^2$ projection of g_0 onto Λ_r^F given by $\Pi_r^F g_0 = E(g_0\mathcal{V}_r^T)\left\{E(\mathcal{V}_r\mathcal{V}_r^T)\right\}^{-1}\mathcal{V}_r$. Then via standard $\mathcal{L}_2$ projection theory (Small & McLeish, 1994), $g_r = g_0 - \Pi_r^F g_0$ is the estimating function that is maximally correlated with the original g_0, yet uncorrelated with $V_1,\ldots V_r$. That g_r is rth order locally ancillary is a direct result of (3). Setting $g_0 = U_0 = (\partial f/\partial\theta)/f$, the projected scores $U_r = U_0 - \Pi_r^F U_0$, are the efficient rth-order locally ancillary estimating functions, $r = 1, 2, \ldots$ (WL, 1996).

The importance of this result is that we can begin with a relatively θ-efficient kernel g_0, and render it insensitive to ϕ to an arbitrary order r. Since the V_rs are naturally arranged in descending order of information content for ϕ, Λ_r^F is "automatically" the best finite basis of dimension r in the sense of preserving efficiency while gaining robustness. For $r = 2$, WL (1996) show that the θ-information in U_2 is very close to that in the limiting U_∞. Also for $r = 2$, consider the "plug-in" estimating function $\hat{g}_2(\theta) = g_2(\theta, \hat{\phi}_\theta)$ obtained using the estimator $\hat{\phi}_\theta$ that solves $V_1(\phi) = 0$ for fixed θ. While $\hat{g}_2$ will generally be biased, $\hat{g}_2 - g_2$ converges in law to a mean-zero random variable as $n \to \infty$ (Small & McLeish, 1989; unpublished report by Lindsay & Waterman, Pennsylvania State University, February 26, 1992). These results suggest that important robustness properties already obtain for $r = 2$.

There is a more direct approach to deriving g_r that does not rely on projection theory. First, using the linear operators $b_k(\cdot)$, $k = 1, \ldots, r$, define the $p \times r$ functional $b_{(r)}(g) = \{b_1(g), \ldots, b_r(g)\}$. Then, applying (3) to $\mathcal{V}_r$ and g, $\Pi_r^F g = b_{(r)}(g) \left\{b_{(r)}(\mathcal{V}_r)\right\}^{-1} \mathcal{V}_r$. Now, consider the class of unbiased estimating functions in the span of $(g, V_1, \ldots, V_r)^T$, i.e. of the form

$$g - a\mathcal{V}_r, \tag{6}$$

over all $p \times r$ matrices $a = a(\theta, \phi)$. Since $b_{(r)}(\cdot)$ is a linear operator, $b_{(r)}(g - a\mathcal{V}_r) = b_{(r)}(g) - a\, b_{(r)}(\mathcal{V}_r)$. If $b_{(r)}(\mathcal{V}_r)$ is invertible, then $a = a_r = b_{(r)}(g) \left\{b_{(r)}(\mathcal{V}_r)\right\}^{-1}$ is the unique matrix a that produces a rth locally ancillary estimating function of the form (6); hence g_r. Note, however, that the same procedure would apply to other bases $\mathcal{V}_r^*$, say, provided that one could compute $b_{(r)}(g)$ and $b_{(r)}(\mathcal{V}_r^*)$. This would allow one to avoid specifying the likelihood needed to compute $\mathcal{V}_r$. In the next section, we exploit this idea for the case of $r = 2$, using a quasiscore basis for ϕ instead of $\mathcal{V}_2$.

3 Locally Ancillary Quasiscores

We now extend the projected score methodology of Waterman & Lindsay (1996) to models for the mean and variance of Y, concentrating on the important $r = 2$ case of second-order local ancillarity. In § 3.2, a candidate subspace of ϕ-information based on quasiscores is proposed after noticing that the true subspace $\mathcal{V}_2$ can be emulated using the information-unbiasedness property of the optimal quasiscore. Then, by viewing projection onto a nuisance subspace as simply solving a linear system, the WL procedure is generalized to less efficient representations of the nuisance subspace. This idea is exploited in § 3.3 to construct locally ancillary quasiscores without resorting to projection. § 3.4 contains the quantities necessary for computing the θ-information in S_2. Throughout, we maintain the symbols U and V for the true likelihood scores, and adopt S and T for the corresponding quasiscores.

3.1 The model and assumptions

Let $x_1, \ldots, x_n$ be a sequence of covariate vectors with arbitrary empirical distribution function $G_x(\cdot)$. Let $Y = (Y_1, \ldots, Y_n)^T$, where the Y_js are a sequence of random variables that are independent conditional on the x_js. As a special case of the semiparametric model $\mathcal{F}$ in § 2.1, let $\mu_j = \mu(\theta, \phi; x_j) = E(Y_j \mid x_j; \theta, \phi)$ and $v_j = v(\theta, \phi; x_j) = \mathrm{var}(Y_j \mid x_j; \theta, \phi)$. We assume that μ_j and v_j are finite, continuous, and admit continuous first and second derivatives with respect to (θ, ϕ). We also require that $(Y_j | x_j)$ have a finite fourth moment. For identifiability, assume that for any given one-dimensional subset of Θ, say Θ^*, parameterized by s, the model $\bar{\mu}(\theta, \phi) = (\mu(\theta, \phi; x_1), \ldots, \mu(\theta, \phi; x_n))^T$ is such that distinct values of (s, ϕ) imply a distinct $\bar{\mu}(\theta(s), \phi)$. We require within-stratum identifiability only on the subset Θ^*, assuming that information for identifiability on Θ accumulates across strata. These assumptions are not restrictive; for many models in practical usage, moments and their derivatives to several orders exist. The identifiability assumption simply requires that some joint information on (θ, ϕ) be available. For example, it serves to eliminate strata in Example 1 where only one y_{ij} is observed.

3.2 Quasi-Bhattacharyya basis

If ϕ were known, one might propose the θ-quasiscore (Wedderburn, 1974)

$$S_0 = \sum_j \left(\frac{\partial \mu_j^T}{\partial \theta}\right) \frac{y_j - \mu_j}{v_j} = \sum_j S_{0j}$$

for inferences on θ. Similarly, we define the ϕ-quasiscore

$$T_1 = \sum_j \left(\frac{\partial \mu_j}{\partial \phi}\right) \frac{y_j - \mu_j}{v_j} = \sum_j T_{1j}.$$

Note that T_1 is optimal for ϕ in the class of linear estimating functions and is thereby information unbiased (Crowder, 1987). Exploiting (5), the quasiscore analogue of V_2, namely $T_2 = \partial T_1/\partial\phi + T_1^2$, is another unbiased estimating function for ϕ. Letting prime ($'$) denote differentiation with respect to ϕ,

$$\begin{aligned} T_2 &= T_1' + T_1^2 = \left(\sum_j T_{1j}\right)' + \left(\sum_j T_{1j}\right)^2 \\ &= \sum_j T_{1j}' + \sum_j \sum_k T_{1j} T_{1k} = \sum_j T_{2j} + 2 \sum_{j<k} T_{1j} T_{1k}, \end{aligned}$$

where $T_{2j} = T_{1j}' + T_{1j}^2$. By information-unbiasedness of T_1 and T_{1j}, $E(T_2) = E(T_{2j}) = 0$, and the second ϕ-quasiscore for the jth observation is

$$T_{2j} = \frac{\partial}{\partial \phi}\left(\frac{\partial \mu_j}{\partial \phi} v_j^{-1}\right)(y_j - \mu_j) + \left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-2}\left\{(y_j - \mu_j)^2 - v_j\right\}.$$

A key thesis of this paper is that the quasiscores T_1 and T_2 form a partial basis for the nuisance subspace that shares many of the properties of the second order Bhattacharyya basis $\mathcal{V}_2$ when used for bias correction. Define the quasi-Bhattacharyya bases of order $r = 1$ and $r = 2$ to be $\mathcal{T}_1 = (T_1)$ and $\mathcal{T}_2 = (T_1, T_2)$, and define Λ_r^Q to be the span of $\mathcal{T}_r$. In the next section, we exploit the linear operator $b_{(r)}(\cdot)$ to derive rth-order locally ancillary quasiscores using S_0 and Λ_r^Q.

