Locally Efficient Semiparametric Estimators for Generalized Skew-Elliptical Distributions

Yanyuan Ma* Marc G. Genton† Anastasios A. Tsiatis ‡

April 20, 2004

Abstract

We consider a class of generalized skew-normal distributions which 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. These estimators are shown to be 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 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. The method is applicable to generalized skew-elliptical distributions.

Key words: Generalized skew-elliptical distributions; Influence function; Nuisance tangent space; Selection Models; Semiparametric efficiency.

1 Introduction

Consider the model where a p-dimensional random vector X is distributed with density $g(x;\beta)$, where β is a q-dimensional vector of unknown parameters. In order to make inference

*SAMSI & Center for Research in Scientific Computation, North Carolina State University, Box 8205, Raleigh, NC27695-8205, USA, E-mail: yma@unity.ncsu.edu.

†Department of Statistics, North Carolina State University, Box 8203, Raleigh, NC 27695-8203, USA. E-mail: genton@stat.ncsu.edu

‡Department of Statistics, North Carolina State University, Box 8203, Raleigh, NC 27695-8203, USA. E-mail: tsiatis@stat.ncsu.edu

about β, the usual statistical analysis assumes that a random sample $X_1, \ldots, X_n$ from $g(x; \beta)$ 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 $w(x; \beta, \alpha)$, where α denotes some r-dimensional vector of additional unknown parameters, then the observed data is a random sample from a distribution with density

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

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. For example, this can happen if the observation vector X of characteristics of a certain population is measured only for individuals who manifest a certain disease due to cost or ethical reasons, see the survey article by Bayarri and DeGroot (1992) and references therein. For such problems, the goal is to find consistent and asymptotically normal estimators of β in the presence of the nuisance weight function w.

A slightly different point of view is given by a robustness argument. Effectively, if $g(x; \beta)$ 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 β, that is again to provide consistent and asymptotically normal estimators of β in the presence of a certain class of the nuisance weight function w.

The paper 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 *** and *** (2002). 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*

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

where g is the probability density function of a spherical distribution, ξ is the location parameter, $\Sigma^{-1/2}$ is the Cholesky decomposition of the positive definite scale matrix Σ, i.e. $(\Sigma^{-1/2})^T\Sigma^{-1/2} = \Sigma$, and the function $\pi : \mathbb{R}^p \to [0 \ 1]$ satisfies $\pi(x) + \pi(-x) = 1$ 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 ξ and the scale matrix Σ are not, in general, the expected value and the covariance matrix for f, since GSE distributions may not be symmetric with respect to ξ, but they are for g. In particular, if $g = \phi_p$, the probability density function of the standard p-dimensional multivariate normal distribution, and we choose a parametric model $\pi(x) = \Phi(\alpha^T x)$ for the skewing function, where Φ is the univariate standard normal cumulative distribution function, 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 non-random (biased) samples (Copas and Li, 1997) and are therefore selection models of the form in Equation (1). Representation of a GSE distribution as a selection model is straightforward with $g(x;\beta) = |\Sigma|^{-1/2}g\{\Sigma^{-1/2}(x-\xi)\}$, $w(x;\beta,\alpha) = \pi\{\Sigma^{-1/2}(x-\xi)\}$, $\mathrm{E}\{w(X;\beta,\alpha)\} = 1/2$, $\beta = \{\xi^T, \mathrm{vec}(\Sigma)^T\}^T$, and α is 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 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 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, i.e., 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, 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 above mentioned skew-normal distribution. One example of this specific setting is the distribution of height and weight. Assume 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 the samples obtained are the ones whose weight is larger than the expected weight given their height. This corresponds to a selection criterion $\tilde{x} > E(\tilde{X}|\tilde{y})$, 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 paper, we analyze a data set of body mass index in a group of athletes, which is assumed to be larger than its expected value conditional on an individual's other body characteristics, for example, height, weight, body fat percentage, etc. in a general population, for male and for female. If we assume the BMI in a general population of the same gender follows a normal distribution, without specifying a precise selection method, then the observed data follows 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 we have two random variables X and Y, with the symmetric pdfs $g(x)$ and $h(y)$ respectively. 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 comparing to a general selection model of the form in (1), the GSE models are restricted by the constraint $\pi(x) + \pi(-x) = 1$, the above 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 ξ and Σ 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 ξ and Σ, which we summarize as β, and the nuisance parameter is π. In such setting, regular asymptotically linear (RAL) estimators are studied in Newey (1990). An RAL estimator $\hat{\beta}$ satisfies

