

A Note on Cumulative Mean Estimation

BILIN ZENG¹, ZHOU YU² AND XUERONG MEGGIE WEN³

¹*Department of Mathematics, California State University, Bakersfield, CA, U.S.A.*

²*School of Finance and Statistics, East China Normal University, Shanghai, China*

³*Department of Mathematics and Statistics, Missouri University of Science and Technology, MO, U.S.A.*

Abstract

For many-valued or continuous Y , the standard practice in *sufficient dimension reduction* (Li, 1991; Cook, 1998) of replacing the response Y with a discrete version of Y usually results in the loss of power due to the loss of the intra-slice information. Most of the existing slicing methods highly rely on the choices of the total number of slices h . Zhu et al. (2010) proposed a method called the *cumulative slicing estimation* (CUME) which avoids the otherwise subjective selection of h . In this paper, we revisit CUME from a different perspective to gain more insights, and then refine its performance by incorporating the intra-slice covariances. We prove that our new method, which we call the *covariance cumulative slicing estimation* (COCUM), is more comprehensive than CUME since it captures a larger part of the central subspace. Simulation studies suggest that our method is comparable to CUME, and outperforms CUME when the response is skewed. The asymptotic results of COCUM are also well proved.

KEY WORDS: Cumulative Slicing Estimation, Dimension Reduction, Sliced Inverse Regression, Intraslice Covariance Estimation.

1 Introduction

For a typical regression problem with a univariate random response Y and a p -dimensional random vector \mathbf{X} , sufficient dimension reduction (SDR: Li, 1991; Cook, 1998) aims to

reduce the dimension of \mathbf{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 $Y \perp\!\!\!\perp \mathbf{X} | \boldsymbol{\eta}^T \mathbf{X}$. 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 $\mathcal{S}_{Y|\mathbf{X}}$. The dimension d of $\mathcal{S}_{Y|\mathbf{X}}$ is called the structural dimension of the regression. Under mild conditions (Yin, Li and Cook, 2008), the central subspace exists and is unique. We assume that $\mathcal{S}_{Y|\mathbf{X}}$ exists throughout this article. The goal of sufficient dimension reduction is to estimate and make statistical inferences about $\mathcal{S}_{Y|\mathbf{X}}$ and d .

Many methods have been developed to estimate $\mathcal{S}_{Y|\mathbf{X}}$. 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 $E(Y|\mathbf{X})$, and the central mean subspace $\mathcal{S}_{E(Y|\mathbf{X})}$ is defined as the smallest column spaces spanned by $\boldsymbol{\eta}$ such that $E(Y|\mathbf{X}) \perp\!\!\!\perp \mathbf{X} | \boldsymbol{\eta}^T \mathbf{X}$.

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 $\mathcal{S}_{\check{Y}|\mathbf{X}} = \mathcal{S}_{Y|\mathbf{X}}$. However, this assumption is not always true, and the differences between the working and target regressions can be significant. Moreover, even under the case of equality, we might still face the loss of power since we use only the information retained in \check{Y} , discarding all the intra-slice information.

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 $d + 1$, trying a few different values of h as necessary. However, beyond empirical experiences, how to select

the optimal h still remains unknown in the literature.

Zhu et al. (2010) 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 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 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 estimators 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. For easy of exposition, we defer all proofs to the Appendix.

2 CUME: the Ensemble Estimators' Approach

2.1 CUME

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

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

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

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

where $w(\cdot)$ 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 3.), which is a mild condition imposed on the marginal distribution of \mathbf{X} only, the column space of \mathbf{M} is a subset of $\Sigma\mathcal{S}_{Y|\mathbf{X}}$, where $\Sigma = \text{Cov}(\mathbf{X})$. At the sample level, suppose (\mathbf{X}_i, Y_i) , $i = 1, \dots, n$ are independent copies of (\mathbf{X}, Y) , we can estimate \mathbf{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 $\mathbf{m}_n(Y_i) = \frac{1}{n} \sum_{j=1}^n (\mathbf{X}_j - \bar{\mathbf{X}}) I(Y_j \leq Y_i)$, and $\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i$. Let $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T$ be the sample predictor variance, assuming a known d , the CUME estimator of $\mathcal{S}_{Y|\mathbf{X}}$ is constructed by the d eigenvectors of $\hat{\Sigma}^{-1} \mathbf{M}_n$ corresponding to its d largest eigenvalues. Zhu et al. (2010) studied the asymptotic properties of the CUME estimator, and also provided a data-driven method on the determination of d .

