Supplement to "Aggregate inverse mean estimation for sufficient dimension reduction"

Qin Wang *and Xiangrong Yin †

Abstract

In this supplementary file, we provide additional detailed simulation results to the paper "Aggregate inverse mean estimation for sufficient dimension reduction", including the estimation accuracy and the computation time for the proposed method. For the convenience of readers, the partial results reported in the paper (Figure 2 in section 4.1) are also included here.

*Department of Information Systems, Statistics, and Management Science, The University of Alabama, Tuscaloosa, AL 35487. Email: qwang57@ua.edu

†Department of Statistics, University of Kentucky, Lexington, KY 40536. Email: yinxiangrong@uky.edu

1 Additional Simulation Results

1.1 Comparisons of estimation accuracy

In this section, we evaluate the finite sample performance of AIME through simulations. We compare AIME with several existing methods: SIR, SAVE, CUME, and fSIR. We use the vector correlation coefficient q (Hotelling, 1936; Ye and Weiss, 2003) to measure the estimation accuracy. Let $\mathbf{B} \in \mathbb{R}^d$ be an orthonormal basis of the CS, and $\hat{\mathbf{B}}$ be an estimate of the orthonormal basis. Then the correlation q between the two d-dimensional subspaces, $\mathcal{S}(\hat{\mathbf{B}})$ and $\mathcal{S}(\mathbf{B})$, is defined as

$$q = \sqrt{\mid \hat{\mathbf{B}}^T(\mathbf{B}\mathbf{B}^T)\hat{\mathbf{B}} \mid} = \sqrt{\prod_{i=1}^{d} \rho_i^2},$$

where $0 \leq \rho_d \leq \cdots \leq \rho_1 \leq 1$ are the eigenvalues of matrix $\hat{\mathbf{B}}^T(\mathbf{B}\mathbf{B}^T)\hat{\mathbf{B}}$. The larger q indicates the closer of $\mathcal{S}(\hat{\mathbf{B}})$ to $\mathcal{S}(\mathbf{B})$.

The following six models in our paper, extensively studied in the SDR literature, were considered in our simulation.

$$\text{Model I:} \quad Y = 0.5(\beta_1^T X)^3 + 0.5(1 + \beta_2^T X)^2 + 0.2\epsilon_1,$$

$$\text{Model II:} \quad Y = \text{sgn}(2\beta_1^T X + \epsilon_1) \times \log \mid 2\beta_2^T X + 3 + \epsilon_2 \mid,$$

$$\text{Model III:} \quad Y = 2(\beta_1^T X)^2 + 2\exp(\beta_2^T X)\epsilon_1,$$

$$\text{Model IV:} \quad Y = \frac{\beta_1^T X}{0.5 + (1.5 + \beta_2^T X)^2} + (\beta_3^T X)^2\epsilon_1,$$

$$\text{Model V:} \quad Y = (\beta_1^T X)(\beta_2^T X + 2) + (\beta_3^T X + 2)^3 + 0.5\epsilon_1,$$

$$\text{Model VI:} \quad Y = (\beta_1^T X)(\beta_2^T X)^2 + (\beta_3^T X)(\beta_4^T X) + 0.5\epsilon_1.$$

For the comparison, data were generated in the same way as they appear in literature. The covariates $X \sim N_p(0,\ \Sigma)$, with $\Sigma = (\sigma_{ij}) = (0.5^{|i-j|})$ for models I-V, while $\Sigma = \mathrm{I}_p$ for model VI. The standard Gaussian noise ϵ_1 and ϵ_2 are independent of X.