$$\sqrt{n}(\hat{\beta} - \beta_0) = \sum_{i=1}^{n} \psi(X_i, \beta_0) + o_p(1),$$

where β represents the finite dimensional parameter of interest, with its true value β_0, $\psi(X_i, \beta_0)$ is the i-th influence function of the estimator, which satisfies $\mathrm{E}(\psi) = 0$, and $\mathrm{E}(\psi\psi^T)$ is finite and nonsingular. In addition, an RAL estimator also satisfies regularity conditions, which would exclude estimators that are "superefficient" for some true parameter β_0, see Newey (1990) for detail. Because of the link between an RAL estimator and its influence function, an RAL estimator can be constructed through finding its influence function, namely, $\hat{\beta}$ is given as the solution that solves $\sum_{i=1}^{n} \psi(X_i, \beta) = 0$. Due to such link, it is also clear that the variance of the estimator $\hat{\beta}$ is given by the variance $\mathrm{E}(\psi\psi^T)$. A geometrical point of view is taken in Bickel et al. (1993) where they characterize the set of all influence functions. Consider a Hilbert space $\mathcal{H}$ consisting of all the q-dimensional mean zero random functions, with the inner product defined as the covariance between two functions, where q is the dimension of β. 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 the expectations are taken with respect to the true density $p(X; \beta_0, \pi_0)$. 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 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 $p(X;\beta,\alpha)$ is defined as a linear space spanned by the nuisance score vector, i.e. all the functions of the form BS_α, where B is an $q \times r$ matrix, with r being the dimension of α, and $S_\alpha = \frac{\partial \log p(X;\beta_0,\alpha)}{\partial \alpha}|_{\alpha_0}$ is the nuisance score vector, where $p(X;\beta_0,\alpha_0)$ 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. Any function in the nuisance tangent space orthogonal complement, after being properly normalized, yields an influence function; on the contrary, any influence function can be obtained through properly normalizing 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 S_β must equal to the identity, where $S_\beta = \frac{\partial \log p(X;\beta,\pi_0)}{\partial \beta}|_{\beta_0}$.

We will use the tools mentioned above to derive RAL estimators and the efficient RAL estimator for the GSE distributions. To remain specific and focused, in this section, all our results are developed in the special case where $g = \phi$, the univariate standard normal probability density function, in which case we use the generalized skew-normal (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 we use can be extended in a straightforward manner to more general cases, see the discussion in Section 6. In the sequel, β 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 cumulative distribution function 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 cumulative distribution function. Throughout the text, parameters or functions with index 0 refer to the true values of the parameters or the true functions. Since we are considering a two dimensional parameter of interest, the Hilbert space $\mathcal{H}$ 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* $\Gamma_\pi = \{u\{(x-\xi_0)/\sigma_0\} : \mathbb{R} \to \mathbb{R}^2$ *: each component of* $\pi_0(x)u(x)$ *is an odd function.*$\}$.

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

$$\frac{\partial \log f(x;\beta,\alpha)}{\partial\alpha}\Big|_{(\beta_0,\alpha_0)} = \frac{\partial\pi\{(x-\xi_0)/\sigma_0,\alpha\}}{\partial\alpha}\Big|_{\alpha_0} \Big/ \pi_0\left(\frac{x-\xi_0}{\sigma_0}\right).$$

Since $\pi\{(x-\xi_0)/\sigma_0,\alpha\} + \pi\{(x+\xi_0)/\sigma_0,\alpha\} = 1$, $\partial\pi\{(x-\xi_0)/\sigma_0,\alpha\}/\partial\alpha + \partial\pi\{(-x+\xi_0)/\sigma_0,\alpha\}/\partial\alpha = 0$, i.e.

$$\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right)\frac{\partial \log f(x;\beta,\alpha)}{\partial\alpha}\Big|_{(\beta_0,\alpha_0)}$$

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

$$B\frac{\partial \log f(x;\beta,\alpha)}{\partial\alpha}\Big|_{(\beta_0,\alpha_0)}$$

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

$$u(x) = B_1\frac{\partial f_1(x;\beta,\alpha_1)}{\partial\alpha_1} + B_2\frac{\partial f_2(x;\beta,\alpha_2)}{\partial\alpha_2},$$

