The Annals of Statistics
2013, Vol. 41, No. 1, 250–268
DOI: 10.1214/12-AOS1072

EFFICIENT ESTIMATION IN SUFFICIENT DIMENSION REDUCTION

BY YANYUAN MA[1] AND LIPING ZHU[2]

Texas A&M University and Shanghai University of Finance and Economics

We develop an efficient estimation procedure for identifying and estimating the central subspace. Using a new way of parameterization, we convert the problem of identifying the central subspace to the problem of estimating a finite dimensional parameter in a semiparametric model. This conversion allows us to derive an efficient estimator which reaches the optimal semiparametric efficiency bound. The resulting efficient estimator can exhaustively estimate the central subspace without imposing any distributional assumptions. Our proposed efficient estimation also provides a possibility for making inference of parameters that uniquely identify the central subspace. We conduct simulation studies and a real data analysis to demonstrate the finite sample performance in comparison with several existing methods.

1. Introduction. Consider a general model in which the univariate response variable Y is assumed to depend on the p-dimensional covariate vector $\mathbf{x}$ only through a small number of linear combinations $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$, where $\boldsymbol{\beta}$ is a $p \times d$ matrix with $d < p$. In this model, how Y depends on $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$ is left unspecified. It is not difficult to see that $\boldsymbol{\beta}$ is not identifiable. The quantity of general interest is usually the column space of $\boldsymbol{\beta}$, which is termed the central subspace if d is the smallest possible value to satisfy the model assumption [5].

This general model was proposed by Li [12] and has attracted much attention in the last two decades. It generated the field of sufficient dimension reduction [5], in which the main interest is to estimate the central subspace consistently. Influential works in this area include, but are not limited to, sliced inverse regression [12], sliced average variance estimation [6], directional regression [10], the generalization of the aforementioned methods to nonelliptically distributed predictors [7, 9], Fourier transformation [30], cumulative slicing estimators [29] and conditional density based minimum average variance estimation [26], etc.

Received November 2011; revised August 2012.

[1]Supported by NSF Grants DMS-12-06693 and DMS-10-00354 and the National Institute of Neurological Disorders and Stroke Grant R01-NS073671.

[2]Supported by Natural Science Foundation of China (11071077), Innovation Program of Shanghai Municipal Education Commission (13ZZ055), Pujiang Project of Science and Technology Commission of Shanghai Municipality (12PJ1403200) and Grants for New Century Excellent Talents in University, Ministry of Education (NCET-12-0901). All the correspondence should be directed to Liping Zhu at zhu.liping@mail.shufe.edu.cn.

MSC2010 subject classifications. Primary 62H12, 62J02; secondary 62F12.

Key words and phrases. Central subspace, dimension reduction, estimating equations, semiparametric efficiency, sliced inverse regression.

Despite the various estimation methods, it is unclear if any of these estimators are optimal in the sense that they can exhaustively estimate the entire central subspace and have the minimum possible asymptotic estimation variance. To the best of our knowledge, the efficiency issue has never been discussed in the context of sufficient dimension reduction.

In this paper we study the estimation and inference in sufficient dimension reduction. We propose a simple parameterization so that the central subspace is uniquely identified by a $(p-d)d$-dimensional parameter that is not subject to any constraints. Thus we convert the problem of identifying the central subspace into a problem of estimating a finite dimensional parameter in a semiparametric model. This allows us to derive the estimation procedures and perform inference using semiparametric tools. How to make inference about the central subspace is a challenging issue. This is partially caused by the complexity of estimating a space rather than a parameter. Our new parameterization overcomes this complexity and permits a relatively straightforward calculation of the estimation variability.

We further construct an efficient estimator, which reaches the minimum asymptotic estimation variance bound among all possible consistent estimators. Efficiency bounds are of fundamental importance to the theoretical consideration. Such bounds quantify the minimum efficiency loss that results from generalizing one restrictive model to a more flexible one, and hence they can be important in making the decision of which model to use. The efficiency bounds also provide a gold standard by which the asymptotic efficiency of any particular semiparametric estimator can be measured [22]. Generally speaking, a semiparametric efficient estimator is usually the ultimate destination when searching for consistent estimators or trying to improve existing procedures. When an efficient estimator is obtained, the procedure of estimation can be considered to have reached certain optimality.

In the literature, vast and significant effort has been devoted to studying the semiparametric efficiency bounds for consistent estimators in semiparametric models. The simplest and most familiar examples are the ordinary and weighted least square estimators in the linear regression setting. Efficiency issues are also considered in more complex semiparametric problems such as regressions with missing covariates [23], skewed distribution families [18, 19], measurement error models [15, 25], partially linear models [16], the Cox model [24], page 113, accelerated failure model [27] or other general survival models [28] and latent variable models [17].

One typical semiparametric tool is to obtain estimators through obtaining the corresponding influence functions. In deriving the influence function family and its efficient member, we use the geometric technique illustrated in [2] and [24]. All our derivations are performed without using the linearity or constant variance condition that is often assumed in the dimension reduction literature. Our analysis is thus readily applicable when some covariates are discrete or categorical. In summary, we provide an efficient estimator which can exhaustively estimate the central subspace without imposing any distributional assumptions on the covariate $\mathbf{x}$.

The rest of this paper is organized as follows. In Section 2, we propose a simple parameterization of the central subspace and highlight the semiparametric approach to estimating the central subspace. We also derive the efficient score function. In Section 3, we present a class of locally efficient estimators and identify the efficient member. We illustrate how to implement the efficient estimator to reach the optimal efficiency bound. Simulation studies are conducted in Section 4 to demonstrate the finite sample performance and the method is implemented in a real data example in Section 5. We finish the paper with a brief discussion in Section 6. All the technical derivations are given in a supplementary material [21].

2. The semiparametric formulation.

2.1. *Parameterization of central subspace.* In the context of sufficient dimension reduction [5, 12], one often assumes

$$F(y|\mathbf{x}) = F(y|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) \qquad \text{for } y \in \mathrm{R}, \tag{2.1}$$

where $F(y|\mathbf{x}) \stackrel{\text{def}}{=} \Pr(Y \le y|\mathbf{x})$ is the conditional distribution function of the response Y given the covariates $\mathbf{x}$, and $\boldsymbol{\beta}$ is a $p \times d$ matrix as defined previously. The goal of sufficient dimension reduction is to estimate the column space of $\boldsymbol{\beta}$, which is termed the dimension reduction subspace. Because a dimension reduction subspace is not necessarily unique, the primary interest is usually the central subspace $\mathcal{S}_{Y|\mathbf{x}}$, which is defined as the minimum dimension reduction subspace if it exists and is unique [5]. The dimension of $\mathcal{S}_{Y|\mathbf{x}}$, denoted with d, is commonly referred to as the structural dimension. Similarly to [4], we exclude a pathological case where there exists a vector $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}$ is a deterministic function of $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$ while $\boldsymbol{\alpha}$ does not belong to the column space of $\boldsymbol{\beta}$.