Models I comes from Zhu, Zhu, and Feng (2010) in the study of CUME, where $\beta_1 = (1, 1, 1, 0, \ldots, 0)^T$, and $\beta_2 = (1, 0, 0, 0, 1, 3, 0, \ldots, 0)^T$. Model II was adopted from Chen and Li (1998), where $\beta_1 = (0.5, 0.5, 0.5, 0.5, 0, \ldots, 0)^T$, and $\beta_2 = (0, \ldots, 0, 0.5, 0.5, 0.5, 0.5)^T$. Model III come from Xia (2007), the CS in this model is contained in both the regression mean and variance functions, where the first 10 elements of β_1 and β_2 are $(1, 2, 0, \ldots, 0, 2)^T/2$ and $(0, 0, 3, 4, 0, \ldots, 0)^T/5$, respectively, and the rest elements are 0's. Model IV comes rom Wang and Xia (2008) for sliced regression (SR), where $\beta_1 = (1, 0, \ldots, 0)^T$, $\beta_2 = (0, 1, 0, \ldots, 0)^T$, and $\beta_3 = (0, 0, 1, 0, \ldots, 0)^T$. Model V is from Zhu, Miao, and Peng (2006) in assessing the performance of SIR in high-dimensional setting, where the CS is spanned by $\beta_1 = (1, 0, \ldots, 0)^T$, $\beta_2 = (0, 1, 1, 0, \ldots, 0)^T$, and $\beta_3 = (0, 0, 0, 1, 1, 0, \ldots, 0)^T$. Model VI was used in Xia et al. (2002) with $d_{Y|\mathbf{X}} = 4$, and the four directions are $\beta_1 = (1, 2, 3, 4, 0, \ldots, 0)^T/\sqrt{30}$, $\beta_2 = (-2, 1, -4, 3, 1, 2, 0, \ldots, 0)^T/\sqrt{35}$, $\beta_3 = (0, \ldots, 0, 2, -1, 2, 1, 2, 1)^T/\sqrt{15}$, and $\beta_4 = (0, \ldots, 0, -1, -1, 1, 1)^T/2$.

Extensive numerical experiments were carried out with different sample sizes n and predictor dimensions p. For each parameter setting, 500 simulation replications were conducted. The results are summarized in the following figures and tables.

