Statistica Sinica **29** (2019), 567-588
doi:https://doi.org/10.5705/ss.202017.0084

# SEMIPARAMETRIC ESTIMATION AND INFERENCE OF VARIANCE FUNCTION WITH LARGE DIMENSIONAL COVARIATES

Yanyuan Ma and Liping Zhu
*Penn State University and Renmin University of China*

*Abstract:* We investigate the simultaneous estimation and inference of the central mean subspace and central variance subspace to reduce the effective number of covariates that predict, respectively, the mean and variability of the response variable. We study the estimation, inference and efficiency properties under different scenarios, and further propose a class of locally efficient estimators when the truly efficient estimator is not practically available. This partially explains the necessity of some dimension-reduction assumptions that are commonly imposed on the conditional mean function in estimating the central variance subspace. Comprehensive simulation studies and a data analysis are performed to demonstrate the finite sample performance and efficiency gain of the locally efficient estimators in comparison with existing estimation procedures.

*Key words and phrases:* Central mean subspace, central variance subspace, dimension reduction, location-scale family, semiparametric efficiency.

## 1. Introduction

In many statistical studies, variance functions are treated as nuisance parameters (Carroll (2003)). They are solely used to improve the estimation of the mean functions. However, there are many other statistical studies where variance functions are important and are the main interest of these studies. Important applications of variance functions include, but are not limited to, description of volatility or risk in a stock market and identification of homoscedastic transformations in regression. For more classical applications of variance functions, one can refer to Box and Hill (1974), Box and Meyer (1986), Carroll and Ruppert (1988), Davidian, Carroll and Smith (1988), Davidian and Carroll (1987). In the recent study of social inequality (Western and Bloome (2009)), variance function estimation is the main quantity to characterize the income insecurity. More recently, it is further recognized that variability can also serve as a predictor of other outcomes. For example, that large variability of weight presents a hazard to heart health. Thomas, Stefanski and Davidian (2012) showed that individual variability in longitudinal measurements can predict certain health outcomes. Teschendorff and Widschwendter (2012) argued that, in cancer genomics, differential variability is important in predicting disease phenotypes. Even when the mean function is the sole quantity of interest, the variance function is still needed in inference for the mean (Cai and Wang (2008), Ma and Zhu (2014)). See Lian, Liang and Carroll (2014) for a review of the importance of variance functions in statistical models.

Modeling and estimating the variance function is not always easy. The location-scale family is probably among the most familiar in modeling the variance function together with the mean (Meyer (1987)). But, since both the mean and variance functions are parametrically modeled, this approach is restrictive and only suits the case of low dimensional covariates. In fact, when the covariate is univariate, the variance function can be estimated nonparametrically without ever modeling or estimating the mean function (Tong and Wang (2005), Tong, Ma and Wang (2013) and references therein). In this sense, variance function estimation is well studied when covariates are of low dimension. However, things are quite different when the covariate dimension is high, and mean estimation can no longer be avoided. In this territory, Cai, Levine and Wang (2009) explored the issue of variance estimation in nonparametric regression, Zhu and Zhu (2009) proposed the central variance subspace to describe the variance, Lian, Liang and Carroll (2014) adopted a partially linear structure in modeling the variance function.

In this work, we adopt the modeling strategy of the central variance subspace (Zhu and Zhu (2009)). However, our work is different in that we simultaneously consider modeling the mean structure via central mean subspace (Cook and Li (2002)). This turns out to be crucial, partly because, as we have pointed out, mean estimation is unavoidable in the presence of high dimensional covariates even if our sole interest is in the variance. Specifically, let $\mathbf{x} \in \mathbb{R}^p$ be a $p$-dimensional covariate vector and $Y \in \mathbb{R}$ be the associated univariate response variable. For large $p$, we assume that there exist $\boldsymbol{\alpha} \in \mathbb{R}^{p\times d_\alpha}$, $\boldsymbol{\beta} \in \mathbb{R}^{p\times d_\beta}$, for some smallest possible $d_\alpha$ and $d_\beta$ much smaller than $p$, such that

$$E(Y|\mathbf{x}) = E(Y|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}), \qquad \mathrm{var}(\varepsilon|\mathbf{x}) = \mathrm{var}(\varepsilon|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}), \tag{1.1}$$

where $\varepsilon \overset{\text{def}}{=} Y - E(Y|\mathbf{x})$. This assumption essentially reduces the effective number of covariates from $p$ to $d_\alpha$ in estimating mean and to $d_\beta$ in estimating variance. That is, it suffices to replace $\mathbf{x}$ with $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}$ and $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}$ respectively in understanding how the conditional mean and variance vary with $\mathbf{x}$. If $d_\alpha$, $d_\beta$ are sufficiently small and we can identify $\boldsymbol{\alpha}, \boldsymbol{\beta}$ or their column subspaces, we can then change the problem of studying $E(Y|\mathbf{x})$ and $\text{var}(\varepsilon|\mathbf{x})$ to the problem of studying $E(Y|\boldsymbol{\alpha}^{\rm T}\mathbf{x})$ and $\text{var}(\varepsilon|\boldsymbol{\beta}^{\rm T}\mathbf{x})$, which subsequently facilitates the implementation of nonparametric regression techniques such as local polynomial regression or spline approximation. In (1.1), we do not require $\boldsymbol{\alpha} = \boldsymbol{\beta}$ or $d_\alpha = d_\beta$, which is different from the conditional $k$th moment subspace defined by Yin and Cook (2002). If $\boldsymbol{\alpha} = \boldsymbol{\beta}$, then (1.1) coincides with their second moment subspace, hence we can view (1.1) as its generalization.

Although our main interest is in estimating the central variance subspace, or equivalently the parameter $\boldsymbol{\beta}$ if a unique parameterization is decided a priori, we study the estimation of the central mean subspace, or $\boldsymbol{\alpha}$ simultaneously due to the tight connection between the two. Obviously, model (1.1) can be equivalently written as

$$Y = m(\boldsymbol{\alpha}^{\rm T}\mathbf{x}) + \sigma(\boldsymbol{\beta}^{\rm T}\mathbf{x})\epsilon, \tag{1.2}$$

where $m(\cdot)$ and $\sigma(\cdot) \geq 0$ are unspecified functions, and $\epsilon$ satisfies $E(\epsilon|\mathbf{x}) = 0$, $E(\epsilon^2|\mathbf{x}) = 1$. In contrast with Lian, Liang and Carroll (2014), we do not further require $\epsilon$, or equivalently $\varepsilon/\sigma(\boldsymbol{\beta}^{\rm T}\mathbf{x})$, to be independent of $\mathbf{x}$, hence our model is more flexible in this aspect. Model (1.1), or equivalently model (1.2), is also much less stringent than the central subspace model considered in Ma and Zhu (2013b) in that it only specifies some dimension-reduction forms for the means of $Y$ and $\varepsilon^2$ on $\mathbf{x}$, hence only the first two conditional moments of $Y$ on $\mathbf{x}$. The moments of orders higher than two can be arbitrary functions of $\mathbf{x}$. In contrast, the central subspace model assumes the entire distributional function of $Y$ depends on $\mathbf{x}$ only through a few linear combinations of $\mathbf{x}$ or, equivalently, all the conditional moments of $Y$ given $\mathbf{x}$ admit dimension-reduction structures. In addition, model (1.1) allows us to investigate how the covariates affect the mean and the variance individually, while the central second moment subspace model in Yin and Cook (2002) and the central subspace model in Ma and Zhu (2013b) require a common dimension reduction form for both the mean and the variance. For completeness, we will also study the estimation and inference issues when the mean and the variance subspaces coincide.

To estimate the variance or the central variance subspace, a common approach is to obtain residuals and then work with the residual squares and the covariates. See, for example, Zhu and Zhu (2009), Zhu, Dong and Li (2013) and Luo, Li and Yin (2014) for such two step estimation procedures. Obtaining residuals requires consistent estimation of the conditional mean or the central mean subspace, where many existing methods apply (Li and Duan (1989), Li (1992), Ichimura (1993), Cook and Li (2002), Xia et al. (2002), Ma and Zhu (2014), Luo, Li and Yin (2014)). The two-step procedure of estimating mean and variance separately may not be the most efficient approach. In fact, for the model described in (1.1) or (1.2), efficiency or even inference properties of these procedures have not been studied rigorously in the literature. We conjecture that one reason for this gap in the literature is the subtlety of space estimation, in that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are not identifiable, only the space spanned by their columns is identifiable. The other reason is that the separation of the estimation of the two subspaces breaks the natural bond of the two problems and hides the complete picture.

Here, we direct our interest to both subspaces. We investigate the simultaneous estimation and inference of the central mean and the central variance subspaces, and further study the estimation efficiency. Our work is different from Yin and Cook (2002) in that we estimate two generally different spaces, the central mean subspace and central variance subspace, while they estimate a single space which simultaneously satisfies the central mean and variance requirement. Our work is also different from a recent work by Luo, Li and Yin (2014), in that they estimate each subspace separately without taking into account the dimension reduction property of the other component of the model. We first parameterize the central mean and the central variance subspaces so that estimating these two subspaces is equivalent to estimating a vector of free parameters. Such a parameterization allows us to derive the semiparametric efficient score for simultaneously estimating the central mean and the central variance subspaces, to understand the efficiency properties in this problem, and to construct a class of locally efficient estimators that perform satisfactorily in practice. We further consider a special case in which the central mean subspace and the central variance subspace coincide, and perform the parallel studies. Estimation and inference results of the two situations turn out to be very different.

## 2. The Efficient and Locally Efficient Estimators

### 2.1. Some preliminaries

In this section we investigate efficient and locally efficient estimators of the central mean and the central variance subspaces. Although in the classical semiparametric analysis, approaches and tools have been developed (Bickel et al. (1993)), these tools are applicable only when the quantities under investigation are parameters, not spaces, as we encounter here. The lack of inference tools for space estimation leads us to convert the problem of space estimation and inference to that of the parameter estimation and inference. As long as we can characterize each space with a unique set of parameters, then analysis of the parameters is equivalent to the analysis of the spaces. To simultaneously study the central mean and central variance subspaces, we are obliged to parameterize the two spaces simultaneously.