The central subspace $\mathcal{S}_{Y|\mathbf{x}}$ has a well-known invariance property [5], page 106, that is, $\mathcal{S}_{Y|\mathbf{x}} = \mathbf{D}\mathcal{S}_{Y|\mathbf{z}}$, where $\mathbf{z} = \mathbf{D}^{\mathrm{T}}\mathbf{x} + \mathbf{b}$ for any $p \times p$ nonsingular matrix $\mathbf{D}$ and any length p vector $\mathbf{b}$. This allows us to assume throughout that the covariate vector $\mathbf{x}$ satisfies $E(\mathbf{x}) = \mathbf{0}$ and $\operatorname{cov}(\mathbf{x}) = \mathbf{I}_p$. Identifying $\mathcal{S}_{Y|\mathbf{x}}$ is the essential interest of sufficient dimension reduction for model (2.1). Typically, $\mathcal{S}_{Y|\mathbf{x}}$ is identified through estimating a basis matrix $\boldsymbol{\beta} \in \mathrm{R}^{p\times d}$ of minimal dimension that satisfies (2.1). Although $\mathcal{S}_{Y|\mathbf{x}}$ is unique, the basis matrix $\boldsymbol{\beta}$ is clearly not. In fact, for any $d \times d$ full rank matrix $\mathbf{A}$, $\boldsymbol{\beta}\mathbf{A}$ generates the same column space as $\boldsymbol{\beta}$. Thus, to uniquely map one central subspace $\mathcal{S}_{Y|\mathbf{x}}$ to one basis matrix, we need to focus on one representative member of all the $\boldsymbol{\beta}\mathbf{A}$ matrices generated by different $\mathbf{A}$'s. We write $\boldsymbol{\beta} = (\boldsymbol{\beta}_u^{\mathrm{T}}, \boldsymbol{\beta}_l^{\mathrm{T}})^{\mathrm{T}}$, where the upper submatrix $\boldsymbol{\beta}_u$ has size $d \times d$ and the lower submatrix $\boldsymbol{\beta}_l$ has size $(p-d) \times d$. Because $\boldsymbol{\beta}$ has rank d, we can assume without loss of generality that $\boldsymbol{\beta}_u$ is invertible. The advantage of using $\boldsymbol{\beta}\boldsymbol{\beta}_u^{-1}$ is that its upper $d \times d$ submatrix is the identity matrix, while the lower $(p-d) \times d$ matrix can be any matrix. In addition, two matrices $\boldsymbol{\beta}_1\boldsymbol{\beta}_{1u}^{-1}$ and $\boldsymbol{\beta}_2\boldsymbol{\beta}_{2u}^{-1}$ are different if and only if the column spaces of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are different. Therefore, if we consider the set of all the $p \times d$ matrices $\boldsymbol{\beta}$ where the upper $d \times d$ submatrix is the identity matrix $\mathbf{I}_d$, it has a one-to-one mapping with the set of all the different central subspaces. Thus, as long as we restrict our attention to the set of all such matrices, the problem of identifying $\mathcal{S}_{Y|\mathbf{x}}$ is converted to the problem of estimating $\boldsymbol{\beta}_l$, which contains $p_t = (p-d)d$ free parameters. Note that p_t is the dimension of the Grassmann manifold formed by the column spaces of all different $\boldsymbol{\beta}$ matrices. Thus, we can view $\boldsymbol{\beta}_l$ as a unique parameterization of the manifold. Here the subscript "$_t$" stands for total. For notational convenience in the remainder of the text, for an arbitrary $p \times d$ matrix $\boldsymbol{\beta} = (\boldsymbol{\beta}_u^{\mathrm{T}}, \boldsymbol{\beta}_l^{\mathrm{T}})^{\mathrm{T}}$, we define the concatenation of the columns contained in the lower $p-d$ rows of $\boldsymbol{\beta}$ as $\operatorname{vecl}(\boldsymbol{\beta}) = \operatorname{vec}(\boldsymbol{\beta}_l) = (\beta_{d+1,1}, \ldots, \beta_{p,1}, \ldots, \beta_{d+1,d}, \ldots, \beta_{p,d})^{\mathrm{T}}$, where in the notation vecl, "vec" stands for vectorization, and "l" stands for the lower part of the original matrix. We then can write the concatenation of the parameters in $\boldsymbol{\beta}$ as $\operatorname{vecl}(\boldsymbol{\beta})$. Thus, from now on, we only consider basis matrix of $\mathcal{S}_{Y|\mathbf{x}}$ that has the form $\boldsymbol{\beta} = (\mathbf{I}_d, \boldsymbol{\beta}_l^{\mathrm{T}})^{\mathrm{T}}$, where $\boldsymbol{\beta}_l$ is a $(p-d) \times d$ matrix. Estimating the parameters in $\boldsymbol{\beta}$ is a typical semiparametric estimation problem, in which the parameter of interest is $\operatorname{vecl}(\boldsymbol{\beta})$. Therefore we have converted the problem of estimating the central space $\mathcal{S}_{Y|\mathbf{x}}$ into a problem of semiparametric estimation.

REMARK 1. The above parameterization of $\mathcal{S}_{Y|\mathbf{x}}$ excludes the pathological case where one or more of the first d covariates do not contribute to the model or contribute to the model through a fixed linear combination. When this happens, $\boldsymbol{\beta}_u$ will be singular. However, because $\boldsymbol{\beta}$ has rank d, hence if this happens, one can always rotate the order of the covariates (hence rotate the rows of $\boldsymbol{\beta}$) to ensure that after rotation, the resulting $\boldsymbol{\beta}_u$ has full rank.

2.2. *Efficient score.* In this section we derive the efficient score for estimating $\boldsymbol{\beta}$ under the above parameterization. That is, we now consider model (2.1), where $\boldsymbol{\beta} = (\mathbf{I}_d, \boldsymbol{\beta}_l^{\mathrm{T}})^{\mathrm{T}}$ and $\mathbf{x}$ satisfies $E(\mathbf{x}) = \mathbf{0}$ and $\operatorname{var}(\mathbf{x}) = \mathbf{I}_p$. The general semiparametric technique we use is originated from [2] and is wonderfully presented in [24]. Using this approach, we obtain the main result of this section, that we can use (2.2) to obtain an efficient estimation of $\boldsymbol{\beta}$.

The likelihood of one random observation $(\mathbf{x}, Y)$ in (2.1) is $\eta_1(\mathbf{x})\eta_2(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, where η_1 is a probability mass function (p.m.f.) or a probability density function (p.d.f.) of $\mathbf{x}$, or a mixture, depending on whether $\mathbf{x}$ contains discrete variables, and η_2 is the conditional p.m.f./p.d.f. of Y on $\mathbf{x}$. We view η_1, η_2 as infinite dimensional nuisance parameters and $\operatorname{vecl}(\boldsymbol{\beta})$ as the p_t-dimensional parameter of interest. Following the semiparametric analysis procedure, we first derive the nuisance tangent space $\Lambda = \Lambda_1 \oplus \Lambda_2$, where

$$\Lambda_1 = \{\mathbf{f}(\mathbf{x}) : \forall \mathbf{f} \text{ such that } E(\mathbf{f}) = \mathbf{0}\},$$

$$\Lambda_2 = \{\mathbf{f}(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) : \forall \mathbf{f} \text{ such that } E(\mathbf{f}|\mathbf{x}) = E(\mathbf{f}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = \mathbf{0}\}.$$

Here, the notation $\oplus$ means the usual addition of the two spaces Λ_1, Λ_2, while Λ_1 and Λ_2 have the extra property that they are orthogonal to each other. This means the inner product of two arbitrary functions from Λ_1 and Λ_2, respectively, calculated as the covariance between them, is zero. We then obtain its orthogonal complement

$$\Lambda^{\perp} = \{\mathbf{f}(Y, \mathbf{x}) - E(\mathbf{f}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}, Y) : E(\mathbf{f}|\mathbf{x}) = E(\mathbf{f}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}), \forall \mathbf{f}\}.$$

The detailed derivation of Λ and $\Lambda^{\perp}$ is given in Appendix A.2 of [20]. The form of $\Lambda^{\perp}$ permits many possibilities for constructing estimating equations. For example, for arbitrary functions $\mathbf{g}_i$ and $\boldsymbol{\alpha}_i$, the linear combination

$$\sum_{i=1}^{k} \{\mathbf{g}_i(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) - E(\mathbf{g}_i|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}\{\boldsymbol{\alpha}_i(\mathbf{x}) - E(\boldsymbol{\alpha}_i|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}$$

will provide a consistent semiparametric estimator since it is a valid element in $\Lambda^{\perp}$. This form is exploited extensively in [20] to establish links between the semiparametric approach and various inverse regression methods. Among all elements in $\Lambda^{\perp}$, the most interesting one is the efficient score, defined as the orthogonal projection of the score vector $\mathbf{S}_\beta$ onto $\Lambda^{\perp}$. We write the efficient score as $\mathbf{S}_{\mathrm{eff}} = \Pi(\mathbf{S}_\beta|\Lambda^{\perp})$. Because the efficient score can be normalized to the efficient influence function, it enables us to construct an efficient estimator of $\mathrm{vecl}(\boldsymbol{\beta})$ which reaches the optimal semiparametric efficiency bound in the sense of [2]. In the supplementary document [21], we derive the efficient score function to be

$$\mathbf{S}_{\mathrm{eff}}(Y, \mathbf{x}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}, \eta_2) = \mathrm{vecl}\left[\{\mathbf{x} - E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}\frac{\partial \log\{\eta_2(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}}{\partial(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})}\right]. \tag{2.2}$$

Hypothetically, the efficient estimator can be obtained through implementing

$$\sum_{i=1}^{n} \mathbf{S}_{\mathrm{eff}}(Y_i, \mathbf{x}_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i, \eta_2) = \mathbf{0}.$$

However, $\mathbf{S}_{\mathrm{eff}}$ is not readily implementable because it contains the unknown quantities $E(\mathbf{x}|\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})$ and $\partial \log \eta_2(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})/\partial(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})$. For this reason, we first discuss a simpler alternative in the following section.

