Locally Efficient Semiparametric Estimators for Generalized Skew-Elliptical Distributions

Yanyuan MA, Marc G. GENTON, and Anastasios A. TSIATIS

We consider a class of generalized skew-normal distributions that is useful for selection modeling and robustness analysis and derive a class of semiparametric estimators for the location and scale parameters of the central part of the model. We show that these estimators are consistent and asymptotically normal. We present the semiparametric efficiency bound and derive the locally efficient estimator that achieves this bound if the model for the skewing function is correctly specified. The estimators that we propose are consistent and asymptotically normal even if the model for the skewing function is misspecified, and we compute the loss of efficiency in such cases. We conduct a simulation study and provide an illustrative example. Our method is applicable to generalized skew-elliptical distributions.

KEY WORDS: Generalized skew-elliptical distribution; Influence function; Nuisance tangent space; Selection model; Semiparametric efficiency.

1. INTRODUCTION

Consider the model where a p-dimensional random vector $\mathbf{X}$ is distributed with density $g(\mathbf{x}; \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a q-dimensional vector of unknown parameters. To make inference about $\boldsymbol{\beta}$, the usual statistical analysis assumes that a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ from $g(\mathbf{x}; \boldsymbol{\beta})$ can be observed. However, there are many situations where such a random sample might not be available, for instance, if it is too difficult or too costly to obtain. If the probability density function is distorted by some multiplicative nonnegative weight function $w(\mathbf{x}; \boldsymbol{\beta}, \boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ denotes some r-dimensional vector of additional unknown parameters, then the observed data is a random sample from a distribution with density

$$f(\mathbf{x}; \boldsymbol{\beta}, \boldsymbol{\alpha}) = g(\mathbf{x}; \boldsymbol{\beta}) \frac{w(\mathbf{x}; \boldsymbol{\beta}, \boldsymbol{\alpha})}{\mathrm{E}\{w(\mathbf{X}; \boldsymbol{\beta}, \boldsymbol{\alpha})\}}, \tag{1}$$

where f is said to be the probability density function of a weighted distribution (see Rao 1985 and references therein). In particular, if the observed data are obtained only from a selected portion of the population of interest, then (1) is called a "selection model." This can happen if, for example, the observation vector $\mathbf{X}$ of characteristics of a certain population is measured only for individuals who manifest a certain disease due to cost or ethical reasons (see Bayarri and DeGroot 1992 and references therein). For such problems, the goal is to find consistent and asymptotically normal estimators of $\boldsymbol{\beta}$ in the presence of the nuisance weight function w.

A slightly different point of view is given by a robustness argument. Effectively, if $g(\mathbf{x}; \boldsymbol{\beta})$ is the central model of interest, then the weight function w in (1) can be seen as a contaminating function. For instance, if g is an elliptical probability density function, then w generates asymmetric outliers in the observed sample from f. The goal is then to derive robust estimators of $\boldsymbol{\beta}$, that is, again to provide consistent and asymptotically normal estimators of $\boldsymbol{\beta}$ in the presence of a certain class of the nuisance weight function w.

Yanyuan Ma is Assistant Professor, Department of Statistics, Texas A&M University, College Station, TX 77843 (E-mail: *ma@stat.tamu.edu*). Marc G. Genton is Associate Professor, Department of Statistics, Texas A&M University, College Station, TX 77843 (E-mail: *genton@stat.tamu.edu*). Anastasios A. Tsiatis is Professor, Department of Statistics, North Carolina State University, Box 8203, Raleigh, NC 27695 (E-mail: *tsiatis@stat.ncsu.edu*).
The work of Yanyuan Ma is supported by grant NIGMS 1 R01 Gm67299-01. The authors thank the editor, the associate editor, and three referees for helpful comments that greatly improved the manuscript.

The article is organized as follows. In Section 2 we describe a class of generalized skew-elliptical distributions which is useful for selection modeling and robustness analysis. We present our main results in Section 3 for a univariate location-scale normal central model. In particular, we derive semiparametric location and scale estimators that are consistent and asymptotically normal regardless of the possible misspecification of the weight function. In addition, we will show that estimators within this class achieve the semiparametric efficiency bound. We present a simulation study in Section 4 and an illustrative example of Australian athletes' body mass index (BMI) data in Section 5. We discuss the extension of the procedure to generalized skew-elliptical/skew-symmetric distributions in Section 6.

2. GENERALIZED SKEW–ELLIPTICAL DISTRIBUTIONS

Generalized skew-elliptical (GSE) distributions have been introduced by Genton and Loperfido (2005). The density of a random vector with a GSE distribution is defined through an elliptical density and a skewing function as follows.

Definition 1. A p-dimensional generalized skew-elliptical (GSE) distribution is a distribution whose probability density function is of the form

$$f(\mathbf{x}) = 2|\boldsymbol{\Sigma}|^{-1/2} g\{\boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\xi})\} \times \pi\{\boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\xi})\}, \qquad \mathbf{x} \in \mathbb{R}^p, \tag{2}$$

where g is the probability density function of a spherical distribution, $\boldsymbol{\xi}$ is the location parameter, $\boldsymbol{\Sigma}^{-1/2}$ is the Cholesky decomposition of the inverse of the positive definite scale matrix $\boldsymbol{\Sigma}$, that is, $(\boldsymbol{\Sigma}^{-1/2})^T \boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1}$, and the function $\pi : \mathbb{R}^p \to [0\ 1]$ satisfies $\pi(\mathbf{x}) + \pi(-\mathbf{x}) = 1$ and π is continuous. We refer to π as the skewing function.

In this paper, we restrict our attention to the situation where the skewing function is differentiable, in order to accommodate the application of semiparametric theories. Note that the location vector $\boldsymbol{\xi}$ and the scale matrix $\boldsymbol{\Sigma}$ are not in general the expected value and the covariance matrix for f, because GSE distributions may not be symmetric with respect to $\boldsymbol{\xi}$, but they are for g. In particular, if $g = \phi_p$, the probability density function (pdf) of the standard p-dimensional multivariate normal

Journal of the American Statistical Association
September 2005, Vol. 100, No. 471, Theory and Methods
DOI 10.1198/016214505000000079

distribution, and we choose a parametric model $\pi(\mathbf{x}) = \Phi(\boldsymbol{\alpha}^T\mathbf{x})$ for the skewing function, where Φ is the univariate standard normal cumulative distribution function (cdf), then (2) is the probability density function of the multivariate skew-normal distribution (Azzalini and Dalla Valle 1996).

From Definition 1, it is clear that the GSE distributions arise in inference from nonrandom (biased) samples (Copas and Li 1997) and thus are selection models of the form in (1). Representation of a GSE distribution as a selection model is straightforward with $g(\mathbf{x};\boldsymbol{\beta}) = |\boldsymbol{\Sigma}|^{-1/2}g\{\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\xi})\}$, $w(\mathbf{x};\boldsymbol{\beta},\boldsymbol{\alpha}) = \pi\{\boldsymbol{\Sigma}^{-1/2}(\mathbf{x}-\boldsymbol{\xi})\}$, $\mathrm{E}\{w(\mathbf{X};\boldsymbol{\beta},\boldsymbol{\alpha})\} = 1/2$, $\boldsymbol{\beta} = \{\boldsymbol{\xi}^T, \mathrm{vec}(\boldsymbol{\Sigma})^T\}^T$, and $\boldsymbol{\alpha}$ embedded in the skewing function π. A weight function w with such property can naturally occur when the selection criterion is that a certain component of the measurement is larger than its expected value given the other measurement components (see Arnold and Beaver 2002). Assume that there are two random variables X and Y, where X follows a symmetric distribution with pdf $g(x)$ and the pdf of the conditional distribution of X given Y, $p(x|y)$, is a function of $x - cy$, denoted by $u(x - cy)$, where u is a symmetric function and c is a constant. We can verify that the expectation of X conditional on Y is cY and that the selection criterion $x > \mathrm{E}(X|y)$ yields a weight function $w(x) = H(x/c)$, where H is the corresponding cdf of the marginal density of Y, say h; that is, h satisfies $\int u(x - cy)h(y)\,dy = g(x)$. For a variety of functions u, a unique solution h can be obtained through deconvolution. In addition, such h is guaranteed to be symmetric, and hence the resulting weight function satisfies the requirement $H(x/c) + H(-x/c) = 1$. A special case is when u and g are both normal. The resulting pdf of the selected samples is then the aforementioned skew-normal distribution. One example of this specific setting is the distribution of height and weight. Assume that the weight (X) and height (Y) follow a bivariate normal distribution in a general population. After centering and normalizing, we obtain two standard normal distributions for $\tilde{X}$ and $\tilde{Y}$ with correlation c. Yet in a clinic treating obesity, one would expect that all of the samples obtained would be the ones whose weight is larger than the expected weight given their height. This corresponds to a selection criterion $\tilde{x} > \mathrm{E}(\tilde{X}|\tilde{y})$, with $g(\tilde{x}) = \phi_1(\tilde{x})$, and $p(\tilde{x}|\tilde{y}) = \phi_1(\tilde{x}; c\tilde{y}, \sqrt{1-c^2})$. Here we use the notation $\phi_1(x;\xi,\sigma)$ to denote the normal pdf with mean ξ and standard deviation σ. It can be verified that $h(\tilde{y}) = \phi_1(\tilde{y})$ and the pdf of the distribution of the patients' weight is given by $2\phi_1(\tilde{x})\Phi\{\tilde{x}/c\}$, which translates to the pdf of the observed weight X with the form $2\phi_1(x;\xi,\sigma)\Phi\{\alpha(x-\xi)\}$. Similarly, in the example presented in this article, we analyze a dataset of BMI in a group of athletes, which is assumed to be larger than its expected value conditional on an individual's other body characteristics, including height, weight, and body fat percentage in a general population, for males and for females. If we assume that the BMI in a general population of the same gender follows a normal distribution without specifying a precise selection method, then the observed data follow a generalized skew-normal (GSN) distribution with unspecified skewing function.