2.2 The Ensemble Estimators' Approach

In this section, we revisit CUME via the ensemble estimator's 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 \mathcal{J} be a family of functions $f : \Omega_Y \rightarrow \mathbb{F}$, where \mathbb{F} can be the set of real numbers or complex numbers; 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 $\mathcal{S}_{\mathbb{E}[f(Y)|\mathbf{X}]}$ be the central mean subspace for the conditional mean $\mathbb{E}[f(Y)|\mathbf{X}]$, as defined in Cook and Li (2002). If \mathcal{J} is a subset of $L_2(F_Y)$ that is dense in \mathcal{B} , where $\mathcal{B} = \{I_B : B \text{ is a Borel set in } \Omega_Y\}$, then we have:*

$$\text{Span}\{\mathcal{S}_{\mathbb{E}[f(Y)|\mathbf{X}]} : f \in \mathcal{J}\} = \mathcal{S}_{Y|\mathbf{X}}. \quad (2.3)$$

Hence, for a sufficiently rich family of $f(Y)$, the conditional mean subspaces $\mathcal{S}_{\mathbb{E}[f(Y)|\mathbf{X}]}$, when put together, can recover the central subspace $\mathcal{S}_{Y|\mathbf{X}}$. 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 $\mathcal{J} = \{f_t(y) = e^{ity} : t \in \mathbb{R}^1\}$,

where i is the imaginary unit; while Wang and Xia (2008) used $\mathcal{J} = \{f_t(y) = I_{(-\infty, t)}(y) : t \in \mathbb{R}^1\}$, and MAVE (Xia et al., 2002) to estimate the individual central mean subspace $\mathcal{S}_{E[f_t(Y)|\mathbf{X}]}$.

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

Proposition 1 *Let $\mathcal{J} = \{f_t(y) = I_{(-\infty, t)}(y) : t \in \mathbb{R}^1\}$, for $t = Y_1, \dots, Y_n$, $\boldsymbol{\Sigma}^{-1}m(t) \in \mathcal{S}_{E[f_t(Y)|\mathbf{X}]}$, and the column space of $\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{M}_n$ provides a consistent estimator of $\mathcal{S}_{Y|\mathbf{X}}$.*

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 $\boldsymbol{\eta}_t$ 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 $\mathcal{S}_{f_t(Y)|\mathbf{X}}$.

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

$$\mathbf{m}_c(\tilde{y}) = E(\mathbf{X}YI(Y \leq \tilde{y})) \quad (3.1)$$

for $\tilde{y} \in \mathbb{R}^1$. The kernel matrix for COCUM, \mathbf{M}_c , is constructed similar as (2.2) except replacing $\mathbf{m}(\tilde{y})$ with $\mathbf{m}_c(\tilde{y})$.

Denote $\beta_t = \Sigma^{-1}E(\mathbf{X}YI(Y \leq t))$, it is easy to show that β_t can be decomposed as $F_t\Sigma^{-1}\text{Cov}(\mathbf{X}, Y|I(Y \leq t) = 1) + E(Y|I(Y \leq t) = 1)\boldsymbol{\eta}_t$, where $F_t = P(Y \leq t)$.

Theorem 1 *Assuming the common linearity condition, then $\beta_t \in \mathcal{S}_{Y|\mathbf{X}}$, for $t \in \mathbb{R}^1$. Furthermore, assuming the coverage condition (Cook, 1998), comparing with CUME, the column space of $\Sigma^{-1}\mathbf{M}_c$ always encloses that of $\Sigma^{-1}\mathbf{M}$.*

Theorem 1 suggests that theoretically Cocum always wins over CUME 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, Miao and Peng (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, Miao and Peng (2006). The proof is similar to that of Theorem 2 of CUME (Zhu et al., 2010) and is omitted.

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

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

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

Theorem 3 *Assuming the following regularity conditions:*

1. $\max_{1 \leq i \leq p} E(X_i^8 Y^8) < \infty$ uniformly for p ,
2. The minimum eigenvalue of Σ 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 γ ,
5. $p = o(n^{1/2})$;

then

$$\sqrt{n} \gamma^T (\Sigma_n^{-1} \mathbf{M}_n - \Sigma^{-1} \mathbf{M}) \gamma \xrightarrow{\mathcal{D}} N(0, G).$$

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, Miao and Peng (2006), Zhu, Wang, Zhu and Ferré (2010). Following Zhu, Zhu and Feng (2010), we use a modified BIC-type method for COCUM. Define

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

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

$$\hat{K} = \arg \max_{1 \leq k \leq p} G(k). \quad (3.3)$$

As pointed out in Zhu, Zhu and Feng (2010), smaller value of the penalty constant C_n tends to overestimate the dimension K ; while larger C_n tends to underestimate the dimension K . 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} .