3. Locally efficient and efficient estimators.

3.1. *Locally efficient estimators*. We now discuss how to construct a locally efficient estimator. This is an estimator that contains some subjectively chosen components. If the components are "well" chosen, the resulting estimator is efficient. Otherwise, it is not efficient, but still consistent. The efficient estimator defined in (2.2) requires one to estimate η_2, the conditional p.d.f. of Y on $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$, and its first derivative with respect to $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$. Although this is feasible, as we will describe in detail in Section 3.2, it certainly is not a trivial task as it involves several nonparametric estimations. Because of this, a compromise is to consider an estimator that depends on a posited model of η_2. Specifically, we would choose some favorite form for η_2, denoted $\eta_2^*(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, and utilize it in place of η_2 to construct an estimating equation. If the posited model is correct (i.e., $\eta_2^* = \eta_2$), then we would have the optimal efficiency using the corresponding $\mathbf{S}_{\mathrm{eff}}^*$. However, even if the posited model is incorrect (i.e., $\eta_2^* \neq \eta_2$), we would still have consistency using the corresponding $\mathbf{S}_{\mathrm{eff}}^*$. A valid choice of $\mathbf{S}_{\mathrm{eff}}^*$ that indeed guarantees such property is

$$\begin{aligned}
&\mathbf{S}_{\mathrm{eff}}^*(Y_i, \mathbf{x}_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i, \eta_2^*) \\
&\quad = \mathrm{vecl}\Biggl(\{\mathbf{x}_i - E(\mathbf{x}_i | \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\} \\
&\qquad\qquad \times \biggl[\frac{\partial \log\{\eta_2^*(Y_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\}}{\partial(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})} - E\biggl\{\frac{\partial \log \eta_2^*(Y_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)}{\partial(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})}\Big|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i\biggr\}\biggr]\Biggr).
\end{aligned}$$

When $\eta_2^* = \eta_2$, $E\{\partial \log \eta_2^*(Y_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)/\partial(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i\} = \mathbf{0}$, hence $\mathbf{S}_{\mathrm{eff}}^* = \mathbf{S}_{\mathrm{eff}}$. The construction of a locally efficient estimator is often useful in practice due to its relative simplicity. $\mathbf{S}_{\mathrm{eff}}^*$ is almost readily applicable except that the two expectations $E(\mathbf{x}_i|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)$ and $E\{\partial \log \eta_2^*(Y_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)/\partial(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i\}$ need to be estimated nonparametrically. One can use the familiar kernel or local polynomial estimators. In Theorem 1, we show that under mild conditions, with the two expectations estimated via the Nadaraya–Watson kernel estimators, the local efficiency property indeed holds and estimating the two expectations does not cause any difference from knowing them in terms of its first order asymptotic property.

We first present the regularity conditions needed for the theoretical development.

(A1) (*The posited conditional density* η_2^*). Denote $\mathbf{u} = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$. The posited conditional density $\eta_2^*(Y, \mathbf{u})$ of Y given $\mathbf{u}$ is bounded away from 0 and infinity on its support $\mathcal{Y}$. The second derivative of $\log \eta_2^*(Y, \mathbf{u})$ with respect to $\mathbf{u}$ is continuous, positive definite and bounded. In addition, there is an open set $\boldsymbol{\Omega} \in \mathrm{R}^{p_t}$ which contains the true parameter $\mathrm{vecl}(\boldsymbol{\beta})$, such that the third derivative of $\eta_2(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ satisfies

$$\bigl|\partial^3\{\eta_2^*(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}/\bigl(\partial\,\mathrm{vecl}(\boldsymbol{\beta})_j\,\partial\,\mathrm{vecl}(\boldsymbol{\beta})_k\,\partial\,\mathrm{vecl}(\boldsymbol{\beta})_l\bigr)\bigr| \leq M_{jkl}^*(Y, \mathbf{x})$$

for all $\mathrm{vecl}(\boldsymbol{\beta}) \in \boldsymbol{\Omega}$ and $1 \leq j, k, l \leq p_t$, where $M_{jkl}^*(Y, \mathbf{x})$ satisfies $E\{{M_{jkl}^*}^2(Y, \mathbf{x})\} < \infty$, and β_j is the jth component of $\mathrm{vecl}(\boldsymbol{\beta})$.

(A2) (*The nonparametric estimation*). $E\{\partial \log \eta_2^*(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})/\partial(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\}$ and $E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ are estimated via the Nadaraya–Watson kernel estimator. For simplicity, a common bandwidth h is used which satisfies $nh^8 \to 0$ and $nh^{2d} \to \infty$ as $n \to \infty$.

(B1) (*The true conditional density* η_2). The true conditional density $\eta_2(Y, \mathbf{u})$ of Y given $\mathbf{u}$ is bounded away from 0 and infinity on its support $\mathcal{Y}$. The first and second derivatives of $\log \eta_2$ satisfy

$$E\left[\frac{\partial\{\log \eta_2(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}}{\partial\,\mathrm{vecl}(\boldsymbol{\beta})}\right] = \mathbf{0}$$

and

$$E\left[\frac{\partial\{\log \eta_2(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}}{\partial\,\mathrm{vecl}(\boldsymbol{\beta})}\frac{\partial\{\log \eta_2(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}}{\partial\,\mathrm{vecl}(\boldsymbol{\beta})^{\mathrm{T}}}\right] = -E\left[\frac{\partial^2\{\log \eta_2(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}}{\partial\,\mathrm{vecl}(\boldsymbol{\beta})\,\partial\,\mathrm{vecl}(\boldsymbol{\beta})^{\mathrm{T}}}\right]$$

is positive definite and bounded. In addition, there is an open set $\boldsymbol{\Omega} \in \mathrm{R}^{p_t}$ which contains the true parameter $\mathrm{vecl}(\boldsymbol{\beta})$, such that the third derivative of $\eta_2(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ satisfies

$$\left|\partial^3\{\eta_2(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}/(\partial\,\mathrm{vecl}(\boldsymbol{\beta})_j\,\partial\,\mathrm{vecl}(\boldsymbol{\beta})_k\,\partial\,\mathrm{vecl}(\boldsymbol{\beta})_l)\right| \le M_{jkl}(Y, \mathbf{x})$$

for all $\mathrm{vecl}(\boldsymbol{\beta}) \in \boldsymbol{\Omega}$ and $1 \le j, k, l \le p_t$, where $M_{jkl}(Y, \mathbf{x})$ satisfies $E\{M_{jkl}^2(Y, \mathbf{x})\} < \infty$, and β_j is the jth component of $\mathrm{vecl}(\boldsymbol{\beta})$.

(B2) (*The bandwidths*). The bandwidths satisfy $h_y \to 0$, $b \to 0$ and $h_x \to 0$, and $nh_y^{d+2}b \to \infty$, $n^{1/2}\{h_x^2 + (nh_x^d)^{-1/2}\}\{h_y^2 + b^2 + (nh_y^{d+2}b)^{-1/2}\} \to 0$.

(C1) (*The density functions of covariates*). Let $\mathbf{u} = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$. The density functions of $\mathbf{u}$ and $\mathbf{x}$ are bounded away from 0 and infinity on their support $\mathcal{U}$ and $\mathcal{X}$ where $\mathcal{U} = \{\mathbf{u} = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x} : \mathbf{x} \in \mathcal{X}\}$ and $\mathcal{X}$ is a compact support set of $\mathbf{x}$. Their second derivatives are finite on their supports.

(C2) (*The smoothness*). The regression functions $E(\mathbf{x}|\mathbf{u})$ has a bounded and continuous derivative on $\mathcal{U}$.

(C3) (*The kernel function*). The univariate kernel function $K(\cdot)$ is a bounded symmetric probability density function, has a bounded derivative and compact support $[-1, 1]$, and satisfies $\mu_2 = \int u^2 K(u)\,du \neq 0$. The d-dimensional kernel function is a product of d univariate kernel functions, that is, $K(\mathbf{u}) = \prod_{j=1}^d K(u_j)$, and $K_h(\mathbf{u}) = \prod_{j=1}^d K_h(u_j) = h^{-d}\prod_{j=1}^d K(u_j/h)$ for $\mathbf{u} = (u_1, \ldots, u_d)^{\mathrm{T}}$ and any bandwidth h.

