Contents lists available at ScienceDirect

# Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

# A note on cumulative mean estimation

# Bilin Zeng<sup>a</sup>, Zhou Yu<sup>b,\*</sup>, Xuerong Meggie Wen<sup>c</sup>

<sup>a</sup> Department of Mathematics, California State University, Bakersfield, CA, USA

<sup>b</sup> School of Finance and Statistics, East China Normal University, Shanghai, China

<sup>c</sup> Department of Mathematics and Statistics, Missouri University of Science and Technology, MO, USA

#### ARTICLE INFO

#### Article history: Received 23 January 2014 Received in revised form 11 September 2014 Accepted 12 September 2014 Available online 26 September 2014

Keywords: Cumulative slicing estimation Dimension reduction Sliced inverse regression Intraslice covariance estimation Ensemble estimator approach

#### 1. Introduction

# ABSTRACT

We revisit *cumulative slicing estimation* (CUME; Zhu et al., 2010) from a different perspective to gain more insights, then refine its performance by incorporating the intra-slice covariances. We also prove that our new method, under some conditions, is more comprehensive than CUME.

© 2014 Elsevier B.V. All rights reserved.

For a typical regression problem with a univariate random response *Y* and a *p*-dimensional random vector **X**, sufficient dimension reduction (SDR: Li, 1991; Cook, 1998) aims to reduce the dimension of **X** without loss of information on the regression and without requiring a pre-specified parametric model. The basic idea of sufficient dimension reduction is to replace the predictors  $\mathbf{X} \in \mathbb{R}^p$  with a *d*-dimensional linear vector  $\boldsymbol{\eta}^T \mathbf{X}$ , where  $\boldsymbol{\eta}$  is a  $p \times d$  matrix with  $d \leq p$ , such that  $Y \perp \mathbf{X} \mid \boldsymbol{\eta}^T \mathbf{X}$ . The column space of  $\boldsymbol{\eta}$  is called a dimension reduction subspace, and the intersection of all such subspaces is called the *central subspace*, denoted by  $\mathscr{S}_{Y|\mathbf{X}}$ . The dimension *d* of  $\mathscr{S}_{Y|\mathbf{X}}$  is called the structural dimension of the regression. Under mild conditions (Yin et al., 2008), the central subspace exists and is unique. We assume that  $\mathscr{S}_{Y|\mathbf{X}}$  exists throughout this article. The goal of sufficient dimension reduction is to estimate and make statistical inferences about  $\mathscr{S}_{Y|\mathbf{X}}$  and *d*.

Many methods have been developed to estimate  $\delta_{Y|X}$ . Among them, sliced inverse regression (SIR; Li, 1991), sliced average variance estimation (SAVE; Cook and Weisberg, 1991), minimum average variance estimation (MAVE; Xia et al., 2002), and directional regression (DR; Li and Wang, 2007) are perhaps the most widely investigated methods in the literature. Cook and Li (2002) proposed the central mean subspace where the interest of dimension reduction is restricted to the conditional mean function E(Y|X), and the central mean subspace  $\delta_{E(Y|X)}$  is defined as the smallest column spaces spanned by  $\eta$  such that  $E(Y|X) \perp X | \eta^T X$ .

For many-valued or continuous *Y*, the standard practice in SDR is to replace the response *Y* with a discrete version  $\check{Y}$  by partitioning the range of *Y* into *h* non-overlapping slices, then work on  $\check{Y}$  and assume that  $\delta_{\check{Y}|\mathbf{X}} = \delta_{Y|\mathbf{X}}$ . However, this assumption does not always hold, and the difference between the working and target regressions can be significant. Moreover, even when this assumption holds, we might still face the loss of power since we use only the information retained in  $\check{Y}$ , discarding all the intra-slice information.

http://dx.doi.org/10.1016/j.spl.2014.09.014 0167-7152/© 2014 Elsevier B.V. All rights reserved.







<sup>\*</sup> Corresponding author. E-mail addresses: yz19830224@gmail.com, zyu@stat.ecnu.edu.cn (Z. Yu).

The number of slices h is a tuning parameter much like the tuning parameter encountered in the smoothing literature (Li, 1987; Härdle et al., 1988). Experience indicates that good results are often obtained by choosing h to be somewhat larger than d + 1, and sometimes requires several trials of h. However, beyond empirical experiences, how to select the optimal h still remains unknown in the literature.

Zhu et al. (2010a,b) proposed a method called the *cumulative slicing estimation* (CUME) which sums up all possible estimations relating to  $E(\mathbf{X}I(Y \leq \tilde{y}))$  for all  $\tilde{y}$  in the support of Y to avoid the subjective selection of h. They showed that the estimator of CUME enjoys the common  $\sqrt{n}$  convergence rate and is more efficient comparing to SIR and other first-moment slicing estimation methods.