3.3 Construction of locally ancillary quasiscores

For $r = 1, 2$, let $\Pi_r^Q : \mathcal{L}^2 \to \Lambda_r^Q$ be the linear operator such that $(g - \Pi_r^Q g) \perp \mathcal{V}_r$. Then, by (3), $(g - \Pi_r^Q g)$ is rth-order locally ancillary. Note that Π_r^Q is not a true projection operator onto Λ_r^Q, since it does not minimize the distance between g and $\Pi_r^Q g$. Setting $g = S_0$, let a_1 be the $(p \times 1)$ vector such that $\Pi_1^Q S_0 = a_1 \mathcal{T}_1$ and let a_2 be the $(p \times 2)$ matrix such that $\Pi_2^Q S_0 = a_2 \mathcal{T}_2$.

Define $S_1 = S_0 - \Pi_1^Q S_0$ and $S_2 = S_0 - \Pi_2^Q S_0$. For S_1, set $b_1(S_1) = b_1(S_0) - a_1 \, b_1(T_1) = 0$ to obtain $a_1 = b_1(S_0)/b_1(T_1)$. Similarly, let $b_{(2)}(\cdot) = \{b_1(\cdot), b_2(\cdot)\}$, and set $b_{(2)}(S_2) = b_{(2)}(S_0) - a_2 \, b_{(2)}(\mathcal{T}_2) = 0$ to obtain $a_2 = b_{(2)}(S_0) \left\{ b_{(2)}(\mathcal{T}_2) \right\}^{-1}$. Two lemmas, proved in the appendix, now provide explicit expressions for the maps of S_0, T_1, and T_2 via the functionals $b_1(\cdot)$ and $b_{(2)}(\cdot)$.

LEMMA 2 *For the quasiscores S_0, T_1 and T_2 and the functional $b_1(\cdot)$ defined above,*

$$\begin{aligned} b_1(S_0) &= \sum_j \left(\frac{\partial \mu_j^T}{\partial \theta} \right) v_j^{-1} \left(\frac{\partial \mu_j}{\partial \phi} \right) \equiv \sum_j D_{01j} \equiv D_{01} \\ b_1(T_1) &= \sum_j \left(\frac{\partial \mu_j}{\partial \phi} \right) v_j^{-1} \left(\frac{\partial \mu_j}{\partial \phi} \right) \equiv \sum_j D_{11j} \equiv D_{11} \\ b_1(T_2) &= \sum_j \left(\frac{\partial \mu_j}{\partial \phi} \right) v_j^{-1} \left(\frac{\partial^2 \mu_j}{\partial \phi^2} \right) \equiv \sum_j D_{21j} \equiv D_{21}, \end{aligned}$$

where $D_{rs} = \sum_j D_{rsj}$, and the D_{rsj}s are defined implicitly.

LEMMA 3 *For the quasiscores S_0, T_1 and T_2 and the functional $b_2(\cdot)$ defined above,*

$$\begin{aligned}
b_2(S_0) &= \sum_j \left(\frac{\partial \mu_j^T}{\partial \theta}\right) v_j^{-1} \left(\frac{\partial^2 \mu_j}{\partial \phi^2}\right) \equiv \sum_j D_{02j} \equiv D_{02} \\
b_2(T_1) &= \sum_j \left(\frac{\partial \mu_j}{\partial \phi}\right) v_j^{-1} \left(\frac{\partial^2 \mu_j}{\partial \phi^2}\right) \equiv \sum_j D_{12j} \equiv D_{12} \\
b_2(T_2) &= \sum_j \left\{ \left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-2} \left(\frac{\partial^2 v_j}{\partial \phi^2}\right) + 2\left(\frac{\partial \mu_j}{\partial \phi}\right)^4 v_j^{-2} + \left(\frac{\partial^2 \mu_j}{\partial \phi^2}\right)^2 v_j^{-1} \right. \\
&\qquad \left. - \left(\frac{\partial \mu_j}{\partial \phi}\right)\left(\frac{\partial v_j}{\partial \phi}\right) v_j^{-2} \left(\frac{\partial^2 \mu_j}{\partial \phi^2}\right) \right\} + 4K_{11} \\
&\equiv \sum_j D_{22j} + 4K_{11} \equiv D_{22} + 4K_{11},
\end{aligned}$$

where $K_{11} = \sum_{j<j'} D_{11j} D_{11j'}$, and the D_{rsj}s are defined implicitly.

With Lemmas 2 and 3, it is established directly that the matrices a_1 and a_2 for $\Pi_1^Q S_0$ and $\Pi_2^Q S_0$ are given by $a_1 = D_{01}/D_{11}$ and

$$a_2 = (D_{01}, D_{02}) \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} + 4K_{11} \end{pmatrix}^{-1}. \tag{7}$$

The new quasiscores S_1 and S_2 are thus bias-corrected versions of the original quasiscore S_0. The following theorem, proved by the foregoing argument, establishes that S_1 and S_2 are first- and second-order locally ancillary, respectively.

THEOREM 4 *For the quasiscores S_1 and S_2, the matrices a_1 and a_2, and the operators $b_1(\cdot)$ and $b_2(\cdot)$ defined above, $b_1(S_1) = b_1(S_2) = b_2(S_2) = 0$ and*

$$b_2(S_1) = D_{02} - \frac{D_{01}}{D_{11}} D_{12} = O(n).$$

3.4 Precision, variance and information matrices

The quantities D_{rs} can be collected together into an "extended" precision matrix

$$D = E\left\{ \begin{pmatrix} S_0 \\ T_1 \\ T_2 \end{pmatrix} \left(U_0^T, V_1, V_2\right) \right\} = \left[E\left\{ -\frac{\partial}{\partial \theta} \begin{pmatrix} S_0 \\ T_1 \\ T_2 \end{pmatrix} \right\}, b_{(2)} \begin{pmatrix} S_0 \\ T_1 \\ T_2 \end{pmatrix} \right].$$

For a single observation j,

$$D_j = E\left\{\begin{pmatrix} S_{0j} \\ T_{1j} \\ T_{2j} \end{pmatrix}\left(U_{0j}^T, V_{1j}, V_{2j}\right)\right\} = \begin{pmatrix} D_{00j} & D_{01j} & D_{02j} \\ D_{10j} & D_{11j} & D_{12j} \\ D_{20j} & D_{21j} & D_{22j} \end{pmatrix}.$$

Then

$$D = \begin{pmatrix} D_{00} & D_{01} & D_{02} \\ D_{10} & D_{11} & D_{22} \\ D_{20} & D_{21} & D_{22} + 4K_{11} \end{pmatrix}, \tag{8}$$

where, for $r, s = 0, 1, 2$, $D_{rs} = \sum_j D_{rsj}$, $K_{11} = \sum_{j<j'} D_{11j} D_{11j'}$, D_{rsj} is given in Lemmas 2 and 3, and $D_{00j} = (\partial\mu_j^T/\partial\theta) v_j^{-1} (\partial\mu_j^T/\partial\theta)^T$. Note that all components of D are $O(n)$, except the lower right, which is $O(n) + O(n(n-1)) = O(n^2)$. Remarkably, D is a symmetric matrix. Analogously, define the variance matrices

$$M = E\left\{\begin{pmatrix} S_0 \\ T_1 \\ T_2 \end{pmatrix}\left(S_0^T, T_1, T_2\right)\right\} = \begin{pmatrix} D_{00} & D_{01} & M_{02} \\ D_{10} & D_{11} & M_{22} \\ M_{20} & M_{21} & M_{22} + 4K_{11} \end{pmatrix},$$

and M_j. Details of the components of M_j are in the appendix. The matrices D and M are useful in computing the θ-information in S_1 and S_2. For S_2, define: $D_{+0} = (D_{00}, D_{01}, D_{02})^T$, the "$\theta$-column" of D; and $L_2 = (I_p, -a_2) = (I_p, -a_{21}, -a_{22})$, a $p \times (p+2)$ matrix. Then, using (7), $E\{-(\partial S_2/\partial\theta)\} = E(S_2 U_0^T) = L_2 D_{+0}$ and $E(S_2 S_2^T) = L_2 M L_2^T$, so that the θ information in S_2 is

$$\mathcal{I}_{S_2} = D_{+0}^T L_2^T \left\{L_2 M L_2^T\right\}^{-1} L_2 D_{+0}. \tag{9}$$

Remark 1. While higher moments $\rho_j = E\{(Y_j - \mu_j)^3 \mid x_j; \theta, \phi\}$ and $\kappa_j = E\{(Y_j - \mu_j)^4 \mid x_j; \theta, \phi\}$ are in general required for the quasiscore variance M, S_2 itself in no way depends on ρ_j or κ_j. The roles of ρ_j and κ_j may be further reduced by using a robust estimator for the variance of S_2 (Royall, 1986; Liang & Zeger, 1986).