THEOREM 1. *Under conditions* (A1)–(A2) *and* (C1)–(C3), *the estimator obtained from the estimating equation*

$$\sum_{i=1}^n \mathbf{S}_{\mathrm{eff}}^*(Y_i, \mathbf{x}_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i, \eta_2^*, \widehat{E}) = \mathbf{0}$$

is locally efficient. Specifically, the estimator is consistent if $\eta_2^* \neq \eta_2$, *and is efficient if* $\eta_2^* = \eta_2$. *In addition, using the estimated* $\widehat{E}(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ *results in the same estimation variance for* $\mathrm{vecl}(\boldsymbol{\beta})$ *as using the true* $E(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$. *Specifically, the estimate* $\widehat{\boldsymbol{\beta}}$ *satisfies*

$$\sqrt{n}\{\mathrm{vecl}(\widehat{\boldsymbol{\beta}}) - \mathrm{vecl}(\boldsymbol{\beta})\} \to N\{\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}(\mathbf{A}^{-1})^{\mathrm{T}}\}$$

when $n \to \infty$, *where*

$$\mathbf{A} = E\left\{\frac{\partial \mathbf{S}^*_{\mathrm{eff}}(Y_i, \mathbf{x}_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i, \eta_2^*)}{\partial\, \mathrm{vecl}(\boldsymbol{\beta})^{\mathrm{T}}}\right\}, \qquad \mathbf{B} = E\{\mathbf{S}^*_{\mathrm{eff}}(Y_i, \mathbf{x}_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i, \eta_2^*)^{\otimes 2}\}.$$

In Theorem 1 and thereafter, we use $\mathbf{v}^{\otimes 2}$ to denote $\mathbf{v}\mathbf{v}^{\mathrm{T}}$ for any matrix or vector $\mathbf{v}$, and use $\widehat{E}$ to denote the nonparametrically estimated expectation.

We describe how to implement the locally efficient estimator in several specific cases. For example, when Y is continuous, we can propose a simple conditional normal model for η_2 and hence obtain the locally efficient estimator based on summing terms of the form

$$\begin{aligned} &\mathbf{S}^*_{\mathrm{eff}}(Y, \mathbf{x}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}, \eta_2^*) \\ &\quad = \mathrm{vecl}\left(\{\mathbf{x} - E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}\left[\{Y - E(Y|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}\frac{\partial E^*(Y|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}{\partial(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})}\right]\right) \end{aligned} \tag{3.1}$$

evaluated at different observations. Here $E^*(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ is computed using the model η_2^*. When Y is binary, a common model to posit for η_2 is a logistic model. The summation of the terms of form (3.1) evaluated at different observations also provides a locally efficient estimator. When Y is a counting response variable, the Poisson model is a popular choice for η_2. This choice also yields an identical locally efficient estimator formed by the sum of (3.1). The benefits of these locally efficient estimators are two-fold. The first benefit lies in the robustness property, in that they guarantee the consistency of the resulting estimators regardless of the proposed model. The second benefit is their computational simplicity gained through avoiding estimating the conditional density η_2 and its derivative. In addition, if, by luck, the posited model happens to be correct, then the estimator is efficient.

REMARK 2. We have restricted the posited model η_2^* to be a completely known model in order to illustrate the local efficiency concept. In fact, one can also posit a model η_2^* that contains an additional unknown parameter vector, say $\boldsymbol{\gamma}$. As long as $\boldsymbol{\gamma}$ can be estimated at the root-n rate, the resulting estimator with the estimator $\widehat{\boldsymbol{\gamma}}$ plugged in is also referred to as a locally efficient estimator. In addition, if model η_2^* contains the true η_2, say $\eta_2^*(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}, \boldsymbol{\gamma}_0) = \eta_2(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, and $\boldsymbol{\gamma}_0$ is estimated consistently by $\widehat{\boldsymbol{\gamma}}$ at the root-n rate, then the resulting estimator $\mathbf{S}^*_{\mathrm{eff}}$ with $\eta_2^*(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}, \widehat{\boldsymbol{\gamma}})$ plugged in is efficient.

REMARK 3. Even if efficiency is not sought after and consistency is the sole purpose, at least one nonparametric operation, such as one that relates to estimating $E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, is needed. Thus, to completely avoid nonparametric procedures, the only option is to impose additional assumptions. The most popular linearity condition in the literature assumes $E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = \boldsymbol{\beta}(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$. Since Theorem 1 allows an arbitrary η^*, the most obvious choice in practice is probably the exponential link functions. For example, if we choose η_2^* to be the normal link function when $d = 1$, then the locally efficient estimator degenerates to a simple form, where

$$\mathbf{S}_{\mathrm{eff}}^* = \mathrm{vecl}\big[\{\mathbf{x} - \boldsymbol{\beta}(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\}(Y - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\big].$$

If we are even bolder and decide to replace $Y - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$ with Y, which is still valid given that the first term alone already guarantees consistency under the linearity condition, then we obtain the ordinary least square estimator [13]. Further connections to other existing methods are elaborated in [20].

3.2. *The efficient estimator.* Now we pursue the truly efficient estimator that reaches the semiparametric efficiency bound. This is important because in terms of reaching the optimal efficiency, relying on a posited model η_2^* to be true or to contain the true η_2 is not a satisfying practice. Intuitively, it is easy to imagine that in constructing the locally efficient estimator, if we posit a larger model η_2^*, the chance of it containing the true model η_2 becomes larger, hence the chance of reaching the optimal efficiency also increases. Thus, if we can propose the "largest" possible model for η_2^*, we will guarantee to have η_2^* containing η_2. If we can also estimate the parameters in η_2^* "correctly," we will then guarantee the efficiency. This "largest" model with a "correctly" estimated parameter turns out to be what the nonparametric estimation is able to provide. This amounts to estimating $E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, η_2 and its first derivative nonparametrically in (2.2).

We first discuss how to estimate η_2 and its first derivative, based on $(Y_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)$, $i = 1, \dots, n$. This is a problem of estimating conditional density and its derivative. We use the idea of the "double-kernel" local linear smoothing method studied in [8]. Consider $K_b(Y - y) = b^{-1}K\{(Y - y)/b\}$ with y running through all possible values, where $K(\cdot)$ is a symmetric density function, and $b > 0$ is a bandwidth. Then $E\{K_b(Y - y)|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\}$ converges to $\eta_2(y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ as b tends to 0. This observation motivates us to estimate η_2 and its first derivative, evaluated at $(y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ through minimizing the following weighted least squares:

$$\sum_{i=1}^{n}\{K_b(Y_i - y) - a - \mathbf{b}^{\mathrm{T}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}^2 K_{h_y}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}),$$

where h_y is a bandwidth, and K_{h_y} is a multivariate kernel function. The minimizers $\widehat{a}$ and $\widehat{\mathbf{b}}$ are the estimators of η_2 and $\partial\eta_2/\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$. Let the resulting estimators be $\widehat{\eta}_2(\cdot)$ and $\widehat{\eta}_2'(\cdot)$.

It remains to estimate $E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$. Using the Nadaraya–Watson kernel estimator, we have

$$\widehat{E}(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = \frac{\sum_{i=1}^{n}\mathbf{x}_i K_{h_x}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}{\sum_{i=1}^{n} K_{h_x}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})},$$

where h_x is a bandwidth, and K_{h_x} is a multivariate kernel function. The algorithm for obtaining the efficient estimator is the following:

- *Step* 1. Obtain an initial root-n consistent estimator of $\boldsymbol{\beta}$, denoted as $\widetilde{\boldsymbol{\beta}}$, through, for example, a simple locally efficient estimation procedure from Section 3.1.
- *Step* 2. Perform nonparametric estimation of $\eta_2(Y, \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x})$ and its first derivative $\partial\{\eta_2(Y, \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x})\}/\partial(\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x})$. Write the resulting estimators as $\widehat{\eta}_2(\cdot)$ and $\widehat{\eta}_2'(\cdot)$.
- *Step* 3. Perform nonparametric estimation of $E(\mathbf{x}|\widetilde{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x})$. Write the resulting estimator as $\widehat{E}(\cdot)$.
- *Step* 4. Plug $\widehat{\eta}_2(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, $\widehat{\eta}_2'(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ and $\widehat{E}(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ into $\mathbf{S}_{\mathrm{eff}}$ and solve the estimating equation

$$\sum_{i=1}^{n} \mathbf{S}_{\mathrm{eff}}(Y_i, \mathbf{x}_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i, \widehat{\eta}_2, \widehat{\eta}_2', \widehat{E}) = \mathbf{0}$$

to obtain the efficient estimator $\widehat{\boldsymbol{\beta}}$.