Another way to view such data is through a hidden truncation model (Arnold and Beaver 2002). Assume that we have two random variables X and Y, with the symmetric pdf's $g(x)$ and $h(y)$. If we select the sample of X based on the criterion $x/y > c$, then the selected samples X follow the distribution with pdf $2g(x)H(x/c)$, which is of the form of a GSE distribution. Although compared with a general selection model of the form in (1) the GSE models are restricted by the constraint $\pi(x) + \pi(-x) = 1$, the foregoing scenario shows that it is of practical use for a number of situations.

3. MAIN RESULTS

As described in the previous section, we are interested in inference on the parameters $\boldsymbol{\xi}$ and $\boldsymbol{\Sigma}$ in (2), which represent the mean and the covariance matrix of the population of which only samples from a particular subpopulation are available. We make no additional assumptions regarding the skewing function other than that π is a nonnegative differentiable function and $\pi(x) + \pi(-x) = 1$. Consequently, we are considering a semiparametric model where the parameters of interest are $\boldsymbol{\xi}$ and $\boldsymbol{\Sigma}$, which we summarize as $\boldsymbol{\beta}$, and the nuisance parameter is π. In such a setting, regular asymptotically linear (RAL) estimators have been studied by Newey (1990). An RAL estimator $\hat{\boldsymbol{\beta}}$ satisfies

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\psi}(\mathbf{X}_i, \boldsymbol{\beta}_0) + o_p(1),$$

where $\boldsymbol{\beta}$ represents the finite-dimensional parameter of interest, with its true value $\boldsymbol{\beta}_0$, $\boldsymbol{\psi}(\mathbf{X}_i, \boldsymbol{\beta}_0)$ is the ith influence function of the estimator, which satisfies $\mathrm{E}(\boldsymbol{\psi}) = \mathbf{0}$, and $\mathrm{E}(\boldsymbol{\psi}\boldsymbol{\psi}^T)$ is finite and nonsingular. In addition, an RAL estimator also satisfies regularity conditions, which would exclude estimators that are "superefficient" for some true parameter $\boldsymbol{\beta}_0$ (see Newey 1990 for details). Because of the link between an RAL estimator and its influence function, an RAL estimator can be constructed through finding its influence function, namely $\hat{\boldsymbol{\beta}}$ is given as a solution that solves $\sum_{i=1}^{n}\boldsymbol{\psi}(\mathbf{X}_i, \boldsymbol{\beta}) = \mathbf{0}$. Due to such a link, it is also clear that the variance of the estimator $\hat{\boldsymbol{\beta}}$ is given by the variance $\mathrm{E}(\boldsymbol{\psi}\boldsymbol{\psi}^T)$. A geometrical point of view was taken by Bickel, Klaassen, Ritov, and Wellner (1993), who characterized the set of all influence functions. Consider a Hilbert space $\mathcal{H}$ consisting of all the q-dimensional mean-0 random functions, with the inner product defined as the covariance between two functions, where q is the dimension of $\boldsymbol{\beta}$. Subsequently, the norm of a function is defined as the variance of the function, and functions in $\mathcal{H}$ must have finite norm. Note that all of the expectations are taken with respect to the true density $p(\mathbf{X};\boldsymbol{\beta}_0,\pi_0)$. In $\mathcal{H}$, a nuisance tangent space with respect to the semiparametric model is defined as the mean square closure of the nuisance tangent spaces with respect to all of the parametric submodels. Here a parametric submodel is a parametric model that is included in the original semiparametric model and contains the truth. A nuisance tangent space with respect to a parametric model $p(\mathbf{X};\boldsymbol{\beta},\boldsymbol{\alpha})$ is defined as a linear space spanned by the nuisance score vector, that is, all of the functions of the form $\mathbf{B}\mathbf{S}_{\boldsymbol{\alpha}}$, where $\mathbf{B}$ is a $q \times r$ matrix, with r being the dimension of $\boldsymbol{\alpha}$, and $\mathbf{S}_{\boldsymbol{\alpha}} = \frac{\partial \log p(\mathbf{X};\boldsymbol{\beta}_0,\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}|_{\boldsymbol{\alpha}_0}$ is the nuisance score vector, where $p(\mathbf{X};\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)$ gives the true density. The orthogonal complement of the nuisance tangent space in $\mathcal{H}$ is referred to as the nuisance tangent space orthogonal complement. Proper normalization of any function in the nuisance tangent space orthogonal complement yields an influence function; in contrast, any influence function can be obtained through properly nor-

malizing a function in the nuisance tangent space orthogonal complement. The normalization is such that the inner product between an influence function and the score vector $\mathbf{S}_{\boldsymbol{\beta}}$ must be equal to the identity, where $\mathbf{S}_{\boldsymbol{\beta}} = \frac{\partial \log p(\mathbf{X};\boldsymbol{\beta},\pi_0)}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}_0}$.

We use the aforementioned tools to derive RAL estimators and the efficient RAL estimator for the GSE distributions. To remain specific and focused, in this section all of our results are developed in the special case where $g = \phi$, the univariate standard normal probability density function, in which case we use the GSN distributions

$$f(x) = \frac{2}{\sigma}\phi\left(\frac{x-\xi}{\sigma}\right)\pi\left(\frac{x-\xi}{\sigma}\right). \tag{3}$$

The methods that we use can be extended in a straightforward manner to more general cases; see the discussion in Section 6. In the sequel, $\boldsymbol{\beta}$ represents the vector $(\xi, \sigma)^T$. Notice that an arbitrary skewing function $\pi(x)$ can always be written as $H\{m(x)\}$, where H is an arbitrarily chosen symmetric cdf and m is an odd function. In particular, an arbitrary $\pi(x)$ can be written as $\Phi\{m(x)\}$, where Φ is the univariate normal cdf. Throughout the article, parameters or functions with index 0 refer to the true values of the parameters or the true functions. Because we are considering a two-dimensional parameter of interest, the Hilbert space $\mathcal{H}$ that we work in consists of the two-dimensional mean-0 functions with finite variance. We begin by deriving the nuisance tangent space and its orthogonal complement.

Proposition 1. The nuisance tangent space is $\Gamma_\pi = \{\mathbf{u}\{(x-\xi_0)/\sigma_0\}: \mathbb{R} \to \mathbb{R}^2$: each component of $\pi_0(x)\mathbf{u}(x)$ is an odd function$\}$.

Proof. Suppose that $f(x;\boldsymbol{\beta},\boldsymbol{\alpha}) = \frac{2}{\sigma}\phi\{(x-\xi)/\sigma\}\pi\{(x-\xi)/\sigma, \boldsymbol{\alpha}\}$ is a parametric submodel of the GSN model (3); then

$$\left.\frac{\partial \log f(x;\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right|_{(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)} = \left.\frac{\partial \pi\{(x-\xi_0)/\sigma_0, \boldsymbol{\alpha}\}}{\partial \boldsymbol{\alpha}}\right|_{\boldsymbol{\alpha}_0} \Big/ \pi_0\left(\frac{x-\xi_0}{\sigma_0}\right).$$

Because $\pi\{(x-\xi_0)/\sigma_0, \boldsymbol{\alpha}\} + \pi\{(x+\xi_0)/\sigma_0, \boldsymbol{\alpha}\} = 1$, $\partial\pi\{(x-\xi_0)/\sigma_0, \boldsymbol{\alpha}\}/\partial\boldsymbol{\alpha} + \partial\pi\{(-x+\xi_0)/\sigma_0, \boldsymbol{\alpha}\}/\partial\boldsymbol{\alpha} = 0$; that is,

$$\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right)\left.\frac{\partial \log f(x;\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right|_{(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)}$$

is an odd function of $(x-\xi_0)/\sigma_0$. For any $2 \times r$ matrix $\mathbf{B}$, writing

$$\mathbf{B}\left.\frac{\partial \log f(x;\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right|_{(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)}$$

as $\mathbf{u}\{(x-\xi_0)/\sigma_0\}$, we obtain that $\pi_0(x)\mathbf{u}(x)$ is an odd function. In fact, for any linear combination of such $\partial \log f(x;\boldsymbol{\beta},\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}$ resulting from different parametric submodels, for example, for

$$\mathbf{u}(x) = \mathbf{B}_1\frac{\partial f_1(x;\boldsymbol{\beta},\boldsymbol{\alpha}_1)}{\partial \boldsymbol{\alpha}_1} + \mathbf{B}_2\frac{\partial f_2(x;\boldsymbol{\beta},\boldsymbol{\alpha}_2)}{\partial \boldsymbol{\alpha}_2},$$