Remark 2. The stochastic order of the terms in S_1 and S_2 is of interest to understand the nature of the bias correction. Note that (a_{21}, a_{22}) can be written as

$$(a_{21}, a_{22}) = \left(\frac{D_{01}}{D_{11}} + O\left(n^{-1}\right), O\left(n^{-1}\right)\right)$$

and the ϕ-scores

$$\begin{pmatrix} T_1 \\ T_2 \end{pmatrix} = \begin{pmatrix} O_p(\sqrt{n}) \\ O_p(\sqrt{n}) + O_p(n) \end{pmatrix},$$

so that

$$(a_{21}, a_{22}) \begin{pmatrix} T_1 \\ T_2 \end{pmatrix} = \Pi_2^Q S_0 = \frac{D_{01}}{D_{11}} T_1 + O_p(1) = \Pi_1^Q S_0 + O_p(1).$$

Since S_0, like T_1, is $O_p(\sqrt{n})$ the difference between S_2 and S_1 becomes negligible as n gets large. For small n, when information for ϕ is sparse, the additional term $a_{22}T_2$ in $\Pi_2^Q S_0$ provides a bias-correction for imprecision in estimates of ϕ.

Remark 3. Since the forms of S_0, T_1, T_2, D and M are quite general, conversion of computational software from one model to another is simplified. A general fitting routine can be linked to separate subroutines for μ_j, v_j, ρ_j, v_j and their derivatives. New models only require new subroutines for μ_j, v_j, etc.

4 Asymptotics of the plug-in quasiscore $\hat{S}_2$

While the quasiscore S_2 is more robust to misspecification of ϕ than S_0 and S_1, ϕ still must be eliminated from the problem. This can be accomplished by using a plug-in score $\hat{S}_2(\theta) = S_2(\theta, \hat{\phi}_\theta)$, where $\hat{\phi}_\theta$ solves $T_1 = 0$ for fixed θ. § 4.1 establishes that as $n \to \infty$ the asymptotic bias in $\hat{S}_2$ is zero. This development also serves to illustrate the special roles that second-order local ancillarity and the optimality of the quasiscores S_0 and T_1 play in the elimination of ϕ. § 4.2 gives an approximation for the Fisher's information for θ in $\hat{S}_2$ that improves upon (9). Throughout this section, we assume that the covariates X_j are identically distributed, so that (Y_j, X_j) are jointly independent and identically distributed random variables. Matrices D and M are computed by taking expectation over X_j. The detailed proofs for this section follow similar arguments in Lindsay & Waterman (unpublished report, Pennsylvania State University, February 26, 1992) and WL (1996, appendix) and are in the appendix.

4.1 Asymptotic bias in $\hat{S}_2$

From standard asymptotic arguments,

$$n^{1/2}(\hat{\phi}_\theta - \phi) = n^{1/2} D_{11}^{-1} T_1 + O_p(n^{-1/2}) = O_p(1). \tag{10}$$

Then, letting prime ($'$) denote differentiation with respect to ϕ,

$$\begin{aligned} \hat{S}_2 - S_2 &= (\hat{\phi}_\theta - \phi) S_2' + \frac{1}{2}(\hat{\phi}_\theta - \phi)^2 S_2'' + O_p(n^{-1/2}) \\ &= D_{11}^{-1} S_2' T_1 + S_2' O_p(n^{-1}) + \frac{1}{2}\{n(\hat{\phi}_\theta - \phi)^2\}(n^{-1} S_2'') + O_p(n^{-1/2}). \end{aligned} \tag{11}$$

Now, to study equation (11), define the following: $S_0' = \sum_j S_{0j}'$; $T_1' = \sum_j T_{1j}'$; $L_3 = (I_p, -(a_1 - 2a_{22}D_{11}), -a_1')$, a $p \times (p+2)$ matrix; $P = \text{var}\{(S_0'^T, T_1', T_1)^T\}$. Note that a_1 and a_{21} are $O(1)$, that a_{22} is $O(n^{-1})$, and that their derivatives with respect to ϕ are of the same orders. Define $a_{22\bullet} = \lim_{n\to\infty} n a_{22}$. Now, note that due to Lemma 1 and Theorem 4, $E(S_2') = E(S_2'') = 0$. In the following lemma, we approximate S_2' as a sum of independent mean-zero random variables.

LEMMA 5 *The ϕ-derivative S_2' of S_2 is expressible as*

$$\begin{aligned} S_2' &= (S_0' + D_{01}) - a_1(T_1' + D_{11}) - (a_1' - 2a_{22\bullet}D_{111})T_1 + o_p(n^{1/2}) \\ &= \tilde{S}_2' + o_p(n^{1/2}) = \sum_j \tilde{S}_{2j}' + o_p(n^{1/2}), \end{aligned}$$

where $\tilde{S}_{2j}' = (S_{0j}' + D_{011}) - a_1(T_{1j}' + D_{111}) - (a_1' - 2a_{22\bullet}D_{111})T_{1j}$.

Since $(S_0' + D_{01})$, $(T_1' + D_{11})$, and T_1 are independent and identically distributed mean-zero sums, $S_2' = O_p(n^{1/2})$. Similarly, $S_2'' = O_p(n^{1/2})$. Since second-order local ancillarity of S_2 implies that $E(S_2' V_1) = 0$, we might ask if a similar result holds for $E(S_2' T_1)$ or $E(\tilde{S}_2' T_1)$. To that end, two remarkable information equalities obtain:

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*

$$E\{-(S_0' + D_{01})'\} = E\{(S_0' + D_{01})T_1\} \quad \textit{and} \quad E\{-(T_1' + D_{11})'\} = E\{(T_1' + D_{11})T_1\}.$$

Then, by Lemmas 1, 5 and 6 and the fact that $a_{22\bullet} = (D_{02}D_{11} - D_{01}D_{12})/(2D_{11}^2 D_{111})$, $D_{11}^{-1}E(\tilde{S}_2' T_1) = 0$. This second main result of the paper is synthesized in Theorem 7:

THEOREM 7 *Let* $\hat{\phi}_\theta$ *be the solution to* $T_1 = 0$ *for fixed* θ. *Then*

$$\hat{S}_2 = S_2(\hat{\phi}_\theta) = S_2 + D_{11}^{-1}\tilde{S}_2' T_1 + o_p(1),$$

where $D_{11}^{-1}\tilde{S}_2' T_1 = O_p(1)$ *and is unbiased. Furthermore, under uniform integrability,* $E(\hat{S}_2 - S_2) = o(1)$.

This is the same order of bias obtained by Small & McLeish (1989) using $V_1 = 0$ to estimate ϕ. The comparable performance of $T_1 = 0$ and S_2 is due to the approximate orthogonality of S_2' and T_1, and reflects the optimality of S_0 and T_1. We note as a point of comparison that $(\hat{S}_1 - S_1)$ is also $O_p(1)$, but with bias of order $O(1)$.