In performing the various nonparametric estimations in steps 2 and 3, as well as in obtaining the locally efficient estimator in Section 3.1, bandwidths need to be selected. Because the final estimator is very insensitive to the bandwidths, as indicated by conditions (A2), (B2) and Theorems 1, 2, where a range of different bandwidths all lead to the same asymptotic property of the final estimator, we suggest that one should select the corresponding bandwidths by taking the sample size n to its suitable power to satisfy (B2), and then multiply a constant to scale it, instead of performing a full-scale cross validation procedure. For example, when $d = 1$, we let $h = n^{-1/5}$, $h_x = n^{-1/5}$, $h_y = n^{-1/6}$, $b = n^{-1/7}$, and when $d = 2$, we let $h = n^{-1/6}$, $h_x = n^{-1/6}$, $h_y = n^{-1/7}$, $b = n^{-1/8}$, each multiplied by the standard deviation of the regressors calculated at the current $\widehat{\boldsymbol{\beta}}$ value.

The estimator from the above algorithm, $\widehat{\boldsymbol{\beta}}$, with its upper $d \times d$ submatrix being $\mathbf{I}_d$, reaches the optimal semiparametric efficiency bound. We present this result in Theorem 2.

THEOREM 2. *Under conditions* (B1)–(B2) *and* (C1)–(C3), *the estimator obtained from the estimating equation*

$$\sum_{i=1}^{n} \mathbf{S}_{\mathrm{eff}}(Y_i, \mathbf{x}_i, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i, \widehat{\eta}_2, \widehat{\eta}_2', \widehat{E}) = \mathbf{0}$$

is efficient. Specifically, when $n \to \infty$, the estimator of $\mathrm{vecl}(\boldsymbol{\beta})$ *satisfies*

$$\sqrt{n}\{\mathrm{vecl}(\widehat{\boldsymbol{\beta}}) - \mathrm{vecl}(\boldsymbol{\beta})\} \to N(\mathbf{0}, [E\{\mathbf{S}_{\mathrm{eff}}(Y, \mathbf{x}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}, \eta_2)^{\otimes 2}\}]^{-1})$$

in distribution.

REMARK 4. It is discovered that for certain p.d.f. η_2, such as when the inverse mean function $E(\mathbf{x}|Y)$ degenerates, some inverse, regression-based methods, such as SIR, would fail to exhaustively recover $\mathcal{S}_{Y|\mathbf{x}}$. However, this is not the case for the efficient estimator proposed here. That is, our proposed efficient estimator, similar to dMAVE [26], has the exhaustiveness property [11]. In fact, as it is listed in the regularity conditions, as long as the asymptotic covariance matrix is not singular and is bounded away from infinity, our method is always able to produce the efficient estimator.

REMARK 5. It can be easily verified that the above efficient asymptotic variance-covariance matrix can be explicitly written out as

$$\begin{aligned} &E\{\mathbf{S}_{\text{eff}}(Y, \mathbf{x}, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}, \eta_2)^{\otimes 2}\} \\ &\quad = E\left(E\left[\left\{\frac{\partial \log \eta_2(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}{\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\right\}^{\otimes 2}\Big|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right] \otimes E[\{\mathbf{x}_l - E(\mathbf{x}_l|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}^{\otimes 2}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}]\right), \end{aligned}$$

where $\mathbf{x}_l$ is the vector formed by the lower $p-d$ components of $\mathbf{x}$. Thus, the asymptotic variance of $\mathrm{vecl}(\widehat{\boldsymbol{\beta}})$ is nonsingular as long as both $E[\{\partial \log \eta_2(Y, \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})/\partial(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}^{\otimes 2}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}]$ and $E[\{\mathbf{x}_l - E(\mathbf{x}_l|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}^{\otimes 2}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}]$ are nonsingular. The nonsingularity of the first matrix is a standard requirement on the information matrix of the true model η_2 and is usually satisfied. On the other hand, $E(E[\{\mathbf{x}_l - E(\mathbf{x}_l|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}^{\otimes 2}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}])$ is always guaranteed to be nonsingular. This is because if it is singular, then there exists a unit vector $\boldsymbol{\alpha}$ with the first d components zero, such that $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}$ is a deterministic function of $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$. This violates our assumption that $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}$ cannot be a deterministic function of $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$ unless $\boldsymbol{\alpha}$ lies within the column space of $\boldsymbol{\beta}$.

4. Simulation study. In this section we conduct simulations to evaluate the finite sample performance of our efficient and locally efficient estimators and compare them with several existing methods.

We consider the following three examples:

(1) We generate Y from a normal population with mean function $\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}$ and variance 1.

(2) We generate Y from a normal population with mean function $\sin(2\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}) + 2\exp(2 + \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})$ and variance function $\log\{2 + (\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})^2\}$.

(3) We generate Y from a normal population with mean function $2(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_1)^2$ and variance function $2\exp(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_2)$.

In the simulated examples 1 and 2, we set $\boldsymbol{\beta} = (1.3, -1.3, 1.0, -0.5, 0.5, -0.5)^{\mathrm{T}}$ and generate $\mathbf{x} = (X_1, \ldots, X_6)^{\mathrm{T}}$ as follows. We generate X_1, X_2, e_1 and e_2 independently from a standard normal distribution, and form $X_3 = 0.2X_1 + 0.2(X_2 + 2)^2 + 0.2e_1$, $X_4 = 0.1 + 0.1(X_1 + X_2) + 0.3(X_1 + 1.5)^2 + 0.2e_2$. We generate X_5 and X_6 independently from Bernoulli distributions with success probability $\exp(X_1)/\{1 + \exp(X_1)\}$ and $\exp(X_2)/\{1 + \exp(X_2)\}$, respectively.

Example 3 follows the setup of Example 4.2 in [26]. In this example, we set $\boldsymbol{\beta}_1 = (1, 2/3, 2/3, 0, -1/3, 2/3)^{\mathrm{T}}$ and $\boldsymbol{\beta}_2 = (0.8, 0.8, -0.3, 0.3, 0, 0)^{\mathrm{T}}$. We form the covariates $\mathbf{x}$ by setting $X_1 = U_1 - U_2$, $X_2 = U_2 - U_3 - U_4$, $X_3 = U_3 + U_4$, $X_4 = 2U_4$, $X_5 = U_5 + 0.5U_6$ and $X_6 = U_6$, where U_1 is generated from a Bernoulli distribution with probability 0.5 to be 1 or -1, U_2 is also generated from Bernoulli distribution, with probability 0.7 to be $\sqrt{3/7}$ and probability 0.3 to be $-\sqrt{7/3}$. The remaining four components of $\mathbf{u}$ are generated from a uniform distribution between $-\sqrt{3}$ and $\sqrt{3}$. The six components of $\mathbf{u} = (U_1, \dots, U_6)^{\mathrm{T}}$ are independent, marginally having zero mean and unit variance. We construct $\mathbf{x}$ through $\mathbf{u}$ in this way to allow the components of $\mathbf{x}$ to be correlated.

For the purpose of comparison, we implement six estimators: "Oracle," "Eff," "Local," "dMAVE," "SIR" and "DR." The names of the estimators suggest the nature of these estimators, while we briefly explain them in the following:

Oracle: the oracle estimate which correctly specifies η_2 in (2.2), but we estimate $E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ through kernel regressions. We remark here that the oracle estimator is not a realistic estimator because η_2 is usually unknown. We include the oracle estimator here to provide a benchmark since this is the best performance one could hope for.

Eff: the efficient estimator which estimates $E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, η_2 and η_2' through nonparametric regressions. See Section 3.2 for a description about this efficient estimator.

Local: the locally efficient estimate which mis-specifies the model η_2, and estimates $E(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ through nonparametric regression. This is an implementation of (3.1).

dMAVE: the conditional density based minimum average variance estimation proposed by [26].

SIR: the sliced inverse regression [12] which estimates $\boldsymbol{\beta}$ as the first d principal eigenvectors of $\boldsymbol{\Sigma}^{-1}\operatorname{cov}\{E(\mathbf{x}|Y)\}\boldsymbol{\Sigma}^{-1}$, where $\boldsymbol{\Sigma} = \operatorname{cov}(\mathbf{x})$.

