

Intensive numerical studies of optimal sufficient dimension reduction with singularity

Jae Keun Yoo^{1,a}, Da-Hae Gwak^a, Min-Sun Kim^a

^aDepartment of Statistics, Ewha Womans University, Korea

Abstract

Yoo (2015, *Statistics and Probability Letters*, 99, 109–113) derives theoretical results in an optimal sufficient dimension reduction with singular inner-product matrix. The results are promising, but Yoo (2015) only presents one simulation study. So, an evaluation of its practical usefulness is necessary based on numerical studies. This paper studies the asymptotic behaviors of Yoo (2015) through various simulation models and presents a real data example that focuses on ordinary least squares. Intensive numerical studies show that the χ^2 test by Yoo (2015) outperforms the existing optimal sufficient dimension reduction method. The basis estimation by the former can be theoretically sub-optimal; however, there are no notable differences from that by the latter. This investigation confirms the practical usefulness of Yoo (2015).

Keywords: chi-square test, optimality, singularity, sufficient dimension reduction

1. Introduction

Sufficient dimension reduction (SDR) in regression of $Y \in \mathbb{R}^r | X \in \mathbb{R}^p$ seeks to replace the original p -dimensional predictors X by its lower-dimensional linear projection without loss of information on $Y|X$, which is equivalently expressed as:

$$Y \perp\!\!\!\perp X | M^T X, \quad (1.1)$$

where $\perp\!\!\!\perp$ stands for independence, M is a $p \times q$ matrix, and $q \leq p$.

Define $S(A)$ as a column subspace of $A \in \mathbb{R}^{p \times q}$. For M satisfying equation (1.1), $S(M)$ is called a dimension reduction subspace. If the intersection of all possible dimension reduction subspaces exists and it is a dimension reduction subspace, it is unique and minimal. The intersection is then called the *central subspace* $S_{Y|X}$. SDR is a useful dimension reduction tool in high-dimensional data analysis such as gene expression data (Chiaromonte and Martinelli, 2002), microarray data analysis (Li, 2006) and survival analysis (Cook, 2003).

In SDR literature, the inference on $S_{Y|X}$ is the main stream. Hereafter η and $d (< p)$ will be represented as a true orthonormal basis matrix and the true dimension of $S_{Y|X}$, respectively. The goal of SDR is placed on to the estimation of η and d .

Suppose that there exists a matrix $\theta = (\theta_1, \dots, \theta_q) \in \mathbb{R}^{p \times q}$ such that $S(\theta) = S_{Y|X}$. Since the columns of η span $S_{Y|X}$, we have a relation that $\theta = \eta\gamma$ for $\gamma \in \mathbb{R}^{d \times q}$. Let $\hat{\theta}$ be a known consistent estimator of θ in sense that $\sqrt{n}\{\text{vec}(\hat{\theta}) - \text{vec}(\theta)\}$ is asymptotically normally distributed with mean 0 and covariance matrix Γ , where $\text{vec}(\theta) = (\theta_1^T, \dots, \theta_q^T)^T$.

¹ Corresponding author: Department of Statistics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea. E-mail: peter.yoo@ewha.ac.kr

Now consider the estimation of $\boldsymbol{\eta}$ and $\boldsymbol{\gamma}$ by arguments $\hat{\boldsymbol{B}}$ and $\hat{\boldsymbol{C}}$ that minimize the following quadratic objective function over \boldsymbol{B} and \boldsymbol{C} :

$$F_d(\boldsymbol{B}, \boldsymbol{C}) = \{\text{vec}(\hat{\boldsymbol{\theta}}) - \text{vec}(\boldsymbol{BC})\}^T \boldsymbol{V} \{\text{vec}(\hat{\boldsymbol{\theta}}) - \text{vec}(\boldsymbol{BC})\}, \quad (1.2)$$

where $\boldsymbol{B} \in \mathbb{R}^{p \times d}$, $\boldsymbol{C} \in \mathbb{R}^{d \times q}$, and $\boldsymbol{V} > 0$ is a $pq \times pq$ positive definite inner product matrix.

Cook and Ni (2005) and Shapiro (1986) and show that the minimizer $\hat{\boldsymbol{B}}$ of equation (1.2) is a consistent estimator of $\boldsymbol{\eta}$ for any given $\boldsymbol{V} > 0$, in the sense that $\mathcal{S}(\hat{\boldsymbol{B}})$ converges to $\mathcal{S}(\boldsymbol{\eta})$, which is equal to $\mathcal{S}_{Y|X}$. Since different choices of $\boldsymbol{V} > 0$ provides different estimators of $\boldsymbol{\eta}$, the estimation of $\boldsymbol{\eta}$ by minimizing (1.2) forms a class of estimators called the *minimum discrepancy family* (MDF). It is unpopular to select \boldsymbol{V} to have the optimal properties among the MDF, which will be discussed later. According to Shapiro (1986), it is a consistent estimator of the inverse of the covariance matrix $\boldsymbol{\Gamma}$, $\hat{\boldsymbol{\Gamma}}^{-1}$ and the corresponding optimal quadratic discrepancy function is

$$F_d^{\text{opt}}(\boldsymbol{B}, \boldsymbol{C}) = \{\text{vec}(\hat{\boldsymbol{\theta}}) - \text{vec}(\boldsymbol{BC})\}^T \hat{\boldsymbol{\Gamma}}^{-1} \{\text{vec}(\hat{\boldsymbol{\theta}}) - \text{vec}(\boldsymbol{BC})\}. \quad (1.3)$$

Let $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\gamma}}$ be the solutions to minimize F_d^{opt} in (1.3) and n be the sample size. Then the large sample distribution of $n\hat{F}_d^{\text{opt}} = n\hat{F}(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\gamma}})$ is $\chi_{(p-d)(q-d)}^2$. In addition $\hat{\boldsymbol{\eta}}$ is asymptotically efficient, in the sense that $\text{vec}(\hat{\boldsymbol{\eta}}\hat{\boldsymbol{\gamma}})$ has the minimum variance among all elements in the MDF. The minimum variance of $\text{vec}(\hat{\boldsymbol{\eta}}\hat{\boldsymbol{\gamma}})$ is described in the proof of Theorem 2 in Yoo (2015).