4.2 Approximate information in $\hat{S}_2$

Theorem 7 can be further exploited to obtain an improved estimator for the θ-information in $\hat{S}_2$. Write:

$$\begin{aligned} \{\hat{S}_2 - E(\hat{S}_2)\}\{\hat{S}_2 - E(\hat{S}_2)\}^T &= S_2 S_2^T + S_2 D_{11}^{-1}(S_2' T_1)^T + D_{11}^{-1}(S_2' T_1) S_2^T \\ &\quad + D_{11}^{-2}(S_2' T_1)(S_2' T_1)^T + o_p(1), \end{aligned} \tag{12}$$

where the first four terms on the right hand side are $O_p(n)$, $o_p(n)$, $o_p(n)$, and $O_p(1)$, respectively. Also,

$$\{\hat{S}_2 - E(\hat{S}_2)\}U_0^T = S_2 U_0^T + D_{11}^{-1}(S_2' T_1)U_0^T + o_p(1), \tag{13}$$

with terms of order $O_p(n)$ and $o_p(n)$. Taking expectations in (12) and (13) to obtain asymptotic variances and covariances gives $\text{avar}(S_2) = L_2 M L_2^T = O(n)$, $\text{avar}(D_{11}^{-1} S_2' T_1) = \text{avar}(S_2') D_{11}^{-1} = L_3 P L_3^T D_{11}^{-1} = O(1)$ and $\text{acov}(S_2, U_0^T) = L_2 D_{+0} = O(n)$. Furthermore, $\text{acov}(S_2, D_{11}^{-1} S_2' T_1) \approx \text{acov}(D_{11}^{-1} S_2' T_1, U_0^T) \approx 0$, since the multivariate normal distribution has zero skewness. Note that these are all quantities which, when appropriately scaled, can be consistently estimated. Finally, these results combine so that asymptotically (and under uniform integrability), $E\left(-\partial \hat{S}_2/\partial\theta\right) = E(\hat{S}_2 U_0) = E(S_2 U_0^T) + o(1)$ and and $\text{avar}(\hat{S}_2) = L_2 M L_2^T + L_3 P L_3^T D_{11}^{-1} + o(1)$. Therefore, the Fisher's information $\mathcal{I}_{\hat{S}_2}$ for θ in $\hat{S}_2$ might better approximated by

$$\mathcal{I}^*_{\hat{S}_2} = D_{+0}^T L_2^T \left\{ L_2 M L_2^T + L_3 P L_3^T D_{11}^{-1} \right\}^{-1} L_2 D_{+0},$$

which can be compared with $\mathcal{I}_{S_2}$ in (9). The inverse of these quantities can be used to estimate the variance of the estimator $\hat{\theta}_{\hat{\phi}}$ obtained by solving $\hat{S}_2 = 0$.

5 Example Models and Simulations

We now consider two example models, each characterized by the presence of a nuisance parameter ϕ_i in the mean function for each of strata $i = 1, \ldots, 10$. For each model, the empirical means of the plug-in scores $\hat{S}_1 = \sum_{i=1}^{10} \hat{S}_{1i}$ and $\hat{S}_2 = \sum_{i=1}^{10} \hat{S}_{2i}$ were assessed for departures from unbiasedness over 5000 replicates. The performance of two theoretical approximations for $\mathcal{I}_{\hat{S}_2}$ was also examined: the naive quantity $\mathcal{I}_{S_2}$, and the approximation $\mathcal{I}^*_{\hat{S}_2}$ developed in § 4.2. These values were compared with the empirical (simulated) information, $\mathcal{I}^{\text{emp}}_{\hat{S}_2}$, computed as the ratio of the square of the empirical mean of the numerical derivative of $\hat{S}_2$, and the empirical variance of $\hat{S}_2$.

Consider a paired data model (Example 1, §1) with constant coefficient of variation. For $j = 1, 2$, let y_{ij} be independently distributed with mean

$$\mu_{ij} = \{\phi_i + \theta_0 + \theta_1 I(j = 1)\}^{-1}, \tag{14}$$

and variance $\nu\mu_{ij}^2$, where $I(j = 1) = 1$ if $j = 1$ and zero otherwise, and θ_0 is a fixed and known offset. We considered parameter values $\theta_0 = (0.4, 0.1, -0.05)$, $\theta_1 = (0.1, 0.2, 0.5)$, and dispersions $\nu = (0.2, 0.6, 2.0)$. For the ϕ_is, we generated ten values from a uniform distribution on the $(0.1, 1.0)$ interval and fixed these over all replicates. Model (14) has the canonical link and the mean-variance relationship of the gamma regression model, which is taken as a working model for the third and fourth moments of y_{ij}. However, to confirm that our method relies only weakly on the specification of skewness and kurtosis, the data were generated using a log-normal distribution, with mean μ_{ij} and variance $\nu\mu_{ij}^2$. The results are presented in Table 1. First, $\hat{S}_1$ is negatively biased in almost all cases, indicating negative bias in the estimator $\hat{\theta}_1$. For many cases, $\hat{S}_2$ is not detectably biased ($|Z| \leq 1.96$), and in all cases, $\hat{S}_2$ is considerably less biased than $\hat{S}_1$. The quantities $\mathcal{I}_{S_2}$ or $\mathcal{I}^*_{\hat{S}_2}$ do not differ greatly from the empirical information. These results (not shown) did not vary considerably from those obtained with data generated under the true gamma regression model.

As a second example, consider an overdispersed (Williams, 1982) Poisson regression model (McCullagh & Nelder, 1989, Ch. 6) with a mismeasured covariate ϕ_i (Example 2, §1). Assume that surrogate $x_i = \phi_i + \delta_i$ where $\delta_i \sim N(0, \sigma_\delta^2)$. Let the mean and variance of the response y_i given the true covariate ϕ_i be:

$$\log(\mu_i) = \theta_0 + \theta_1\phi_i \quad \text{and} \quad v_i = \nu\mu_i. \tag{15}$$

Again holding θ_0 fixed and focusing on θ_1, $S_{0i} = \phi_i\nu^{-1}(y_i - \mu_i)$ and $T_{1i} = \theta_1\nu^{-1}(y_i - \mu_i) + \sigma_\delta^{-2}(x_i - \phi_i)$. In computing the information to account for overdispersion, working (but incorrect) models for the skewness and kurtosis of y_i were obtained by multiplying the standardized cumulants of the Poisson distribution by the appropriate power of v_i. We generated ten ϕ_is from a standard normal population and fixed them over all replicates. In model (15), we fixed $\theta_0 = 0$, and considered relative risk values $\exp(\theta_1) = (1.5, 2.0, 3.0)$, measurement error variance $\sigma_\delta^2 = (0.2, 0.5, 0.7)$, and overdispersion $\nu = (1.5, 2.0, 3.0)$. Overdispersion was introduced by generating Y_i as a $(3:7)$ Bernoulli mixture of two Poisson random variables such that the marginal mean and variance are μ_i and $\nu\mu_i$. The results (Table 2) again show that the considerable bias in $\hat{S}_1$ is almost completely eliminated through the use of $\hat{S}_2$. In comparison with the paired gamma example, however, $\mathcal{I}_{S_2}$ overestimates the θ_1-information by about 22%, on average. By contrast, the corrected information approximation $\mathcal{I}^*_{\hat{S}_2}$ underestimates the empirical information by about 6%, which would lead to slightly conservative interval inferences on θ_1.

6 Concluding Remarks

We make two brief comments. First, the quasi-Bhattacharyya basis and the asymptotic results strongly reflect the information optimality of the θ- and ϕ-quasiscores. Second, replacing the projection operator with a solution operator suggests a general procedure which could be useful in more complicated problems, including those involving multivariate data. To wit, if one has available estimating functions g_θ, and $h_\phi = (h_{1,\phi}, \ldots, h_{r,\phi})^T$, then one can obtain an rth-order locally ancillary estimating function $g_\theta - b_{(r)}(g_\theta)\{b_{(r)}(h_\phi)\}^{-1}h_\phi$ for robust inferences on θ.

APPENDIX

Proofs and Technical Complements

Proof of equation 3. For an unbiased estimating function $g = g(y; \theta, \phi)$,

$$
\begin{aligned}
b_k(g) &= \left\{ \frac{\partial^k}{\partial \phi^{*k}} \int g(\phi) f(\phi^*)\, dx \right\}_{\phi^* = \phi} \\
&= \left\{ \int g(\phi) \frac{\partial^k f(\phi^*)/\partial \phi^{*k}}{f(\phi^*)} f(\phi^*)\, dx \right\}_{\phi^* = \phi} \\
&= E\left\{ g \left(\frac{\partial^k f / \partial \phi^k}{f} \right) \right\}, \qquad \text{(A1)}
\end{aligned}
$$

the first two lines being equal by regularity.