DR: the directional regression [10] which estimates $\boldsymbol{\beta}$ as the first d principal eigenvectors of the kernel matrix $\boldsymbol{\Sigma}^{-1/2}E\{2\mathbf{I}_p - \mathbf{A}(Y, \widetilde{Y})\}^2\boldsymbol{\Sigma}^{-1/2}$, where $\mathbf{A}(Y, \widetilde{Y}) = \boldsymbol{\Sigma}^{-1/2}E\{(\mathbf{x} - \widetilde{\mathbf{x}})(\mathbf{x} - \widetilde{\mathbf{x}})^{\mathrm{T}}|Y, \widetilde{Y}\}\boldsymbol{\Sigma}^{-1/2}$, and $(\widetilde{\mathbf{x}}, \widetilde{Y})$ is an independent copy of $(\mathbf{x}, Y)$.

We repeat each experiment 1000 times with sample size $n = 500$. The results are summarized in Table 1 for example 1, Table 2 for example 2 and Table 3 for example 3. Because the estimators we propose here use a different parameterization of the central subspace $\mathcal{S}_{Y|\mathbf{x}}$ from the existing methods such as SIR, DR or dMAVE, we transform the results from all the estimation procedures to the original $\boldsymbol{\beta}$ used to generate the data for a fair and intuitive comparison.

From the results in Table 1, we can see that Oracle, Eff, Local, dMAVE provide estimators with small bias, while SIR and DR have substantial bias in some of the elements in $\boldsymbol{\beta}$. For example, the average of the second estimated component of $\boldsymbol{\beta}$ obtained by DR is -0.2217, in contrast to the true value -1.3. This is because the covariate $\mathbf{x}$ does not satisfy the linearity or the constant variance condition,

TABLE 1

The average ("ave") and the sample standard errors ("std") for various estimates, and the inference results, respectively, the average of the estimated standard deviation ("$\widehat{\text{std}}$") and the coverage of the estimated 95% confidence interval ("95%"), of the oracle estimator and the efficient estimator, of $\boldsymbol{\beta}$ in simulated example 1

		β_1 **1.3**	β_2 **−1.3**	β_3 **1**	β_4 **−0.5**	β_5 **0.5**	β_6 **−0.5**
Oracle	ave	1.2978	−1.3036	1.0049	−0.4985	0.5033	−0.4943
	std	0.1221	0.1477	0.1505	0.1169	0.0966	0.1049
	$\widehat{\text{std}}$	0.1264	0.1510	0.1527	0.1212	0.0983	0.1052
	95%	0.9510	0.9540	0.9440	0.9540	0.9520	0.9450
Eff	ave	1.2980	−1.3046	1.0064	−0.4990	0.5040	−0.4936
	std	0.1280	0.1546	0.1567	0.1221	0.1000	0.1075
	$\widehat{\text{std}}$	0.1317	0.1588	0.1602	0.1264	0.1011	0.1084
	95%	0.9480	0.9380	0.9380	0.9440	0.9480	0.9510
Local	ave	1.3052	−1.2629	0.9687	−0.4988	0.5023	−0.4897
	std	0.1478	0.1736	0.1715	0.1393	0.1069	0.1153
dMAVE	ave	1.2599	−1.2933	1.0014	−0.4763	0.4984	−0.4935
	std	0.1932	0.1427	0.1550	0.1701	0.1368	0.1378
SIR	ave	1.3881	−1.1930	0.9261	−0.5968	0.4793	−0.4724
	std	0.1696	0.1522	0.1414	0.1489	0.0976	0.0995
DR	ave	0.9935	−0.2217	0.1930	−0.6863	0.1245	−0.1071
	std	0.6567	1.2305	1.0107	0.6411	0.3069	0.2999

and hence violates the requirement of SIR and DR. Although Local and dMAVE both appear consistent, they have much larger variance in some components than Eff. For example, in estimating β_1, the asymptotic variance of dMAVE is 0.1932, whereas that of Eff is as small as 0.1264. This is not surprising since Eff is asymptotically efficient. In fact, for this very simple setting, the estimation variance of Eff is almost as good as Oracle, which indicates that the asymptotic efficiency already exhibits for $n = 500$.

We also provide the average of the estimated standard error using the results in Theorem 2 and the 95% coverage in Table 1. The numbers show a close approximation of the sample and estimated standard error and 95% coverage is reasonable close to the nominal value.

Similar phenomena are observed for the simulated example 2 from Table 2, where SIR and DR are biased, Local and dMAVE are consistent but have larger variability than Eff and Oracle. In this more complex model where the mean function is highly nonlinear and the error is heteroscedastic, we lose the proximity between the oracle performance and the Eff performance. This is probably because $n = 500$ is still too small for this model. The inference results in Table 2, however,

TABLE 2

The average ("ave") and the sample standard errors ("std") for various estimates, and the inference results, respectively, the average of the estimated standard deviation ("$\widehat{\text{std}}$") and the coverage of the estimated 95% confidence interval ("95%"), of the oracle estimator and the efficient estimator, of $\boldsymbol{\beta}$ in simulated example 2

		β_1 **1.3**	β_2 **−1.3**	β_3 **1**	β_4 **−0.5**	β_5 **0.5**	β_6 **−0.5**
Oracle	ave	1.2999	−1.3001	1.0001	−0.4999	0.5002	−0.4999
	std	0.0023	0.0025	0.0028	0.0022	0.0023	0.0024
	$\widehat{\text{std}}$	0.0021	0.0020	0.0026	0.0020	0.0021	0.0023
	95%	0.9260	0.9070	0.9270	0.9220	0.9210	0.9380
Eff	ave	1.2996	−1.2999	0.9998	−0.4996	0.5002	−0.5000
	std	0.0116	0.0116	0.0117	0.0111	0.0068	0.0079
	$\widehat{\text{std}}$	0.0123	0.0124	0.0124	0.0120	0.0075	0.0081
	95%	0.9480	0.9550	0.9570	0.9450	0.9630	0.9520
Local	ave	1.2992	−1.3010	1.0007	−0.4993	0.5011	−0.5001
	std	0.0155	0.0210	0.0209	0.0140	0.0142	0.0147
dMAVE	ave	1.2405	−1.3422	1.0303	−0.4490	0.5114	−0.5134
	std	0.0229	0.0151	0.0133	0.0153	0.0081	0.0082
SIR	ave	0.3064	−1.6387	1.2390	0.2477	0.4697	−0.4743
	std	0.1248	0.3965	0.3149	0.1057	0.1135	0.1141
DR	ave	0.3424	0.8686	−0.6620	−0.6895	−0.1923	0.1912
	std	0.2550	1.2518	0.9653	0.6938	0.3360	0.3410

are still satisfactory, indicating that although we cannot achieve the theoretical optimality, inference is still sufficiently reliable.

What we observe in Table 3, for the simulated example 3, tells a completely different story. For this case with $d = 2$, both the linearity and the constant variance condition are violated. In addition, $\mathbf{x}$ contains categorical variables. dMAVE, SIR and DR all fail to provide good estimators in terms of estimation bias. Local and Eff remain to be consistent, although like in the simulated example 2, we can no longer hope to see the optimality as the estimation standard error is much larger than the Oracle estimator. Inference results presented in Table 3 still show satisfactory 95% coverage values, while the average estimated estimation standard error can deviate away from the sample standard error. This is caused by some numerical instability of a small proportion of the simulation repetitions. In fact, if we replace the average with the median estimated standard error, the results are closer.