We now describe the parameterization we propose. For convenience, we assume $d_\beta$ and $d_\alpha$ are fixed numbers, and the issue of deciding the suitable $d_\beta$ and $d_\alpha$ will be discussed in Section 6. Simultaneously parameterizing two spaces is much more complex than parameterizing a single space, the latter was studied in Ma and Zhu (2013a). We first assume the upper block of $\boldsymbol{\beta}$ is the $d_\beta$-dimensional identity matrix $\mathbf{I}_{d_\beta \times d_\beta}$, while its lower block is an arbitrary matrix of size $(p - d_\beta) \times d_\beta$, denoted $\mathbf{B}$. Thus,

$$\boldsymbol{\beta}_{p \times d_\beta} = \begin{pmatrix} \mathbf{I}_{d_\beta \times d_\beta} \\ \mathbf{B}_{(p-d_\beta) \times d_\beta} \end{pmatrix}.$$

This parameterization implies that we know $d_\beta$ useful covariates and arrange them as the beginning $d_\beta$ components of $\mathbf{x}$. This is not a strong implication since usually each covariate is included because it is useful. When $d_\beta = 1$, the parameterization reduces to the familiar parameterization in single index models where the first parameter is assumed to be 1 (hence the first component is assumed to be useful). Unfortunately, these $d_\beta$ components in the conditional variance function may not coincide with the $d_\alpha$ useful components for the conditional mean function part, hence it does not necessarily lead to a convenient parameterization of $\boldsymbol{\alpha}$. We further identify $d_\alpha$ variables that are known to be useful for the conditional mean component. Assume the intersection of the $d_\beta$ variable set and $d_\alpha$ variable set contains $d_0$ variables. We arrange these $d_0$ variables as the first $d_0$ components in $\mathbf{x}$. We then arrange the remaining $d_\beta - d_0$ variables from the conditional variance set as the next $d_\beta - d_0$ components in $\mathbf{x}$ and arrange the remaining $d_\alpha - d_0$ variables from the conditional mean set as the last $d_\alpha - d_0$ components in $\mathbf{x}$. There are $p - d_\beta - d_\alpha + d_0$ variables left and we arrange them arbitrarily as the remaining middle components of $\mathbf{x}$. This allows us to use the original parameterization of $\boldsymbol{\beta}$ as we described, and at the same time allows us to require $\boldsymbol{\alpha}$ to satisfy the following requirements. The upper $d_0 \times d_\alpha$ block consists of a $d_0$-dimensional identity matrix $\mathbf{I}_{d_0 \times d_0}$ on the left and a $d_0 \times (d_\alpha - d_0)$ matrix of zeros on the right. The middle $(p - d_\alpha) \times d_\alpha$ block, denoted $\mathbf{A}$, is an arbitrary matrix. The last $(d_\alpha - d_0) \times d_\alpha$ block consists of a $(d_\alpha - d_0) \times d_0$ matrix of zeros on the left and a $(d_\alpha - d_0)$-dimensional identity matrix $\mathbf{I}_{(d_\alpha - d_0)\times(d_\alpha - d_0)}$ on the right. Thus, $\boldsymbol{\alpha}$ is of the form

$$\boldsymbol{\alpha}_{p\times d_\alpha} = \begin{pmatrix} \mathbf{I}_{d_0\times d_0} & \mathbf{0}_{d_0\times(d_\alpha - d_0)} \\ \mathbf{A}_{(p-d_\alpha)\times d_\alpha} & \\ \mathbf{0}_{(d_\alpha - d_0)\times d_0} & \mathbf{I}_{(d_\alpha - d_0)\times(d_\alpha - d_0)} \end{pmatrix},$$

where $\mathbf{A}$ is an arbitrary $(p - d_\alpha) \times d_\alpha$ matrix. Under this parameterization, we estimate the central mean and the central variance subspaces via estimating $\mathbf{A}$ and $\mathbf{B}$. The parameterization via $\mathbf{A}$ and $\mathbf{B}$ is a one-to-one mapping to these two subspaces. Recall that to insure the identifiability of a single-index model, one convention is to fix the first entry of the index parameter to be exactly one (Ichimura (1993)). Our proposal here is a generalization of the conventional parameterization used in single-index models. For notational convenience, we write $\mathrm{vecm}(\boldsymbol{\alpha})$ as the concatenation of the columns of $\mathbf{A}$ and $\mathrm{vecl}(\boldsymbol{\beta})$ as the concatenation of the columns of $\mathbf{B}$ in our subsequent exposition.

**Example 1.** We consider model (1.2) with $d_\alpha = 2$ and $d_\beta = 1$. Suppose we know in advance that the last two components of $\mathbf{x}$ contribute to the mean part and the first component of $\mathbf{x}$ contributes to the variance part. We then parameterize $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ as follows:

$$\boldsymbol{\beta}_{p\times 1} = \begin{pmatrix} 1 \\ \mathbf{B}_{(p-1)\times 1} \end{pmatrix}, \text{ and } \boldsymbol{\alpha}_{p\times 2} = \begin{pmatrix} \mathbf{A}_{(p-2)\times 2} \\ \mathbf{I}_{2\times 2} \end{pmatrix}.$$

If we also know that the first component of $\mathbf{x}$ contributes to the mean part as well, then $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are parameterized as follows:

$$\boldsymbol{\beta}_{p\times 1} = \begin{pmatrix} 1 \\ \mathbf{B}_{(p-1)\times 1} \end{pmatrix}, \text{ and } \boldsymbol{\alpha}_{p\times 2} = \begin{pmatrix} 1 & 0 \\ \multicolumn{2}{c}{\mathbf{A}_{(p-2)\times 2}} \\ 0 & 1 \end{pmatrix}.$$

We point out that the familiar parameterization where both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are required to have orthonormal columns does not yield identification of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, hence is not suitable for further estimation and inference analysis.

### 2.2. The efficient score function

From model (1.2), it is easy to see that the joint probability density of $(\mathbf{x}, Y)$ is $f_{\mathbf{x},Y}(\mathbf{x}, Y) = \eta_1(\mathbf{x})\eta_2(\epsilon, \mathbf{x})/\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ where $\epsilon = \{Y - m(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})\}/\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$. Here, $\eta_1(\mathbf{x}) \ge 0$ is the marginal density function of $\mathbf{x}$ that satisfies $\int \eta_1(\mathbf{x})d\mu(\mathbf{x}) = 1$, $\eta_2(\epsilon, \mathbf{x}) \ge 0$ is the conditional density of $Y$ on $\mathbf{x}$ and satisfies $\int \eta_2(\epsilon, \mathbf{x})d\mu(\epsilon) = 1$,

$\int \epsilon\eta_2(\epsilon,\mathbf{x})d\mu(\epsilon)=0$, $\int \epsilon^2\eta_2(\epsilon,\mathbf{x})d\mu(\epsilon)=1$. To estimate $\boldsymbol{\alpha},\boldsymbol{\beta}$, we view $\text{vecm}(\boldsymbol{\alpha})$ and $\text{vecl}(\boldsymbol{\beta})$ as the parameters of interest, with total number of parameters $d_t=(p-d_\alpha)d_\alpha+(p-d_\beta)d_\beta$, and $\eta_1,\eta_2,m,\sigma$ as the infinite dimensional nuisance parameters. From the geometrical approach (Bickel et al. (1993) and Tsiatis (2006)), we can obtain the efficient score function. It is unfortunately very complex, hence we first introduce some notations to simplify its expression. Let $\otimes$ represent Kronecker product, and take

$$\begin{aligned}
\mu_3 &\equiv \mu_3(\mathbf{x}) = E(\epsilon^3|\mathbf{x}),\\
c &\equiv c(\mathbf{x}) = \{E(\epsilon^4|\mathbf{x})-E(\epsilon^3|\mathbf{x})^2-1\}^{-1/2},\\
u &\equiv u(\epsilon,\mathbf{x}) = c(\mathbf{x})\left\{\epsilon^2-1-E(\epsilon^3|\mathbf{x})\epsilon\right\},\\
k_1 &\equiv k_1(\boldsymbol{\alpha}^{\mathrm T}\mathbf{x},\boldsymbol{\beta}^{\mathrm T}\mathbf{x}) = \sigma^{-1}E\{c^2(\mathbf{x})\mu_3(\mathbf{x})|\boldsymbol{\alpha}^{\mathrm T}\mathbf{x},\boldsymbol{\beta}^{\mathrm T}\mathbf{x}\},\\
k_2 &\equiv k_2(\boldsymbol{\beta}^{\mathrm T}\mathbf{x}) = E\{c^2(\mathbf{x})|\boldsymbol{\beta}^{\mathrm T}\mathbf{x}\},\\
k_3 &\equiv k_3(\boldsymbol{\alpha}^{\mathrm T}\mathbf{x}) = E[\{1+c^2(\mathbf{x})\mu_3^2(\mathbf{x})\}\sigma^{-2}|\boldsymbol{\alpha}^{\mathrm T}\mathbf{x}],\\
\mathbf{g}_1 &\equiv \mathbf{g}_1(\boldsymbol{\beta}^{\mathrm T}\mathbf{x}) = \sigma^{-1}E\left[c^2(\mathbf{x})\mu_3(\mathbf{x})\text{vecm}\left\{\mathbf{x}\otimes m'(\boldsymbol{\alpha}^{\mathrm T}\mathbf{x})^{\mathrm T}\right\}|\boldsymbol{\beta}^{\mathrm T}\mathbf{x}\right],\\
\mathbf{g}_2 &\equiv \mathbf{g}_2(\boldsymbol{\alpha}^{\mathrm T}\mathbf{x}) = E\left[\sigma^{-2}\left\{1+c^2(\mathbf{x})\mu_3^2(\mathbf{x})\right\}\text{vecm}\left\{\mathbf{x}\otimes m'(\boldsymbol{\alpha}^{\mathrm T}\mathbf{x})^{\mathrm T}\right\}|\boldsymbol{\alpha}^{\mathrm T}\mathbf{x}\right],\\
\mathbf{f}_1 &\equiv \mathbf{f}_1(\boldsymbol{\alpha}^{\mathrm T}\mathbf{x}) = 2E\left\{\sigma^{-2}c^2(\mathbf{x})\mu_3(\mathbf{x})\text{vecl}\left(\mathbf{x}\otimes \sigma'^{\mathrm T}\right)|\boldsymbol{\alpha}^{\mathrm T}\mathbf{x}\right\},\\
\mathbf{f}_2 &\equiv \mathbf{f}_2(\boldsymbol{\beta}^{\mathrm T}\mathbf{x}) = 2\sigma^{-1}E\left\{c^2(\mathbf{x})\text{vecl}\left(\mathbf{x}\otimes \sigma'^{\mathrm T}\right)|\boldsymbol{\beta}^{\mathrm T}\mathbf{x}\right\}.
\end{aligned}\tag{2.1}$$