$\pi_0(x)\mathbf{u}(x)$ is still an odd function. In contrast, for any $\mathbf{u}(x)$: $\mathbb{R} \to \mathbb{R}^2$, such that each component of $\pi_0(x)\mathbf{u}(x)$ is odd, let $\mathbf{h}(x) = \pi_0(x)\mathbf{u}(x)/[m_0(x)\phi\{m_0(x)\}]$, where $\pi_0(x) = \Phi\{m_0(x)\}$. Then $\mathbf{h}(x): \mathbb{R} \to \mathbb{R}^2$ is an even function and for $\boldsymbol{\alpha} \in \mathbb{R}^2$,

$$f(x;\boldsymbol{\beta},\boldsymbol{\alpha}) = \frac{2}{\sigma}\phi\left(\frac{x-\xi}{\sigma}\right)\Phi\left\{m_0\left(\frac{x-\xi}{\sigma}\right)e^{\boldsymbol{\alpha}^T\mathbf{h}((x-\xi)/\sigma)}\right\}$$

is a parametric submodel, where $\boldsymbol{\alpha} = \mathbf{0}$ yields the true model. Notice that

$$\begin{aligned}
&\left.\frac{\partial \log f(x;\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right|_{(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)} \\
&= m_0\left(\frac{x-\xi_0}{\sigma_0}\right)e^{\boldsymbol{\alpha}^T\mathbf{h}((x-\xi_0)/\sigma_0)}\mathbf{h}\left(\frac{x-\xi_0}{\sigma_0}\right) \\
&\quad \times \phi\left\{m_0\left(\frac{x-\xi_0}{\sigma_0}\right)e^{\boldsymbol{\alpha}^T\mathbf{h}((x-\xi_0)/\sigma_0)}\right\}\Big/\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right)\Big|_{\boldsymbol{\alpha}=\mathbf{0}} \\
&= \mathbf{u}\left(\frac{x-\xi_0}{\sigma_0}\right).
\end{aligned}$$

In the special case when $\pi_0(x) \equiv 1/2$, and thus $m_0(x) \equiv 0$, we can set the parametric submodel to be

$$f(x;\boldsymbol{\beta},\boldsymbol{\alpha}) = \frac{2}{\sigma}\phi\left(\frac{x-\xi}{\sigma}\right)\Phi\left\{\frac{\boldsymbol{\alpha}^T}{2}\mathbf{u}\left(\frac{x-\xi}{\sigma}\right)\Big/\phi(0)\right\}.$$

It can be easily verified that

$$\left.\frac{\partial \log f(x;\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right|_{(\boldsymbol{\beta}_0,\boldsymbol{\alpha}_0)} = \mathbf{u}\left(\frac{x-\xi_0}{\sigma_0}\right),$$

and hence $\mathbf{u}\{(x-\xi_0)/\sigma_0\} \in \Gamma_\pi$; that is, $\mathbf{u}\{(x-\xi_0)/\sigma_0\}$ is really an element in the nuisance tangent space.

In the proof of Proposition 1, to show that $\pi_0(x)\mathbf{u}(x)$ being odd is a sufficient condition for $\mathbf{u}(x)$ to be in Γ_π, we demonstrate that there exists a parametric submodel with nuisance score vector $\mathbf{u}(x)$. The existence of such a parametric submodel was proved by constructing one specific example. This does not mean that the constructed example is the only parametric submodel that will have $\mathbf{u}(x)$ as its nuisance score vector. In fact, many different parametric submodels could have the same nuisance score vector $\mathbf{u}(x)$. As long as the proposition is concerned, finding one such parametric submodel is sufficient.

Proposition 2. The orthogonal complement of the nuisance tangent space is $\Gamma_\pi^\perp = \{\mathbf{v}\{(x-\xi_0)/\sigma_0\}: \mathbb{R} \to \mathbb{R}^2$ is an even function (each component is even) that satisfies $\int \mathbf{v}(x)\phi(x)\,d\mu(x) = \mathbf{0}\}$, where $\mu(x)$ is the Lebesgue measure for which densities are defined.

Proof. Elements in $\Gamma_\pi^\perp$ satisfy

$$\int \mathbf{v}\left(\frac{x-\xi_0}{\sigma_0}\right)\mathbf{u}^T\left(\frac{x-\xi_0}{\sigma_0}\right) \times \frac{2}{\sigma_0}\phi\left(\frac{x-\xi_0}{\sigma_0}\right)\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right)d\mu(x) = \mathbf{0} \tag{4}$$

for any $\mathbf{u}\{(x-\xi_0)/\sigma_0\} \in \Gamma_\pi$ and

$$\int \mathbf{v}\left(\frac{x-\xi_0}{\sigma_0}\right)\frac{2}{\sigma_0}\phi\left(\frac{x-\xi_0}{\sigma_0}\right)\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right)d\mu(x) = \mathbf{0}. \tag{5}$$

Because $2\mathbf{u}(x)\phi(x)\pi_0(x)/\sigma_0$ is an arbitrary odd function, $\mathbf{v}\{(x-\xi_0)/\sigma_0\}$ has to be an even function of $(x-\xi_0)/\sigma_0$ to ensure (4). Notice that $\pi_0(x) - 1/2$ is in fact an odd function, so we get $\int \mathbf{v}(x)\phi(x)\,d\mu(x) = \mathbf{0}$ from (5). Likewise, for any even

function $\mathbf{v}(x)$, where $\int \mathbf{v}(x)\phi(x)\,d\mu(x) = \mathbf{0}$, we can verify that (4) and (5) are satisfied, and hence $\mathbf{v}\{(x-\xi_0)/\sigma_0\} \in \Gamma_\pi^\perp$.

Because influence functions for RAL estimators belong to the nuisance tangent space orthogonal complement derived in Proposition 2, this motivates estimators obtained by solving the following estimating equations.

Proposition 3. For any even function $\mathbf{v}(x):\mathbb{R} \to \mathbb{R}^2$ s.t. $\int \mathbf{v}(x)\phi(x)\,d\mu(x) = \mathbf{0}$, $\sum_{i=1}^n \mathbf{v}\{(X_i-\xi)/\sigma\} = \mathbf{0}$ defines an RAL estimator for $\boldsymbol{\beta} = (\xi, \sigma)^T$.

Proposition 3 provides us a way of constructing RAL estimators as long as we can find a suitable function $\mathbf{v} = (v_1, v_2)^T$. For example, we can take any even function $h(x)$ and construct v_1 or v_2 to be $h(x) - \int h(x)\phi(x)\,d\mu(x)$. If we take h to be x^{2k}, then the corresponding components of the $\mathbf{v}$ functions are $v_i(x) = x^2 - 1$, $v_i(x) = x^4 - 3$, $v_i(x) = x^6 - 15$, and so on, for $i = 1, 2$.

Because the functions v_i are unbiased, regularity conditions will ensure the existence and uniqueness of a consistent sequence of estimators for ξ and σ (see Foutz 1977 for the regularity conditions in detail). However, in practice when we solve the estimating equation for a single dataset, the solution is not necessarily unique. Care must be taken in selecting a suitable estimate. Although there is no definite solution to this problem, we recommend the following action when multiple roots occur. If only one solution $\hat{\xi}$ is within the range of all of the observed X_i's, $i = 1, \ldots, n$, then pick this one as $\hat{\xi}$ and its accompanying $\hat{\sigma}$. If one has certain understanding of the selection criterion (say, the population mean would tend to be smaller than the sample mean as in the example of the clinic), then pick the solution that makes practical sense ($\hat{\xi} < \bar{X}$). In general, the sensible choice needs to be determined on a case-by-case basis, and no universal rule is available.

As mentioned earlier, the variance of an RAL estimator is the variance of its influence function. The RAL estimator with the smallest variance is referred to as the *semiparametric efficient* estimator. It is known (Bickel et al. 1993) that the semiparametric efficient estimator is the RAL estimator that has an influence function proportional to the efficient score. The efficient score $\mathbf{S}_{\text{eff}}$ is the residual after projecting the score vector with respect to $\boldsymbol{\beta}$ onto the nuisance tangent space, that is, $\mathbf{S}_{\text{eff}} = \mathbf{S}_\beta - \Pi(\mathbf{S}_{\boldsymbol{\beta}}|\Gamma_\pi)$. The corresponding influence function is given by $\boldsymbol{\psi}_{\text{eff}} = \text{cov}(\mathbf{S}_{\boldsymbol{\beta}}, \mathbf{S}_{\text{eff}})^{-1}\mathbf{S}_{\text{eff}} = \text{var}(\mathbf{S}_{\text{eff}})^{-1}\mathbf{S}_{\text{eff}}$, whose variance $\text{var}(\mathbf{S}_{\text{eff}})^{-1}$ is smallest among all of the influence functions. Here by smallest, we mean that the difference $\text{var}(\boldsymbol{\psi}) - \text{var}(\boldsymbol{\psi}_{\text{eff}})$ is nonnegative definite for any influence function $\boldsymbol{\psi}$. We derive $\mathbf{S}_{\text{eff}}$ and calculate the optimal variance in the following propositions.

Proposition 4. The efficient score function is

$$\mathbf{S}_{\text{eff}} = \left[\frac{x-\xi_0}{\sigma_0^2}\left\{2\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right) - 1\right\} - \frac{2}{\sigma_0}\pi_{01}\left(\frac{x-\xi_0}{\sigma_0}\right), \frac{(x-\xi_0)^2}{\sigma_0^3} - \frac{1}{\sigma_0}\right]^T,$$

where $\pi_{01}(x) = d\pi_0(x)/dx$.