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

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

is a parametric submodel where $\alpha=0$ yields the true model. Notice that

$$\begin{aligned}
&\frac{\partial \log f(x;\beta,\alpha)}{\partial\alpha}\Big|_{(\beta_0,\alpha_0)} \\
=\ & m_0\left(\frac{x-\xi_0}{\sigma_0}\right)e^{\alpha^T h\left(\frac{x-\xi_0}{\sigma_0}\right)} h\left(\frac{x-\xi_0}{\sigma_0}\right)\phi\left\{m_0\left(\frac{x-\xi_0}{\sigma_0}\right)e^{\alpha^T h\left(\frac{x-\xi_0}{\sigma_0}\right)}\right\}\Big/\pi_0\left(\frac{x-\xi_0}{\sigma_0}\right)\Big|_{\alpha=0} \\
=\ & u\left(\frac{x-\xi_0}{\sigma_0}\right).
\end{aligned}$$

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

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

It can be easily verified that

$$\frac{\partial \log f(x;\beta,\alpha)}{\partial \alpha}\Big|_{(\beta_0,\alpha_0)} = u\left(\frac{x-\xi_0}{\sigma_0}\right),$$

hence $u\{(x-\xi_0)/\sigma_0\} \in \Gamma_\pi$, i.e. $u\{(x-\xi_0)/\sigma_0\}$ is really an element in the nuisance tangent space. □

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

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

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

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

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

Because $2u(x)\phi(x)\pi_0(x)/\sigma_0$ is an arbitrary odd function, $v\{(x-\xi_0)/\sigma_0\}$ has to be an even function of $(x-\xi_0)/\sigma_0$ to ensure Equation (4). Notice that $\pi_0(x) - 1/2$ is in fact an odd function, so we get $\int v(x)\phi(x)d\mu(x) = 0$ from Equation (5). Likewise, for any even function $v(x)$, where $\int v(x)\phi(x)d\mu(x) = 0$, we can verify that Equations (4) and (5) are satisfied, hence $v\{(x-\xi_0)/\sigma_0\} \in \Gamma_\pi^\perp$. □

Since 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* $v(x) : \mathbb{R} \to \mathbb{R}^2$ *s.t.* $\int v(x)\phi(x)d\mu(x) = 0$, $\sum_{i=1}^n v\{(X_i-\xi)/\sigma\} = 0$ *defines a regular asymptotically linear (RAL) estimator for* $\beta = (\xi,\sigma)^T$.

Proposition 3 provides us a way of constructing RAL estimators as long as we can find a suitable function $v = (v_1, v_2)^T$. 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 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 $i = 1, 2$.

As we mentioned, 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 which has influence function proportional to the efficient score. The efficient score S_{eff} is the residual after projecting the score vector with respect to β onto the nuisance tangent space, i.e, $S_{eff} = S_\beta - \Pi(S_\beta|\Gamma_\pi)$. The corresponding influence function is given by $\psi_{eff} = \text{cov}(S_\beta, S_{eff})^{-1} S_{eff} = \text{var}(S_{eff})^{-1} S_{eff}$, whose variance $\text{var}(S_{eff})^{-1}$ is smallest among all the influence functions. Here by smallest we mean the difference $\text{var}(\psi) - \text{var}(\psi_{eff})$ is nonnegative definite for any influence function ψ. We derive S_{eff} and calculate the optimal variance in the following propositions.

Proposition 4 *The efficient score function is*

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

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

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

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

We calculate the projection of S_β onto $\Gamma_\pi^\perp$ through using the fact that the difference between S_β and its projection onto $\Gamma_\pi^\perp$ is an element in Γ_π. Assume 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

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

and

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

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* $\beta = (\xi, \sigma)^T$ *is given by*

$$\sum_{i=1}^{n} F_0(X_i; \xi, \sigma) = 0 \tag{6}$$

where

$$F_0(X_i; \xi, \sigma) = \left(\left[\frac{X_i - \xi}{\sigma} \left\{ 2\pi_0 \left(\frac{X_i - \xi}{\sigma} \right) - 1 \right\} - 2\pi_{01} \left(\frac{X_i - \xi}{\sigma} \right) \right], \{(X_i - \xi)^2 - \sigma^2\} \right)^T.$$

Assume the solution to Equation (6) is $\hat{\beta}$, *then* $n^{1/2}(\hat{\beta} - \beta_0) \rightarrow N_2(0, \{E(S_{eff} S_{eff}^T)\}^{-1})$ *in distribution. Here, the smallest variance of the estimate given by* $\{E(S_{eff} S_{eff}^T)\}^{-1}$ *has the form*