The display in the curly brackets in the definition of $c\equiv c(\mathbf{x})$ is always positive. It is used to normalize $u\equiv u(\epsilon,\mathbf{x})$ so that $u$ has unit variance. Let $\mathbf{a}_1(\boldsymbol{\alpha}^{\mathrm T}\mathbf{x})$, $\mathbf{a}_2(\boldsymbol{\alpha}^{\mathrm T}\mathbf{x})$, respectively solve the equations

$$\begin{aligned}
k_3\mathbf{a}_1 - E\{k_1k_2^{-1}E(k_1\mathbf{a}_1|\boldsymbol{\beta}^{\mathrm T}\mathbf{x})|\boldsymbol{\alpha}^{\mathrm T}\mathbf{x}\} &= \mathbf{g}_2 - E(k_1k_2^{-1}\mathbf{g}_1|\boldsymbol{\alpha}^{\mathrm T}\mathbf{x}),\\
k_3\mathbf{a}_2 - E\{k_1k_2^{-1}E(k_1\mathbf{a}_2|\boldsymbol{\beta}^{\mathrm T}\mathbf{x})|\boldsymbol{\alpha}^{\mathrm T}\mathbf{x}\} &= E(k_1k_2^{-1}\mathbf{f}_2|\boldsymbol{\alpha}^{\mathrm T}\mathbf{x}) - \mathbf{f}_1,
\end{aligned}\tag{2.2}$$

and define

$$\begin{aligned}
\mathbf{b}_1(\boldsymbol{\beta}^{\mathrm T}\mathbf{x}) &= k_2^{-1}\{E(k_1\mathbf{a}_1|\boldsymbol{\beta}^{\mathrm T}\mathbf{x})-\mathbf{g}_1\},\\
\mathbf{b}_2(\boldsymbol{\beta}^{\mathrm T}\mathbf{x}) &= k_2^{-1}\{\mathbf{f}_2+E(k_1\mathbf{a}_2|\boldsymbol{\beta}^{\mathrm T}\mathbf{x})\}.
\end{aligned}$$

Finally, let $\mathbf{a}=(\mathbf{a}_1^{\mathrm T},\mathbf{a}_2^{\mathrm T})^{\mathrm T}$ and $\mathbf{b}=(\mathbf{b}_1^{\mathrm T},\mathbf{b}_2^{\mathrm T})^{\mathrm T}$. Then the efficient score for simultaneously estimating the central mean and central variance subspaces, derived in the Supplement, is

$$\mathbf{S}_{\text{eff}}(\mathbf{x},Y)=\begin{bmatrix}\{\epsilon-uc(\mathbf{x})\mu_3(\mathbf{x})\}\,\text{vecm}\left\{\mathbf{x}\otimes\dfrac{m'(\boldsymbol{\alpha}^{\mathrm T}\mathbf{x})^{\mathrm T}}{\sigma(\boldsymbol{\beta}^{\mathrm T}\mathbf{x})}\right\}\\ 2uc(\mathbf{x})\text{vecl}\left\{\mathbf{x}\otimes\dfrac{\sigma'(\boldsymbol{\beta}^{\mathrm T}\mathbf{x})^{\mathrm T}}{\sigma(\boldsymbol{\beta}^{\mathrm T}\mathbf{x})}\right\}\end{bmatrix}-uc\mathbf{b}-(\epsilon-uc\mu_3)\sigma^{-1}\mathbf{a}.$$

### 2.3. Locally efficient estimation

The efficient score suffers from some practical difficulties. First of all, (2.2) contains two integral equations that typically have to be solved numerically. While this is feasible and it has already been done in the literature (see, for example, Tsiatis and Ma (2004), Ma and Carroll (2006)), it slows down the implementation. A more serious issue is that the efficient score contains the quantities $\mu_3(\mathbf{x}) \equiv E(\epsilon^3|\mathbf{x})$ and $\mu_4(\mathbf{x}) \equiv E(\epsilon^4|\mathbf{x})$. This is an obstacle because estimating these quantities is subject to the curse of dimensionality, which is the original reason that motivated the literature of dimension reduction. We emphasize that the knowledge of $\mu_3(\mathbf{x})$ and $\mu_4(\mathbf{x})$ in constructing the efficient estimator is determined by the structure of the model. This is a fact that will not change if the efficient estimator were derived from any different approach. The difficulty of estimating $\mu_3(\mathbf{x})$ and $\mu_4(\mathbf{x})$ is also inherent to the problem as a direct consequence of curse of dimensionality. Thus, the difficulty in obtaining an efficient estimator in this problem is universal.

One could brave the estimation under the curse of dimensionality to achieve efficiency, however, a practical compromise is to seek local efficiency, where we replace quantities such as $\mu_3(\mathbf{x})$, $\mu_4(\mathbf{x})$, and possibly some other quantities, by known functions or models that do not necessarily reflect the truth. To this end, a popular choice is to set $\mu_3(\mathbf{x}) = 0$ and set $\mu_4(\mathbf{x})$ to be some known fourth moment function such as $\mu_4(\mathbf{x}) = 3$ if $\epsilon$ is treated as an independent normal random variable.This treatment is not technically necessary and does not have to reflect the true nature of $\epsilon$, but it substantially eases the computation in the estimation of the central mean and central variance subspaces. Any choices of $\mu_3(\mathbf{x})$, $\mu_4(\mathbf{x})$ calculated from some other working models for the error distribution are equally valid. We choose to work out the details under the normal working model only. If one suspects a different model might be more appropriate, then one can choose a suitable model in each problem. Under the choice of the current vanishing $\mu_3(\mathbf{x})$ and prespecified $\mu_4(\mathbf{x})$, $c(\mathbf{x})$ is a fully specified function. Further simplification yields $u = c(\epsilon^2 - 1)$, $k_1 = 0$, $k_2 = E(c^2|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, $k_3 = E(\sigma^{-2}|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$, $\mathbf{g}_1 = \mathbf{0}$, $\mathbf{g}_2 = \mathrm{vecm}\{E(\mathbf{x}\sigma^{-2}|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}) \otimes m'^{\mathrm{T}}\}$, $\mathbf{f}_1 = \mathbf{0}$, $\mathbf{f}_2 = 2\sigma^{-1}\mathrm{vecl}\{E(\mathbf{x}c^2|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) \otimes \sigma'^{\mathrm{T}}\}$. From the first equation of (2.2), we obtain $\mathbf{a}_1 = E(\sigma^{-2}|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})^{-1}\mathrm{vecm}\{E(\mathbf{x}\sigma^{-2}|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}) \otimes m'^{\mathrm{T}}\}$, and $\mathbf{b}_1 = \mathbf{0}$. From the second equation of (2.2), we obtain $\mathbf{a}_2 = \mathbf{0}$ and $\mathbf{b}_2 = 2\sigma^{-1}\mathrm{vecl}\{E(\mathbf{x}c^2|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})/E(c^2|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) \otimes \sigma'^{\mathrm{T}}\}$. Hence we have an explicit expression of the locally efficient score as

$$
\begin{aligned}
&\mathbf{S}^{\star}_{\text{eff}}(\mathbf{x}, Y) \qquad\qquad (2.3)\\
&= \begin{pmatrix} \{Y - m(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})\} \operatorname{vecm}\left[\sigma^{-2}\left\{\mathbf{x} - \dfrac{E(\mathbf{x}\sigma^{-2}|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})}{E(\sigma^{-2}|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})}\right\} \otimes m'^{\mathrm{T}}\right] \\ 2\sigma^{-1}(\epsilon^2 - 1)\operatorname{vecl}\left[c^2\left\{\mathbf{x} - \dfrac{E(\mathbf{x}c^2|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}{E(c^2|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\right\} \otimes \sigma'^{\mathrm{T}}\right] \end{pmatrix}.
\end{aligned}
$$