In this article, we first revisit CUME from a different perspective to gain more insights, and then refine its performance by incorporating the intra-slice covariances. The rest of this article is organized as follows. In Section 2, we reinterpret CUME via the ensemble estimator approach (Yin and Li, 2011). Our new estimation method, which we call the *covariance cumulative slicing estimation* (COCUM), along with its asymptotic properties are investigated in Section 3. We illustrate the performances of our method via simulation studies in Section 4. A brief summary is given in Section 5. For easy of exposition, we defer all proofs to the Appendix.

# 2. CUME: the ensemble estimator approach

## 2.1. CUME

In this subsection, we give a brief review of CUME (Zhu et al., 2010a,b). For ease of exposition, we assume hereafter that  $E(\mathbf{X}) = 0$ . Define

$$\mathbf{m}(\tilde{y}) = \mathbf{E}(\mathbf{X}I(Y \le \tilde{y})) \tag{2.1}$$

for  $\tilde{y} \in \mathbb{R}^1$ . To preserve the integrity of  $\mathscr{S}_{Y|X}$ , let  $\tilde{Y}$  be an independent copy of Y and the kernel matrix for CUME be defined by:

$$\mathbf{M} = \mathbf{E}[\mathbf{m}(\tilde{\mathbf{Y}})\mathbf{m}^{T}(\tilde{\mathbf{Y}})w(\tilde{\mathbf{Y}})], \tag{2.2}$$

where w(.) is a nonnegative weight function which is often set as 1.

Assuming the linearity condition (see Feng et al., 2013 more details will be given in Section 2.2), which is a mild condition imposed on the marginal distribution of **X** only, the column space of **M** is a subset of  $\Sigma \mathscr{S}_{Y|X}$ , where  $\Sigma = \text{Cov}(X)$ . At the sample level, suppose  $(\mathbf{X}_i, Y_i)$ , i = 1, ..., n are independent copies of  $(\mathbf{X}, Y)$ , we can estimate **M** by  $\mathbf{M}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbf{m}_n(Y_i) \mathbf{m}_n^T(Y_i) w(Y_i)$ , where  $\mathbf{m}_n(Y_i) = \frac{1}{n} \sum_{j=1}^{n} (\mathbf{X}_j - \bar{\mathbf{X}}) I(Y_j \le Y_i)$ , and  $\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{X}_i$ . Let  $\hat{\boldsymbol{\Sigma}} = \frac{1}{n} \sum_{i=1}^{n} (\mathbf{X}_i - \bar{\mathbf{X}}) (\mathbf{X}_i - \bar{\mathbf{X}})^T$  be the sample predictor variance, assuming a known *d*, the CUME estimator of  $\mathscr{S}_{Y|\mathbf{X}}$  is constructed by the *d* eigenvectors of  $\hat{\boldsymbol{\Sigma}}^{-1} \mathbf{M}_n$  corresponding to its *d* largest eigenvalues. Zhu et al. (2010a,b) studied the asymptotic properties of the CUME estimator, and also provided a data-driven method on the determination of *d*.

#### 2.2. The ensemble estimator approach

In this section, we revisit CUME via the ensemble estimator approach (Yin and Li, 2011). The main result of Yin and Li (2011) is summarized by the following lemma.

**Lemma 1** (*Yin and Li, 2011*). Let  $\mathscr{J}$  be a family of functions  $f : \Omega_Y \to \mathbb{F}$ , where  $\mathbb{F}$  can be the set of real or complex numbers; let  $F_Y$  be the distribution function of Y,  $L_2(F_Y)$  be the class of functions f(Y) with finite variances and  $(f_1, f_2) = \mathbb{E}[f_1(Y)f_2(Y)]$  as the inner product. Let  $\mathscr{S}_{\mathbb{E}[f(Y)|\mathbf{X}]}$  be the central mean subspace for the conditional mean  $\mathbb{E}[f(Y)|\mathbf{X}]$ , as defined in Cook and Li (2002). If  $\mathscr{J}$  is a subset of  $L_2(F_Y)$  that is dense in  $\mathscr{B}$ , where  $\mathscr{B} = \{I_B : B \text{ is a Borel set in } \Omega_Y\}$ , then we have:

$$\operatorname{Span}\{\delta_{[\mathrm{E}(f(Y)|\mathbf{X})]}: f \in \mathscr{J}\} = \delta_{Y|\mathbf{X}}.$$
(2.3)