Proof. Calculating $\partial \log f(x;\boldsymbol{\beta},\boldsymbol{\alpha})/\partial\xi$ and $\partial \log f(x;\boldsymbol{\beta},\boldsymbol{\alpha})/\partial\sigma$, evaluating at ξ_0 and σ_0 yields the score vector

$$\mathbf{S}_{\boldsymbol{\beta}} = \left\{\frac{x-\xi_0}{\sigma_0^2} - \frac{\pi_{01}\{(x-\xi_0)/\sigma_0\}}{\sigma_0\pi_0\{(x-\xi_0)/\sigma_0\}}, -\frac{1}{\sigma_0} + \frac{(x-\xi_0)^2}{\sigma_0^3} - \frac{(x-\xi_0)\pi_{01}\{(x-\xi_0)/\sigma_0\}}{\sigma_0^2\pi_0\{(x-\xi_0)/\sigma_0\}}\right\}^T.$$

We calculate the projection of $\mathbf{S}_{\boldsymbol{\beta}}$ onto $\Gamma_\pi^\perp$ through using the fact that the difference between $\mathbf{S}_{\boldsymbol{\beta}}$ and its projection onto $\Gamma_\pi^\perp$ is an element in Γ_π. Assume that the projection is $[v_1\{(x-\xi_0)/\sigma_0\}, v_2\{(x-\xi_0)/\sigma_0\}]$, where both v_1 and v_2 are even functions; then

$$\begin{aligned}
&\left\{\frac{x-\xi_0}{\sigma_0^2} - \frac{1}{\sigma_0}\pi_{01}\left(\frac{x-\xi_0}{\sigma_0}\right)\Big/\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right) - v_1\left(\frac{x-\xi_0}{\sigma_0}\right)\right\} \times \pi_0\left(\frac{x-\xi_0}{\sigma_0}\right) \\
&\quad + \left[\frac{-x+\xi_0}{\sigma_0^2} - \frac{1}{\sigma_0}\pi_{01}\left(\frac{x-\xi_0}{\sigma_0}\right)\Big/\left\{1-\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right)\right\} - v_1\left(\frac{x-\xi_0}{\sigma_0}\right)\right] \\
&\quad \times \left\{1-\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right)\right\} = 0
\end{aligned}$$

and

$$\begin{aligned}
&\left\{-\frac{1}{\sigma_0} + \frac{(x-\xi_0)^2}{\sigma_0^3} - \frac{(x-\xi_0)}{\sigma_0^2}\pi_{01}\left(\frac{x-\xi_0}{\sigma_0}\right)\Big/\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right) - v_2\left(\frac{x-\xi_0}{\sigma_0}\right)\right\}\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right) \\
&\quad + \left[-\frac{1}{\sigma_0} + \frac{(x-\xi_0)^2}{\sigma_0^3} + \frac{(x-\xi_0)}{\sigma_0^2}\pi_{01}\left(\frac{x-\xi_0}{\sigma_0}\right)\Big/\left\{1-\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right)\right\} - v_2\left(\frac{x-\xi_0}{\sigma_0}\right)\right] \\
&\quad \times \left\{1-\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right)\right\} = 0.
\end{aligned}$$

Notice that we used the fact that $\pi_{01}(x)$ is an even function of x. Solving the two equations yields the result.

Proposition 5. A semiparametric efficient estimator of $\boldsymbol{\beta} = (\xi, \sigma)^T$ is given by

$$\sum_{i=1}^n \mathbf{F}_0(X_i;\xi,\sigma) = \mathbf{0}, \tag{6}$$

where

$$\begin{aligned}
&\mathbf{F}_0(X_i;\xi,\sigma) \\
&\quad = \left(\left[\frac{X_i-\xi}{\sigma}\left\{2\pi_0\left(\frac{X_i-\xi}{\sigma}\right) - 1\right\} - 2\pi_{01}\left(\frac{X_i-\xi}{\sigma}\right)\right], \{(X_i-\xi)^2 - \sigma^2\}\right)^T.
\end{aligned}$$

Assume that the estimator obtained through solving (6) is $\hat{\boldsymbol{\beta}}$; then $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to \mathrm{N}_2(\mathbf{0}, \{\mathrm{E}(\mathbf{S}_{\mathrm{eff}}\mathbf{S}_{\mathrm{eff}}^T)\}^{-1})$ in distribution. Here the smallest variance of the estimate given by $\{\mathrm{E}(\mathbf{S}_{\mathrm{eff}}\mathbf{S}_{\mathrm{eff}}^T)\}^{-1}$ has the form

$$\mathbf{A} = \sigma_0^2 \begin{pmatrix} \int [\{2\pi_0(x) - 1\}^2 + 4\pi_{01}(x)^2]\phi(x)\,d\mu(x) & 4\int \pi_{01}(x)\phi(x)\,d\mu(x) \\ 4\int \pi_{01}(x)\phi(x)\,d\mu(x) & 2 \end{pmatrix}^{-1}. \quad (7)$$

Remark 1. Notice that when $\pi_0(x) \equiv 1/2$, the first component of the efficient score vector is 0, in which case an efficient semiparametric estimator does not exist. Similar phenomena have been observed in Bayesian analysis of selection models, where a constant weight function [corresponding to $\pi_0(x) \equiv 1/2$ in our case] has to be ruled out a priori to any analysis (see Lee and Berger 2001).

Remark 2. The only situation for the semiparametric efficient estimator to degenerate is when $\pi_0(x) \equiv 1/2$. This can be verified by inspecting the differential equation $x\{2\pi_0(x) - 1\} - 2\pi_{01}(x) = c(x^2 - 1)$, for an arbitrary constant c. The solution to this equation is of the form $\pi_0(x) = (cx + 1)/2 + de^{x^2/2}$, where d is a constant. Subject to the constraint that $\pi_0(x) + \pi_0(-x) = 1$ and $\pi_0(x)$ is nonnegative, both c and d are 0 and $\pi_0(x) \equiv 1/2$ is the only legitimate solution. Thus, as long as the true model has a nontrivial skewing function, a semiparametric efficient estimator always exists.

Remark 3. As long as π_0 is differentiable, regardless of whether or not it is a constant, a consistent estimator always exists, and hence the problem is always identifiable. For example, one consistent estimator is given by adopting $\mathbf{v}(x) = (x^4 - 3, x^2 - 1)^T$.

We omit the proof of Proposition 5, which involves only straightforward algebra. The efficient estimator defined by (6) depends on using the true skewing function π_0, which is unknown to us. However, any choice of a differentiable skewing function in (6) will lead to a consistent asymptotically normal estimator for $\boldsymbol{\beta}$, as long as we are not using $\pi(x) \equiv 1/2$. This can be shown by noticing that $\mathbf{v}(x) = [x\{2\pi(x) - 1\} - 2\pi_1(x), x^2 - 1]^T$ satisfies the requirement in Proposition 3, where $\pi_1(x) = d\pi(x)/dx$. In fact, such an estimator is guaranteed to be nondegenerate, that is, $x\{2\pi(x) - 1\} - 2\pi_1(x) \not\propto x^2 - 1$. This is because had $\pi(x)$ been the correct skewing function, then $\mathbf{v}(x)$ would have been the efficient estimator, and hence it is at least nondegenerate. In practice, we generally posit a model for $\pi(\cdot)$ in terms of a finite set of parameters $\boldsymbol{\alpha}$, say, $\pi(x/\sigma - \xi/\sigma, \boldsymbol{\alpha})$, and then estimate $\boldsymbol{\alpha}$ using an estimator $\hat{\boldsymbol{\alpha}}$. We use

$$\sum_{i=1}^{n} \mathbf{F}(X_i; \xi, \sigma, \hat{\boldsymbol{\alpha}}) = \mathbf{0} \quad (8)$$

to denote estimators of the form in (6) with $\pi_0\{(x - \xi)/\sigma\}$ replaced by $\pi\{(x-\xi)/\sigma, \hat{\boldsymbol{\alpha}}\}$. Notice that $\mathrm{E}\{\mathbf{F}(X_i; \xi, \sigma, \boldsymbol{\alpha})\} = \mathbf{0}$ for all values of $\boldsymbol{\alpha}$, and hence $\mathrm{E}\{\partial\mathbf{F}(X_i; \xi, \sigma, \boldsymbol{\alpha})/\partial\boldsymbol{\alpha}\} = \mathbf{0}$ assuming sufficiently smooth conditions on $\mathbf{F}$ to interchange the expectation and the partial derivative. If the true skewing function belongs to this parametric model, then $\pi(\cdot, \hat{\boldsymbol{\alpha}})$ will converge to $\pi_0(\cdot)$. But even if the parametric model does not contain the true $\pi_0(\cdot)$, the estimate $\hat{\boldsymbol{\alpha}}$ will generally converge to a constant $\boldsymbol{\alpha}^*$ and $\pi(\cdot, \hat{\boldsymbol{\alpha}})$ will converge to some skewing function $\pi(\cdot, \boldsymbol{\alpha}^*)$. As long as $n^{1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)$ is bounded in probability, as we show in the next proposition, the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ obtained by using $\pi(\cdot, \hat{\boldsymbol{\alpha}})$ is asymptotically the same as that which uses $\pi(\cdot, \boldsymbol{\alpha}^*)$, which we have argued is consistent and asymptotically normal. However, if the parametric model does contain the truth, then the estimator for $\boldsymbol{\beta}$ in (8) is semiparametric efficient. Such estimators are referred to as *locally efficient*.

Similar to the discussion after Proposition 3, (6) and (8) could have multiple solutions. We use the same rule as mentioned after Proposition 3 to decide which solution to choose in practice.

Proposition 6. Assume that $\frac{2}{\sigma}\phi\{(x - \xi)/\sigma\}\pi\{(x - \xi)/\sigma, \boldsymbol{\alpha}\}$ is a parametric model and that $n^{1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)$ is bounded in probability. Then the two RAL estimators resulting from solving the two estimating equations $\sum_{i=1}^n \mathbf{F}(X_i; \xi, \sigma, \boldsymbol{\alpha}^*) = \mathbf{0}$ and $\sum_{i=1}^n \mathbf{F}(X_i; \xi, \sigma, \hat{\boldsymbol{\alpha}}) = \mathbf{0}$ are asymptotically equivalent; that is, if $(\hat{\xi}_1, \hat{\sigma}_1)$ is the estimator obtained through solving the first equation and $(\hat{\xi}_2, \hat{\sigma}_2)$ is the estimator obtained through solving the second equation, then $n^{1/2}(\hat{\xi}_1 - \hat{\xi}_2) \to 0$ and $n^{1/2}(\hat{\sigma}_1 - \hat{\sigma}_2) \to 0$ in probability.