Proof of lemma 1. For $r = 1$,

$$
0 = \frac{\partial}{\partial \phi} \int g f\, dx = \int g \frac{\partial f}{\partial \phi}\, dx + \int \frac{\partial g}{\partial \phi}\, dx = b_1(g) - E\left(-\frac{\partial g}{\partial \phi} \right). \qquad \text{(A2)}
$$

The first equality holds since $\int f g\, dx = 0$, the second is due to regularity, and the third is due to (A1). For the second equality of the lemma,

$$
\frac{\partial b_1(g)}{\partial \phi} = \frac{\partial}{\partial \phi} \int g \frac{\partial f}{\partial \phi}\, dx = \int \frac{\partial g}{\partial \phi} \frac{\partial f}{\partial \phi}\, dx + \int g \frac{\partial^2 f}{\partial \phi^2}\, dx,
$$

so

$$
E\left(\frac{\partial g}{\partial \phi} \frac{\partial \log f}{\partial \phi} \right) = \frac{\partial b_1(g)}{\partial \phi} - b_2(g).
$$

Now

$$
\begin{aligned}
0 = \frac{\partial^2}{\partial \phi^2} \int g f\, dx &= \int \left[\frac{\partial^2 g}{\partial \phi^2} f + \frac{\partial g}{\partial \phi} \frac{\partial \log f}{\partial \phi} f + \frac{\partial g}{\partial \phi} \frac{\partial \log f}{\partial \phi} f + g \frac{\partial^2 f}{\partial \phi^2} \right] dx \\
&= E\left(\frac{\partial^2 g}{\partial \phi^2} \right) + 2E\left(\frac{\partial g}{\partial \phi} \frac{\partial \log f}{\partial \phi} \right) + b_2(g) \\
&= E\left(\frac{\partial^2 g}{\partial \phi^2} \right) + 2\frac{\partial b_1(g)}{\partial \phi} - 2b_2(g) + b_2(g).
\end{aligned}
$$

The first equality is due to the fact that $\int df\, dx = 0$ for all ϕ; the second results from regularity. The result follows.

Proof of Lemma 3.

$$E[S_{0j}(\phi)\,;\phi^*] = \left(\frac{\partial \mu_j^T}{\partial \theta}\right) v_j^{-1}[\mu_j(\phi^*) - \mu_j(\phi)],$$

which implies

$$b_2(S_{0j}) = \left(\frac{\partial \mu_j^T}{\partial \theta}\right) v_j^{-1} \frac{\partial^2 \mu_j}{\partial \phi^2}.$$

The result follows for $S_0 = \sum_j S_{0j}$. A similar argument gives the expression for $b_2(T_1)$. For $b_2(T_2)$, first write

$$T_{2j} = \frac{\partial}{\partial \phi}\left(\frac{\partial \mu_j}{\partial \phi} v_j^{-1}\right)(y_j - \mu_j) + \left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-2}(y_j - \mu_j)^2 - \left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-1}, \tag{A3}$$

and note that

$$\frac{\partial}{\partial \phi}\left(\frac{\partial \mu_j}{\partial \phi} v_j^{-1}\right) = \frac{\partial^2 \mu_j}{\partial \phi^2} v_j^{-1} - \left(\frac{\partial \mu_j}{\partial \phi}\right)\left(\frac{\partial v_j}{\partial \phi}\right) v_j^{-2}. \tag{A4}$$

Then,

$$\begin{aligned}
E[T_{2j}(\phi)\,;\phi^*] &= \frac{\partial}{\partial \phi}\left(\frac{\partial \mu_j}{\partial \phi} v_j^{-1}\right)[\mu_j(\phi^*) - \mu_j(\phi)] \\
&\quad + \left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-2}[v_j(\phi^*) - v_j(\phi) + (\mu_j(\phi^*) - \mu_j(\phi))^2], \\
b_2(T_{2j}) &= \frac{\partial}{\partial \phi}\left(\frac{\partial \mu_j}{\partial \phi} v_j^{-1}\right)\frac{\partial^2 \mu_j}{\partial \phi^2} \\
&\quad + \left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-2}\frac{\partial^2 v_j}{\partial \phi^2} + 2\left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-2}\left(\frac{\partial \mu_j}{\partial \phi}\right)^2 \\
&= \left(\frac{\partial^2 \mu_j}{\partial \phi^2}\right)^2 v_j^{-1} - \left(\frac{\partial \mu_j}{\partial \phi}\right) v_j^{-2}\left(\frac{\partial v_j}{\partial \phi}\right)\frac{\partial^2 \mu_j}{\partial \phi^2} \\
&\quad + \left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-2}\frac{\partial^2 v_j}{\partial \phi^2} + 2\left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-2}\left(\frac{\partial \mu_j}{\partial \phi}\right)^2.
\end{aligned}$$

Also,

$$\begin{aligned}
E[T_{1j}(\phi)T_{1k}(\phi)\,;\phi^*] &= \left(\frac{\partial \mu_j}{\partial \phi}\right) v_j^{-1}[\mu_j(\phi^*) - \mu_j(\phi)] \\
&\quad \bullet \left(\frac{\partial \mu_k}{\partial \phi}\right) v_k^{-1}[\mu_k(\phi^*) - \mu_k(\phi)] \\
b_2(T_{1j}T_{1k}) &= 2\left(\frac{\partial \mu_j}{\partial \phi}\right) v_j^{-1}\left(\frac{\partial \mu_j}{\partial \phi}\right)\left(\frac{\partial \mu_k}{\partial \phi}\right) v_k^{-1}\left(\frac{\partial \mu_k}{\partial \phi}\right) \\
&= 2D_{11j}D_{11k}.
\end{aligned}$$

Now, write $T_2 = \sum_j T_{2j} + 2\sum_{j<k} T_{1j}T_{1k}$ and the result follows.

Derivation of variance matrix M. Consider the variance-covariance matrix M as a function of M_j for the jth observation. For the first three terms, $E(S_0S_0^T) = \sum_j M_{00j}$, $E(S_0T_1) = \sum_j M_{01j}$, and $E(T_1^2) = \sum_j M_{11j}$ by independence. Then

$$\begin{aligned} E(S_0T_2) &= E\left\{\left(\sum_j S_{0j}\right)\left(\sum_k T_{2k} + \sum_{k\neq k'} T_{1k}T_{1k'}\right)\right\} \\ &= \sum_j E(S_{0j}T_{2j}) + \sum_{j\neq k} E(S_{0j}T_{2k}) + \sum_{j,k\neq k'} E(S_{0j}T_{1k}T_{1k'}) \\ &= \sum_j M_{02j}, \end{aligned}$$

the last two terms in the second line being zero by independence. Similarly, $E(T_1T_2) = \sum_j M_{12j}$. Finally,

$$\begin{aligned} E(T_2^2) &= E\left\{\left(\sum_j T_{2j} + \sum_{k\neq k'} T_{1k}T_{1k'}\right)^2\right\} \\ &= \sum_j E(T_{2j}^2) + 2\sum_{j,k\neq k'} E(T_{2j}T_{1k}T_{1k'}) \\ &\quad + \sum_A E(T_{1j}T_{1j'}T_{1k}T_{1k'}) + 2\sum_{k\neq k'} E(T_{1k}^2T_{1k'}^2) \\ &= \sum_j M_{22j} + 4\sum_{j<j'} M_{11j}M_{11j'}, \end{aligned}$$

where $A = \{(j,j'),(k,k') : j\neq j', k\neq k', (j,j')\neq(k,k'), (j,j')\neq(k',k)\}$. In the expansion (second and third lines), the second and third terms are zero by independence. The 2 in the fourth term results from the product of $(T_{1k}T_{1k'})$ with both $(T_{1k}T_{1k'})$ and $(T_{1k'}T_{1k})$.

For the variance-covariance quantities in the matrix M_j for the jth observation, the expressions for M_{00j}, $M_{01j} = M_{10j}$ and M_{11j} are well-known and equal to the corresponding expressions from the matrix D_j, due to the information-unbiasedness of the quasi-score. For the components related to T_{2j}, recall (A3) and (A4). Then,

$$\begin{aligned} M_{02j} &= E(S_{0j}T_{2j}) \\ &= E\left\{\left(\frac{\partial\mu_j^T}{\partial\theta}\right)\frac{y_j-\mu_j}{v_j}\frac{\partial}{\partial\phi}\left(\frac{\partial\mu_j}{\partial\phi}v_j^{-1}\right)(y_j-\mu_j)\right. \end{aligned}$$

$$
\begin{aligned}
& + \left(\frac{\partial \mu_j^T}{\partial \theta}\right) \frac{y_j - \mu_j}{v_j} \left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-2} (y_j - \mu_j)^2 \\
& - \left(\frac{\partial \mu_j^T}{\partial \theta}\right) \frac{y_j - \mu_j}{v_j} \left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-1} \Bigg\} \\
= & \left(\frac{\partial \mu_j^T}{\partial \theta}\right) \frac{1}{v_j} \left(\frac{\partial^2 \mu_j}{\partial \phi^2} v_j^{-1} - \left(\frac{\partial \mu_j}{\partial \phi}\right)\left(\frac{\partial v_j}{\partial \phi}\right) v_j^{-2}\right) v_j \\
& + \left(\frac{\partial \mu_j^T}{\partial \theta}\right) \frac{1}{v_j} \left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-2} \rho_j - 0. \\
= & \; D_{02j} + \left(\frac{\partial \mu_j^T}{\partial \theta}\right) \left(\frac{\partial \mu_j}{\partial \phi}\right) v_j^{-3} \left[\left(\frac{\partial \mu_j}{\partial \phi}\right) \rho_j - \left(\frac{\partial v_j}{\partial \phi}\right) v_j\right].
\end{aligned}
$$