5. An application. We use the proposed efficient estimator to analyze a dataset concerning the employees' salary in the Fifth National Bank of Springfield [1]. The aim of the study is to understand how an employee's salary associates

TABLE 3

The average ("ave") and the sample standard errors ("std") for various estimates, and the inference results, respectively, the average of the estimated standard deviation ("$\widehat{\text{std}}$") and the coverage of the estimated 95% confidence interval ("95%"), of the oracle estimator and the efficient estimator, of $\boldsymbol{\beta}$ in simulated example 3

		$\boldsymbol{\beta_{11}}$ **1**	$\boldsymbol{\beta_{21}}$ **0.6667**	$\boldsymbol{\beta_{31}}$ **0.6667**	$\boldsymbol{\beta_{41}}$ **0**	$\boldsymbol{\beta_{51}}$ **−0.3333**	$\boldsymbol{\beta_{61}}$ **0.6667**	$\boldsymbol{\beta_{12}}$ **0.8**	$\boldsymbol{\beta_{22}}$ **0.8**	$\boldsymbol{\beta_{32}}$ **−0.3**	$\boldsymbol{\beta_{42}}$ **0.3**	$\boldsymbol{\beta_{52}}$ **0**	$\boldsymbol{\beta_{62}}$ **0**
Oracle	ave	1.0009	0.6676	0.6674	0.0002	−0.3339	0.6675	0.8064	0.8064	−0.2905	0.2969	−0.0047	0.0053
	std	0.0305	0.0305	0.0325	0.0099	0.0198	0.0314	0.0860	0.0860	0.0902	0.0291	0.0550	0.0854
	$\widehat{\text{std}}$	0.0275	0.0275	0.0295	0.0109	0.0178	0.0276	0.0828	0.0828	0.0876	0.0296	0.0547	0.0826
	95%	0.9270	0.9270	0.9300	0.9590	0.9200	0.9110	0.9410	0.9410	0.9320	0.9450	0.9520	0.9430
Eff	ave	1.0097	0.6763	0.6764	−0.0000	−0.3384	0.6752	0.8038	0.8038	−0.3067	0.3105	0.0022	−0.0003
	std	0.0714	0.0714	0.0745	0.0162	0.0434	0.0740	0.1737	0.1737	0.1993	0.0485	0.1511	0.1895
	$\widehat{\text{std}}$	0.0709	0.0709	0.0734	0.0175	0.0454	0.0702	0.1439	0.1439	0.1490	0.0381	0.0973	0.1439
	95%	0.9280	0.9280	0.9350	0.9530	0.9460	0.9430	0.9230	0.9230	0.9240	0.9410	0.9150	0.9080
local	ave	1.0633	0.7300	0.7372	−0.0072	−0.3701	0.7468	0.7689	0.7689	−0.3066	0.2754	−0.0116	−0.0042
	std	1.8783	1.8783	2.1273	0.2493	1.0694	2.3913	1.1281	1.1281	1.5767	0.4517	0.2192	0.2516
dMAVE	ave	0.8884	0.6079	−0.1703	0.2119	−0.2498	0.5065	0.8282	0.7722	−0.0901	0.2371	−0.0153	0.0354
	std	0.0748	0.1021	0.0951	0.0569	0.0888	0.1155	0.0379	0.0378	0.1188	0.0731	0.0761	0.0489
SIR	ave	0.5443	0.3781	−0.3301	0.1816	−0.0944	0.1976	0.7768	0.6849	−0.4083	0.2908	0.0441	−0.0828
	std	0.1514	0.1414	0.0863	0.0586	0.1257	0.2022	0.0650	0.0808	0.1098	0.0748	0.1059	0.0831
DR	ave	0.6332	0.2753	−0.2968	0.0939	−0.2701	0.5422	0.7004	0.6823	−0.4512	0.1498	0.0013	−0.0151
	std	0.1813	0.2009	0.1003	0.0739	0.1288	0.1567	0.1063	0.1446	0.1688	0.0880	0.1639	0.0945

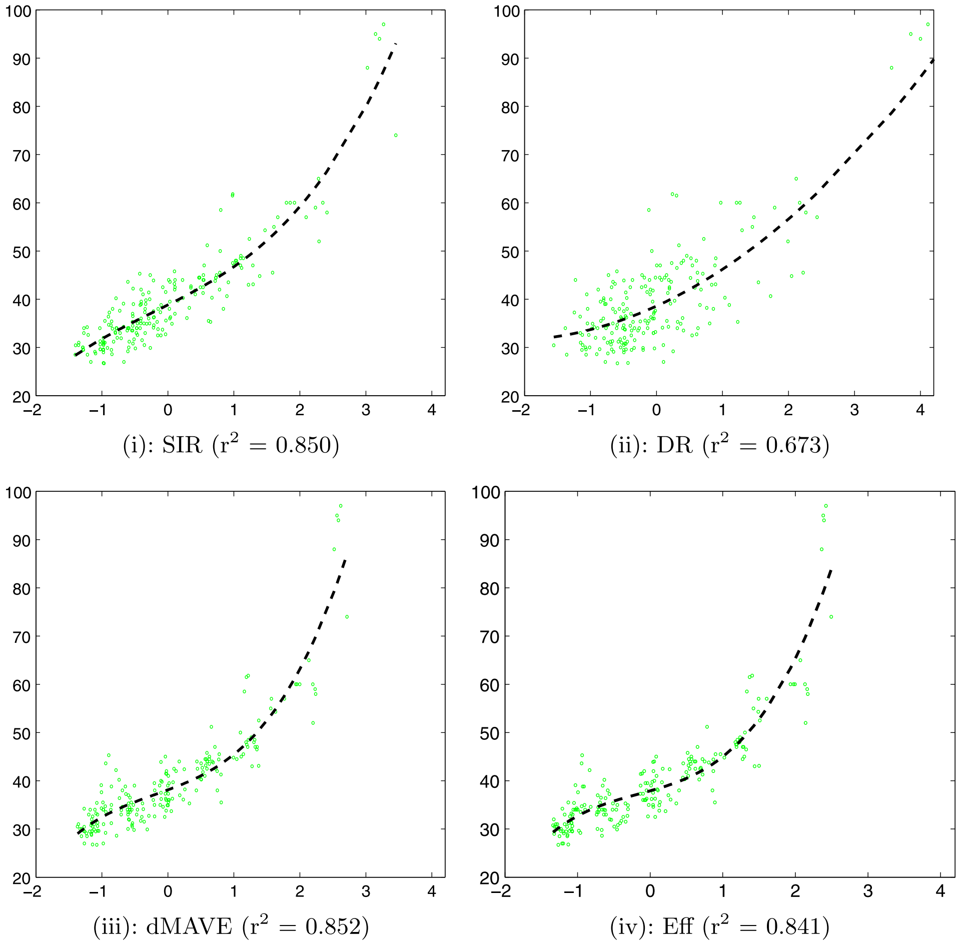

FIG. 1. *The scatter plot of Y versus $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}$, with $\widehat{\boldsymbol{\beta}}$ obtained from SIR, DR, dMAVE and Eff, respectively. The fitted cubic regression curves (–) and the adjusted r^2 values are shown.*

with his/her social characteristics. We regard an employee's annual salary as the response variable Y, and several social characteristics as the associated covariates. These covariates are, specifically, current job level (X_1); number of years working at the bank (X_2); age (X_3); number of years working at other banks (X_4); gender (X_5); whether the job is computer related (X_6). After removing an obvious outlier, the dataset contains 207 observations.

We calculated the Pearson correlation coefficients and found the current job level (X_1) has the largest correlation with his/her annual salary (Y) [$\mathrm{corr}(X_1, Y) = 0.614$]. This implies that the current job level is possibly an important factor and thus we fix the coefficient of X_1 to be 1 in our subsequent analysis. We applied SIR, DR, dMAVE and Eff methods to estimate the remaining coefficients. In Figure 1

TABLE 4
The estimated coefficients and standard errors obtained by Eff

		$\widehat{\beta}_2$	$\widehat{\beta}_3$	$\widehat{\beta}_4$	$\widehat{\beta}_5$	$\widehat{\beta}_6$
Eff	coef.	0.477	0.265	0.024	0.050	0.146
	std.	0.021	0.031	0.030	0.037	0.031
	p-value	$< 10^{-4}$	$< 10^{-4}$	0.427	0.176	$< 10^{-4}$

we present the scatter plots of Y versus a single linear combination $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}$, where $\mathbf{x} = (X_1, \ldots, X_6)^{\mathrm{T}}$ and $\widehat{\boldsymbol{\beta}}$ denote the estimate obtained from the four estimation procedures. The scatter plots exhibit similar monotone patterns in that the annual salary increases with the value of $\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}$. Except for DR, the data cloud of all other three proposals looks very compact. To quantify this visual difference, we fit a cubic model by regressing Y on 1, $(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}), (\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x})^2$ and $(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x})^3$. The adjusted r^2 values are also reported in Figure 1. The r^2 value of DR is much smaller than that of the other estimators, which suggests worse performance of DR. This is not a surprise because DR requires the most stringent conditions on the covariate vector $\mathbf{x}$, which are violated here because of the categorical covariates. The r^2 values of all other estimators including Eff are satisfactory, indicating that $\mathcal{S}_{Y|\mathbf{x}}$ is possibly one dimensional. We would also like to point out that because the r^2 value factors in the goodness-of-fit of the cubic model, hence it only provides a reference.