Proof. Write $(\xi, \sigma)^T$ as $\boldsymbol{\beta}$ and $\mathbf{F}(X_i; \xi, \sigma, \boldsymbol{\alpha})$ as $\mathbf{F}(X_i; \boldsymbol{\beta}, \boldsymbol{\alpha})$. A Taylor expansion of $\sum_{i=1}^n \mathbf{F}(X_i; \hat{\boldsymbol{\beta}}_2, \boldsymbol{\alpha}^*)$ at $\hat{\boldsymbol{\alpha}}$ yields $\sum_{i=1}^n \mathbf{F}(X_i; \hat{\boldsymbol{\beta}}_2, \boldsymbol{\alpha}^*) = \sum_{i=1}^n \mathbf{F}(X_i; \hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\alpha}}) + \{\sum_{i=1}^n \partial\mathbf{F}(X_i; \hat{\boldsymbol{\beta}}_2, \tilde{\boldsymbol{\alpha}})/\partial\boldsymbol{\alpha}^T\}(\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}})$, where $\tilde{\boldsymbol{\alpha}}$ is between $\boldsymbol{\alpha}^*$ and $\hat{\boldsymbol{\alpha}}$. Denoting $\{\sum_{i=1}^n \partial\mathbf{F}(X_i; \hat{\boldsymbol{\beta}}_2, \tilde{\boldsymbol{\alpha}})/\partial\boldsymbol{\alpha}^T\}/n$ by $\boldsymbol{\Lambda}_n$, we obtain $\sum_{i=1}^n \mathbf{F}(X_i; \hat{\boldsymbol{\beta}}_2, \boldsymbol{\alpha}^*) = n\boldsymbol{\Lambda}_n(\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}})$. Notice that when $n \to \infty$, because of the convergence of $\hat{\boldsymbol{\alpha}}$ to $\boldsymbol{\alpha}^*$ and the consistency property of ξ_2 and σ_2, $\boldsymbol{\Lambda}_n \to \mathrm{E}\{\partial\mathbf{F}(X_i; \boldsymbol{\beta}_0, \boldsymbol{\alpha}^*)/\partial\boldsymbol{\alpha}^T\} = \mathbf{0}$ in probability.

A Taylor expansion of $\sum_{i=1}^n \mathbf{F}(X_i; \hat{\boldsymbol{\beta}}_2, \boldsymbol{\alpha}^*)$ at $\hat{\boldsymbol{\beta}}_1$ yields

$$\begin{aligned} \hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1 &= \left\{\sum_{i=1}^n \frac{\partial\mathbf{F}(X_i; \tilde{\boldsymbol{\beta}}, \boldsymbol{\alpha}^*)}{\partial\boldsymbol{\beta}^T}\right\}^{-1} \left\{\sum_{i=1}^n \mathbf{F}(X_i; \hat{\boldsymbol{\beta}}_2, \boldsymbol{\alpha}^*) - 0\right\} \\ &= \left\{\frac{1}{n}\sum_{i=1}^n \frac{\partial\mathbf{F}(X_i; \tilde{\boldsymbol{\beta}}, \boldsymbol{\alpha}^*)}{\partial\boldsymbol{\beta}^T}\right\}^{-1} \boldsymbol{\Lambda}_n(\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}}), \end{aligned}$$

where $\tilde{\boldsymbol{\beta}}$ is a quantity between $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$.

When $n \to \infty$,

$$\mathbf{J}_n = \frac{1}{n}\sum_{i=1}^n \frac{\partial\mathbf{F}(X_i; \tilde{\boldsymbol{\beta}}, \boldsymbol{\alpha}^*)}{\partial\boldsymbol{\beta}^T} \to \mathrm{E}\left\{\frac{\partial\mathbf{F}(X_i; \boldsymbol{\beta}_0, \boldsymbol{\alpha}^*)}{\partial\boldsymbol{\beta}^T}\right\}$$

in probability. For parametric models, $\mathrm{E}\{\partial\mathbf{F}(X_i; \boldsymbol{\beta}_0, \boldsymbol{\alpha}^*)/\partial\boldsymbol{\beta}^T\}$ is the matrix related to the Fisher information matrix, which is generally nonsingular, and we denote it by $\mathbf{J}$. Combining the results, we have $n^{1/2}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2) = n^{1/2}\mathbf{J}_n^{-1}\boldsymbol{\Lambda}_n(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)$. Because $n^{1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^*)$ is bounded in probability, $\mathbf{J}_n^{-1} \to \mathbf{J}^{-1}$ in probability and $\boldsymbol{\Lambda}_n \to \mathbf{0}$ in probability, which implies that $\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2 \to \mathbf{0}$ in probability.

In practice, $\boldsymbol{\alpha}$ is estimated using a specific estimator, and its convergence rate is determined by the corresponding estimator implemented. For example, if the maximum likelihood estimator (MLE) is used to estimate $\boldsymbol{\alpha}$, then we would know

that the boundedness condition of Proposition 6 is automatically satisfied. More important, Proposition 6 indicates that how efficiently we estimate the nuisance parameter $\boldsymbol{\alpha}$ does not influence how efficiently we can estimate ξ and σ. In fact, as long as we can estimate $\boldsymbol{\alpha}$ consistently, using the estimated value of $\boldsymbol{\alpha}$, $\hat{\boldsymbol{\alpha}}$, or the true value of $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}_0$, will yield the same efficiency for ξ and σ.

The efficiency of an estimator depends on how close the true π_0 is to the parametric family $\{\pi(x, \boldsymbol{\alpha})\}$. One way to construct the parametric model proposed by Ma and Genton (2004) is to use $\Phi\{P_K(x)\}$ to approximate $\pi_0(x)$, where $P_K(x)$ is an odd polynomial of order K. Because an odd polynomial can approximate a continuous odd function arbitrarily well, $\Phi\{P_K(x)\}$ will approximate $\pi_0(x) = \Phi\{m_0(x)\}$ well, and hence will make the "distance" between $\Phi\{P_K(x)\}$ and π_0 arbitrarily small. In general, the relationship between the efficiency loss and the "distance" between π_0 and the parametric family $\{\pi(x, \boldsymbol{\alpha})\}$ that approximates π_0 is given in the following proposition.

Proposition 7. Let $\nu(x) = \pi(x, \boldsymbol{\alpha}) - \pi_0(x)$, $\theta = \int 4[\partial\{\nu(x) \times \phi(x)\}/\partial x]^2/\phi(x)\, d\mu(x)$. The most efficient semiparametric estimator of the form in (8) has efficiency $\mathbf{A} + \min_{\boldsymbol{\alpha}}(\theta)\mathbf{B}$, where $\mathbf{A}$ is given by (7), and

$$\mathbf{B} = \frac{\sigma_0^2}{[\mathrm{E}\{2\pi_0(X) - 1 + 2X\pi_{01}(X) - 2\pi_{02}(X)\} - 2\mathrm{E}(X)^2]^2} \times \begin{Bmatrix} 1 & -\mathrm{E}(X) \\ -\mathrm{E}(X) & \mathrm{E}(X)^2 \end{Bmatrix},$$

which does not depend on the estimator in (8). Here π_{02} denotes $d^2\pi_0(x)/dx^2$, and the expectations are taken with respect to $2\phi(x)\pi_0(x)$.

Proof. Assume that the estimating equation $\sum_{i=1}^n \mathbf{F}(X_i; \boldsymbol{\beta}, \hat{\boldsymbol{\alpha}}) = \mathbf{0}$ yields the estimate $\hat{\boldsymbol{\beta}}_1 = (\hat{\xi}_1, \hat{\sigma}_1)^T$, the estimating equation $\sum_{i=1}^n \mathbf{F}_0(X_i; \boldsymbol{\beta}) = \mathbf{0}$ yields the estimate $\hat{\boldsymbol{\beta}} = (\hat{\xi}, \hat{\sigma})^T$. Then

$$\begin{aligned}
\mathbf{0} &= \sum_{i=1}^n \mathbf{F}(X_i; \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\alpha}}) \\
&= \sum_{i=1}^n \mathbf{F}_0(X_i; \hat{\boldsymbol{\beta}}) + \sum_{i=1}^n \{\mathbf{F}_0(X_i; \hat{\boldsymbol{\beta}}_1) - \mathbf{F}_0(X_i; \hat{\boldsymbol{\beta}})\} \\
&\quad + \sum_{i=1}^n \{\mathbf{F}(X_i; \hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\alpha}}) - \mathbf{F}_0(X_i; \hat{\boldsymbol{\beta}}_1)\} \\
&= \sum_{i=1}^n \frac{\partial \mathbf{F}_0(X_i; \tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^T}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}) \\
&\quad + \sum_{i=1}^n \left\{ \frac{X_i - \hat{\xi}_1}{\hat{\sigma}_1} 2\nu\left(\frac{X_i - \hat{\xi}_1}{\hat{\sigma}_1}\right) - 2\nu_1\left(\frac{X_i - \hat{\xi}_1}{\hat{\sigma}_1}\right), 0 \right\}^T,
\end{aligned}$$

where $\tilde{\boldsymbol{\beta}}$ is a quantity between $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_1$, $\nu_1(x) = d\nu(x)/dx$. Notice that when $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^n \frac{\partial \mathbf{F}_0(X_i; \tilde{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}^T} \to \mathrm{E}\left\{ \frac{\partial \mathbf{F}_0(X_i; \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T} \right\}$$