The form of M_{12j} is derived similarly as

$$
M_{12j} = D_{12j} + \left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-3} \left[\left(\frac{\partial \mu_j}{\partial \phi}\right) \rho_j - \left(\frac{\partial v_j}{\partial \phi}\right) v_j\right].
$$

Finally,

$$
\begin{aligned}
M_{22j} = & \; E(T_{2j}^2) \\
= & \left[\frac{\partial}{\partial \phi}\left(\frac{\partial \mu_j}{\partial \phi} v_j^{-1}\right)\right]^2 v_j + \left(\frac{\partial \mu_j}{\partial \phi}\right)^4 v_j^{-4} \kappa_j + \left(\frac{\partial \mu_j}{\partial \phi}\right)^4 v_j^{-2} \\
& + 2 \frac{\partial}{\partial \phi}\left(\frac{\partial \mu_j}{\partial \phi} v_j^{-1}\right) \left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-2} \rho_j + 0 \\
& - 2 \left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-2} \left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-1} v_j \\
= & \left(\frac{\partial \mu_j}{\partial \phi}\right)^4 v_j^{-4} (\kappa_j - v_j^2) \\
& + \frac{\partial}{\partial \phi}\left(\frac{\partial \mu_j}{\partial \phi} v_j^{-1}\right) \left[\frac{\partial}{\partial \phi}\left(\frac{\partial \mu_j}{\partial \phi} v_j^{-1}\right) v_j + 2 \left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-2} \rho_j\right] \\
= & \left(\frac{\partial \mu_j}{\partial \phi}\right)^4 v_j^{-4} \left(\kappa_j - v_j^2\right) + \left[\left(\frac{\partial^2 \mu_j}{\partial \phi^2}\right) v_j^{-1} - \left(\frac{\partial \mu_j}{\partial \phi}\right) v_j^{-2} \left(\frac{\partial v_j}{\partial \phi}\right)\right] \\
& \times \left[\left(\frac{\partial^2 \mu_j}{\partial \phi^2}\right) - \left(\frac{\partial \mu_j}{\partial \phi}\right) v_j^{-1} \left(\frac{\partial v_j}{\partial \phi}\right) + 2 \left(\frac{\partial \mu_j}{\partial \phi}\right)^2 v_j^{-2} \rho_j\right].
\end{aligned}
$$

Proof of Lemma 5. The proof follows that of Lindsay and Waterman (unpublished report, Pennsylvania State University, February 26, 1992). Write $S_2 = S_0 - (a_1 - b\, a_{22}) T_1 - a_{22} T_2$, where $b = D_{12}/D_{11}$. Consider the first two terms $S_0 - (a_1 - b\, a_{22}) T_1$. Recall that $S_0' = \sum_j S_{0j}'$ and $T_1' = \sum_j T_{1j}'$, that $E(S_{0j}') = -D_{01j}$ and $E(T_{1j}') = -D_{11j}$, that the order of a_1, b, and their derivatives is $O(1)$, and that the order of a_{22} and its derivatives is $O(n^{-1})$. In fact, standard matrix inversion and limits of polynomials shows that

$$a_{22\bullet} = \lim_{n\to\infty} n a_{22} = \frac{D_{02}D_{11} - D_{01}D_{12}}{2D_{11}^2 D_{111}}, \tag{A5}$$

where D_{111} is $E(T_{1j})$ for some j. Then via straightforward asymptotic arguments,

$$\{S_0 - (a_1 - b\,a_{22})T_1\}' = \sum_j Q_{1j} + O_p(1),$$

where $Q_{1j} = S_{0j}' - a_1 T_1' - a_1' T_{1j}$. Note that the leading term in the remainder is $b\,a_{22}\sum_j T_{1j}' = O_p(1)$. Now, since $a_1 = D_{10}/D_{11} = D_{10j}/D_{11j}$ for any j, we can rewrite $Q_{1j} = (S_{0j}' + D_{01j}) - a_1(T_1' + D_{11j}) - a_1' T_{1j}$, which is obviously unbiased for each j. Therefore, $\sum_j Q_{1j} = O_p(n^{1/2})$, and the second term in the above expression is negligible.

For the term $a_{22}T_2$, recall that $T_2 = \sum_j T_{2j} + 2\sum_{j<j'} T_{1j}T_{1j'}$, and write

$$(a_{22}T_2)' = a_{22}'\sum_j T_{2j} + a_{22}\sum_j T_{2j}' + 2a_{22}'\sum_{j<j'} T_{1j}T_{1j'} + 2a_{22}\sum_{j<j'}(T_{1j}'T_{1j'} + T_{1j}T_{1j'}'). \tag{A6}$$

Since a_{22} and a_{22}' are both $O(n^{-1})$, it follows that $a_{22}'\sum_j T_{2j} = o_p(1)$, since T_{2j} is unbiased, and that $a_{22}\sum_j T_{2j}' = O_p(1)$. For the other terms, we use a U-statistic representation of $2\sum_{j<j'} T_{1j}T_{1j'}$ and its derivative (Serfling, 1980, p. 188). Define

$$W_1 = \frac{2}{n(n-1)}\sum_{j<j'} T_{1j}T_{1j'}$$

$$W_2 = \frac{2}{n(n-1)}\sum_{j<j'}(T_{1j}'T_{1j'} + T_{1j}T_{1j'}').$$

Now, for $k = 1, 2$, $W_k = \hat{W}_k + o_p(n^{-1/2})$, where $\hat{W}_k = \sum_j E(W_k \mid T_{1j})$. It is then easy to see that $\hat{W}_1 = 0$, and

$$\hat{W}_2 = -\frac{2}{n}\sum_j T_{1j}D_{111} = -\frac{2}{n}D_{111}T_1.$$

Now,

$$2a_{22}'\sum_{j<j'} T_{1j}T_{1j'} = a_{22}'n(n-1)W_1 = O(n^{-1})n(n-1)o_p(n^{-1/2}) = o_p(n^{1/2}),$$

and

$$
\begin{aligned}
2a_{22}\sum_{j<j'}(T'_{1j}T_{1j'}+T_{1j}T'_{1j'}) &= a_{22}n(n-1)W_2 \\
&= -a_{22}n(n-1)\frac{2}{n}D_{111}T_1+O(n^{-1})n(n-1)o_p(n^{-1/2}) \\
&= -2a_{22\bullet}D_{111}T_1+o(1)O_p(n^{1/2})+o_p(n^{1/2}).
\end{aligned}
$$

So, equation (A6) can be written

$$
(a_{22}T_2)' = -2a_{22\bullet}D_{111}T_1+o_p(n^{1/2}),
$$

where the leading term is $O_p(n^{1/2})$. Finally,

$$
S'_2 = \sum_j Q_{1j}+O_p(1)+2a_{22\bullet}D_{111}T_1+o_p(n^{1/2}),
$$

completing the proof.