Hence, for a sufficiently rich family of f(Y), the conditional mean subspaces  $\delta_{[E(f(Y)|X)]}$ , when put together, can recover the central subspace  $\delta_{Y|X}$ . Yin and Li (2011) showed that both the Fourier transformation method proposed by Zhu and Zeng (2006), and the sliced regression (SR) proposed by Wang and Xia (2008) are special examples of the above ensemble estimators. Specifically, Zhu and Zeng (2006) used  $\mathscr{J} = \{f_t(y) = e^{ity} : t \in \mathbb{R}^1\}$ , where  $\iota$  is the imaginary unit; while Wang and Xia (2008) used  $\mathscr{J} = \{f_t(y) = I_{(-\infty,t)}(y) : t \in \mathbb{R}^1\}$ , and in practice taking those slicing points as the values for t, then use MAVE (Xia et al., 2002) to estimate the individual central mean subspace  $\delta_{E[f_t(Y)|X]}$ .

Assuming the following two conditions which are commonly used in the sufficient dimension reduction literature (Cook, 2004):

(a) linearity condition:  $E(\mathbf{X}|\boldsymbol{\eta}^T\mathbf{X}) = \mathbf{P}_{\boldsymbol{\eta}}^T(\boldsymbol{\Sigma})\mathbf{X}$ , where  $\mathbf{P}_{\boldsymbol{\eta}}(\boldsymbol{\Sigma}) = \boldsymbol{\eta}(\boldsymbol{\eta}^T\boldsymbol{\Sigma}\boldsymbol{\eta})^{-1}\boldsymbol{\eta}^T\boldsymbol{\Sigma}$ .

(b) coverage condition: Span( $\Sigma^{-1}\mathbf{m}(t)$ ) =  $\mathscr{S}_{E[f_t(Y)|\mathbf{X}]}$ .

We are now ready to demonstrate that CUME also belongs to the family of the ensemble estimators. Let  $\mathscr{J} = \{f_t(y) = I_{(-\infty,t)}(y) : t \in \mathbb{R}^1\}$ . The population coefficient vector from the ordinary least squares fit of  $f_t(Y)$  on **X** is  $\eta_t = \Sigma^{-1} E(f_t(Y) \mathbf{X}) = \Sigma^{-1} m(t)$  without the intercept term. Following the result of Duan and Li (1991), it is not hard to prove that, assuming condition (a), the above OLS coefficient falls into the central mean subspace  $\mathscr{S}_{E[f_t(Y)|\mathbf{X}]}$ . The following proposition concludes our discussion of CUME in this section.

**Proposition 1.** Let  $\mathscr{J} = \{f_t(y) = I_{(-\infty,t)}(y) : t \in \mathbb{R}^1\}$ , then assuming conditions (a) and (b),  $\Sigma^{-1}m(t) = \mathscr{S}_{E[f_t(Y)|\mathbf{X}]}$ , and the column space of  $\hat{\Sigma}^{-1}\mathbf{M}_n$  provides a consistent estimator of  $\mathscr{S}_{Y|\mathbf{X}}$ .

According to the above proposition, we can see that CUME is also a special example of the family of the ensemble estimators.

## 3. Covariance cumulative slicing estimation (COCUM)

#### 3.1. The method

As Cook and Ni (2006) pointed out, the use of  $\eta_t$  discards the intra-slice information which might result in a loss of power. Instead, in this section, we propose a method called the *covariance cumulative slicing estimation* (COCUM) which incorporates the intra-slice information into the estimation of the central subspace  $\vartheta_{Y|X}$ .

To recover the intra-slice covariance information, following Cook and Ni (2006), we take

$$\mathbf{m}_{c}(\tilde{y}) = \mathrm{E}(\mathbf{X}YI(Y \leq \tilde{y}))$$

Denote  $p_t = 2$   $E(\mathbf{x} H(T \le t))$ , it is easy to show that  $p_t$  can be decomposed as  $F_t 2$   $Cov(\mathbf{x}, T) (T \le t) = 1$  $1) + E(Y|I(Y \le t) = 1)\eta_t$ , where  $F_t = P(Y \le t)$ .

**Theorem 1.** Assuming condition (a), then  $\boldsymbol{\beta}_t \in \mathscr{S}_{Y|\mathbf{X}}$ , for  $t \in \mathbb{R}^1$ . Furthermore, assuming condition (b) and a similar coverage condition  $\text{Span}(\boldsymbol{\Sigma}^{-1}\mathbf{m}_c(t)) = S_{\text{E}[g_t(Y)|\mathbf{X}]}$ , where  $g_t(y) = yI(y \leq t)$ , then, comparing with CUME, the column space of  $\boldsymbol{\Sigma}^{-1}\mathbf{M}_c$  always encloses that of  $\boldsymbol{\Sigma}^{-1}\mathbf{M}$ .