in probability,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \left\{ \frac{X_i - \hat{\xi}_1}{\hat{\sigma}_1} 2\nu\left(\frac{X_i - \hat{\xi}_1}{\hat{\sigma}_1}\right) - 2\nu_1\left(\frac{X_i - \hat{\xi}_1}{\hat{\sigma}_1}\right) \right\} \\
&\quad \to \mathrm{E}\left\{ 2\frac{X_i - \hat{\xi}_1}{\hat{\sigma}_1}\nu\left(\frac{X_i - \hat{\xi}_1}{\hat{\sigma}_1}\right) - 2\nu_1\left(\frac{X_i - \hat{\xi}_1}{\hat{\sigma}_1}\right) \right\} \\
&\quad \to \mathrm{E}\left\{ 2\frac{X_i - \xi_0}{\sigma_0}\nu\left(\frac{X_i - \xi_0}{\sigma_0}\right) - 2\nu_1\left(\frac{X_i - \xi_0}{\sigma_0}\right) \right\} \\
&\quad = 0
\end{aligned}$$

in probability due to the consistency of $\hat{\xi}_1$ and $\hat{\sigma}_1$. We calculate the variance of $2\{(X_i - \xi_0)/\sigma_0\}\nu\{(X_i - \xi_0)/\sigma_0\} - 2\nu_1\{(X_i - \xi_0)/\sigma_0\}$, which is an even function of $(X_i - \xi_0)/\sigma_0$,

$$\begin{aligned}
&\mathrm{E}\left[\left\{ 2\frac{X_i - \xi_0}{\sigma_0}\nu\left(\frac{X_i - \xi_0}{\sigma}\right) - 2\nu_1\left(\frac{X_i - \xi_0}{\sigma_0}\right) \right\}^2 \right] \\
&\quad = 4\int \{x\nu(x) - \nu_1(x)\}^2 2\phi(x)\pi_0(x)\, d\mu(x) \\
&\quad = 4\int \{x\nu(x) - \nu_1(x)\}^2 \phi(x)\, d\mu(x) \\
&\quad = \int \frac{4}{\phi(x)}\left[\frac{\partial}{\partial x}\{\nu(x)\phi(x)\} \right]^2 d\mu(x) \\
&\quad = \theta.
\end{aligned}$$

Thus

$$n^{1/2}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}) \to \mathrm{N}_2\left(\mathbf{0}, \left[\mathrm{E}\left\{ \frac{\partial \mathbf{F}_0(X_i; \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T} \right\}\right]^{-1} \begin{pmatrix} \theta & 0 \\ 0 & 0 \end{pmatrix} \times \left[\mathrm{E}\left\{ \frac{\partial \mathbf{F}_0(X_i; \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T} \right\}\right]^{-T}\right)$$

in distribution. It can be verified that

$$\mathrm{E}\left\{ \frac{\partial \mathbf{F}_0(X_i; \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}^T} \right\} = -\begin{bmatrix} \frac{\mathrm{E}\{2\pi_0(X) - 1 + 2X\pi_{01}(X) - 2\pi_{02}(X)\}}{\sigma_0} & \frac{2\mathrm{E}(X)}{\sigma_0} \\ 2\sigma_0\mathrm{E}(X) & 2\sigma_0 \end{bmatrix},$$

where expectation E on the right side is taken with respect to $2\phi(x)\pi_0(x)$. Putting these together, we get $n^{1/2}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}) \to \mathrm{N}_2(\mathbf{0}, \theta\mathbf{B})$ in distribution, and thus $n^{1/2}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_0) \to \mathrm{N}_2(\mathbf{0}, \mathbf{A} + \theta\mathbf{B})$ in distribution. With an $\boldsymbol{\alpha}$ that minimizes θ, we will get the most efficient estimator given the parametric model $\pi(x, \boldsymbol{\alpha})$. The variance of $n^{1/2}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_0)$ is $\mathbf{A} + \min_{\boldsymbol{\alpha}}(\theta)\mathbf{B}$.

In Proposition 7 we deliberately avoided specifying how to find the $\boldsymbol{\alpha}$ that minimizes θ, because this depends on the true π_0 that is unknown to us. In practice, we can always estimate $\boldsymbol{\beta}$ and calculate its variance for any fixed $\boldsymbol{\alpha}$ and select the $\boldsymbol{\alpha}$ that yields the smallest estimation variance. Thus the $\boldsymbol{\alpha}$ that minimizes θ can be found numerically. Often, a parametric model is assumed in terms of both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, and the MLE is used to estimate both sets of parameters. But if the model for the skewing function is not correct, then the MLE for $\boldsymbol{\beta}$ will be biased. A correction procedure should follow, where after obtaining the MLE $\hat{\boldsymbol{\alpha}}$ in the π function, we need to proceed to estimate ξ and σ using the semiparametric estimating equation in (6) with π_0 replaced by π. Notice that the $\hat{\boldsymbol{\alpha}}$ obtained through the MLE need not be the $\boldsymbol{\alpha}$ that minimizes θ. However, the resulting estimator will

be consistent and asymptotically normal even if the model for π_0 is incorrectly specified and will be semiparametric efficient if it is correctly specified.

4. SIMULATION RESULTS

We carried out a simulation study with a sample size of 500. We generated the datasets from the distribution $\frac{2}{\sigma}\phi\{(x-\xi)/\sigma\}\Phi[(\sin\{c(x-\xi)/\sigma\}]$ with $\sigma = 1$, $\xi = 3$, and $c = -2$. We approximated the true $\pi_0(x) = \Phi[(\sin\{c(x-\xi)/\sigma\}]$ with $\pi_K(x) = H[P_K\{(x-\xi)/\sigma\}]$, where H is the logit link function [i.e., $H(x) = 1/\{1+\exp(-x)\}$] and P_K is an odd polynomial of order K. We generated 1,000 datasets and calculated the empirical variances of the estimates and also the average of the estimated variances. We calculated the estimated variance via the standard sandwich matrix of M-estimators; that is, we calculated

$$\left\{\sum_{i=1}^{n} D\mathbf{F}(X_i;\xi,\sigma,\hat{\boldsymbol{\alpha}})\right\}^{-1} \times \left\{\sum_{i=1}^{n} \mathbf{F}(X_i;\xi,\sigma,\hat{\boldsymbol{\alpha}})\mathbf{F}(X_i;\xi,\sigma,\hat{\boldsymbol{\alpha}})^T\right\} \times \left\{\sum_{i=1}^{n} D\mathbf{F}(X_i;\xi,\sigma,\hat{\boldsymbol{\alpha}})\right\}^{-T} \tag{9}$$

as the estimated variance matrix, where $\mathbf{F}(X_i;\xi,\sigma,\hat{\boldsymbol{\alpha}})$ is the same as in (8), $D\mathbf{F}(X_i;\xi,\sigma,\hat{\boldsymbol{\alpha}})$ is the Jacobian of $\mathbf{F}(X_i;\xi,\sigma,\hat{\boldsymbol{\alpha}})$ with respect to ξ and σ, and $\hat{\boldsymbol{\alpha}}$ is the MLE of the polynomial coefficients $\boldsymbol{\alpha}$ when fitting the data with $\frac{2}{\sigma}\phi\{(x-\xi)/\sigma\}H[P_K\{(x-\xi)/\sigma\}]$. Notice that the variance resulting from estimating the parameters in the skewing function is not taken into account; however, the final average estimated variance still agrees with the empirical variance, which is exactly what we expected due to the result in Proposition 6. The simulation results are given in the upper half of Table 1.

For comparison, we also adopted the correct model for $\pi_0(x)$; that is, we set $\pi_t(x) = \Phi[(\sin\{\alpha(x-\xi)/\sigma\}]$, with α being the nuisance parameter. We can verify that all three estimators are unbiased, whereas the estimator with the true posited model for $\pi_0(x)$ has the smallest variance. The variance for $\pi_3(x)$ is smaller than that for $\pi_1(x)$, because $\pi_3(x)$ approximates $\pi_0(x)$ better. In fact, as shown by Ma and Genton (2004), $\pi_0(x)$ can be approximated arbitrarily well if we allow the order of the odd polynomial to increase sufficiently, and hence the estimator will approach the most efficient one.

We also estimated ξ and σ using the $\hat{\boldsymbol{\alpha}}$ that minimizes the "distance" θ between $\pi_0(x)$ and $\pi_K(x)$, that is, minimizes the variance in (9). The results are tabulated in the lower half of Table 1. It is clear that the variance of the estimators in the lower part of the table is improved compared with the corresponding estimators in the upper part of the table, where $\hat{\boldsymbol{\alpha}}$ is simply obtained through MLE. In fact, the estimation variance when using $\pi_3(x)$ is so close to that when using the correct model $\pi_t(x)$ that as far as estimation of ξ and σ is concerned, there is hardly any need to go for an approximation of a higher order polynomial. We plotted the resulting average estimated pdf's $\frac{2}{\sigma}\phi\{(x-\xi)/\sigma\}\pi_1(x)$, $\frac{2}{\sigma}\phi\{(x-\xi)/\sigma\}\pi_3(x)$, and $\frac{2}{\sigma}\phi\{(x-\xi)/\sigma\}\pi_t(x)$ in Figure 1. The difference between these curves and the true pdf indicates that the consistency property is not a result of the similarity between these pdf's. It also indicates that being able to estimate the population parameters ξ and σ does not necessarily mean being able to estimate the skewed distribution of a biased subsample.

To support the result in the example in Section 5, we performed a simulation study with the data generated in a "similar" way to the example. Specifically, we generated the datasets from a distribution $\frac{2}{\sigma}\phi\{(x-\xi)/\sigma\}\Phi\{c(x-\xi)/\sigma\}$ with $\sigma = 4$, $\xi = 20$, and $c = 5$. The same estimators as in the previous simulation were implemented here. The sample size was 100 and 1,000 datasets were generated and analyzed. The simulation results are tabulated in Table 2, and the average of estimated pdf's are plotted in Figure 2. As one would expect, the asymptotic properties do not *exhibit as clearly* as in the previous simulation, due to the relatively small sample size. In this specific simulation, we find that the estimations of ξ and σ are still unbiased and that most of the estimation variances for ξ and σ match with the sample variances reasonably well. However, for the estimation of ξ, when we use a third-order polynomial and derive the results through minimizing the resulting estimated variance, the estimated variance tends to be smaller than the sample variance. Thus when we are dealing with smaller sample sizes, it is helpful to implement and refer to different estimators to reach more sensible conclusions.