The first and second components in (2.3) are, respectively, the efficient score of the central mean model without variance structure and the efficient score of the central variance model without mean structure (Ma and Zhu (2014), Luo, Li and Yin (2014)). Intuitively, this is because $\mu_3 = 0$ implies the uncorrelation between $\epsilon$ and $\epsilon^2$ conditional on $\mathbf{x}$, hence the two moment models do not affect one another. There are many interesting aspects of (2.3). First of all, in using this locally efficient score to construct estimating equations, we need to estimate the conditional expectations $E(\cdot|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, $E(\cdot|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$ and $m(\cdot), m'(\cdot), \sigma(\cdot), \sigma'(\cdot)$. Fortunately, all of these are low dimensional problems and can be handled via traditional nonparametric methods with moderate sample sizes. For example, $E(\mathbf{x}|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}), E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ can be replaced by

$$
\begin{aligned}
\widehat{E}(\mathbf{x}|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}) &= \frac{\sum_{i=1}^n \mathbf{x}_i K_{h_0}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})}{\sum_{i=1}^n K_{h_0}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})},\\
\widehat{E}(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) &= \frac{\sum_{i=1}^n \mathbf{x}_i K_{h_1}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}{\sum_{i=1}^n K_{h_1}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})},
\end{aligned}
$$

where $h_0$ and $h_1$ are bandwidths, $K_{h_0}(\cdot) = K(\cdot/h_0)/{h_0}^{d_\alpha}$ and $K_{h_1}(\cdot) = K(\cdot/h_1)/{h_1}^{d_\beta}$, and $K$ is the multiplication of $d_\alpha$ or $d_\beta$ univariate kernel functions, denoted by $K$. Similarly, we can use

$$
\begin{aligned}
&\left[\widehat{m}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}\right), \{\widehat{m}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}\right)\}'\right]\\
&\stackrel{\text{def}}{=} \arg\min_{a,b} \sum_{i=1}^n \left\{Y_i - a - b^{\mathrm{T}}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}\right)\right\}^2 K_{h_2}\left(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}\right),\\
&\left[\widehat{\sigma}^2\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right), \{\widehat{\sigma}^2\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\}'\right]\\
&\stackrel{\text{def}}{=} \arg\min_{a,b} \sum_{i=1}^n \left\{\widehat{\varepsilon}_i^2 - a - b^{\mathrm{T}}\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)\right\}^2 K_{h_3}\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}\right)
\end{aligned}
$$

to replace $m(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$, $m'(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$ and $\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, $\{\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}'$ respectively, where $\widehat{\varepsilon}_i = Y_i - \widehat{m}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}_i)$. As a known function of $\mathbf{x}$, $c$ may or may not represent the truth. The resulting estimating equation is always consistent due to how $c$ appears in $\mathbf{S}^{\star}_{\text{eff}}$. One phenomenon that is quite unique here is that, even when $c(\mathbf{x})$ happens to be the truth, $\mathbf{S}_{\text{eff}}^{\star}$ may still be inefficient. This is because the true efficiency requires the correct specification of both $\mu_3$ and $\mu_4$, instead of simply a true $c$ as a combination of them. Here $\sigma$ in the first equation of (2.3) plays the same role as $c$ in the second equation. Its mis-specification in the first equation does not affect the consistency. Hence, if desired, we can replace $\sigma$ using a known form for simplicity. For example, we can let $\sigma = 1$ in the first equation and $c = 1$ in the second equation to obtain

$$\mathbf{S}_{\text{eff}}^{\star}(\mathbf{x}, Y) = \begin{pmatrix} \{Y - m(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})\} \operatorname{vecm}\left[\{\mathbf{x} - E(\mathbf{x}|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})\} \otimes m'^{\mathrm{T}}\right] \\ 2\sigma^{-1}(\epsilon^2 - 1)\operatorname{vecl}\left[\{\mathbf{x} - E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\} \otimes \sigma'^{\mathrm{T}}\right] \end{pmatrix}.$$

Of course, further simplification is still possible. For example, in the first equation, we can specify a form of $m, m'$ and estimate $E(\mathbf{x}|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$ only, or specify $E(\mathbf{x}|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}), m'$ and estimate $m$ only. Similarly, in the second equation, we can choose to specify $\sigma, \sigma'$ and estimate $E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ only or specify $E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}), \sigma'$ and estimate $\sigma$ only.

Iteratively solving for the parameters in $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, denoted as $\boldsymbol{\theta}$, from the estimating equation

$$\sum_{i=1}^{n} \mathbf{S}_{\text{eff}}^{\star}(\mathbf{x}_i, Y_i) = \mathbf{0}, \tag{2.4}$$

through Newton-Raphson method similarly as done in Ma and Zhu (2013b), where $\mathbf{S}_{\text{eff}}^{\star}$ is given via (2.3) with the unknown functions replaced by their estimates, provides a locally efficient estimator.

**Theorem 1.** *If $\widehat{\boldsymbol{\theta}}$ solves* (2.4)*, then under the regularity conditions stated in the Supplement,*

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \longrightarrow N\left\{\mathbf{0}, E\left(-\frac{\partial \mathbf{S}_{\text{eff}}^{\star}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right)^{-1} E\left(\mathbf{S}_{\text{eff}}^{\star}\mathbf{S}_{\text{eff}}^{\star\mathrm{T}}\right) E\left(-\frac{\partial \mathbf{S}_{\text{eff}}^{\star\mathrm{T}}}{\partial \boldsymbol{\theta}}\right)^{-1}\right\},$$

*in distribution when $n \to \infty$.*

If we specify a local model $\eta_2^{\star}(\epsilon, \mathbf{x})$ with the first four moments, then the resulting $E(-\partial \mathbf{S}_{\text{eff}}^{\star}/\partial \boldsymbol{\theta}^{\mathrm{T}})^{-1}\mathbf{S}_{\text{eff}}^{\star}$ is a valid influence function, which implies that $E\left(-\partial \mathbf{S}_{\text{eff}}^{\star}/\partial \boldsymbol{\theta}^{\mathrm{T}}\right)$ is always invertible. From Theorem 1, taking into consideration that the efficient estimation variance of $\widehat{\boldsymbol{\theta}}$ is $\{E(\mathbf{S}_{\text{eff}}\mathbf{S}_{\text{eff}}^{\mathrm{T}})\}^{-1}$, it is clear that because of the difficulty in obtaining the true $\mu_3(\mathbf{x}), \mu_4(\mathbf{x})$, our local estimator has a potential efficiency loss quantified by

$$n\{\operatorname{var}(\widehat{\boldsymbol{\theta}}) - \operatorname{var}(\widehat{\boldsymbol{\theta}}_{\text{eff}})\} = \operatorname{var}\left\{E\left(-\frac{\partial \mathbf{S}_{\text{eff}}^{\star}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right)^{-1}\mathbf{S}_{\text{eff}}^{\star} - E(\mathbf{S}_{\text{eff}}\mathbf{S}_{\text{eff}}^{\mathrm{T}})^{-1}\mathbf{S}_{\text{eff}}\right\}.$$

We have been estimating $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ and treating them as equally important. If only the index $\boldsymbol{\alpha}$ of the mean component or $\boldsymbol{\beta}$ of the variance component is of interest, we can easily extract the sole information about $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ by retaining the first or second component of $\widehat{\boldsymbol{\theta}}$ as $\widehat{\boldsymbol{\alpha}}$ or $\widehat{\boldsymbol{\beta}}$, and extracting the upper-left $(p-d_\alpha)d_\alpha \times (p-d_\alpha)d_\alpha$ or lower-right $(p-d_\beta)d_\beta \times (p-d_\beta)d_\beta$ matrix from $\mathrm{var}(\widehat{\boldsymbol{\theta}})$ as the corresponding asymptotic variance matrix. Usually, a simplification can be obtained through noting that $E(\partial \mathbf{S}^\star_{\mathrm{eff},\alpha}/\partial \boldsymbol{\beta}^{\mathrm{T}}) = \mathbf{0}$ and $E(\partial \mathbf{S}^\star_{\mathrm{eff},\beta}/\partial \boldsymbol{\alpha}^{\mathrm{T}}) = \mathbf{0}$. Here, we use $\mathbf{S}^\star_{\mathrm{eff},\alpha}$ and $\mathbf{S}^\star_{\mathrm{eff},\beta}$ to denote, respectively, the first $(p-d_\alpha)d_\alpha$ and the last $(p-d_\beta)d_\beta$ components of $\mathbf{S}^\star_{\mathrm{eff}}$. Thus, we have

$$\mathrm{var}(\widehat{\boldsymbol{\beta}}) = n^{-1} E\left(-\frac{\partial \mathbf{S}^\star_{\mathrm{eff},\beta}}{\partial \boldsymbol{\beta}^{\mathrm{T}}}\right)^{-1} E\left(\mathbf{S}^\star_{\mathrm{eff},\beta} \mathbf{S}^{\star\mathrm{T}}_{\mathrm{eff},\beta}\right) E\left(-\frac{\partial \mathbf{S}^{\star\mathrm{T}}_{\mathrm{eff},\beta}}{\partial \boldsymbol{\beta}}\right)^{-1}$$

asymptotically, where all the functions are evaluated at the truth. One observation then is that the estimation variance of $\widehat{\boldsymbol{\alpha}}$ has no effect on the estimation variance of $\widehat{\boldsymbol{\beta}}$ asymptotically. Hence, in terms of the quality of the $\boldsymbol{\beta}$ estimation measured by its asymptotic variance, plugging in any consistent estimator $\widehat{\boldsymbol{\alpha}}$ to the estimating equation $\sum_{i=1}^n \mathbf{S}^\star_{\mathrm{eff},\beta}(\mathbf{x}_i, Y_i) = \mathbf{0}$ has the same consequence.

## 3. Numerical Studies

### 3.1. Simulation

We illustrate our proposed methodology through a simulated example. We fixed $n = 800$ and $p = 6$. We generated $X_1$, $X_2$, $X_5$ and $X_6$ independently from the standard normal distribution, and $X_3$ and $X_4$ from the Bernoulli distribution with success probability 0.5. Given $\mathbf{x} = (X_1, \ldots, X_6)^{\mathrm{T}}$, we generated $Y$ from a normal distribution with mean function $m(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x}) = (\boldsymbol{\alpha}_1^{\mathrm{T}}\mathbf{x}+1)^2 + (\boldsymbol{\alpha}_2^{\mathrm{T}}\mathbf{x}+1)^2$ and standard deviation $\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = 0.5/\{0.1 + (\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{x})^2 + (\boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{x})^2\}$. Here $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2)$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, $\boldsymbol{\alpha}_1 = (1, 0, -0.2, -0.2, 0.2, 0.2)^{\mathrm{T}}$, $\boldsymbol{\alpha}_2 = (0, 1, -0.5, 0.2, -0.2, 0.2)^{\mathrm{T}}$, $\boldsymbol{\beta}_1 = (1, 0, -0.5, -0.2, -0.5, -0.2)^{\mathrm{T}}$ and $\boldsymbol{\beta}_2 = (0, 1, -0.2, -0.5, -0.2, -0.5)^{\mathrm{T}}$. Thus, $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2)$, $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2)$, $\mathbf{A}_1 = (-0.2, -0.2, 0.2, 0.2)^{\mathrm{T}}$, $\mathbf{A}_2 = (-0.5, 0.2, -0.2, 0.2)^{\mathrm{T}}$, $\mathbf{B}_1 = (-0.5, -0.2, -0.5, -0.2)^{\mathrm{T}}$ and $\mathbf{B}_2 = (-0.2, -0.5, -0.2, -0.5)^{\mathrm{T}}$.