Figure 1: Model I: comparison of q over 500 replicates

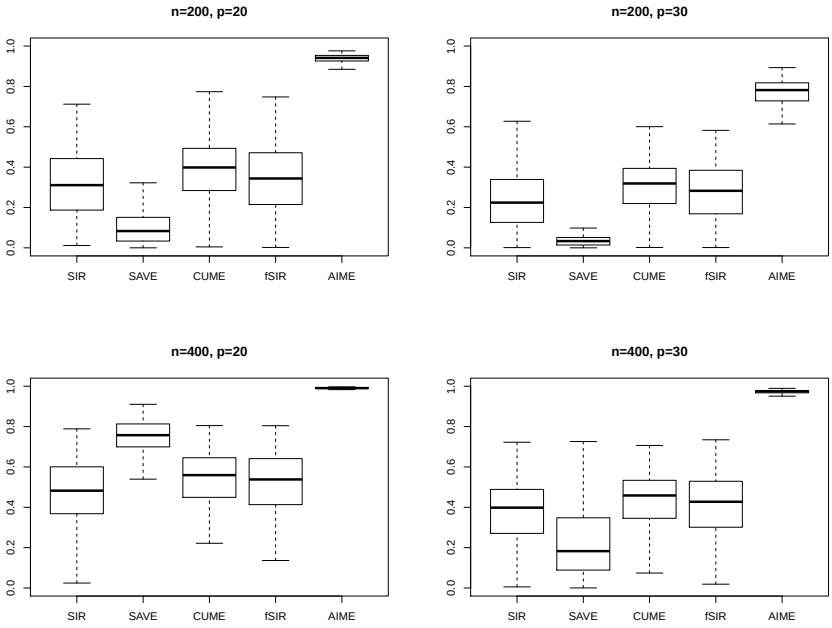

Figure 2: Model II: comparison of q over 500 replicates

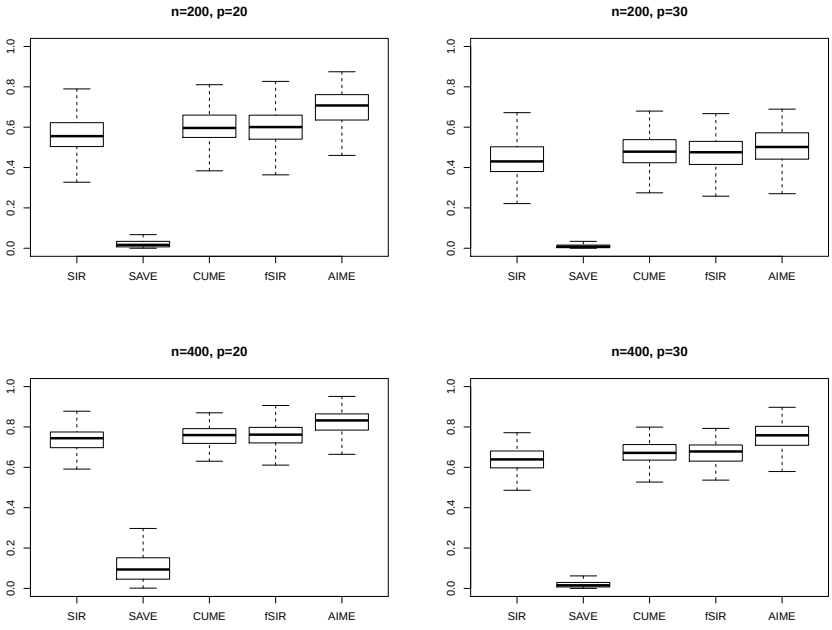

Figure 3: Model III: comparison of q over 500 replicates

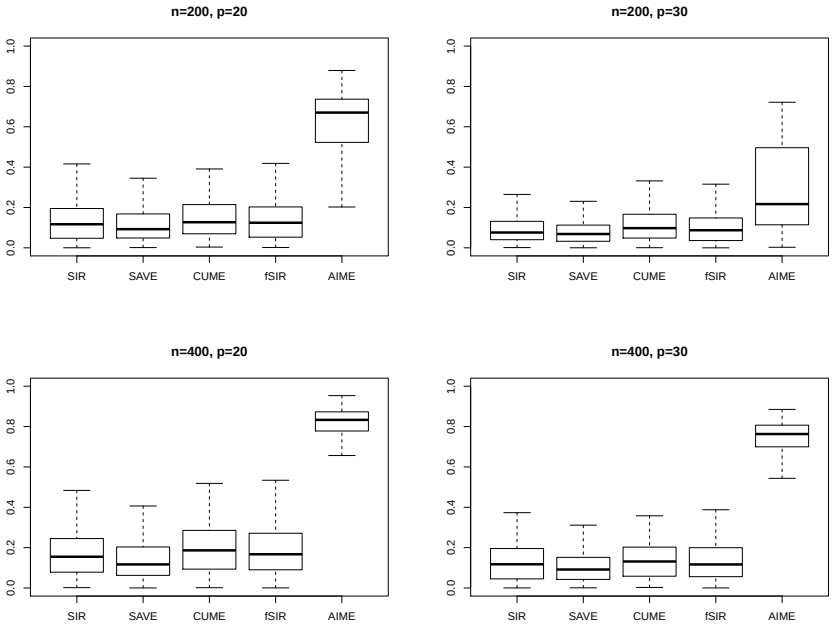

Figure 4: Model IV: comparison of q over 500 replicates

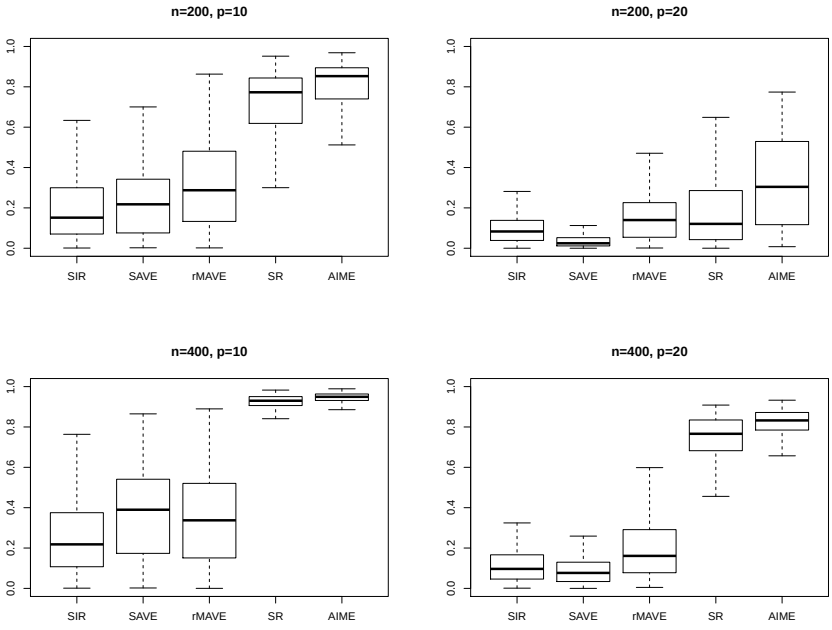

Figure 5: Model V: comparison of q over 500 replicates

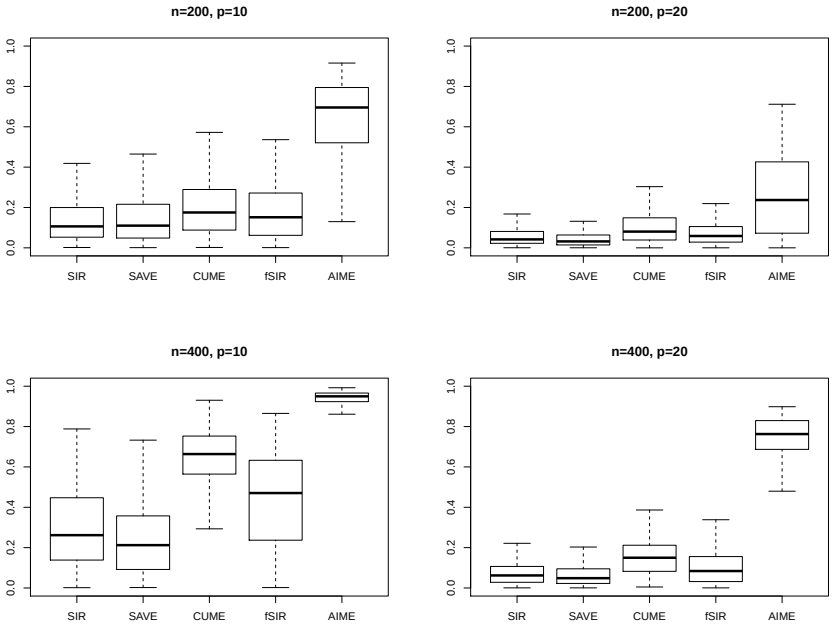

Figure 6: Model VI: comparison of q over 500 replicates

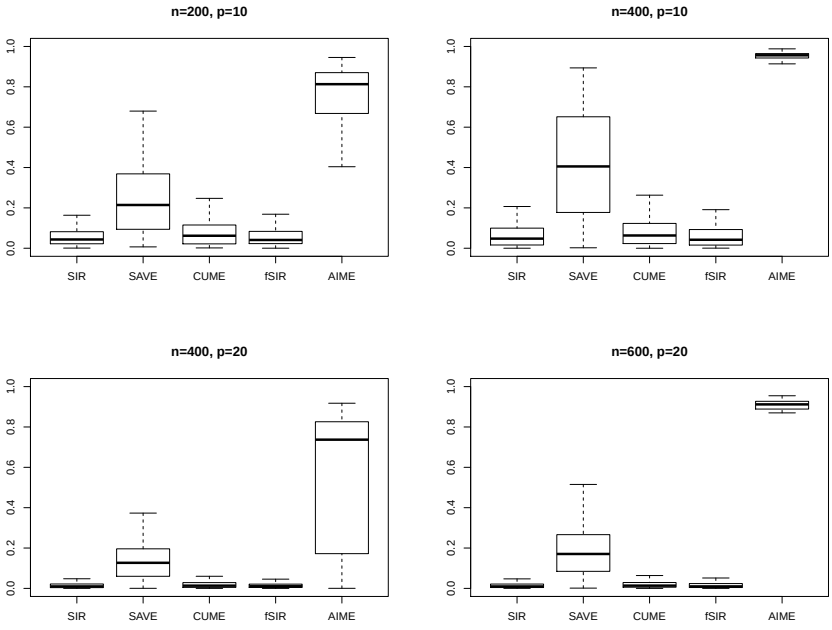

Table 1: Model I, the mean (standard deviation) of q over 500 replicates

		SIR (H=5)	SIR (H=10)	SAVE (H=5)	SAVE (H=10)	CUME	fSIR (k=10)	aSIR	AIME
$n = 200$	$p = 10$	0.450 (0.228)	0.401 (0.223)	0.732 (0.164)	0.368 (0.230)	0.537 (0.193)	0.474 (0.218)	0.981 (0.010)	0.984 (0.008)
	$p = 20$	0.280 (0.170)	0.255 (0.168)	0.112 (0.099)	0.050 (0.041)	0.396 (0.144)	0.330 (0.175)	0.889 (0.042)	0.914 (0.042)