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

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) have 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 non-trivial skewing function, a semiparametric efficient estimator always exists.*

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

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

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

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

Proposition 6 *Assume $\frac{2}{\sigma}\phi\{(x-\xi)/\sigma\}\pi\{(x-\xi)/\sigma, \alpha\}$ is a parametric model and $n^{1/2}(\hat{\alpha} - \alpha^*)$ is bounded in probability. Then the two RAL estimators, resulting from solving the two estimating equations $\sum_{i=1}^n F(X_i; \xi, \sigma, \alpha^*) = 0$ and $\sum_{i=1}^n F(X_i; \xi, \sigma, \hat{\alpha}) = 0$, are asymptotically equivalent, i.e. if $(\hat{\xi}_1, \hat{\sigma}_1)$ is the solution to the first equation, and $(\hat{\xi}_2, \hat{\sigma}_2)$ is the solution to 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 β, $F(X_i; \xi, \sigma, \alpha)$ as $F(X_i; \beta, \alpha)$. A Taylor expansion of $\sum_{i=1}^n F(X_i; \hat{\beta}_2, \alpha^*)$ at $\hat{\alpha}$ yields $\sum_{i=1}^n F(X_i; \hat{\beta}_2, \alpha^*) = \sum_{i=1}^n F(X_i; \hat{\beta}_2, \hat{\alpha}) + \{\sum_{i=1}^n \partial F(X_i; \hat{\beta}_2, \tilde{\alpha})/\partial\alpha^T\}(\alpha^* - \hat{\alpha})$, where $\tilde{\alpha}$ is between α^* and $\hat{\alpha}$. Denoting $\{\sum_{i=1}^n \partial F(X_i; \hat{\beta}_2, \tilde{\alpha})/\partial\alpha^T\}/n$ by Λ_n, we obtain

$\sum_{i=1}^n F(X_i;\hat\beta_2,\alpha^*) = n\Lambda_n(\alpha^* - \hat\alpha)$. Notice that when $n\to\infty$, because of the convergence of $\hat\alpha$ to α^* and the consistency property of ξ_2 and σ_2, $\Lambda_n \to \mathrm{E}\{\partial F(X_i;\beta_0,\alpha^*)/\partial\alpha^T\} = 0$ in probability.

A Taylor expansion of $\sum_{i=1}^n F(X_i;\hat\beta_2,\alpha^*)$ at $\hat\beta_1$ yields

$$\begin{aligned}\hat\beta_2 - \hat\beta_1 &= \left\{\sum_{i=1}^n \frac{\partial F(X_i;\tilde\beta,\alpha^*)}{\partial\beta^T}\right\}^{-1}\left\{\sum_{i=1}^n F(X_i;\hat\beta_2,\alpha^*) - 0\right\}\\ &= \left\{\frac{1}{n}\sum_{i=1}^n \frac{\partial F(X_i;\tilde\beta,\alpha^*)}{\partial\beta^T}\right\}^{-1}\Lambda_n\left(\alpha^* - \hat\alpha\right),\end{aligned}$$

where $\tilde\beta$ is a quantity between $\hat\beta_1$ and $\hat\beta_2$.

When $n\to\infty$,

$$J_n = \frac{1}{n}\sum_{i=1}^n \frac{\partial F(X_i;\tilde\beta,\alpha^*)}{\partial\beta^T} \to \mathrm{E}\left\{\frac{\partial F(X_i;\beta_0,\alpha^*)}{\partial\beta^T}\right\}$$

in probability. For parametric models, $\mathrm{E}\{\partial F(X_i;\beta_0,\alpha^*)/\partial\beta^T\}$ is the matrix related to the Fisher information matrix, which is generally nonsingular and we denote it by J. Combining the results, we have $n^{1/2}(\hat\beta_1 - \hat\beta_2) = n^{1/2}J_n^{-1}\Lambda_n(\hat\alpha - \alpha^*)$. Because $n^{1/2}(\hat\alpha-\alpha^*)$ is bounded in probability, $J_n^{-1}\to J^{-1}$ in probability and $\Lambda_n \to 0$ in probability, which implies that $\hat\beta_1 - \hat\beta_2 \to 0$ in probability. □

Proposition 6 indicates that how efficient we estimate the nuisance parameter α does not influence how efficient we can estimate ξ and σ. In fact, as long as we can estimate α consistently, using the estimated value of α, $\hat\alpha$, or the true value of α, α_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 $\{\pi(x,\alpha)\}$. One way proposed by *** and *** (2004a) to construct the parametric model 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, hence make the "distance" between $\Phi\{P_K(x)\}$ and π_0 arbitrarily small. In general, the relation between the efficiency loss and the "distance" between π_0 and the parametric family $\{\pi(x,\alpha)\}$ that approximates π_0 is given in the following proposition.