If $\varepsilon = Y - m(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$, we solved the estimating equation (2.4), where

$$\begin{aligned}
&\mathbf{S}^\star_{\mathrm{eff}}(\mathbf{x}, Y)\\
&= \begin{pmatrix} \{Y - m(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})\}\mathrm{vecm}\left[\left\{\mathbf{x} - E(\mathbf{x}|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})\right\} \otimes \{m(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})\}'^{\mathrm{T}}\right] \\ \{\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}^{-2}\{\varepsilon^2 - \sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}\mathrm{vecl}\left[\left\{\mathbf{x} - E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\right\} \otimes \{\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}'^{\mathrm{T}}\right] \end{pmatrix},
\end{aligned}$$

to simultaneously estimate both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. For comparison purpose, we imple-

Table 1. The bias ("bias") and the sample standard errors ("std") for rMAVE, SEE, ECS, C2MS, and our estimating equations estimators (EEE), and the inference results, respectively the average of the estimated standard deviation ("$\widehat{\text{std}}$") and the coverage of the estimated 95% confidence interval ("cp"), of our proposals. All numbers reported below are multiplied by 100.

| | | $\alpha_{1,3}$ | $\alpha_{1,4}$ | $\alpha_{1,5}$ | $\alpha_{1,6}$ | $\alpha_{2,3}$ | $\alpha_{2,4}$ | $\alpha_{2,5}$ | $\alpha_{2,6}$ |
|---|---|---|---|---|---|---|---|---|---|
| | true | −0.20 | −0.20 | 0.20 | 0.20 | −0.50 | 0.20 | −0.20 | 0.20 |
| rMAVE | bias | −0.07 | 0.17 | 0.07 | −0.00 | 0.18 | −0.04 | −0.10 | 0.06 |
| | std | 3.04 | 2.97 | 1.40 | 1.45 | 3.16 | 2.97 | 1.50 | 1.29 |
| SEE | bias | 0.13 | 0.16 | 0.02 | −0.19 | 0.28 | −0.31 | 0.11 | −0.05 |
| | std | 3.65 | 2.91 | 2.25 | 4.87 | 9.71 | 4.27 | 1.90 | 2.95 |
| ECS | bias | 0.26 | 0.17 | −0.23 | −0.01 | 0.05 | −0.36 | −0.32 | 0.62 |
| | std | 2.69 | 2.16 | 2.65 | 3.19 | 2.89 | 2.51 | 2.95 | 3.39 |
| C2MS | bias | 1.75 | 2.74 | −3.78 | −1.46 | 11.03 | −2.73 | 2.74 | 2.34 |
| | std | 19.67 | 19.61 | 27.07 | 29.27 | 20.20 | 20.63 | 28.09 | 29.75 |
| EEE | bias | 0.03 | 0.31 | −0.06 | −0.09 | 0.20 | −0.09 | 0.09 | 0.07 |
| | std | 2.32 | 2.22 | 1.15 | 1.00 | 2.44 | 2.18 | 1.13 | 0.99 |
| | $\widehat{\text{std}}$ | 2.23 | 2.14 | 1.03 | 0.93 | 2.32 | 2.11 | 1.00 | 0.93 |
| | cp | 93.00 | 94.50 | 94.30 | 94.10 | 94.00 | 95.00 | 94.10 | 93.90 |
| | | $\beta_{1,3}$ | $\beta_{1,4}$ | $\beta_{1,5}$ | $\beta_{1,6}$ | $\beta_{2,3}$ | $\beta_{2,4}$ | $\beta_{2,5}$ | $\beta_{2,6}$ |
| | true | −0.50 | −0.20 | −0.50 | −0.20 | −0.20 | −0.50 | −0.20 | −0.50 |
| rMAVE | bias | 1.94 | −1.12 | 1.80 | −1.09 | −1.20 | 0.98 | −0.85 | 1.75 |
| | std | 38.46 | 34.33 | 20.85 | 20.24 | 42.64 | 33.69 | 19.62 | 21.03 |
| SEE | bias | 0.53 | −0.39 | 0.27 | −0.09 | 0.69 | 0.22 | 0.07 | 0.36 |
| | std | 3.49 | 3.36 | 3.11 | 3.05 | 5.87 | 5.90 | 2.83 | 2.47 |
| ECS | bias | 0.80 | 0.11 | −0.67 | 0.22 | −0.27 | 1.16 | 1.00 | −1.49 |
| | std | 3.11 | 2.71 | 3.34 | 4.00 | 4.97 | 4.19 | 5.11 | 6.04 |
| C2MS | bias | 14.55 | 1.19 | 29.54 | 6.47 | −0.08 | 17.56 | −6.82 | 29.86 |
| | std | 27.34 | 24.20 | 39.07 | 37.66 | 29.85 | 30.20 | 38.48 | 43.22 |
| EEE | bias | 0.04 | −0.11 | 0.14 | −0.31 | −0.35 | 0.21 | 0.07 | 0.11 |
| | std | 8.76 | 8.43 | 4.89 | 5.23 | 8.82 | 8.49 | 5.01 | 4.85 |
| | $\widehat{\text{std}}$ | 11.39 | 11.34 | 5.61 | 5.74 | 11.43 | 11.47 | 5.66 | 5.66 |
| | cp | 97.10 | 97.20 | 94.80 | 94.90 | 97.50 | 97.90 | 95.10 | 94.90 |

mented the refined minimum average variance estimation (rMAVE, Xia et al. (2002) with $Y$ and the residual squares $\varepsilon^2$ as response variables, respectively, to estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. These rMAVE estimators are used as initial values in solving (2.4) throughout our numerical studies.

We summarized the simulation results from 1,000 data sets in Table 1. In estimating $\boldsymbol{\alpha}$, our proposal has slightly better performance than rMAVE in that it has slightly smaller standard deviations. In estimating $\boldsymbol{\beta}$, our proposal is an obvious winner since both the estimation biases and the Monte Carlo standard deviations are significantly smaller than those from rMAVE. We also compared with the efficient central space (ECS) method of Ma and Zhu (2013b) and semiparametric estimating equation (SEE) based estimator of Luo, Li and Yin (2014). The performances of SEE and ECS appear similar. We found that in estimating $\alpha$, our proposal has slightly better performance with smaller standard deviations, while in estimating $\beta$, our estimating equation estimators yield smaller biases but larger standard deviations. We further compared with the conditional 2nd moment subspace (C2MS) estimator of Yin and Cook (2002), and found that our results are significantly better.

In Table 1, we report the averages of the estimated standard deviations ("$\widehat{\text{std}}$") and the empirical coverage probabilities ("cp") at the nominal level 95%. The standard deviations are estimated using the asymptotic results in Theorem 1. The averages of the estimated standard deviations approximate the corresponding Monte Carlo standard deviations ("std") well, and the empirical coverage probabilities are fairly close to the nominal level 95%, indicating that the inference results of our proposal are reasonably precise.

### 3.2. Extra simulation

Following the request of a referee, we performed additional simulation studies. Specifically, we kept the mean and variance model of the simulation in Section 3.1, and generated $\epsilon$ from the standard $t$ with $(\mathbf{x}^{\mathrm{T}}\mathbf{x}+4)$ degrees of freedom. The true values of $\mu_3$ and $\mu_4$ are 0 and $6/(\mathbf{x}^{\mathrm{T}}\mathbf{x})+3$, respectively, in this case. We still implemented (2.3) to estimate $\alpha$ and $\beta$. From (2.3), the change of $\mu_4$ only affects the efficient score of the central variance space model. It does not affect the efficient score of the central mean space model. The simulation results of the locally efficient estimators and the oracle estimators are given in Table S1 in the supplement. The two efficient estimators yield identical results in estimating $\alpha$, but different in estimating $\beta$. The oracle estimator appears to have smaller biases than the locally efficient estimator, while it has slightly larger variances.

### 3.3. Analysis of bank data

We further demonstrate the performance of our estimating equation based estimators through a gender discrimination data set. The Fifth National Bank of Springfield (Albright, Winston and Zappe (1999)) faced a lawsuit for paying substantially lower salaries to its female employees. To investigate whether this is the fact, the bank collected annual salaries ($Y$) of 207 employees, and some other personal characteristics such as an employee's current job level ($X_1$), working

Table 2. Analysis of the bank data.

| | rMAVE | | EEE from (2.4) | | | | EEE from (4.2) | |
|---|---|---|---|---|---|---|---|---|
| | $\widehat{\boldsymbol{\alpha}}$ | $\widehat{\boldsymbol{\beta}}$ | $\widehat{\boldsymbol{\alpha}}$ | $\widehat{\mathrm{std}}(\widehat{\boldsymbol{\alpha}})$ | $\widehat{\boldsymbol{\beta}}$ | $\widehat{\mathrm{std}}(\widehat{\boldsymbol{\beta}})$ | $\widehat{\boldsymbol{\beta}}$ | $\widehat{\mathrm{std}}(\widehat{\boldsymbol{\beta}})$ |
| $X_2$ | 0.324 | 0.142 | 0.310 | 0.058 | 0.320 | 0.114 | 0.205 | 0.037 |
| $X_3$ | −0.083 | 0.164 | −0.075 | 0.065 | −0.085 | 0.187 | −0.013 | 0.023 |
| $X_4$ | 0.070 | 0.038 | 0.074 | 0.045 | 0.066 | 0.074 | 0.073 | 0.009 |
| $X_5$ | 0.096 | −1.676 | 0.086 | 0.077 | 0.100 | 0.124 | −0.011 | 0.050 |
| $X_6$ | 0.689 | −0.967 | 0.689 | 0.113 | 0.689 | 0.219 | 0.604 | 0.042 |

experience at current bank ($X_2$), age ($X_3$), prior experience at other banks ($X_4$), gender ($X_5$) and a binary variable indicating whether a job is computer related ($X_6$).