	$p = 30$	0.220 (0.137)	0.188 (0.130)	0.042 (0.039)	0.027 (0.026)	0.289 (0.124)	0.264 (0.146)	0.521 (0.195)	0.748 (0.148)
$n = 300$	$p = 10$	0.582 (0.225)	0.523 (0.239)	0.886 (0.060)	0.791 (0.147)	0.604 (0.179)	0.596 (0.214)	0.993 (0.004)	0.994 (0.003)
	$p = 20$	0.411 (0.174)	0.370 (0.200)	0.441 (0.215)	0.078 (0.065)	0.458 (0.149)	0.451 (0.180)	0.964 (0.012)	0.978 (0.008)
	$p = 30$	0.308 (0.153)	0.273 (0.170)	0.071 (0.059)	0.035 (0.032)	0.375 (0.128)	0.343 (0.163)	0.890 (0.031)	0.936 (0.019)
$n = 400$	$p = 10$	0.653 (0.176)	0.607 (0.211)	0.925 (0.030)	0.899 (0.050)	0.686 (0.181)	0.664 (0.184)	0.996 (0.002)	0.997 (0.002)
	$p = 20$	0.466 (0.187)	0.435 (0.200)	0.725 (0.126)	0.197 (0.167)	0.534 (0.139)	0.505 (0.173)	0.981 (0.006)	0.989 (0.004)
	$p = 30$	0.372 (0.159)	0.330 (0.174)	0.229 (0.156)	0.047 (0.038)	0.451 (0.119)	0.405 (0.163)	0.950 (0.014)	0.971 (0.009)

Table 2: Model II, the mean (standard deviation) of q over 500 replicates

		SIR (H=5)	SIR (H=10)	SAVE (H=5)	SAVE (H=10)	CUME	fSIR (k=10)	aSIR	AIME
$n = 200$	$p = 10$	0.742 (0.086)	0.750 (0.085)	0.221 (0.158)	0.099 (0.100)	0.759 (0.085)	0.768 (0.078)	0.802 (0.091)	0.829 (0.078)