Theorem 4 *If $C_n/n \rightarrow 0$ as $n \rightarrow \infty$ and $C_n \rightarrow \infty$, then $\hat{K} - K = O(1)$.*

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 distribution. 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 :

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

as evaluation measurements. Here $\hat{\beta}$ is the estimate for true β and Σ is the covariance matrix for predictor \mathbf{X} . ρ^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 addition, the frequencies of estimated structural dimension over the 1000 trials are also given to illustrate the performance of the modified BIC method. 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 ϵ . The univariate response Y is constructed as:

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

In this case, $\beta = (\beta_1, \beta_2)$ where $\beta_1 = (1, 0, \dots, 0)^T$ and $\beta_2 = (0, 1, 1, 0, \dots, 0)^T$. 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 $\text{Gamma}(0.1, 10)$, and standard normal ϵ . The univariate response Y is constructed as:

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

In this case, $\beta = (\beta_1, \beta_2)$ where $\beta_1 = (1, 0, \dots, 0)^T$ and $\beta_2 = (0, 1, 0, \dots, 0)^T$. Table 4.3 provides the averages of the ratios and standard deviation 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 above 0.9 even when $p = 20$. Table 4.4 again shows that CUME tends to underestimate the structural dimension.

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 \mathbf{X} without the intercept term. By Theorem 1 of Cook and Li (2002), we can show that $\Sigma^{-1}m(t) \in \mathcal{S}_{\mathbb{E}[f_t(Y)|\mathbf{X}]}$.

Since $\mathcal{J} = \{f_t(y) = I_{(-\infty, t)}(y) : t \in \mathbb{R}^1\}$ is dense in $L_2(F_Y)$, by Theorem 1, we have $\text{Span}\{\mathcal{S}_{\mathbb{E}[f(Y)|\mathbf{X}]} : f \in \mathcal{J}\} = \mathcal{S}_{Y|\mathbf{X}}$. Let $\hat{m}(t)$ denote the corresponding sample estimator of $m(t)$, since $\hat{\Sigma}^{-1}\hat{m}(t)$ is a consistent estimator of $\mathcal{S}_{\mathbb{E}[f_t(Y)|\mathbf{X}]}$, $\hat{\Sigma}^{-1}\mathbf{M}_n$ is also a consistent estimator of $\mathcal{S}_{Y|\mathbf{X}}$. \square

Proof of Theorem 1:

Let $h_t(y) = I(y \leq t)$, and $g_t(y) = yI(y \leq t)$, then $m(t) = \Sigma^{-1}\text{Cov}\{h_t(y), \mathbf{X}\}$, and $m_c(t) = \Sigma^{-1}\text{Cov}\{g_t(y), \mathbf{X}\}$. Following similar arguments of Theorem 1 of Cook and Li (2002), we can show that $m(t) \in S_{h_t(Y)|X}$, and $m_c(t) \in S_{g_t(Y)|X}$.

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

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

which implies that

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

The coverage condition implies that $\Sigma^{-1}\mathbf{M}_c = S_{g_t(Y)|\mathbf{X}}$ and $\Sigma^{-1}\mathbf{M} = S_{h_t(Y)|\mathbf{X}}$. Therefore, $\Sigma^{-1}\mathbf{M} \subseteq \Sigma^{-1}\mathbf{M}_c$. \square