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

$$B = \frac{\sigma_0^2}{[E\{2\pi_0(X)-1+2X\pi_{01}(X)-2\pi_{02}(X)\}-2E(X)^2]^2}\left\{\begin{array}{cc} 1 & -E(X) \\ -E(X) & E(X)^2 \end{array}\right\},$$

which does not depend on the estimator in Equation (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 the estimating equation $\sum_{i=1}^n F(X_i;\beta,\hat\alpha) = 0$ yields the estimate $\hat\beta_1 = (\hat\xi_1, \hat\sigma_1)^T$, the estimating equation $\sum_{i=1}^n F_0(X_i;\beta) = 0$ yields the estimate $\hat\beta = (\hat\xi, \hat\sigma)^T$. Then

$$\begin{aligned} 0 &= \sum_{i=1}^n F(X_i;\hat\beta_1,\hat\alpha) \\ &= \sum_{i=1}^n F_0(X_i;\hat\beta) + \sum_{i=1}^n \{F_0(X_i;\hat\beta_1) - F_0(X_i;\hat\beta)\} + \sum_{i=1}^n \{F(X_i;\hat\beta_1,\hat\alpha) - F_0(X_i;\hat\beta_1)\} \\ &= \sum_{i=1}^n (\hat\beta_1 - \hat\beta)\frac{\partial F_0(X_i;\tilde\beta)}{\partial\beta^T} + \sum_{i=1}^n \left\{\frac{X_i-\hat\xi_1}{\hat\sigma_1}2\nu\left(\frac{X_i-\hat\xi_1}{\hat\sigma_1}\right) - 2\nu_1\left(\frac{X_i-\hat\xi_1}{\hat\sigma_1}\right), 0\right\}, \end{aligned}$$

where $\tilde\beta$ is a quantity between $\hat\beta$ and $\hat\beta_1$, $\nu_1(x) = d\nu(x)/dx$. Notice that when $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^n \frac{\partial F_0(X_i;\tilde\beta)}{\partial\beta^T} \to \mathrm{E}\left\{\frac{\partial F_0(X_i;\beta_0)}{\partial\beta^T}\right\}$$

in probability,

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^n \left\{\frac{X_i-\hat\xi_1}{\hat\sigma_1}2\nu\left(\frac{X_i-\hat\xi_1}{\hat\sigma_1}\right) - 2\nu_1\left(\frac{X_i-\hat\xi_1}{\hat\sigma_1}\right)\right\} \to \mathrm{E}\left\{2\frac{X_i-\hat\xi_1}{\hat\sigma_1}\nu\left(\frac{X_i-\hat\xi_1}{\hat\sigma_1}\right) - 2\nu_1\left(\frac{X_i-\hat\xi_1}{\hat\sigma_1}\right)\right\} \\ &\to \mathrm{E}\left\{2\frac{X_i-\xi_0}{\sigma_0}\nu\left(\frac{X_i-\xi_0}{\sigma_0}\right) - 2\nu_1\left(\frac{X_i-\xi_0}{\sigma_0}\right)\right\} = 0 \end{aligned}$$

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 $(X_i - \xi_0)/\sigma_0$.

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

Thus,

$$n^{1/2}(\hat{\beta}_1 - \hat{\beta}) \to N_2 \left(0, \left[\mathrm{E}\left\{\frac{\partial F_0(X_i;\beta_0)}{\partial \beta^T}\right\}\right]^{-1} \begin{pmatrix} \theta & 0 \\ 0 & 0 \end{pmatrix} \left[\mathrm{E}\left\{\frac{\partial F_0(X_i;\beta_0)}{\partial \beta^T}\right\}\right]^{-T}\right)$$

in distribution. It can be verified that

$$\mathrm{E}\left\{\frac{\partial F_0(X_i;\beta_0)}{\partial \beta^T}\right\} = -\begin{bmatrix} \frac{\mathrm{E}\{2\pi_0(X)-1+2X\pi_{01}(X)-2\pi_{02}(X)\}}{\sigma_0} & \frac{2\mathrm{E}(X)}{\sigma_0} \\ 2\sigma_0\mathrm{E}(X) & 2\sigma_0 \end{bmatrix},$$