	$p = 20$	0.562 (0.091)	0.565 (0.099)	0.024 (0.022)	0.015 (0.017)	0.599 (0.095)	0.592 (0.090)	0.653 (0.093)	0.675 (0.102)
	$p = 30$	0.434 (0.082)	0.438 (0.091)	0.011 (0.011)	0.008 (0.009)	0.479 (0.085)	0.472 (0.081)	0.415 (0.101)	0.491 (0.099)
$n = 300$	$p = 10$	0.819 (0.061)	0.831 (0.059)	0.398 (0.200)	0.206 (0.151)	0.837 (0.066)	0.842 (0.055)	0.872 (0.063)	0.892 (0.046)
	$p = 20$	0.676 (0.073)	0.689 (0.066)	0.050 (0.048)	0.023 (0.025)	0.684 (0.075)	0.703 (0.067)	0.772 (0.070)	0.778 (0.071)
	$p = 30$	0.555 (0.073)	0.575 (0.075)	0.014 (0.015)	0.009 (0.009)	0.590 (0.072)	0.592 (0.070)	0.669 (0.091)	0.678 (0.092)
$n = 400$	$p = 10$	0.852 (0.054)	0.862 (0.052)	0.547 (0.208)	0.330 (0.203)	0.869 (0.057)	0.868 (0.049)	0.897 (0.053)	0.915 (0.043)
	$p = 20$	0.733 (0.061)	0.749 (0.060)	0.103 (0.085)	0.038 (0.040)	0.756 (0.067)	0.758 (0.059)	0.824 (0.052)	0.826 (0.073)
	$p = 30$	0.633 (0.058)	0.654 (0.059)	0.026 (0.027)	0.011 (0.012)	0.664 (0.064)	0.669 (0.057)	0.753 (0.061)	0.754 (0.076)