References

- [1] Bura, E. and Cook, R. D. (2001). Extending sliced inverse regression: The weighted chi-squared test. *Journal of the American Statistical Association*, **96**, 996–1003.
- [2] Cook, R. D. (1998). *Regression Graphics*. Wiley, New York.
- [3] Cook, R. D. and Li, B. (2002). Dimension Reduction for Conditional Mean in Regression. *The Annals of Statistics*, **30**, No. 2, 455–474.
- [4] Cook, R. D. and Ni, L. (2006). Using intraslice covariances for improved estimation of the central subspace in regression. *Biometrika*, **93**, No. 1, 65–74.
- [5] Cook, R. D. and Weisberg, S. (1991). Sliced Inverse Regression for Dimension Reduction: Comment. *Journal of the American Statistical Association*, **86**, No. 414, 328–332.
- [6] Duan, N. and Li, K. C. (1991). Slicing Regression: A Link-Free Regression Method. *The Annals of Statistics*, **19**, No. 2, 505–530.
- [7] Feng, Z., Wen, X., Yu Z. and Zhu, L. X. (2013). On Partial Sufficient Dimension Reduction With Applications to Partially Linear Multi-Index Models. *Journal of the American Statistical Association*, **108**, No. 501, 237–246.
- [8] Hardle, W., Hall, P. and Marron, J. S. (1988). How Far are Automatically Chosen Regression Smoothing Parameters from their Optimum. *Journal of the American Statistical Association*, **83**, Issue 401, 86–95.
- [9] Li, B. and Dong, Y. (2009). Dimension reduction for non-elliptically distributed predictors. *The Annals of Statistics*, **37**, 1272–1298.
- [10] Li, B. and Wang, S. (2007). On directional regression for dimension reduction. *Journal of the American Statistical Association*, **102**, 997–1008.
- [11] Li, K. C. (1987). Asymptotic optimality for C_p , CL , cross-validation and generalized cross-validation: discrete index set. *The Annals of Statistics*, **15**, 958–975.

- [12] Li, K. C. (1991). Sliced Inverse Regression for Dimension Reduction. *Journal of the American Statistical Association*, **86**, 316–327.
- [13] Schott, J. R. (1994). Determining the dimensionality in sliced inverse regression. *Journal of American Statistical Association*, **89**, No. 425, 141–148.
- [14] Wang, H. and Xia, Y. (2008). Sliced Regression for Dimension Reduction. hansheng.gsm.pku.edu.cn/pdf/2008/SR-Main.pdf, May 2013.
- [15] Xia, Y., Tong, H., Li W. K., and Zhu, L.-X. (2002). An adaptive estimation of dimension reduction space. *Journal of Royal Statistical Society*, **64**, No. 3, 363–410.
- [16] Yin, X. and Li, B. (2011). Sufficient dimension reduction based on a nensemble of minimum average variance estimators. *The Annals of Statistics*, **39**, No. 6, 3392–3416.
- [17] Yin, X., Li, B., and Cook, R. D. (2008). Successive direction extraction for estimating the central subspace in a multiple-index regression. *Journal of Multivariate Analysis*, **99**, 1733–1757.
- [18] Zhu, L. X., Miao, B. Q., and Peng, H. (2006). Sliced Inverse Regression with large dimensional covariates. *Journal of American Statistical Association*, **101**, 630–643.
- [19] Zhu, Y. and Zeng, P. (2006). Fourier methods for estimating the central subspace and the central mean subspace in regression. *Journal of American Statistical Association*, **101**, No. 476, 1638–1651.
- [20] Zhu, L., Zhu, L.-X. and Feng, Z. (2010). Dimension reduction in regression through cumulative slicing estimation. *Journal of American Statistical Association*, 105:492, 1455–1466.

Table 4.1: Ratio Average (standard deviation) from 1000 simulation runs for Model I

	Method	$n = 200$		$n = 400$		$n = 600$	
		ratio	deviation	ratio	deviation	ratio	deviation
$p = 10$	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
$p = 15$	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
$p = 20$	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

Table 4.2: Dimension estimate for Model I

p	n	$d = 1$		$d = 2$		$d > 2$	
		CUME	COCUM	CUME	COCUM	CUME	COCUM
10	200	100	1	0	98.5	0	0.5
10	400	100	0.3	0	99.7	0	0
10	600	100	0.2	0	99.8	0	0
15	200	100	0	0	82.9	0	17.1
15	400	100	0	0	98.1	0	1.9
15	600	100	0	0	99.3	0	0.7
20	400	100	0	0	74	0	26
20	600	100	0	0	91	0	9

Table 4.3: Ratio Average (standard deviation) from 1000 simulation runs for Model II

	Method	$n = 200$		$n = 400$		$n = 600$	
		ratio	deviation	ratio	deviation	ratio	deviation
$p = 10$	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
$p = 15$	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
$p = 20$	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

p	n	$d = 1$		$d = 2$		$d > 2$	
		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	200	90.6	2	9.4	79.2	0	18.8
20	400	100	0.4	0	91.2	0	8.4
20	600	100	0.1	0	94.3	0	5.6