where expectation E on the right side is taken with respect to $2\phi(x)\pi_0(x)$. Putting these together, we get $n^{1/2}(\hat{\beta}_1-\hat{\beta}) \to N_2(0,\theta B)$ in distribution, thus $n^{1/2}(\hat{\beta}_1-\beta_0) \to N_2(0,A+\theta B)$ in distribution. With an α that minimizes θ, we will get the most efficient estimator given the parametric model $\pi(x,\alpha)$. The variance of $n^{1/2}(\hat{\beta}_1 - \beta_0)$ is $A + \min_\alpha(\theta)B$. □

In Proposition 7, we deliberately avoided specifying how to find the α that minimizes θ, since this depends on the true π_0 that is unknown to us. In practice, we can always estimate β and calculate its variance for any fixed α and select the α that yields the smallest estimation variance. Thus, the α that minimizes θ can be found numerically. Often, a parametric model is assumed in terms of both β and α and MLE is used to estimate both sets of parameters. However, if the model for the skewing function is not correct then the MLE for β will be biased. A correction procedure should follow, where after obtaining the MLE estimator $\hat{\alpha}$ 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 $\hat{\alpha}$ obtained through MLE needs not be the α that minimize θ. However, the resulting estimator will be consistent and asymptotically normal even if the model for π_0 was incorrectly specified and will be semiparametric efficient if it is correctly specified.

4 Simulation Results

We carry out a simulation study with a sample size of 500. The data sets are generated 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 approximate 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

Table 1: Simulation results on ξ and σ with different posited skewing function $\pi_1(x)$, $\pi_3(x)$ and $\pi_t(x)$. The true values of ξ and σ are 3 and 1. The sample size is 500, and 1,000 data sets are simulated.

	$\hat{\xi}(3)$			$\hat{\sigma}(1)$		
	mean	est. var.	emp. var.	mean	est. var.	emp. var.
	$\hat{\alpha}$ estimated through MLE					
$\pi_1(x)$	2.9899	0.0042	0.0040	1.0007	0.0013	0.0011
$\pi_3(x)$	3.0006	0.0029	0.0027	1.0018	0.0014	0.0011
$\pi_t(x)$	2.9977	0.0017	0.0017	1.0005	0.0011	0.0011
	$\hat{\alpha}$ minimizes the resulting variance					
$\pi_1(x)$	2.9895	0.0035	0.0036	1.0004	0.0012	0.0011
$\pi_3(x)$	2.9988	0.0018	0.0020	1.0009	0.0010	0.0011
$\pi_t(x)$	2.9976	0.0017	0.0018	1.0005	0.0011	0.0011

order K. We generate 1,000 data sets, and calculate the empirical variances of the estimates and also the average of the estimated variances. The estimated variance is calculated via the standard sandwich matrix of M-estimators. That is, we calculate

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

as the estimated variance matrix, where $F(X_i;\xi,\sigma,\hat{\alpha})$ is the same as in Equation (8) and $DF(X_i;\xi,\sigma,\hat{\alpha})$ is the Jacobian of $F(X_i;\xi,\sigma,\hat{\alpha})$ with respect to ξ and σ, $\hat{\alpha}$ is the MLE of the polynomial coefficients α 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 shown in the upper half of Table 1. For comparison, we also adopted the correct model for $\pi_0(x)$, i.e., 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, while the estimator with the true posited model for $\pi_0(x)$ has the smallest variance. The variance for

$\pi_3(x)$ is smaller than for $\pi_1(x)$ because $\pi_3(x)$ approximates $\pi_0(x)$ better. In fact, as shown in *** and *** (2004a), $\pi_0(x)$ can be approximated arbitrarily well if we allow the order of the odd polynomial to increase sufficiently, hence the estimator will approach the most efficient one.

We also estimated ξ and σ using the $\hat{\alpha}$ that minimizes the "distance" θ between $\pi_0(x)$ and $\pi_K(x)$, i.e., 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 table is improved compared to the corresponding estimators in the upper table where $\hat{\alpha}$ is simply obtained through MLE. In fact, the estimation variance when using $\pi_3(x)$ is so close to that of using the correct model $\pi_t(x)$ that as far as the 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 pdfs $\frac{2}{\sigma}\phi\{(x-\xi)/\sigma\}\pi_1(x), \frac{2}{\sigma}\phi\{(x-\xi)/\sigma\}\pi_3(x)$, 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 pdfs. 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.