5. AN EXAMPLE

We applied the estimator in (8) to a dataset of Australian athletes' BMI data. This dataset comprises the BMIs of 202 ath-

Table 1. Simulation Results on ξ and σ With Different Posited Skewing Functions $\pi_1(x)$, $\pi_3(x)$, and $\pi_t(x)$

	$\hat{\xi}$ (3)			$\hat{\sigma}$ (1)		
	Mean	Estimated var.	Empirical var.	Mean	Estimated var.	Empirical var.
$\hat{\alpha}$ estimated through MLE						
$\pi_1(x)$	2.9899	.0041	.0040	1.0007	.0012	.0011
$\pi_3(x)$	3.0006	.0027	.0027	1.0018	.0012	.0011
$\pi_t(x)$	2.9977	.0017	.0017	1.0005	.0011	.0011
$\hat{\alpha}$ minimizes the resulting variance						
$\pi_1(x)$	2.9895	.0035	.0036	1.0004	.0012	.0011
$\pi_3(x)$	2.9988	.0018	.0020	1.0009	.0010	.0011
$\pi_t(x)$	2.9976	.0017	.0018	1.0005	.0011	.0011

NOTE: The true values of ξ and σ are 3 and 1. The true skewing function is $\pi_0(x) = \Phi\{\sin(-2x)\}$. The sample size is 500, and 1,000 datasets are simulated.

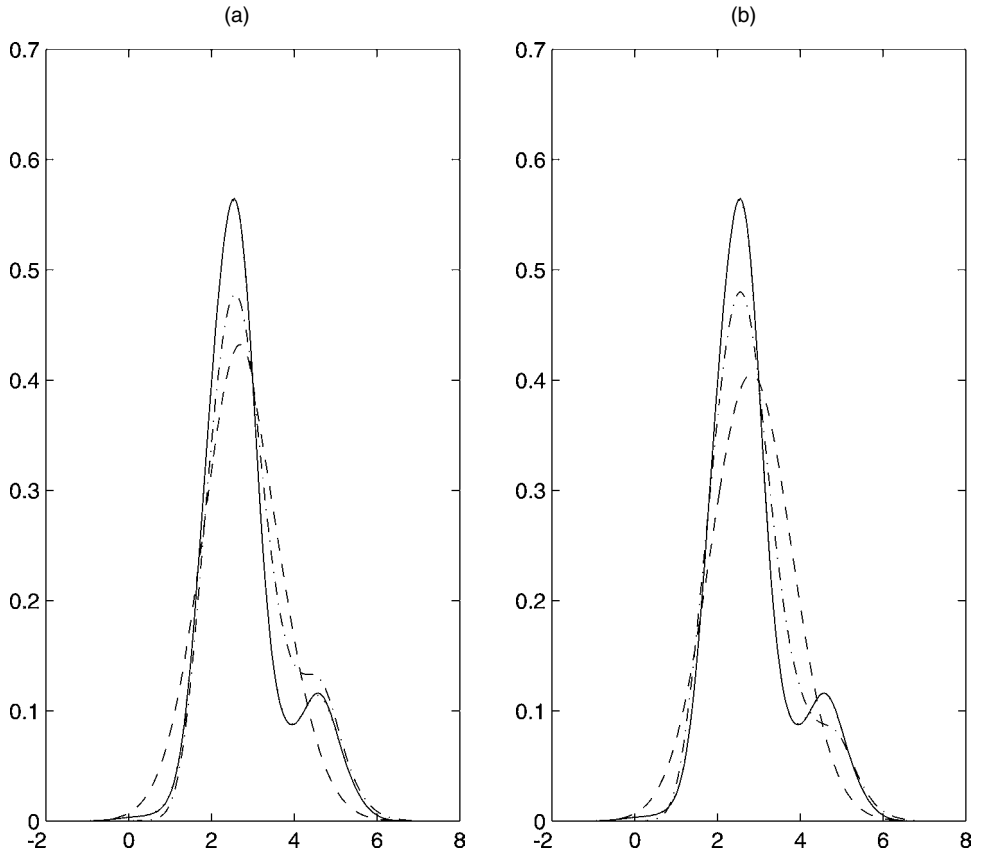

Figure 1. The Average Estimated pdf's Using the Posited Models π_1, π_3, and π_t [—— $\pi_0(x)$; – – $\pi_1(x)$; · – · $\pi_3(x)$; · · · · $\pi_t(x)$]. The true pdf using π_0 and the average estimated pdf using π_t overlay each other and are indistinguishable in these plots. The nuisance parameters in the posited models are selected to minimize the resulting variance (a), and are estimated using MLE (b). The sample size is 500, and 1,000 simulations were done.

letes, including 102 male and 100 female athletes. Histograms of these data are shown in Figure 3.

Assume that we try to infer the mean and variance of the BMI in general Australian male and female adults. Certainly, a simple sample average and variance will give a biased estimate, because athletes would certainly have higher BMIs than the general population. We used a GSN distribution with skewing function $\pi_K(x) = H[P_K\{(x-\xi)/\sigma\}]$ to estimate ξ and σ. Here H is the logit link function and P_K is an odd polynomial of order K. We applied $K = 1$ and $K = 3$. For the nuisance parameters (the coefficients in the polynomial), we estimated them via MLE as well as via minimizing the final total variance of ξ and σ. The results are presented in Table 3. It can be noted that although the estimated variance of ξ using $\pi_3(x)$ appears to be much smaller than the estimated variances for the other three estimators, the simulation study in Section 4 indicates that the estimated variance may be overly optimistic due to the small sample size.

As we expected, the average BMI is lower in the general population than in athletes, and the variance σ^2 is larger in general population than in athletes. The pdf's corresponding to different skewing function π_K's are plotted in Figure 3. Because our estimators are semiparametric, we do not necessarily need to have a good estimate of the skewing function to have a consistent

Table 2. Simulation Results on ξ and σ With Different Posited Skewing Functions $\pi_1(x)$, $\pi_3(x)$, and $\pi_t(x)$

	$\hat{\xi}$ (20)			$\hat{\sigma}$ (4)		
	Mean	*Estimated var.*	*Empirical var.*	*Mean*	*Estimated var.*	*Empirical var.*
$\hat{\alpha}$ estimated through MLE						
$\pi_1(x)$	19.9981	.0958	.0881	3.9954	.1373	.1301
$\pi_3(x)$	19.9946	.0874	.0957	3.9986	.1331	.1333
$\pi_t(x)$	19.9973	.0950	.0890	3.9960	.1360	.1313
$\hat{\alpha}$ minimizes the resulting variance						
$\pi_1(x)$	20.0129	.0856	.0845	3.9839	.1315	.1240
$\pi_3(x)$	20.0239	.0648	.1060	3.9772	.1109	.1372
$\pi_t(x)$	20.0144	.0817	.0844	3.9828	.1294	.1283

NOTE: The true values of ξ and σ are 20 and 4. The true skewing function is $\pi_0(x) = \Phi(5x)$. The sample size is 100, and 1,000 datasets are simulated.

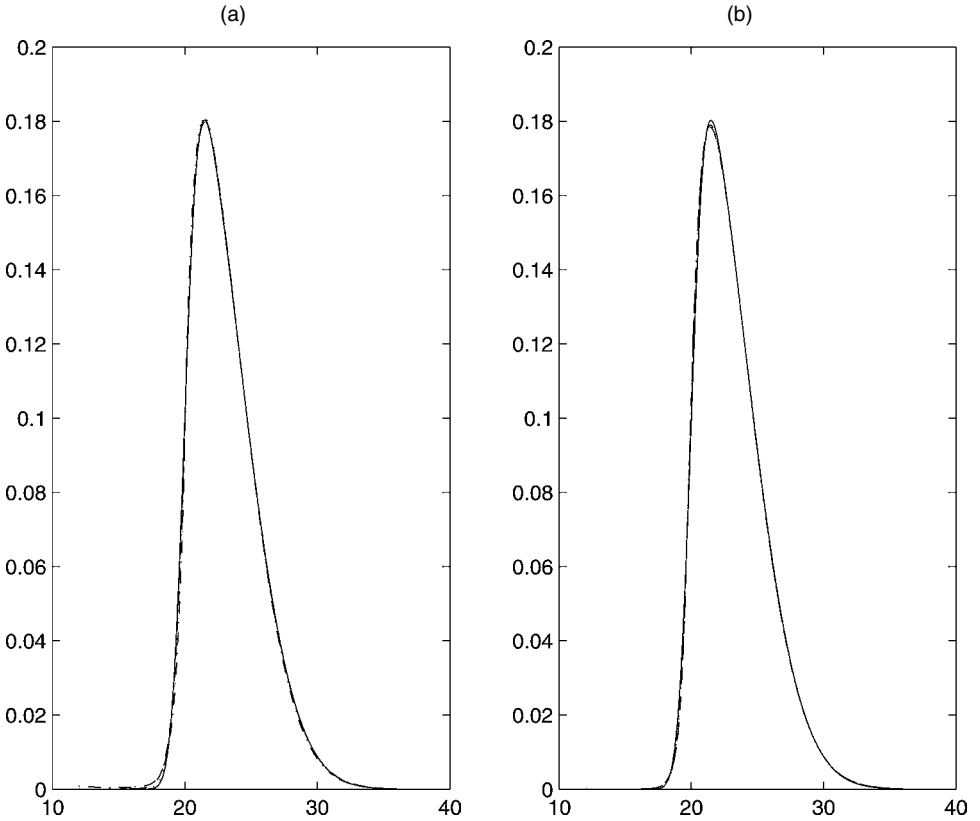

Figure 2. The Average Estimated pdf's Using the Posited Models π_1, π_3, and π_t [—— $\pi_0(x)$; – – $\pi_1(x)$; · — · $\pi_3(x)$; · · · · $\pi_t(x)$]. The curves are very close to each other and are indistinguishable. The nuisance parameters in the posited models are selected to minimize the resulting variance (a), and are estimated using MLE (b). The sample size is 100, and 1,000 simulations were done.

estimator for ξ and σ. Hence the corresponding estimated pdf does not have to fit the observed data. However, as we showed in Proposition 4, the most efficient estimator uses the true skewing function. Consequently, we might expect a good fit of the resulting pdf to the data to be indicative of a more efficient estimator. In our example, we stopped with $K = 3$ because the fit was good, and the dramatic decrease in the estimated standard error from using $\pi_{3m}(x)$ to $\pi_3(x)$, in combination with the simulation result from Section 4, suggests that numerical errors will dominate the resulting estimated standard errors beyond that.