Ma and Zhu (2012) demonstrated through bootstrap that the dependence of $Y$ on the covariates in this data set can be captured by a one-dimensional model. We analyze this data set using (1.2) with $d_\alpha = d_\beta = 1$. We expect that an employee's annual salary is positively correlated with his/her current job level, thus the coefficient of $X_1$ must be nonzero. We fixed the coefficient of $X_1$ at 1 for identifiability, then applied rMAVE to estimate $\boldsymbol{\alpha}$ in the mean function. Treating the squared residual as response, we further applied rMAVE to estimate $\boldsymbol{\beta}$ in the variance function. The results are in the first block of Table 2.

We also applied the estimating equations (2.4) to solve for both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The resulting estimates and their associated standard deviations are in the second block of Table 2. Here $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$ show no evidence of gender effect. In addition, $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}$ are similar. This motivates us to consider

$$H_0 : \boldsymbol{\alpha} = \boldsymbol{\beta}.$$

To formally test this hypothesis, we write $\boldsymbol{\theta} = (\boldsymbol{\alpha}^{\mathrm{T}}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$ and $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}, \widehat{\boldsymbol{\beta}}^{\mathrm{T}})^{\mathrm{T}}$. In addition, we denote $\mathrm{var}(\widehat{\boldsymbol{\theta}})$ the asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\theta}}$. Let $\mathbf{C}$ be a $5 \times 10$ matrix, with the identity matrix $\mathbf{I}_{5\times 5}$ on the left and the negative identity matrix $-\mathbf{I}_{5\times 5}$ on the right. Then the above null hypothesis is equivalent to $H_0 : \mathbf{C}\boldsymbol{\theta} = \mathbf{0}$. Under $H_0$, the test statistic

$$T \equiv \widehat{\boldsymbol{\theta}}^{\mathrm{T}}\mathbf{C}^{\mathrm{T}}\big\{\mathbf{C}\widehat{\mathrm{var}}(\widehat{\boldsymbol{\theta}})\mathbf{C}^{\mathrm{T}}\big\}^{-1}\mathbf{C}\widehat{\boldsymbol{\theta}} \longrightarrow \chi^2_5$$

in distribution, where $\chi^2_5$ denotes a $\chi^2$ distribution with 5 degrees of freedom, and $\widehat{\mathrm{var}}(\widehat{\boldsymbol{\theta}})$ is an estimate of $\mathrm{var}(\widehat{\boldsymbol{\theta}})$, obtained using the results in Theorem 1. Using the bank data, we obtained $T = 0.193$ and the p-value of 0.999. Therefore, we cannot reject the null hypothesis, indicating that the central mean and the central variance subspaces coincide in this data set.

## 4. Analysis When Central Mean and Central Variance Subspaces Coincide

The numerical analysis on the bank data suggests that, in practice, it is not unreasonable for the mean and variance to rely on the same set of indexes. This can be described as model (1.2) with the additional assumption that $\boldsymbol{\alpha} = \boldsymbol{\beta}$ contains $(p-d_\beta)d_\beta$ parameters of interest, and corresponds to the central second moment subspace defined in Yin and Cook (2002). Then

$$Y = m(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) + \sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\epsilon. \tag{4.1}$$

Accordingly, model (1.1) can be simplified to $\mathrm{var}(Y|\mathbf{x}) = \mathrm{var}(Y|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$. This simple additional information, however, drastically changes the model and its subsequent estimation and inference results. The efficient estimation variance decreases as a result of the additional model structure.

Using similar techniques as those used in the general case, in the Supplement, we derive the efficient score to be

$$\begin{aligned}\mathbf{S}_{\mathrm{eff}}(\mathbf{x}, Y) = {}& \epsilon\left\{\frac{\mathbf{x}\otimes m'^{\mathrm{T}}}{\sigma}(1+\mu_3^2c^2) - \frac{2c^2\mu_3\mathrm{vecl}(\mathbf{x}\otimes\sigma'^{\mathrm{T}})}{\sigma} + \mu_3c^2\mathbf{a} - (1+\mu_3^2c^2)\mathbf{b}\right\}\\ &+ (\epsilon^2-1)\left\{\frac{2c^2\mathrm{vecl}(\mathbf{x}\otimes\sigma'^{\mathrm{T}})}{\sigma} - \frac{c^2\mu_3\mathrm{vecl}(\mathbf{x}\otimes m'^{\mathrm{T}})}{\sigma} - c^2\mathbf{a} + c^2\mu_3\mathbf{b}\right\}.\end{aligned}$$

Here, $\mu_3, c$ are defined as before in (2.1), and $\mathbf{a}, \mathbf{b}$ are explicitly given as

$$\begin{aligned}\mathbf{a}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) &= \mathrm{vecl}\left\{\frac{2k_2\mathbf{g}_2\otimes\sigma'^{\mathrm{T}} + k_3\mathbf{g}_2\otimes m'^{\mathrm{T}} - k_2\mathbf{g}_3\otimes m'^{\mathrm{T}} - 2k_3\mathbf{g}_1\otimes\sigma'^{\mathrm{T}}}{\sigma(k_2^2 - k_1k_3)}\right\},\\ \mathbf{b}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) &= \mathrm{vecl}\left\{\frac{2k_1\mathbf{g}_2\otimes\sigma'^{\mathrm{T}} + k_2\mathbf{g}_2\otimes m'^{\mathrm{T}} - k_1\mathbf{g}_3\otimes m'^{\mathrm{T}} - 2k_2\mathbf{g}_1\otimes\sigma'^{\mathrm{T}}}{\sigma(k_2^2 - k_1k_3)}\right\},\end{aligned}$$

where now

$$\begin{aligned}k_1 &\equiv k_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = E(c^2|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}),\\ k_2 &\equiv k_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = E(c^2\mu_3|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}),\\ k_3 &\equiv k_3(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = E(1 + c^2\mu_3^2|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}),\\ \mathbf{g}_1 &\equiv \mathbf{g}_1(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = E(c^2\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}),\\ \mathbf{g}_2 &\equiv \mathbf{g}_2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = E(c^2\mu_3\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}),\\ \mathbf{g}_3 &\equiv \mathbf{g}_3(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = E(\mathbf{x} + c^2\mu_3^2\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}).\end{aligned}$$

Although the form of $\mathbf{S}_{\mathrm{eff}}$ is explicit and no longer involves solving integral equations, it remains complex and involves estimating $\mu_3(\mathbf{x})$ and $\mu_4(\mathbf{x})$, cursed by the possibly high dimensionality $p$. We compromise by looking for local efficiency.

Here we only need to make assumptions on the same $\mu_3$ and $\mu_4$. If we adopt the strategy for the general case where $\boldsymbol{\alpha}$ is not necessarily the same as $\boldsymbol{\beta}$, by setting $\mu_3 = 0$ and pre-specifying $c(\mathbf{x})$ as a known function of $\mathbf{x}$, we have $u = c(\epsilon^2 - 1)$, $k_1 = E(c^2|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, $k_2 = 0$, $k_3 = 1$, $\mathbf{g}_1 = E(c^2\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, $\mathbf{g}_2 = \mathbf{0}$, $\mathbf{g}_3 = E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, $\mathbf{a} = \mathrm{vecl}\{2E(c^2\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) \otimes {\sigma'}^{\mathrm{T}}\}E(c^2|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})^{-1}/\sigma$, $\underline{=}\ \mathrm{vecl}\{E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) \otimes {m'}^{\mathrm{T}}\}/\sigma$. This yields a much simpler form of the locally efficient score

$$\mathbf{S}_{\mathrm{eff}}^*(\mathbf{x}, Y) = \frac{2(\epsilon^2-1)}{\sigma}\mathrm{vecl}\left[\left\{c^2\mathbf{x} - \frac{c^2E(c^2\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}{E(c^2|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\right\} \otimes {\sigma'}^{\mathrm{T}}\right] + \frac{\epsilon}{\sigma}\mathrm{vecl}\left[\left\{\mathbf{x} - E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\right\} \otimes {m'}^{\mathrm{T}}\right].$$

This expression is practically useful for generating estimating equations. To be precise, we can estimate $\boldsymbol{\beta}$ through solving

$$\sum_{i=1}^{n} \mathbf{S}_{\mathrm{eff}}^*(\mathbf{x}_i, Y_i) = \mathbf{0}, \tag{4.2}$$

where $\mathbf{S}_{\mathrm{eff}}^*$ is given above. The resulting estimator is always consistent, and is efficient if indeed $\mu_3 = 0$ and a correct form of $c(\mathbf{x})$ is used.

**Theorem 2.** *If $\widehat{\boldsymbol{\beta}}$ solves* (4.2)*, then under the regularity conditions in the Supplement,*

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to N\left\{\mathbf{0}, E\left(-\frac{\partial \mathbf{S}_{\mathrm{eff}}^*}{\partial \boldsymbol{\beta}^{\mathrm{T}}}\right)^{-1} E\left(\mathbf{S}_{\mathrm{eff}}^* {\mathbf{S}_{\mathrm{eff}}^*}^{\mathrm{T}}\right) E\left(-\frac{\partial {\mathbf{S}_{\mathrm{eff}}^*}^{\mathrm{T}}}{\partial \boldsymbol{\beta}}\right)^{-1}\right\}$$

*in distribution when $n \to \infty$.*

## 5. Further Numerical Studies

### 5.1. Additional simulation

When the central mean and the central variance subspaces coincide, we consider solving (4.2), where

$$\mathbf{S}_{\mathrm{eff}}^*(\mathbf{x}, Y) = \frac{\{\varepsilon^2 - \sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}}{\{\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}^2}\mathrm{vecl}\left[\left\{\mathbf{x} - E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\right\} \otimes {(\sigma^2)'}^{\mathrm{T}}\right] + \frac{\varepsilon}{\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\mathrm{vecl}\left[\left\{\mathbf{x} - E(\mathbf{x}|\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\right\} \otimes {m'}^{\mathrm{T}}\right],$$

and $\varepsilon = Y - m(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, which is indeed $\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\epsilon$ as defined in (4.1). We also compare this estimating equations approach to rMAVE based on $Y$ and $\varepsilon^2$, respectively.