Theorem 1 suggests that theoretically COCUM always wins over CUME under some conditions, since it recovers more of the central subspace.

#### 3.2. The asymptotic properties

The following theorem shows that COCUM possesses the same asymptotic property as CUME. Zhu et al. (2006) derived the strong consistency for the slicing estimation of the SIR matrix when  $p = o(n^{1/4})$ . However, the results from both CUME and COCUM are better than Zhu et al. (2006). The proof is similar to that of Theorem 2 of CUME (Zhu et al., 2010a,b) and is omitted.

**Theorem 2.** Let  $X_i$  be the *i*th coordinate of **X**, suppose  $\max_{1 \le i \le p} E(X_i^8 Y^8) < \infty$  uniformly for *p*, and then

$$\|\boldsymbol{\Sigma}_n^{-1}\mathbf{M}_n^c - \boldsymbol{\Sigma}^{-1}\mathbf{M}_c\| = o(pn^{-1/2}\log n)$$

almost surely where  $\|\cdot\|$  is the Frobenius norm, where  $\mathbf{M}_n^c = \frac{1}{n} \sum_{i=1}^n \mathbf{m}_c(Y_i) \mathbf{m}_c^T(Y_i) w(Y_i)$ .

Let  $\tilde{Y}$  be an independent copy of Y,  $T(\mathbf{X}, \mathbf{Y}) = \Sigma^{-1} \{ \mathbf{X} \mathbf{X}^T - E\mathbf{X} \mathbf{X}^T - (\mathbf{X} - E\mathbf{X})^T + E\mathbf{X} (\mathbf{X} - E\mathbf{X})^T \} \Sigma^{-1} \mathbf{M}_c - \Sigma^{-1} [2\mathbf{m}_c(Y)\mathbf{m}_c^T(Y)\omega(Y) + 2E\{ \mathbf{X}YI(Y \leq \tilde{Y})\mathbf{m}_c^T(\tilde{Y})\omega(\tilde{Y}) | \mathbf{X}, Y \} + 2E\{ \mathbf{m}_c(\tilde{Y})\mathbf{X}^T Y^T I(Y \leq \tilde{Y})\omega(\tilde{Y}) | \mathbf{X}, Y \} - 6\mathbf{M}_c ]$ . The following theorem states the asymptotic normality of our estimator which is similar to that of Theorem 3 of Zhu et al. (2010a,b). The proof is also omitted.

Theorem 3. Assuming the following regularity conditions:

1.  $\max_{1 \le i \le p} E(X_i^8 Y^8) < \infty$  uniformly for p, 2. The minimum eigenvalue of  $\Sigma$  satisfies  $\lambda_{\min}(\Sigma) > 0$ , 3. The largest eigenvalue of  $\mathbf{M}_c$  satisfies  $\lambda_{\max}(\mathbf{M}_c) < \infty$  holds uniformly for p, 4.  $E\{\gamma^T T(\mathbf{X}, Y)\gamma\} \rightarrow G > 0$  for any unit length  $\gamma$ , 5.  $p = o(n^{1/2})$ ; then

$$\sqrt{n}\gamma^{T}(\boldsymbol{\Sigma}_{n}^{-1}\mathbf{M}_{n}-\boldsymbol{\Sigma}^{-1}\mathbf{M})\gamma \stackrel{\mathcal{D}}{\longrightarrow} N(0,G)$$

#### 3.3. The determination of d

One of the goals for sufficient dimension reduction is to estimate the structural dimension. Many methods have been developed to determine the structural dimension, such as Li (1991), Schott (1994), Bura and Cook (2001), Zhu et al. (2006)

Table 4.1Ratio average (standard error) for Model I.