Proof of Lemma 6. Assuming that

$$
E\{-(S'_{0j}+D_{01j})'\} = E\{(S'_{0j}+D_{01j})T_{1j}\}
$$

and

$$
E\{-(T_{1j'}+D_{11j})'\} = E\{(T'_{1j}+D_{11j})T_{1j}\}
$$

hold for each j, the result follows by independence of observations y_j and $y_{j'}$ for $j\neq j'$. Now, for fixed j,

$$
\begin{aligned}
S_{0j} &= \left\{\left(\frac{\partial\mu_j^T}{\partial\theta}\right)v_j^{-1}\right\}(y_j-\mu_j) \\
S'_{0j} &= \left\{\left(\frac{\partial\mu_j^T}{\partial\theta}\right)v_j^{-1}\right\}'(y_j-\mu_j)-\left\{\left(\frac{\partial\mu_j^T}{\partial\theta}\right)v_j^{-1}\right\}\mu'_j \\
S''_{0j} &= \left\{\left(\frac{\partial\mu_j^T}{\partial\theta}\right)v_j^{-1}\right\}''(y_j-\mu_j)-2\left\{\left(\frac{\partial\mu_j^T}{\partial\theta}\right)v_j^{-1}\right\}'\mu'_j-\left\{\left(\frac{\partial\mu_j^T}{\partial\theta}\right)v_j^{-1}\right\}\mu''_j,
\end{aligned}
$$

so

$$
E(S''_{0j}) = -2E\left[\left\{\left(\frac{\partial\mu_j^T}{\partial\theta}\right)v_j^{-1}\right\}'\mu'_j\right]-E\left[\left\{\left(\frac{\partial\mu_j^T}{\partial\theta}\right)v_j^{-1}\right\}\mu''_j\right],
$$

where expectation here is over Y_j, then X_j. Similarly,

$$E(T''_{1j}) = -2E\left[\left\{\left(\frac{\partial \mu_j}{\partial \phi}\right) v_j^{-1}\right\}' \mu'_j\right] - E\left[\left\{\left(\frac{\partial \mu_j}{\partial \phi}\right) v_j^{-1}\right\} \mu''_j\right].$$

Also,

$$\begin{aligned}
D_{01j} &= E\left[\left\{\left(\frac{\partial \mu_j^T}{\partial \theta}\right) v_j^{-1}\right\} \mu'_j\right] = \int \left\{\left(\frac{\partial \mu_j^T}{\partial \theta}\right) v_j^{-1}\right\} \mu'_j \, dG(x_j) \\
D'_{01j} &= \frac{\partial}{\partial \phi} \int \left\{\left(\frac{\partial \mu_j^T}{\partial \theta}\right) v_j^{-1}\right\} \mu'_j \, dG(x_j) \\
&= \int \left\{\left(\frac{\partial \mu_j^T}{\partial \theta}\right) v_j^{-1}\right\}' \mu'_j \, dG(x_j) + \int \left\{\left(\frac{\partial \mu_j^T}{\partial \theta}\right) v_j^{-1}\right\} \mu''_j \, dG(x_j) \\
&= E\left[\left\{\left(\frac{\partial \mu_j^T}{\partial \theta}\right) v_j^{-1}\right\}' \mu'_j\right] + E\left[\left\{\left(\frac{\partial \mu_j^T}{\partial \theta}\right) v_j^{-1}\right\} \mu''_j\right].
\end{aligned}$$

Similarly,

$$D_{11j} = E\left[\left\{\left(\frac{\partial \mu_j}{\partial \phi}\right) v_j^{-1}\right\}' \mu'_j\right] + E\left[\left\{\left(\frac{\partial \mu_j}{\partial \phi}\right) v_j^{-1}\right\} \mu''_j\right].$$

Thus,

$$\begin{aligned}
E\{(S'_{0j} + D_{01j})'\} = E(S''_{0j}) + D'_{01j} &= -E\left[\left\{\left(\frac{\partial \mu_j^T}{\partial \theta}\right) v_j^{-1}\right\}' \mu'_j\right] \\
E\{(T'_{1j} + D_{11j})'\} = E(T''_{1j}) + D'_{11j} &= -E\left[\left\{\left(\frac{\partial \mu_j}{\partial \phi}\right) v_j^{-1}\right\}' \mu'_j\right].
\end{aligned}$$

Now,

$$\begin{aligned}
E\{(S'_{0j} + D_{01j})T_{ij}\} &= E(S'_{0j} T_{ij}) \\
&= E\left[\left\{\left(\frac{\partial \mu_j^T}{\partial \theta}\right) v_j^{-1}\right\}' (y_j - \mu_j)\mu'_j v_j^{-1}(y_j - \mu_j)\right] \\
&\quad - E\left[\left\{\left(\frac{\partial \mu_j^T}{\partial \theta}\right) v_j^{-1}\right\} \mu'_j \mu'_j v_j^{-1}(y_j - \mu_j)\right] \\
&= E\left[\left\{\left(\frac{\partial \mu_j^T}{\partial \theta}\right) v_j^{-1}\right\}' \mu'_j\right],
\end{aligned}$$

and similarly,

$$E\{(T'_{1j} + D_{11j})T_{ij}\} = E\left[\left\{\left(\frac{\partial \mu_j}{\partial \phi}\right) v_j^{-1}\right\}' \mu'_j\right],$$

completing the proof. Note that this development for the identically-distributed X_j case assumes that the distribution $G(\cdot)$ of X_j does not depend upon the nuisance ϕ. However, the analogous lemma holds when the x_js are considered fixed and, as in Section 3, expectations are taken conditional on the x_js.

Proof of Theorem 7. Begin with equation (11). By Lemma 5, $S_2' = O_p(n^{1/2})$. Similarly, $S_2'' = O_p(n^{1/2})$, since $E(S_2'') = 0$ (Small and McLeish, 1989). Since $S_2' = \tilde{S}_2' + o_p(n^{1/2})$ and $M_{11}^{-1}T_1 = O_p(n^{-1/2})$, the first result,

$$\hat{S}_2 - S_2 = M_{11}^{-1}\tilde{S}_2'T_1 + o_p(1) \tag{A7}$$

follows. Also $\tilde{S}_2' = O_p(n^{1/2})$, so $M_{11}^{-1}\tilde{S}_2'T_1 = O_p(1)$.

To establish the unbiasedness of this term, note that from Lemma 6

$$\begin{aligned} E\{(S_0' + D_{01})T_1\} &= E\{-(S_0' + D_{01})'\} = E(-S_0'') - D_{01}' \\ &= 2D_{01}' - D_{02} - D_{01}' = D_{01}' - D_{02}, \end{aligned}$$

where the third equality is due to Lemma 1. Similarly,

$$E\{(T_1' + D_{11})T_1\} = D_{11}' - D_{12}.$$

Recall that $E(T_1^2) = D_{11}$ and note that

$$a_1' = \frac{D_{01}'}{D_{11}} - \frac{D_{01}D_{11}'}{D_{11}^2} = \frac{D_{01}'}{D_{11}} - a_1\frac{D_{11}'}{D_{11}},$$

so $a_1'E(T_1^2) = D_{01}' - a_1D_{11}'$. From expression (A5),

$$2a_{22\bullet}D_{111}E(T_1^2) = \frac{D_{02}D_{11} - D_{01}D_{12}}{D_{11}^2}D_{11} = D_{02} - a_1D_{12}.$$

Finally, from Lemma 5,

$$\begin{aligned} E(\tilde{S}_2'T_1) &= E\{(S_0' + D_{01})T_1\} - a_1E\{(T_1' + D_{11})T_1\} - a_1'E(T_1^2) + 2a_{22\bullet}D_{111}E(T_1^2) \\ &= D_{01}' - D_{02} - a_1(D_{11}' - D_{12}) - (D_{01}' - a_1D_{11}') + D_{02} - a_1D_{12} \\ &= 0. \end{aligned}$$

To complete the proof, uniform integrability of the $o_p(1)$ term in (A7) implies $E(\hat{S}_2 - S_2) = o(1)$.

REFERENCES

CARROLL, R.J. & STEFANSKI, L.A. (1990). Approximate quasi-likelihood estimation in models with surrogate predictors. *J. Am. Statist. Assoc.* **85**, 652-663.

CROWDER, M. (1987). On linear and quadratic estimating functions. *Biometrika* **74**, 591-597.

DESMOND, A.F. (1997). Optimal estimating functions, quasi-likelihood and statistical modelling. *J. Statist. Plan. Inf.* **60**, 70 ff.

DURBIN, J. (1960). Estimation of parameters in time-series regression models. *J. R. Statist. Soc. B.* **22**, 139-153.

GODAMBE, V.P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.* **31**, 1208-1211.

GODAMBE, V.P. (1976). Conditional likelihood and unconditional optimum estimating equations. *Biometrika* **63**, 277-284.

GODAMBE, V.P. (1980). On sufficiency and ancillarity in the presence of a nuisance parameter. *Biometrika* **67**, 155-162.

GODAMBE, V.P. (1984). On ancillarity and Fisher information in the presence of a nuisance parameter. *Biometrika* **71**, 626-629.

JEWELL, N.P. and RAAB, G.M. (1981). Difficulties in obtaining consistent estimators of variance parameters. *Biometrika* **68**, 221-226.

LIANG, K-Y. & ZEGER, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.