Table 4 contains the estimated coefficients $\widehat{\beta}_i$'s, the standard errors and p-values obtained through Eff. It can be seen that in addition to the current job level (X_1), working experience at the current bank (X_2), age (X_3) and whether or not the job is computer related (X_6) are also important factors on salary. While it is not difficult to understand the importance of most of these factors, we believe the age effect is probably caused by its high correlation with the working experience [$\mathrm{corr}(X_2, X_3) = 0.676$].

6. Discussion. We have derived both locally efficient and efficient estimators which exhaust the entire central subspace without imposing any distributional assumptions. We point out here that if the linearity condition holds, the efficiency bound does not change. However, the linearity condition will enable a simplification of the computation because we can simply plug $E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = \boldsymbol{\beta}(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$ into the estimation equation instead of estimating it nonparametrically. However, the constant variance condition does not seem to contribute to the efficiency bound or to the computational simplicity. It is therefore a redundant condition in the efficient estimation of the central subspace.

In this paper we did not discuss how to determine d, the structural dimension of $\mathcal{S}_{Y|\mathbf{x}}$ when an efficient estimation procedure is used, although we agree that this is an important issue in the area of dimension reduction. In the real-data example, we infer the structural dimension through the adjusted r^2 values. This seems a reasonable choice, but the turnout may depend on how to recover the underlying model structure. How to prescribe a rigorous data-driven procedure is needed in future works.

Various model extensions have been considered in the dimensional reduction literature. For example, in partial dimension reduction problems [3], it is assumed that $F(Y|\mathbf{x}) = F(Y|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_1, \mathbf{x}_2)$. Here, $\mathbf{x}_1$ is a covariate sub-vector of $\mathbf{x}$ that the dimension reduction procedure focuses on, while $\mathbf{x}_2$ is a covariate sub-vector that is known to directly enter the model based on scientific understanding or convention. We can see that the semiparametric analysis and the efficient estimation results derived here can be adapted to these models, through changing $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$ to $(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_1, \mathbf{x}_2)$ in all the corresponding functions and expectations while everything else remains unchanged. Another extension is the group-wise dimension reduction [14], where the model $E(Y|\mathbf{x}) = \sum_{i=1}^{k} m_i(Y, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_i)$ is considered. The semiparametric analysis in such models requires separate investigation, and it will be interesting to study the efficient estimation.

SUPPLEMENTARY MATERIAL

Supplement to "Efficient estimation in sufficient dimension reduction" (DOI: 10.1214/12-AOS1072SUPP; .pdf). The supplement file aos1072_supp.pdf is available upon request. It contains derivations of the efficient score for model (2.1) and an outline of proof for Theorems 1 and 2.

REFERENCES

[1] ALBRIGHT, S. C., WINSTON, W. L. and ZAPPE, C. J. (1999). *Data Analysis and Decision Making with Microsoft Excel*. Duxbury, Pacific Grove, CA.

[2] BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins Univ. Press, Baltimore, MD. MR1245941

[3] CHIAROMONTE, F., COOK, R. D. and LI, B. (2002). Sufficient dimension reduction in regressions with categorical predictors. *Ann. Statist.* **30** 475–497. MR1902896

[4] COOK, R. D. (1994). On the interpretation of regression plots. *J. Amer. Statist. Assoc.* **89** 177–189. MR1266295

[5] COOK, R. D. (1998). *Regression Graphics*. Wiley, New York. MR1645673

[6] COOK, R. D. and WEISBERG, S. (1991). Comment on "Sliced inverse regression for dimension reduction," by K.-C. Li. *J. Amer. Statist. Assoc.* **86** 328–332.

[7] DONG, Y. and LI, B. (2010). Dimension reduction for non-elliptically distributed predictors: Second-order methods. *Biometrika* **97** 279–294. MR2650738

[8] FAN, J., YAO, Q. and TONG, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika* **83** 189–206. MR1399164

[9] LI, B. and DONG, Y. (2009). Dimension reduction for nonelliptically distributed predictors. *Ann. Statist.* **37** 1272–1298. MR2509074

[10] LI, B. and WANG, S. (2007). On directional regression for dimension reduction. *J. Amer. Statist. Assoc.* **102** 997–1008. MR2354409

[11] LI, B., ZHA, H. and CHIAROMONTE, F. (2005). Contour regression: A general approach to dimension reduction. *Ann. Statist.* **33** 1580–1616. MR2166556

[12] LI, K.-C. (1991). Sliced inverse regression for dimension reduction (with discussion). *J. Amer. Statist. Assoc.* **86** 316–342. MR1137117

[13] LI, K.-C. and DUAN, N. (1989). Regression analysis under link violation. *Ann. Statist.* **17** 1009–1052. MR1015136

[14] LI, L., LI, B. and ZHU, L.-X. (2010). Groupwise dimension reduction. *J. Amer. Statist. Assoc.* **105** 1188–1201. MR2752614

[15] MA, Y. and CARROLL, R. J. (2006). Locally efficient estimators for semiparametric models with measurement error. *J. Amer. Statist. Assoc.* **101** 1465–1474. MR2279472

[16] MA, Y., CHIOU, J.-M. and WANG, N. (2006). Efficient semiparametric estimator for heteroscedastic partially linear models. *Biometrika* **93** 75–84. MR2277741

[17] MA, Y. and GENTON, M. G. (2010). Explicit estimating equations for semiparametric generalized linear latent variable models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 475–495. MR2758524

[18] MA, Y., GENTON, M. G. and TSIATIS, A. A. (2005). Locally efficient semiparametric estimators for generalized skew-elliptical distributions. *J. Amer. Statist. Assoc.* **100** 980–989. MR2201024

[19] MA, Y. and HART, J. D. (2007). Constrained local likelihood estimators for semiparametric skew-normal distributions. *Biometrika* **94** 119–134. MR2307902

[20] MA, Y. and ZHU, L. (2012). A semiparametric approach to dimension reduction. *J. Amer. Statist. Assoc.* **107** 168–179. MR2949349

[21] MA, Y. and ZHU, L. (2013). Supplement to "Efficient estimation in sufficient dimension reduction." DOI:10.1214/12-AOS1072SUPP.

[22] NEWEY, W. (1990). Semiparametric efficiency bounds. *J. Appl. Econometrics* **5** 99–135.

[23] ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89** 846–866. MR1294730

[24] TSIATIS, A. A. (2006). *Semiparametric Theory and Missing Data*. Springer, New York. MR2233926

[25] TSIATIS, A. A. and MA, Y. (2004). Locally efficient semiparametric estimators for functional measurement error models. *Biometrika* **91** 835–848. MR2126036

[26] XIA, Y. (2007). A constructive approach to the estimation of dimension reduction directions. *Ann. Statist.* **35** 2654–2690. MR2382662

[27] ZENG, D. and LIN, D. Y. (2007). Efficient estimation in the accelerated failure time model. *J. Amer. Statist. Assoc.* **102** 1387–1396.

[28] ZENG, D. and LIN, D. Y. (2007). Maximum likelihood estimation in semiparametric models with censored data (with discussion). *J. Roy. Statist. Soc. Ser. B* **69** 507–564.

[29] ZHU, L.-P., ZHU, L.-X. and FENG, Z.-H. (2010). Dimension reduction in regressions through cumulative slicing estimation. *J. Amer. Statist. Assoc.* **105** 1455–1466. MR2796563

[30] ZHU, Y. and ZENG, P. (2006). Fourier methods for estimating the central subspace and the central mean subspace in regression. *J. Amer. Statist. Assoc.* **101** 1638–1651. MR2279485

DEPARTMENT OF STATISTICS
TEXAS A&M UNIVERSITY
3143 TAMU
COLLEGE STATION, TEXAS 77843-3143
USA
E-MAIL: ma@stat.tamu.edu

SCHOOL OF STATISTICS AND MANAGEMENT
AND THE KEY LABORATORY
OF MATHEMATICAL ECONOMICS
MINISTRY OF EDUCTION
SHANGHAI UNIVERSITY OF FINANCE AND ECONOMICS
777 GUODING ROAD
SHANGHAI 200433
P.R. CHINA
E-MAIL: zhu.liping@mail.shufe.edu.cn