	Method	<i>n</i> = 200		<i>n</i> = 400		n = 600	
		Ratio	Deviation	Ratio	Deviation	Ratio	Deviation
<i>p</i> = 10	CUME	0.71	0.22	0.80	0.18	0.85	0.14
	COCUM	0.75	0.19	0.84	0.15	0.88	0.11
<i>p</i> = 15	CUME	0.66	0.19	0.74	0.17	0.79	0.14
	COCUM	0.69	0.18	0.78	0.14	0.83	0.12
<i>p</i> = 20	CUME	0.63	0.16	0.71	0.16	0.76	0.14
	COCUM	0.64	0.16	0.74	0.14	0.79	0.12

and Zhu et al. (2010a,b). Following Zhu et al. (2010a,b), we use a modified BIC-type method for COCUM. Define

$$G(k) = n \sum_{i=1}^{k} \lambda_{ni}^{2} / \sum_{i=1}^{p} \lambda_{ni}^{2} - C_{n}k(k+1)/2$$
(3.2)

where  $\hat{\lambda}_{n1} \geq \hat{\lambda}_{n2} \geq \cdots \geq \hat{\lambda}_{np}$  are the sample eigenvalues of the kernel matrix. And the estimated dimension  $\hat{K}$  is defined as

$$\tilde{K} = \arg \max_{1 \le k \le p} G(k).$$
(3.3)

As pointed out in Zhu et al. (2010a,b), smaller value of the penalty constant  $C_n$  tends to overestimate the dimension d; while larger  $C_n$  tends to underestimate the dimension d. A data-driven manner is needed to choose an appropriate value for  $C_n$  under a certain method. We choose  $C_n = \frac{1}{2} \log(n)$  which mostly leads to satisfactory results. The following theorem states the consistency of  $\hat{K}$ . The proof is similar to that of Theorem 6 of Zhu et al. (2010a,b) and is omitted.

**Theorem 4.** Assuming the conditions in Theorem 2, if  $\frac{C_n}{n} \to 0$  and  $C_n \to \infty$ , as  $n \to \infty$ , the estimated structural dimension  $\hat{K}$  obtained via (3.3) converges to the true structural dimension d in probability.

# 4. Simulation studies

In this section, we compare the performance of COCUM with CUME. We considered two different models with the design matrix generated from normal and Gamma distributions. To evaluate the performance of different methods, following Li and Dong (2009), we use the ratio of the square multiple correlation coefficient to the dimension *d*:

$$\frac{\rho^2}{d} = \frac{\operatorname{trace}\{(\hat{\boldsymbol{\beta}}^T \boldsymbol{\Sigma} \hat{\boldsymbol{\beta}})^{-1} (\hat{\boldsymbol{\beta}}^T \boldsymbol{\Sigma} \hat{\boldsymbol{\beta}}) (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \hat{\boldsymbol{\beta}}) (\boldsymbol{\beta}^T \boldsymbol{\Sigma} \hat{\boldsymbol{\beta}})^{-1}\}}{d}$$
(4.1)

as evaluation measurements. Here  $\hat{\beta}$  is the estimate for true  $\beta$  and  $\Sigma$  is the covariance matrix for predictor **X**.  $\rho^2$  gets closer to the true dimension d if  $\hat{\beta}^T \mathbf{X}$  and  $\beta^T \mathbf{X}$  have a linear relation and it is 0 if they are uncorrelated. Hence, the closer the ratio is to 1, the better fit of the model. In our simulation studies, the average ratios from 1000 simulation runs are reported. In addition, the frequencies of the estimated structural dimension over 1000 trials are also given to illustrate the performance of the modified BIC method. For CUME, we follow the suggestion of Zhu et al. (2010a,b) and take  $C_n = 2n^{3/4}/p$ . All numbers reported in the frequency table are multiplied by 10.

*Model* I. We first consider a two dimensional model (d = 2) with predictors  $\mathbf{X} = (X_1, X_2, X_3, \dots, X_p)$  generated from  $N(0, I_p)$ , and standard normal  $\epsilon$ . The univariate response Y is constructed as:

$$Y = 1.5(5 + X_1)(2 + X_2 + X_3) + 0.5\epsilon.$$
(4.2)

In this case,  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$  where  $\boldsymbol{\beta}_1 = (1, 0, \dots, 0)^T$  and  $\boldsymbol{\beta}_2 = (0, 1, 1, 0, \dots, 0)^T$ . As shown in Table 4.1, in terms of the average ratios, COCUM slightly outperforms CUME, while the performances of both methods deteriorate as *p* gets larger. Table 4.2 suggests that CUME has serious problem of underestimating the structural dimension *d* = 2.

*Model* II. We now consider a model with the predictor  $\mathbf{X} = (X_1, X_2, X_3, \dots, X_p)$  generated from *Gamma*(0.1, 10), and standard normal  $\epsilon$ . The univariate response *Y* is constructed as:

$$Y = (10 + (X_2 + 0.5)^2)\sqrt{(0.25 + X_1)} + 0.5\epsilon.$$
(4.3)

In this case,  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$  where  $\boldsymbol{\beta}_1 = (1, 0, \dots, 0)^T$  and  $\boldsymbol{\beta}_2 = (0, 1, 0, \dots, 0)^T$ . Table 4.3 provides the averages of the ratios and standard errors of the averages from 1000 simulation runs for Model II. It is obvious that COCUM wins over CUME. With larger sample size (n = 400 or 600), the average ratios are well above 0.9 even when p = 20. Table 4.4 again shows that CUME tends to underestimate the structural dimension, while COCUM does a good job most of the time.