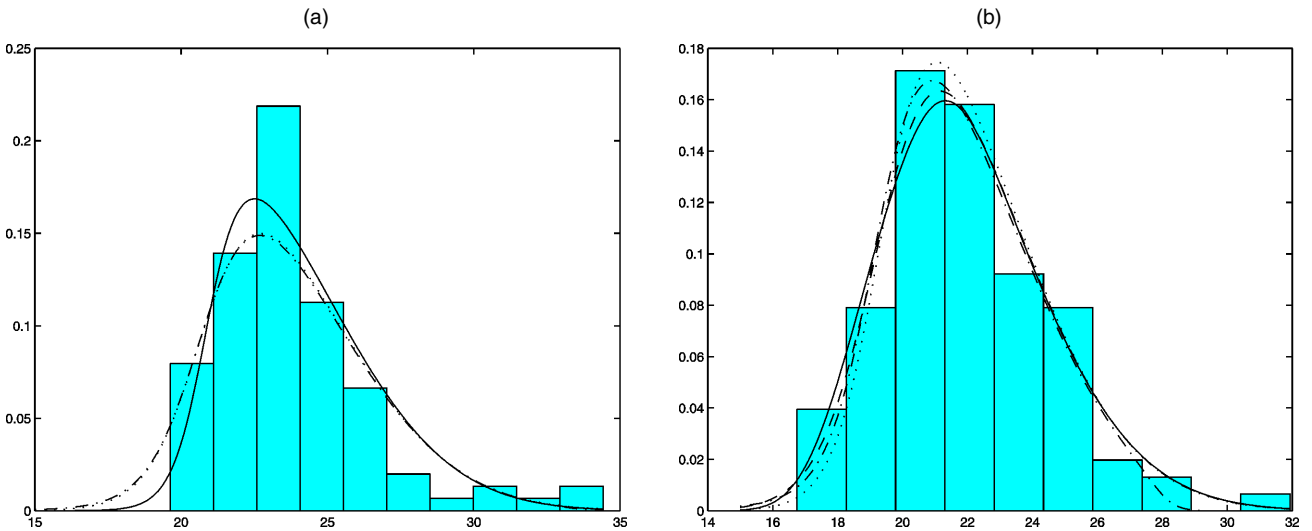

Figure 3. Histograms of 102 Male (a) and 100 Female (b) Australian Athletes' BMIs and pdf's Using Different Posited Skewing Functions. The skewing functions are π_1 and π_3 (whose nuisance parameters are estimated through minimizing estimation variance) and π_{1m} and π_{3m} (whose nuisance parameters are estimated through MLE) [· · · · $\pi_1(x)$; · — · $\pi_3(x)$; – – $\pi_{1m}(x)$; —— $\pi_{3m}(x)$]. In (a), the two pdf's using π_{1m} and π_{3m} overlay each other and are indistinguishable.

Table 3. Estimated Values of ξ and σ and Their Standard Deviations via Different Estimators

	Male				Female			
Skew function	$\hat{\xi}$	*Estimated SD*	$\hat{\sigma}$	*Estimated SD*	$\hat{\xi}$	*Estimated SD*	$\hat{\sigma}$	*Estimated SD*
$\pi_{1m}(x)$	20.8542	.4997	4.1089	.6017	19.2911	.5191	3.7656	.5219
$\pi_{3m}(x)$	20.8542	.4997	4.1089	.6017	19.2225	.5690	3.8151	.5576
$\pi_1(x)$	20.6958	.4276	4.2277	.5801	19.3483	.5036	3.7249	.5061
$\pi_3(x)$	20.6791	.3174	4.2405	.3219	19.1734	.4507	3.8508	.4110
	23.9036	.2727	2.7539	.3044	21.9892	.2627	2.6268	.2341

NOTE: The last estimator is obtained by taking the sample mean and sample standard deviation of the data.

6. DISCUSSION

The derivation of the results in Section 3 also applies to a more general setting of univariate GSE distributions. In fact, the nuisance tangent space Γ_π in that setting remains exactly the same, whereas its orthogonal complement becomes $\Gamma_\pi^\perp = \{\mathbf{v}\{(x-\xi_0)/\sigma_0\} : \mathbf{v}(x)$ is an even function that satisfies $\int \mathbf{v}(x)g(x)\,d\mu(x) = \mathbf{0}\}$, where g is the elliptical part of the GSE distribution. As a result, for such $\mathbf{v}(x)$, $\sum_{i=1}^n \mathbf{v}\{(X_i-\xi)/\sigma\} = \mathbf{0}$ forms an RAL estimator. Similarly, the efficient score function in general is

$$\mathbf{S}_{\text{eff}} = \left[-\frac{g_1(y)}{\sigma_0 g(y)}\{2\pi_0(y)-1\} - \frac{2}{\sigma_0}\pi_{01}(y), -\frac{y g_1(y)}{\sigma_0 g(y)} - \frac{1}{\sigma_0}\right]^T,$$

where $y = (x-\xi_0)/\sigma_0$ and $g_1(y)$ is the first derivative of $g(y)$ with respect to y.

In the multivariate setting, the nuisance tangent space Γ_π still remains exactly the same, whereas its orthogonal complement becomes $\Gamma_\pi^\perp = \{\mathbf{v}\{\boldsymbol{\Sigma}_0^{-1/2}(\mathbf{x}-\boldsymbol{\xi}_0)\} : \mathbf{v}(\mathbf{x})$ is an even function that satisfies $\int \mathbf{v}(\mathbf{x})g(\mathbf{x})\,d\mu(\mathbf{x}) = \mathbf{0}\}$. The efficient score function can be calculated in the same fashion, that is, through projecting the score vector with respect to the parameters in $\boldsymbol{\xi}$ and $\boldsymbol{\Sigma}^{-1/2}$ onto $\Gamma_\pi^\perp$, although the computation becomes much more tedious due to the quick increase of the number of parameters in $\boldsymbol{\Sigma}^{-1/2}$ as the dimension increases.

For certain elliptical distributions, the pdf g may also involve a degree of freedom ν, which is a parameter of interest to us when, for example, g is the pdf of a t-distribution with ν degrees of freedom. In such a case we calculate the score vector with respect to ν and project it onto $\Gamma_\pi^\perp$ to derive the locally efficient estimator.

Finally, it is worth noting that the only property of the central part of the model g that is essential to the procedure is its symmetry, that is, $g(-\mathbf{x}) = g(\mathbf{x})$. Hence the procedure can be applied to the more general skew-symmetric distributions defined by Wang, Boyer, and Genton (2004).

[Received November 2003. Revised December 2004.]

REFERENCES

Arnold, B. C., and Beaver, R. J. (2002), "Skewed Multivariate Models Related to Hidden Truncation and/or Selective Reporting," *Test*, 11, 7–54.

Azzalini, A., and Dalla Valle, A. (1996), "The Multivariate Skew-Normal Distribution," *Biometrika*, 83, 715–726.

Bayarri, M. J., and DeGroot, M. (1992), "A BAD View of Weighted Distributions and Selection Models," in *Bayesian Statistic 4*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Oxford University Press, pp. 17–33.

Bickel, P., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993), *Efficient and Adaptive Inference in Semiparametric Models*, Baltimore: Johns Hopkins University Press.

Copas, J. B., and Li, H. G. (1997), "Inference From Non-Random Samples" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 59, 55–95.

Foutz, R. V. (1977), "On the Unique Solution to the Likelihood Equations," *Journal of the American Statistical Association*, 72, 147–148.

Genton, M. G., and Loperfido, N. (2005), "Generalized Skew-Elliptical Distributions and Their Quadratic Forms," *Annals of the Institute of Statistical Mathematics*, in press.

Lee, J., and Berger, J. O. (2001), "Semiparametric Bayesian Analysis of Selection Models," *Journal of the American Statistical Association*, 96, 1397–1409.

Ma, Y., and Genton, M. G. (2004), "A Flexible Class of Skew-Symmetric Distributions," *Scandinavian Journal of Statistics*, 31, 459–468.

Newey, W. K. (1990), "Semiparametric Efficiency Bounds," *Journal of Applied Economics*, 5, 99–135.

Rao, C. R. (1985), "Weighted Distributions Arising Out of Methods of Ascertainment: What Populations Does a Sample Represent?" in *A Celebration of Statistics: The ISI Centenary Volume*, eds. A. G. Atkinson and S. E. Fienberg, New York: Springer-Verlag, pp. 543–569.

Wang, J., Boyer, J., and Genton, M. G. (2004), "A Skew-Symmetric Representation of Multivariate Distributions," *Statistica Sinica*, 14, 1259–1270.