Table 3: Model III, the mean (standard deviation) of q over 500 replicates

		SIR (H=5)	SIR (H=10)	SAVE (H=5)	SAVE (H=10)	CUME	fSIR (k=10)	aSIR	AIME
$n = 200$	$p = 10$	0.218 (0.160)	0.225 (0.161)	0.221 (0.171)	0.191 (0.138)	0.230 (0.152)	0.223 (0.157)	0.322 (0.205)	0.848 (0.091)
	$p = 20$	0.131 (0.104)	0.135 (0.112)	0.108 (0.081)	0.090 (0.071)	0.142 (0.102)	0.147 (0.114)	0.186 (0.146)	0.615 (0.194)
	$p = 30$	0.092 (0.073)	0.098 (0.076)	0.069 (0.053)	0.048 (0.039)	0.103 (0.076)	0.109 (0.084)	0.060 (0.057)	0.314 (0.188)
$n = 300$	$p = 10$	0.232 (0.169)	0.223 (0.166)	0.305 (0.207)	0.266 (0.187)	0.256 (0.169)	0.260 (0.176)	0.351 (0.219)	0.916 (0.046)
	$p = 20$	0.133 (0.108)	0.133 (0.104)	0.126 (0.096)	0.113 (0.086)	0.166 (0.114)	0.149 (0.111)	0.232 (0.152)	0.807 (0.078)
	$p = 30$	0.103 (0.085)	0.119 (0.095)	0.091 (0.066)	0.073 (0.059)	0.122 (0.087)	0.122 (0.099)	0.162 (0.116)	0.621 (0.166)
$n = 400$	$p = 10$	0.249 (0.175)	0.258 (0.171)	0.405 (0.239)	0.337 (0.221)	0.285 (0.198)	0.284 (0.193)	0.407 (0.229)	0.941 (0.027)
	$p = 20$	0.172 (0.112)	0.166 (0.115)	0.147 (0.114)	0.124 (0.091)	0.186 (0.127)	0.192 (0.128)	0.291 (0.169)	0.871 (0.047)
	$p = 30$	0.122 (0.086)	0.121 (0.092)	0.108 (0.082)	0.086 (0.067)	0.133 (0.092)	0.131 (0.093)	0.212 (0.141)	0.783 (0.076)

1.2 Comparisons of computation time

Indeed, the computation simplicity is a main advantage of the traditional inverse regression based SDR methods, such as SIR, SAVE, CUME, etc. Our proposed AIME requires more computation time due to the local information extraction and aggregation. However, the computational cost of AIME is pretty reasonable based on our numerical studies. Considering the substantial improvements in estimation efficiency and the broadly accessible and affordable computing resources, we think the extra computing time is well worthwhile and a great trade-off. Table 1 summarizes the average running time of one data sample for different sample size n and predictor dimension p. We use Model I for demonstration as results are very similar across all different models.

All numerical studies were done on a laptop with Core i7 processor and 16 GB RAM. As summarized in Table 4, the average computing time for AIME is no more than a half minute for one random sample even with p=30. Despite the computation simplicity of SIR, its estimation accuracy is much worse than that of AIME.

Table 4: Model I, estimation accuracy q and average computing time in seconds over 200 replicates

		Accuracy			Time in Seconds		
		SIR	CUME	AIME	SIR	CUME	AIME
$n = 200$	$p = 10$	0.450 (0.228)	0.537 (0.193)	0.984 (0.008)	0.012	0.013	2.68
	$p = 20$	0.280 (0.170)	0.396 (0.144)	0.914 (0.042)	0.015	0.015	4.45
	$p = 30$	0.220 (0.137)	0.289 (0.124)	0.748 (0.148)	0.016	0.018	6.98
$n = 300$	$p = 10$	0.582 (0.225)	0.604 (0.179)	0.994 (0.003)	0.013	0.014	6.05
	$p = 20$	0.411 (0.174)	0.458 (0.149)	0.978 (0.008)	0.015	0.016	9.76
	$p = 30$	0.308 (0.153)	0.375 (0.128)	0.936 (0.019)	0.017	0.020	14.43
$n = 400$	$p = 10$	0.653 (0.176)	0.686 (0.002)	0.997 (0.002)	0.013	0.016	11.91
	$p = 20$	0.466 (0.187)	0.534 (0.139)	0.989 (0.004)	0.015	0.019	17.78
	$p = 30$	0.372 (0.159)	0.451 (0.119)	0.971 (0.009)	0.017	0.024	25.45

R codes are run on a laptop with Core i7 processor and 16 GB RAM.

References

Chen, C. and K. C. Li (1998). Can sir be as popular as multiple linear regression? *Statistica Sinica 8*, 289–316.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika 28*, 321–377.

Wang, H. and Y. Xia (2008). Sliced regression for dimension reduction. *Journal of the American Statistical Association 103*, 811–821.

Xia, Y. (2007). A constructive approach to the estimation of dimension reduction directions. *Annals of Statistics 35*(6), 2654–2690.

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

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

Zhu, L. P., L. X. Zhu, and Z. Feng (2010). Dimension reduction in regressions through cumulative slicing estimation. *Journal of the American Statistical Association 105*(492), 1455–1466.

Zhu, L. X., B. Q. Miao, and H. Peng (2006). On sliced inverse regression with high-dimensional covariates. *Journal of the American Statistical Association 101*, 630–643.