Table 4.2				
Dimension	estimate	for	Mode	ΙI

- . . . . .

п	d = 1		<i>d</i> = 2		<i>d</i> > 2				
	CUME	COCUM	CUME	COCUM	CUME	COCUM			
200	100	1	0	98.5	0	0.5			
400	100	0.3	0	99.7	0	0			
600	100	0.2	0	99.8	0	0			
200	100	0	0	82.9	0	17.1			
400	100	0	0	98.1	0	1.9			
600	100	0	0	99.3	0	0.7			
400	100	0	0	74	0	26			
600	100	0	0	91	0	9			
	n 200 400 600 200 400 600 400 600	$\begin{array}{c} n & \frac{d=1}{\text{CUME}} \\ \hline \\ 200 & 100 \\ 400 & 100 \\ 600 & 100 \\ 200 & 100 \\ 400 & 100 \\ 600 & 100 \\ 400 & 100 \\ 600 & 100 \\ \hline \end{array}$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$			

#### Table 4.3

Ratio average (standard error) for Model II.

	Method	<i>n</i> = 200		<i>n</i> = 400		n = 600	
		Ratio	Deviation	Ratio	Deviation	Ratio	Deviation
<i>p</i> = 10	CUME	0.52	0.16	0.57	0.25	0.63	0.32
	COCUM	0.93	0.15	0.96	0.08	0.98	0.04
<i>p</i> = 15	CUME	0.49	0.08	0.51	0.11	0.54	0.19
	COCUM	0.89	0.20	0.94	0.10	0.96	0.07
<i>p</i> = 20	CUME	0.48	0.04	0.49	0.07	0.51	0.10
	COCUM	0.85	0.22	0.92	0.14	0.95	0.08

#### Table 4.4

Dimension estimates for Model II.

р	п	d = 1		d = 2	<i>d</i> = 2		<i>d</i> > 2	
		CUME	COCUM	CUME	COCUM	CUME	COCUM	
10	200	100	8.3	0	91	0	0.7	
10	400	100	1.8	0	97.2	0	1	
10	600	100	0.3	0	99	0	0.7	
15	200	100	3.6	0	90.1	0	6.3	
15	400	100	0.8	0	95.4	0	3.8	
15	600	100	0.1	0	95.9	0	4	
20	400	100	0.4	0	91.2	0	8.4	
20	600	100	0.1	0	94.3	0	5.6	

## 5. Conclusion

Zhu et al. (2010a,b) proposed a method called the *cumulative slicing estimation* (CUME) which sums up all possible estimations relating to  $E(XI(Y \le \tilde{y}))$  for all  $\tilde{y}$  in the support of Y to avoid the otherwise subjective selection of h. They showed that the estimator of CUME enjoys the common  $\sqrt{n}$  convergence rate and is more efficient comparing to SIR and other first-moment slicing estimation methods. In this article, we revisit CUME from the ensemble estimator approach (Yin and Li, 2011) which helps us to gain more insights of CUME. We then propose a new method, which we call COCUM which incorporates the intra-slice information into the estimation of the central subspace. The asymptotic properties of COCUM are also discussed.

Our method can also be applied to multi-dimensional responses. For multi-response **Y**, we first adapt the projective resampling method (Li et al., 2008) to our proposed COCUM. Suppose  $\mathbf{Y} \in \mathbb{R}^q$  and let **U** be a random vector uniformly distributed on the unit sphere  $S^q$ . Then we can define  $\mathbf{m}(\mathbf{U}, \mathbf{y}) = E\{\mathbf{X}\mathbf{U}^T\mathbf{Y}I(\mathbf{U}^T\mathbf{Y} \leq \mathbf{U}^T\mathbf{y})\}$  and construct the candidate matrix  $\mathbf{M} = E\{\mathbf{m}(\mathbf{U}, \tilde{\mathbf{Y}})\omega(\tilde{\mathbf{Y}})\}$ . The column space spanned by  $\mathbf{\Sigma}^{-1}\mathbf{M}$  is a subspace of  $S_{\mathbf{Y}|\mathbf{X}}$ .