We set $n = 800$ and $p = 6$, and generated the covariates independently from

Table 3. The bias ("bias") and the sample standard errors ("std") for rMAVE, SEE, ECS, C2MS and our estimating equations estimators (EEE), and the inference results, respectively, the average of the estimated standard deviation ("$\widehat{\text{std}}$") and the coverage of the estimated 95% confidence interval ("cp"), of our proposals. All numbers reported below are multiplied by 100.

| | | $\beta_{1,3}$ | $\beta_{1,4}$ | $\beta_{1,5}$ | $\beta_{1,6}$ | $\beta_{2,3}$ | $\beta_{2,4}$ | $\beta_{2,5}$ | $\beta_{2,6}$ |
|---|---|---|---|---|---|---|---|---|---|
| | true | −0.20 | −0.20 | 0.20 | 0.20 | −0.50 | 0.20 | −0.20 | 0.20 |
| rMAVE | bias | −0.03 | 0.16 | −0.08 | −0.14 | −0.02 | 0.06 | −0.00 | −0.05 |
| | std | 2.41 | 2.21 | 2.19 | 2.30 | 2.70 | 2.61 | 2.46 | 2.53 |
| SEE | bias | 0.01 | 0.07 | −0.11 | −0.16 | 0.06 | 0.02 | 0.06 | 0.12 |
| | std | 1.04 | 1.72 | 2.46 | 5.10 | 2.33 | 1.62 | 2.17 | 1.18 |
| ECS | bias | −0.07 | −0.24 | 0.14 | 0.13 | −0.40 | 0.22 | −0.17 | 0.15 |
| | std | 1.28 | 1.08 | 1.13 | 1.17 | 1.30 | 1.15 | 1.22 | 1.25 |
| C2MS | bias | −3.65 | 8.38 | −8.68 | 0.03 | 3.78 | −9.04 | 8.90 | −0.13 |
| | std | 71.36 | 58.99 | 63.47 | 59.44 | 72.80 | 57.31 | 65.87 | 61.63 |
| EEE | bias | 0.09 | −0.25 | 0.09 | 0.02 | −0.36 | 0.18 | −0.13 | 0.17 |
| | std | 1.68 | 1.50 | 1.57 | 1.44 | 1.81 | 1.57 | 1.56 | 1.58 |
| | $\widehat{\text{std}}$ | 1.79 | 1.61 | 1.50 | 1.62 | 2.03 | 1.73 | 1.71 | 1.72 |
| | cp | 96.10 | 96.60 | 94.20 | 95.50 | 96.90 | 95.30 | 96.10 | 96.60 |

a standard normal, and generated $Y$ from the normal population with mean $m(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = (\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_1 + 1)(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_2 + 1)$ and standard deviation $\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = 0.5/\{0.1 + (\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_1)^2 + (\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_2)^2\}$. Here $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, $\boldsymbol{\beta}_1 = (1, 0, -0.2, -0.2, 0.2, 0.2)^{\mathrm{T}}$ and $\boldsymbol{\beta}_2 = (0, 1, -0.5, 0.2, -0.2, 0.2)^{\mathrm{T}}$. Thus, $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2)$, $\mathbf{B}_1 = (-0.2, -0.2, 0.2, 0.2)^{\mathrm{T}}$ and $\mathbf{B}_2 = (-0.5, 0.2, -0.2, 0.2)^{\mathrm{T}}$.

The simulations are repeated 1,000 times, and the results are summarized in Table 3. There the estimators obtained by solving (4.2) have smaller standard deviations than the rMAVE estimators. This is not surprising because the estimating equation estimator is actually efficient in this case. In comparison with ECS and SEE, our results are slightly worse than those of ECS, understandable since ECS imposes stronger assumption on the whole distribution. The performance of SEE is largely similar to that of ECS, although with worse results for several parameters. In comparison with C2MS, our estiimator performs much better. The estimated standard deviations of our method are close to the Monte Carlo standard deviations, and the empirical coverage probabilities are close to the nominal level 95%.

We further considered estimating $\boldsymbol{\beta}$ by pretending not to know that the central mean and the central variance subspaces are identical. Thus we only assume $d_\alpha = d_\beta = 1$ in model (1.2) and implement (2.4) to estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The re-

Table 4. The bias ("bias") and the sample standard errors ("std") for rMAVE , SEE, ECS, C2MS and the estimators obtained from solving (6), and the inference results, respectively, the average of the estimated standard deviation ("$\widehat{\text{std}}$") and the coverage of the estimated 95% confidence interval ("cp"), of our proposals. All numbers reported below are multiplied by 100.

| | | $\beta_{1,3}$ | $\beta_{1,4}$ | $\beta_{1,5}$ | $\beta_{1,6}$ | $\beta_{2,3}$ | $\beta_{2,4}$ | $\beta_{2,5}$ | $\beta_{2,6}$ |
|---|---|---|---|---|---|---|---|---|---|
| | true | −0.20 | −0.20 | 0.20 | 0.20 | −0.50 | 0.20 | −0.20 | 0.20 |
| rMAVE | bias | −0.01 | 0.17 | −0.08 | −0.16 | −0.02 | 0.10 | 0.01 | −0.06 |
| | std | 2.38 | 2.20 | 2.20 | 2.32 | 2.71 | 2.64 | 2.47 | 2.57 |
| SEE | bias | 0.01 | 0.07 | −0.11 | −0.16 | 0.06 | 0.02 | 0.06 | 0.12 |
| | std | 1.04 | 1.72 | 2.46 | 5.10 | 2.33 | 1.62 | 2.17 | 1.18 |
| ECS | bias | −0.07 | −0.24 | 0.14 | 0.13 | −0.40 | 0.22 | −0.17 | 0.15 |
| | std | 1.28 | 1.08 | 1.13 | 1.17 | 1.30 | 1.15 | 1.22 | 1.25 |
| C2MS | bias | −1.00 | 3.22 | −4.92 | −0.40 | 1.29 | −3.38 | 5.01 | 0.44 |
| | std | 14.94 | 16.07 | 28.60 | 24.97 | 14.99 | 16.34 | 29.02 | 26.08 |
| EEE | bias | 0.04 | 0.13 | −0.08 | −0.05 | −0.01 | 0.08 | −0.08 | −0.00 |
| | std | 1.68 | 1.56 | 1.64 | 1.73 | 1.97 | 1.85 | 1.73 | 1.82 |
| | $\widehat{\text{std}}$ | 1.52 | 1.47 | 1.44 | 1.46 | 1.82 | 1.68 | 1.66 | 1.68 |
| | cp | 93.50 | 94.20 | 93.60 | 92.80 | 92.80 | 93.20 | 93.50 | 92.80 |
| | | $\beta_{1,3}$ | $\beta_{1,4}$ | $\beta_{1,5}$ | $\beta_{1,6}$ | $\beta_{2,3}$ | $\beta_{2,4}$ | $\beta_{2,5}$ | $\beta_{2,6}$ |
| | true | −0.20 | −0.20 | 0.20 | 0.20 | −0.50 | 0.20 | −0.20 | 0.20 |
| rMAVE | bias | −0.01 | 0.63 | −0.90 | −0.51 | 1.07 | −1.21 | 1.02 | −0.26 |
| | std | 9.41 | 9.13 | 8.55 | 8.83 | 9.71 | 9.23 | 8.42 | 8.51 |
| SEE | bias | −0.06 | −1.22 | 0.36 | 0.61 | −0.28 | 0.26 | −0.23 | −0.23 |
| | std | 2.50 | 2.92 | 3.08 | 3.01 | 3.83 | 3.23 | 3.84 | 3.14 |
| ECS | bias | −0.07 | −0.24 | 0.14 | 0.13 | −0.40 | 0.22 | −0.17 | 0.15 |
| | std | 1.28 | 1.08 | 1.13 | 1.17 | 1.30 | 1.15 | 1.22 | 1.25 |
| C2MS | bias | −1.00 | 3.22 | −4.92 | −0.40 | 1.29 | −3.38 | 5.01 | 0.44 |
| | std | 14.94 | 16.07 | 28.60 | 24.97 | 14.99 | 16.34 | 29.02 | 26.08 |
| EEE | bias | 0.10 | −0.22 | 0.10 | −0.18 | −0.20 | 0.06 | 0.09 | 0.01 |
| | std | 4.86 | 4.10 | 4.13 | 4.07 | 4.28 | 4.39 | 4.13 | 3.91 |
| | $\widehat{\text{std}}$ | 5.42 | 4.77 | 4.91 | 4.86 | 5.33 | 4.92 | 4.89 | 4.95 |
| | cp | 94.60 | 95.40 | 95.40 | 95.10 | 95.70 | 95.50 | 96.00 | 96.00 |

sults are summarized in Table 4. It can be seen that here the estimating equation approach still yields consistent estimators of $\boldsymbol{\beta}$. In terms of the estimation bias, rMAVE and the estimating equation estimators are comparable, while in terms of the Monte Carlo standard deviations, the estimating equation estimators are clearly better. In comparison with ECS, SEE and C2MS, the same trend of relative performance is seen as in Table 3, while the difference is larger, especially in terms of estimating the variance parameters. The estimators obtained from (2.4) are not as efficient as those obtained from (4.2), which agrees with our

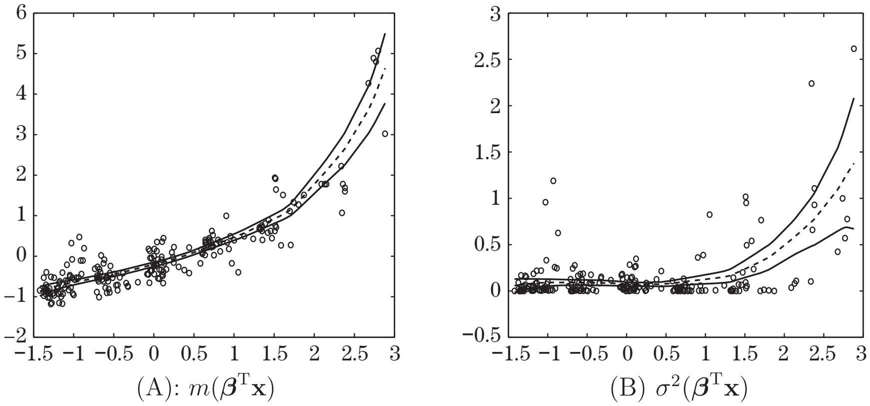


Figure 1. Scatter plot of $Y$ versus $(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x})$, with $\widehat{\boldsymbol{\beta}}$ estimated from (4.2). The dash lines are fitted curves and the solid lines are the 95% pointwise confidence intervals obtained from kernel regression. The $Y$-axis of the plots represents respectively $Y_i$ (left) and $\{Y_i - \widehat{m}(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)\}^2$ (right).

expectation as (4.2) utilizes more model assumptions than (2.4).