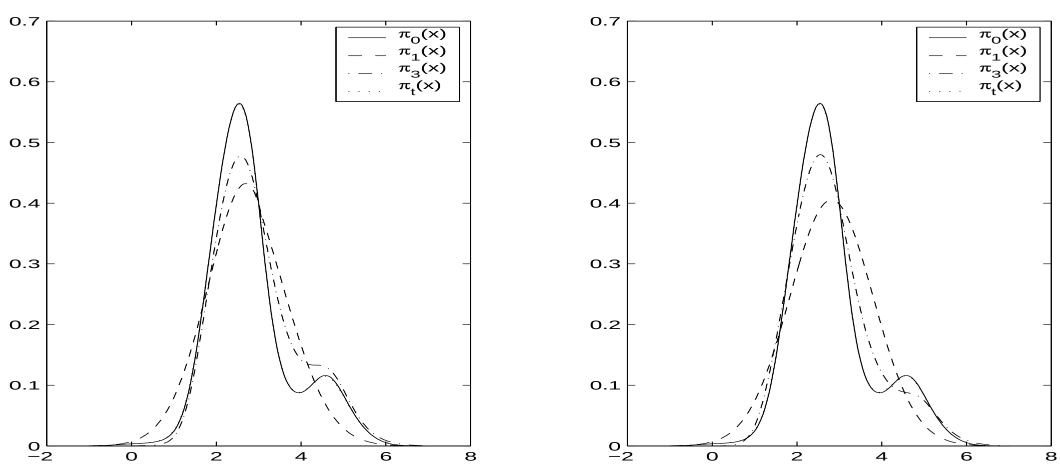

Figure 1: The average estimated pdfs using the posited model π_1, π_3 and π_t respectively. The true pdf using π_0 and the average estimated pdf using π_t overlay each other and are indistinguishable in the picture. The nuisance parameters in the posited models are selected to minimize the resulting variance (left panel), and are estimated using MLE (right panel).

5 An Example

We applied the estimator in Equation (8) on a data set about Australian athletes' body mass index data. In this data set, the BMI of 202 athletes are measured, including 102 male and 100 female athletes. The histograms of the data for men and women are shown in Figure 2. Assume we try to infer the mean and variance of the body mass index in general Australian adults male and female. Certainly, a simple sample average and variance will give a biased estimate, since athletes would certainly have a higher BMI than the population average. We use 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 through minimizing the final total variance of ξ and σ. The results are presented in Table 2. As we expected, the average BMI in general population is lower than in athletes, and the variance σ^2 is also larger than among athletes. The pdfs corresponding to different skewing function π_K's are plotted in

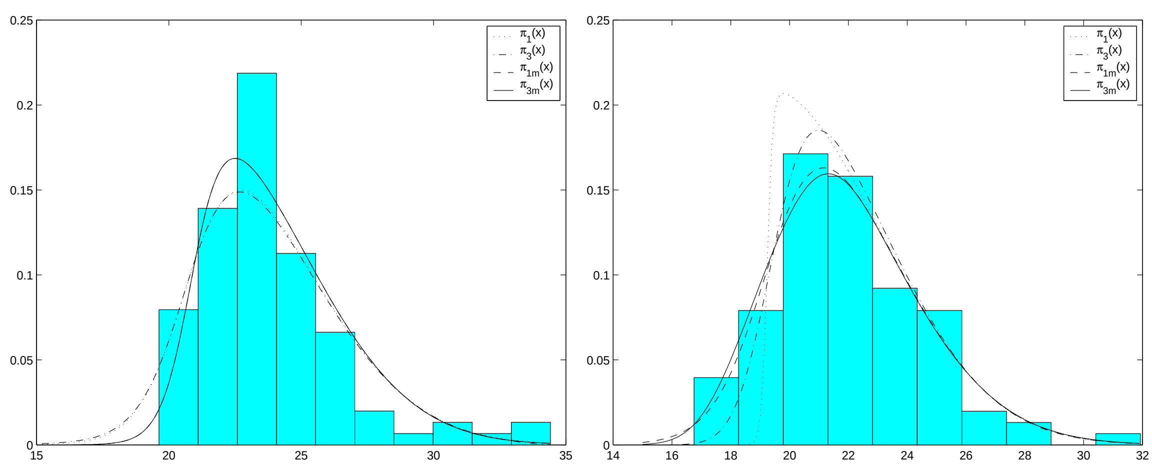

Figure 2: The histograms of 102 male (left panel) and 100 female (right panel) Australian athletes' body mass index and the pdfs utilizing different posited skewing functions. The skewing functions are π_1, π_3, whose nuisance parameters are estimated through minimizing estimation variance, and π_{1m}, π_{3m}, whose nuisance parameters are estimated through MLE. On the left panel, the two pdfs using π_{1m} and π_{3m} overlay each other and are indistinguishable in the picture.