We can further adapt the idea of COCUM to principle Hessian direction (Li, 1992). Define  $\mathbf{m}(y) = E(\mathbf{ZZ}^T YI(Y \le y))$ , and then we can adopt  $\mathbf{M} = E\{\mathbf{m}(\tilde{Y})\mathbf{m}^T(\tilde{Y})\omega(\tilde{Y})\}$  as the candidate matrix as a second order method, which is a potential competitor of CUME–SAVE or CUME–DR.

In our opinion, the ensemble estimator approach (Yin and Li, 2011) sheds lights on the developments of many existing sufficient dimension reduction methods such as SIR (Li, 1991), *K*th moment method (Yin and Cook, 2002), Fourier transformation method (Zhu and Zeng, 2006), SR (Wang and Xia, 2008). It is of special interest to study how this approach can be adopted to help developing new dimension reduction methods with functional data.

# Acknowledgments

Zhou Yu is supported by NSFC grants (No. 11201151) and the 111 project (B14019).

## Appendix

**Proof of Proposition 1.** Since  $f_t(y) = I_{(-\infty,t)}(y)$ ,  $\Sigma^{-1}m(t)$  is the population coefficient vector from the ordinary least square fit of  $f_t(Y)$  on **X** without the intercept term. By Theorem 1 of Cook and Li (2002), we can show that  $\Sigma^{-1}m(t) \in \mathscr{S}_{\text{Elf}_t(Y)|X|}$ assuming condition (a). Take one step further, assuming condition (b), we then have the equality, i.e., Span( $\Sigma^{-1}\mathbf{m}(t)$ ) =  $\mathscr{S}_{E[f_t(Y)|\mathbf{X}]}$ 

Since  $\mathscr{J} = \{f_t(y) = I_{(-\infty,t)}(y) : t \in \mathbb{R}^1\}$  is dense in  $L_2(F_Y)$ , by Lemma 1, we have  $\text{Span}\{\delta_{[E(f_t(Y)|\mathbf{X})]} : f \in \mathscr{J}\} = \delta_{Y|\mathbf{X}}$ . Let  $\hat{\mathbf{m}}(t)$  denote the corresponding sample estimator of  $\mathbf{m}(t)$ , for  $t = Y_1, \ldots, Y_n$ . Hence  $\hat{\mathbf{m}}(Y_i) = \mathbf{m}_n(Y_i)$ . By Theorem 2.2 of Yin and Li (2011), we have Span( $\Sigma^{-1} \sum_{i=1}^{n} \mathbf{m}(Y_i) \mathbf{m}^T(Y_i) w(Y_i) = \delta_{Y|X}$ , a.s. Since Span( $\hat{\Sigma}^{-1} \hat{\mathbf{m}}(t)$ ) is a consistent estimator of Span( $\Sigma^{-1}\mathbf{m}(Y_i)$ ), for  $t = Y_1, \ldots, Y_n$ , Span( $\hat{\Sigma}^{-1}\mathbf{M}_n$ ) is also a consistent estimator of  $\mathscr{J}_{Y|\mathbf{X}}$ .

**Proof of Theorem 1.** Let  $f_t(y) = I(y \le t)$ , and  $g_t(y) = yI(y \le t)$ , then  $\Sigma^{-1}\mathbf{m}(t) = \Sigma^{-1} \operatorname{Cov}\{f_t(y), \mathbf{X}\}$ , and  $\Sigma^{-1}\mathbf{m}_c(t) = \Sigma^{-1} \operatorname{Cov}\{f_t(y), \mathbf{X}\}$ , and  $\Sigma^{-1}\mathbf{m}_c(t) = \Sigma^{-1} \operatorname{Cov}\{f_t(y), \mathbf{X}\}$ .  $\Sigma^{-1}$  Cov $\{g_t(y), X\}$ . Following similar arguments of Theorem 1 of Cook and Li (2002), we can show that  $\Sigma^{-1}\mathbf{m}(t) \in S_{f_t(Y)|X}$ . and  $\Sigma^{-1}\mathbf{m}_{c}(t) \in S_{g_{t}(Y)|X}$ . Assuming that  $Span(\Sigma^{-1}\mathbf{m}(t)) = S_{f_{t}(Y)|X}$ , and  $Span(\Sigma^{-1}\mathbf{m}_{c}(t)) = S_{g_{t}(Y)|X}$ .

For any 
$$\boldsymbol{\alpha} \in \mathbb{R}^{p \times n}$$
,

$$g_t(y) \perp \mathbf{X} \mid \boldsymbol{\alpha}^T \mathbf{X} \Rightarrow f_t(y) \perp \mathbf{X} \mid \boldsymbol{\alpha}^T \mathbf{X}, \text{ a.s.}$$

which implies that

$$S_{f_t(Y)|\mathbf{X}} \subseteq S_{g_t(Y)|\mathbf{X}}$$

By the structure of **M** and **M**<sub>c</sub>, it is straightforward that  $\text{Span}(\Sigma^{-1}\mathbf{M}) \subseteq \text{Span}(\Sigma^{-1}\mathbf{M}_c)$ .