LIANG, K-Y. & ZEGER, S.L. (1995). Inference based on estimating functions in the presence of nuisance parameters (with discussion). *Statist. Sci.* **10**, 158-199.

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. *Ann. Statist.* **13**, 914-931.

MCCULLAGH, P. (1983). Quasi-likelihood functions. *Ann. Statist.* **11**, 59-67.

MCCULLAGH, P. & NELDER, J.A. (1989). *Generalized Linear Models*, 2nd Edition. London: Chapman & Hall.

NEYMAN, J. & SCOTT, E.L. (1948) Consistent estimates based on partially consistent observations. *Econometrica* **16**, 1-32.

ROYALL, R.M. (1986). Model robust confidence intervals using maximum likelihood estimators. *Int. Statist. Rev.* **54**, 221-226.

SERFLING, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: John Wiley & Sons.

SMALL, C.G. & MCLEISH, D.L. (1989). Projection as a method for increasing sensitivity and eliminating nuisance parameters. *Biometrika* **76**, 693-703.

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

STEFANSKI, L.A. & CARROLL, R.J. (1987). Conditional scores and optimal scores for generalized linear measurement error models. *Biometrika* **74**, 703-716.

WATERMAN, R.P. & LINDSAY, B.G. (1996). Projected score methods for approximating conditional scores. *Biometrika* **83**, 1-13.

WEDDERBURN, R.W.M. (1974). Quasi likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika* **61**, 439-447.

WILLIAMS, D.A. (1982). Extra-binomial variation in logistic linear models. *Appl. Statist.* **31**, 144-148.

Table 1. *Paired-data gamma regression model. Means of $\hat{S}_1$ and $\hat{S}_2$, empirical and theoretical information approximations for $\hat{S}_2$. 5000 replicates.*

Model			Z-values		Information Measures				
ν	θ_0	θ_1	$\hat{S}_1$	$\hat{S}_2$	$\mathcal{I}_{\hat{S}_2}^{\text{emp}}$	$\mathcal{I}_{S_2}$	Ratio	$\mathcal{I}_{\hat{S}_2}^{*}$	Ratio
0.2	0.40	0.1	-5.39	-0.91	4.30	4.40	1.02	4.40	1.02
0.2	0.40	0.2	-9.82	-2.01	3.73	3.76	1.01	3.76	1.01
0.2	0.40	0.5	-11.81	1.26	2.55	2.46	0.96	2.46	0.96
0.2	0.10	0.1	-7.51	0.13	15.00	14.74	0.98	14.74	0.98
0.2	0.10	0.2	-10.12	1.20	11.18	10.64	0.95	10.63	0.95
0.2	0.10	0.5	-16.19	-2.00	5.21	5.03	0.97	5.03	0.97
0.2	-0.05	0.1	-11.48	-1.08	52.72	51.35	0.97	51.33	0.97
0.2	-0.05	0.2	-12.69	-1.15	25.31	25.81	1.02	25.80	1.02
0.2	-0.05	0.5	-14.66	-2.87	7.95	8.07	1.02	8.07	1.01
0.6	0.40	0.1	-6.27	1.00	1.51	1.46	0.97	1.46	0.97
0.6	0.40	0.2	-12.52	-0.85	1.34	1.25	0.94	1.25	0.93
0.6	0.40	0.5	-21.64	-4.36	0.89	0.81	0.90	0.80	0.90
0.6	0.10	0.1	-12.63	-1.64	5.17	4.88	0.94	4.88	0.94
0.6	0.10	0.2	-17.87	-2.77	3.82	3.50	0.92	3.49	0.91
0.6	0.10	0.5	-22.47	-5.14	1.62	1.65	1.02	1.64	1.01
0.6	-0.05	0.1	-17.26	-4.09	16.92	16.87	1.00	16.74	0.99
0.6	-0.05	0.2	-20.43	-5.91	7.97	8.49	1.07	8.42	1.06
0.6	-0.05	0.5	-20.06	-5.25	2.49	2.66	1.07	2.65	1.06
2.0	0.40	0.1	-12.49	-2.96	0.67	0.44	0.65	0.44	0.65
2.0	0.40	0.2	-19.10	-7.08	0.48	0.37	0.78	0.36	0.77
2.0	0.40	0.5	-24.46	-10.14	0.23	0.24	1.04	0.22	0.97
2.0	0.10	0.1	-17.03	-5.16	1.94	1.45	0.75	1.41	0.73
2.0	0.10	0.2	-24.34	-9.34	1.34	1.03	0.77	0.97	0.72
2.0	0.10	0.5	-26.34	-10.59	0.48	0.48	1.02	0.44	0.92
2.0	-0.05	0.1	-21.59	-9.08	5.02	4.95	0.98	4.51	0.90
2.0	-0.05	0.2	-23.89	-11.10	2.03	2.50	1.23	2.28	1.12
2.0	-0.05	0.5	-24.35	-10.61	0.68	0.79	1.16	0.73	1.08
average information values					6.76	6.67	0.99	6.63	0.98

Table 2. *Overdispersed Poisson measurement error model model. Means of $\hat{S}_1$ and $\hat{S}_2$, empirical and theoretical information approximations for $\hat{S}_2$. 5000 replicates.*

	Model		Z-values		Information Measures				
ν	$\exp(\theta_1)$	σ^2_δ	$\hat{S}_1$	$\hat{S}_2$	$\mathcal{I}^{\text{emp}}_{\hat{S}_2}$	$\mathcal{I}_{S_2}$	Ratio	$\mathcal{I}^*_{\hat{S}_2}$	Ratio
1.5	1.5	0.2	-0.45	1.13	0.65	0.76	1.18	0.64	0.99
1.5	1.5	0.5	-4.11	-0.54	0.50	0.72	1.43	0.49	0.97
1.5	1.5	0.7	-4.04	0.72	0.43	0.69	1.60	0.42	0.97
1.5	2.0	0.2	-5.48	-0.34	0.88	0.99	1.12	0.86	0.98
1.5	2.0	0.5	-8.71	0.10	0.64	0.79	1.24	0.59	0.92
1.5	2.0	0.7	-12.00	-2.13	0.55	0.70	1.28	0.47	0.86
1.5	3.0	0.2	-8.35	0.60	1.04	1.11	1.06	1.00	0.96
1.5	3.0	0.5	-8.13	0.96	0.56	0.65	1.17	0.52	0.94
1.5	3.0	0.7	-8.30	-0.67	0.40	0.51	1.29	0.39	0.97
2.0	1.5	0.2	-1.83	-0.43	0.50	0.58	1.16	0.48	0.97
2.0	1.5	0.5	-4.38	-1.05	0.39	0.55	1.41	0.37	0.96
2.0	1.5	0.7	-5.56	-1.06	0.35	0.54	1.54	0.32	0.93
2.0	2.0	0.2	-4.52	0.34	0.69	0.77	1.11	0.67	0.97
2.0	2.0	0.5	-9.23	-0.52	0.49	0.64	1.30	0.47	0.96
2.0	2.0	0.7	-11.27	-0.94	0.44	0.58	1.31	0.39	0.88
2.0	3.0	0.2	-7.85	1.39	0.89	0.94	1.06	0.85	0.96
2.0	3.0	0.5	-12.92	-1.99	0.49	0.58	1.17	0.46	0.94
2.0	3.0	0.7	-9.79	0.31	0.40	0.46	1.17	0.35	0.88
3.0	1.5	0.2	-1.42	-0.26	0.33	0.39	1.17	0.33	0.98
3.0	1.5	0.5	-3.05	-0.22	0.26	0.38	1.46	0.25	0.99
3.0	1.5	0.7	-4.60	-0.71	0.24	0.37	1.55	0.22	0.93
3.0	2.0	0.2	-3.43	0.82	0.48	0.54	1.14	0.47	0.99
3.0	2.0	0.5	-10.78	-1.99	0.39	0.47	1.20	0.34	0.88
3.0	2.0	0.7	-12.24	-1.71	0.33	0.43	1.33	0.29	0.88
3.0	3.0	0.2	-8.86	0.70	0.69	0.73	1.05	0.66	0.95
3.0	3.0	0.5	-13.37	-0.56	0.42	0.48	1.14	0.38	0.90
3.0	3.0	0.7	-15.49	-2.52	0.34	0.39	1.14	0.29	0.85
average information values					0.51	0.62	1.22	0.48	0.94