Following a referee's request, we also considered the case when the error term $\epsilon$ was the standard $t$ with $(\mathbf{x}^{\mathrm{T}}\mathbf{x}+4)$ degrees of freedom. We repeated the simulations 1,000 times and report the results in Table S2 in the supplement. There the oracle estimators yield smaller estimation biases and standard deviations than the locally efficient estimators. In Table S3 in the supplement, we provide the simulation results when we pretend not to know that the central mean and the central variance subspaces are identical. In this case, in terms of the estimation biases, the oracle efficient estimator is an obvious winner, while in terms of the standard deviations, the two estimators are comparable.

## 5.2. Bank data revisited

In Section 3, we have shown that the central mean and central variance subspaces are identical in the bank data. Here we revisit this data set by using (4.2) to estimate the parameters $\boldsymbol{\beta}$, a basis of the central second moment subspace (Yin and Cook (2002)). The resulting estimators, and their associated standard deviations are in the last block of Table 2. It again shows that there exists no gender or age effect, while the working experience and whether employee's job is computer related affect the salary significantly. Using the estimates $\widehat{\boldsymbol{\beta}}$ from (4.2), we show estimated mean and variance functions in Figure 1. The curves exhibit obvious increasing patterns, indicating the existence of heteroscedasticity.

## 6. Discussion

If the central mean and the central variance subspaces overlap but are not identical, to compare the efficiency of various estimators in estimating the common part and the difference of these two subspaces is a challenging problem.

An aspect that we have left out is how to decide the dimensions $d_\alpha$ and $d_\beta$. To this end, VIC proposed in Ma and Zhang (2015) can be applied directly to yield consistent estimation. Another approach to this issue is via bootstrap (Ye and Weiss (2003)). Under each candidate $(d_\alpha, d_\beta)$ value, we repeatedly estimate the corresponding subspaces using the bootstrap data and calculate the average correlation between the bootstrap data based subspaces and the original data based subspaces. The $(d_\alpha, d_\beta)$ combination that yields the largest correlation is then selected as the effective dimensions. Similar procedure can be carried out when the two subspaces are identical. Like all bootstrap based procedures, the computational cost of this procedure can be quite high, hence it is worth exploring alternative methods.

The efficiency of an estimator is dependent on the model assumptions, but sometimes the change is surprisingly large. For example, if we have assumed $\epsilon$ to be independent of $\mathbf{x}$, then the results can change quite dramatically. Under such model assumption, quantities such as $f_\epsilon'(\epsilon)/f_\epsilon(\epsilon)$ appear in the efficient score. Hence careful analysis is always needed in deriving efficient estimators, even if the model assumption changes a little. Following this line, if we further assume $\boldsymbol{\beta} = \boldsymbol{\alpha}$, then the central mean, the central variance and the central space unify into the same space spanned by $\boldsymbol{\beta}$. This model has much more structure than the model considered in Ma and Zhu (2013b), hence the efficient result derived there does not apply.

## Supplementary Materials

The regularity conditions, proofs of our main results and some additional simulations can be found in an online supplementary document.

## Acknowledgment

MA's work was supported by grants from National Science Foundation (DMS-1608540) and National Institute of Neurological Disorders and Strokes (R01-NS073671). ZHU's work is supported by grants from National Natural Science Foundation of P. R. China (11371236, 11422107 and 11731011), the Ministry of Education Project of Key Research Institute of Humanities and Social Sciences at Universities (16JJD910002) and National Youth Top-notch Talent Support Program, P. R. China.

## References

Albright, S. C., Winston, W. L. and Zappe, C. J. (1999). Data analysis and decision making with microsoft excel. *Pacific Grove, CA, Duxbury.*

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

Box, G. and Hill, W. (1974). Correcting inhomogeneity of variance with power transformation weighting. *Technometrics* **16**, 385–389.

Box, G. and Meyer, D. (1986). An analysis for unreplicated fractional factorials. *Technometrics* **28**, 11–18.

Cai, T. and Wang, L. (2008). Adaptive variance function estimation in heteroscedastic nonparametric regression. *The Annals of Statistics* **36**, 2025–2054.

Cai, T., Levine, M. and Wang, L. (2009). Variance function estimation in multivariate nonparametric regression. *Journal of Multivariate Analysis* **100**, 126–136.

Carroll, R. J. (2003). Variances are not always nuisance parameters. *Biometrics* **59**, 211–220.

Carroll, R. and Ruppert, D. (1988). *Transformation and Weighting in Regression.* Chapman & Hall: New York.

Cook, R. D. and Li, B. (2002). Dimension reduction for conditional mean in regression. *The Annals of Statistics* **30**, 455–474.

Davidian, M. and Carroll, R. J. (1987). Variance function estimation. *Journal of the American Statistical Association* **82**, 1079–1091.

Davidian, M., Carroll, R. J. and Smith, W. (1988). Variance functions and the minimum detectable concentration in assays. *Biometrika* **75**, 549–556.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* **58**, 71–120.

Li, K. C. (1992). On principal Hessian directions for data visualization and dimension reduction: another application of Stein's lemma. *Journal of the American Statistical Association* **87**, 1025–1039.

Li, K. C. and Duan, N. (1989). Regression analysis under link violation. *The Annals of Statistics* **17**, 1009–1052.

Luo, W., Li, B. and Yin, X. (2014). On efficient dimension reduction with respect to a statistical functional of interest. *The Annals of Statistics* **42**, 382–412.

Lian, H., Liang, H. and Carroll, R. (2014). Variance function partially linear single-index models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77**, 171–194.

Ma, Y. and Carroll, R. J. (2006). Locally efficient estimators for semiparametric models with measurement error. *Journal of the American Statistical Association* **101**, 1465–01474.

Ma, Y and Zhang, X. (2015). A validated information criterion to determine the structural dimension in dimension reduction models. *Biometrika* **102**, 409–420.

Ma, Y. and Zhu, L. P. (2012). A semiparametric approach to dimension reduction. *Journal of the American Statistical Association* **107**, 168–179.

Ma, Y. and Zhu, L. P. (2013a). Efficiency loss caused by linearity condition in dimension reduction. *Biometrika* **100**, 371–383.

Ma, Y. and Zhu, L. P. (2013b). Efficient estimation in sufficient dimension reduction. *The Annals of Statistics* **41**, 250–268.

Ma, Y. and Zhu, L. P. (2014). On estimation efficiency of the central mean subspace. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **76**, 885–901.

Meyer, J. (1987). Two-moment decision models and expected utility maximization. *American Economic Review* **77**, 421–430.

Teschendorff, A. E. and Widschwendter, M. (2012). Differential variability improves the identification of cancer risk markers in DNA methylation studies profiling precursor cancer lesions. *Bioinformatics* **28**, 1487–1494.

Thomas, L., Stefanski, L. A. and Davidian, M. (2012). Measurement error model methods for bias reduction and variance estimation in logistic regression with estimated variance predictors. Tech. rep., North Carolina State University.

Tong, T., Ma, Y. and Wang, Y. (2013). Optimal variance estimation without estimating the mean function. *Bernoulli* **19**, 1839–1854.

Tong, T. and Wang, Y. (2005). Estimating residual variance in nonparametric regression using least squares. *Biometrika* **92**, 821–830.

Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data*. Springer: New York.

Tsiatis, A. A. and Ma, Y. (2004). Locally efficient semiparametric estimators for functional measurement error models. *Biometrika* **91**, 835–848.

Western, B. and Bloome, D. (2009). Variance function regressions for studying inequality. *Sociological Methodology* **39**, 293–326.

Xia, Y., Tong, H., Li, W. K. and Zhu, L. X. (2002). An adaptive estimation of dimension reduction space (with discussion). *Journal of the Royal Statistical Society Series B. (Statistical Methodology)* **64**, 363–410.

Ye, Z. and Weiss, R. E. (2003). Using the bootstrap to select one of a new class of dimension reduction methods. *Journal of the American Statistical Association* **98**, 968–979.

Yin, X. and Cook, D. (2002). Dimension reduction for the conditional $k$th moment in regression. *Journal of the Royal Statistical Society. Series B. (Statistical Methodology)* **64**, 159–175.

Zhu, L. P., Dong, Y. X. and Li, R. (2013). Semiparametric estimation of conditional heteroscedasticitity via single-index modeling. *Statistica Sinica* **23**, 1235–1255.

Zhu, L. P. and Zhu, L. X. (2009). Dimension-reduction for conditional variance in regressions. *Statistica Sinica* **19**, 869–883.

Department of Statistics, Penn State University, University Park, PA 16802, USA.

E-mail: yzm63@psu.edu

Institute of Statistics and Big Data and Center for Applied Statistics, Renmin University of China, 59 Zhongguancun Avenue, Haidian District, Beijing 100872, P. R. China.

E-mail: zhu.liping@ruc.edu.cn

(Received February 2017; accepted November 2017)