Table 2: Estimated value of ξ and σ and their standard deviation via different estimators. The last estimator is obtained by taking the sample mean and sample standard deviation of the data.

	male				female			
skew fun	$\hat{\xi}$	est. sd.	$\hat{\sigma}$	est. sd.	$\hat{\xi}$	est. sd.	$\hat{\sigma}$	est. sd.
$\pi_{1m}(x)$	20.8542	0.4908	4.1089	0.5855	19.2911	0.5159	3.7656	0.5171
$\pi_{3m}(x)$	20.6958	0.4276	4.2277	0.5801	19.3483	0.5036	3.7249	0.5061
$\pi_1(x)$	20.8542	0.4908	4.1089	0.5855	19.2225	0.5674	3.8151	0.5554
$\pi_3(x)$	20.6807	0.3173	4.2392	0.3220	18.9751	0.4778	3.9981	0.4696
	23.9036	0.2727	2.7539	0.3044	21.9892	0.2627	2.6268	0.2341

Figure 2. Because our estimators are semiparametric, we do not have to necessarily have a good estimate of the skewing function in order to have a consistent 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 there was no appreciable decrease in the estimated standard error beyond that.

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, while its orthogonal complement becomes $\Gamma_\pi^\perp = \{v\{(x-\xi_0)/\sigma_0\} : v(x)$ is an even function that satisfies $\int v(x)g(x)d\mu(x) = 0\}$, where g is the elliptical part of the GSE distribution. As a result, for such $v(x)$, $\sum_{i=1}^n v\{(X_i-\xi)/\sigma\} = 0$ forms an RAL estimator. Similarly, the efficient score function in general is

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

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, while its orthogonal complement becomes $\Gamma_\pi^\perp = \{v\{\Sigma_0^{-1/2}(x-\xi_0)\} : v(x)$ is an even function that satisfies $\int v(x)g(x)d\mu(x) = 0\}$. The efficient score function can be calculated in the same fashion, i.e. through projecting the score vector with respect to the parameters in ξ and $\Sigma^{-1/2}$ onto $\Gamma_\pi^\perp$, although the computation becomes much more tedious due to the quick increase of the number of parameters in $\Sigma^{-1/2}$ 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, for example when g is the pdf of a t-distribution with ν degrees of freedom. In such 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, i.e. $g(-x) = g(x)$. Hence the procedure can be applied to the more general skew-symmetric distributions defined in *** et al. (2004b).

7 Reference

Arnold, B. C. & Beaver, R. J. (2002). Skewed multivariate models related to hidden truncation and/or selective reporting. *Test* **11**, 7-54.

Azzalini, A. & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715-726.

Bayarri, M. J. & DeGroot, M. (1992). A BAD view of of weighted distributions and selection models. In *Bayesian Statistic 4*. Ed. J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, pp. 17-33. Oxford: University Press.

Bickel, P., Klaassen, C. A. J., Ritov, Y. & Wellner, J. A. (1993). *Efficient and adaptive inference in semiparametric models.* Baltimore: Johns Hopkins University Press.

Copas, J. B. & Li, H. G. (1997). Inference from non-random samples (with discussion). *J. R. Statist. Soc. B* **59**, 55-95.

Lee, J. & Berger, J. O. (2001). Semiparametric Bayesian Analysis of Selection Models. *J. Am. Statist. Assoc.* **96**, 1397-1409.

Newey, W. K. (1990). Semiparametric Efficiency Bounds. *J. Appl. Econom.* **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.* Ed. A. G. Atkinson & S. E. Fienberg, pp. 543-569. New York:Springer-Verlag.

***. & *** (2002). Generalized skew-elliptical distributions and their quadratic forms.

*** & *** (2004a). A flexible class of skew-symmetric distributions.

***, *** & *** (2004b) A skew-symmetric representation of multivariate distributions.