Also, it is easy to show that both  $\Sigma^{-1}$ **M** and  $\Sigma^{-1}$ **M**<sub>c</sub> are in  $\delta_{Y|X}$ . Hence, under very strong coverage conditions, Span( $\Sigma^{-1}$ **M**) recovers more of the central subspace.  $\Box$ 

#### References

Bura, E., Cook, R.D., 2001, Extending sliced inverse regression: The weighted Chi-squared test, J. Amer. Statist, Assoc, 96, 996–1003.

Cook, R.D., 1998. Regression Graphics. Wiley, New York.

Cook, R.D., 2004. Testing predictor contributions in sufficient dimension reduction. Ann. Statist. 32, 1062–1092.

- Cook, R.D., Li, B., 2002. Dimension reduction for conditional mean in regression. Ann. Statist. 30, 455–474. Cook, R.D., Ni, L. 2006. Using intraslice covariances for improved estimation of the central subspace in regression. Biometrika 93, 65-74.
- Cook, R.D., Weisberg, S., 1991. Discussion of sliced inverse regression for dimension reduction. J. Amer. Statist. Assoc. 86, 328–332.
- Duan, N., Li, K.C., 1991. Slicing regression: A link-free regression method. Ann. Statist. 19, 505–530.

Feng, Z., Wen, X., Yu, Z., Zhu, L.-X., 2013. On partial sufficient dimension reduction with applications to partially linear multi-index models. J. Amer. Statist. Assoc. 108, 237-246.

Härdle, W., Hall, P., Marron, J.S., 1988. How far are automatically chosen regression smoothing parameters from their optimum. J. Amer. Statist. Assoc. 83, 86–95. Li, K.C., 1987. Asymptotic optimality for Cp, CL, cross-validation and generalized cross-validation: discrete index set. Ann. Statist. 15, 958–975. Li, K.C., 1991. Sliced inverse regression for dimension reduction. J. Amer. Statist. Assoc. 86, 316–327.

- Li, K.C., 1992. On principal Hessian directions for data visualization and dimension reduction: another application of Stein's lemma. J. Amer. Statist. Assoc. 87, 1025-1039.
- Li, B., Dong, Y., 2009. Dimension reduction for non-elliptically distributed predictors. Ann. Statist. 37, 1272–1298.
- Li, B., Wang, S., 2007. On directional regression for dimension reduction. J. Amer. Statist. Assoc. 102, 997-1008.
- Li, B., Wen, S.Q., Zhu, L.X., 2008. On a projective resampling method for dimension reduction with multivariate responses. J. Amer. Statist. Assoc. 103, 1177-1186
- Schott, J.R., 1994. Determining the dimensionality in sliced inverse regression. J. Amer. Statist. Assoc. 89, 141-148.
- Wang, H., Xia, Y., 2008. Sliced regression for dimension reduction. J. Amer. Statist. Assoc. 103, 811-821.
- Xia, Y., Tong, H., Li, W.K., Zhu, L.X., 2002. An adaptive estimation of dimension reduction space. J. R. Stat. Soc. 64, 363–410.
- Yin, X., Cook, R.D., 2002. Dimension reduction for the conditional kth moment in regression. J. R. Stat. Soc. 64, 159–175.
- Yin, X., Li, B., 2011. Sufficient dimension reduction based on an ensemble of minimum average variance estimators. Ann. Statist. 39, 3392–3416.
- Yin, X., Li, B., Cook, R.D., 2008. Successive direction extraction for estimating the central subspace in a multiple-index regression. J. Multivariate Anal. 99, 1733-1757
- Zhu, L.-X., Miao, B.Q., Peng, H., 2006. Sliced inverse regression with large dimensional covariates. J. Amer. Statist. Assoc. 101, 630–643.
- Zhu, L., Wang, T., Zhu, L.X., Ferré, L., 2010a. Sufficient dimension reduction through discretization-expectation estimation. Biometrika 97, 295–304. Zhu, Y., Zeng, P., 2006. Fourier methods for estimating the central subspace and the central mean subspace in regression. J. Amer. Statist. Assoc. 101, 1638-1651

Zhu, L., Zhu, L.X., Feng, Z., 2010b. Dimension reduction in regression through cumulative slicing estimation. J. Amer. Statist. Assoc. 105, 1455–1466.