The χ^2 distribution and the asymptotic efficiency are optimal properties among the MDF. Cook and Ni (2005) use the statistic $n\hat{F}_d^{\text{opt}}$ to test $H_0 : d = m$ versus $H_1 : d > m$, $m = 0, \dots, \min(p, q)$ for estimating d . The applications of MDFs in various dimension reduction methods can be found in Cook and Ni (2005), Cook and Zhang (2014), and Yoo and Cook (2007).

In practice, $\hat{\boldsymbol{\Gamma}}^{-1}$ may be singular, and then the use of (1.3) can be problematic. Recently, Yoo (2015) establishes theoretical results in such case. However, only one simulation example is presented in Yoo (2015), so intensive numerical studies and a real data example for the illustration purpose should be necessary to show the practical usefulness of the results of Yoo (2015). This paper investigates the asymptotic behaviors of Yoo (2015) based on various simulated models.

The organization of this article is as follows. In Section 2, the main results of Yoo (2015) are briefly discussed with a focus on ordinary least square estimators (OLS). Intensive numerical studies and a real data example are presented in Section 3, and Section 4 summarizes the work.

2. Review: optimal sufficient dimension reduction (SDR) with singularity

The OLS coefficient vector $\boldsymbol{\beta} \in \mathbb{R}^{p \times r}$ in the regression of $\boldsymbol{Y} \in \mathbb{R}^r | \boldsymbol{X} \in \mathbb{R}^p$ is defined as $\boldsymbol{\beta} = \boldsymbol{\Sigma}^{-1} \text{cov}(\boldsymbol{X}, \boldsymbol{Y})$, where $\boldsymbol{\Sigma} = \text{cov}(\boldsymbol{X})$. Under certain conditions, the columns of $\boldsymbol{\beta}$ span $\mathcal{S}_{Y|X}$, that is, $\mathcal{S}(\boldsymbol{\beta}) = \mathcal{S}_{Y|X}$, so $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ in Introduction are replaced with $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}}$, respectively. The conditions to guarantee that $\mathcal{S}(\boldsymbol{\beta}) = \mathcal{S}_{Y|X}$ will not be discussed here. For more about the conditions, two tutorial papers of Yoo (2016a, 2016b) regarding SDR are recommended. Since $\mathcal{S}(\boldsymbol{\beta}) = \mathcal{S}_{Y|X}$, the relation that $\boldsymbol{\beta} = \boldsymbol{\eta}\boldsymbol{\gamma}$ is induced. Define that $\boldsymbol{\varepsilon} = \boldsymbol{Y} - E(\boldsymbol{Y}) - \boldsymbol{\beta}^T \{\boldsymbol{X} - E(\boldsymbol{X})\}$, which is the OLS residual matrix. Then, according to Yoo and Cook (2007), $\sqrt{n}\{\text{vec}(\hat{\boldsymbol{\beta}}) - \text{vec}(\boldsymbol{\beta})\}$ tends to normal distribution with mean 0 and covariance $\boldsymbol{\Gamma}$:

$$\boldsymbol{\Gamma} = (\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}^{-1/2}) \text{cov}(\boldsymbol{T}_{\text{OLS}}) (\boldsymbol{I}_r \otimes \boldsymbol{\Sigma}^{-1/2}),$$

where $\boldsymbol{T}_{\text{OLS}} \in \mathbb{R}^{pr} = \text{vec}(\boldsymbol{Z}\boldsymbol{\varepsilon}^T)$ and $\boldsymbol{Z} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{X} - E(\boldsymbol{X}))$.

To explain the results of Yoo (2015), it is necessary to define the following Jacobian matrix Δ_β :

$$\Delta_\beta = \left\{ \frac{\partial \text{vec}(\mathbf{BC})}{\partial \text{vec}(\mathbf{B})}, \frac{\partial \text{vec}(\mathbf{BC})}{\partial \text{vec}(\mathbf{C})} \right\} \Big|_{(\mathbf{B}=\eta, \mathbf{C}=\gamma)} = (\boldsymbol{\gamma}^\top \otimes \mathbf{I}_p, \mathbf{I}_q \otimes \boldsymbol{\eta}).$$

Shapiro (1986) discusses that \mathbf{V} in equation (1.2) is required to be positive definite on $\mathcal{S}(\Delta_\beta)$. This is equivalently stated that $\text{rank}(\Delta_\beta) = \text{rank}(\Delta_\beta^\top \mathbf{V} \Delta_\beta)$. Lemma A.3 in Cook and Ni (2005) allows us to use a population quantity $\boldsymbol{\Gamma}$, rather than its sample version $\hat{\boldsymbol{\Gamma}}$, without any loss of generality.

With $\boldsymbol{\Gamma}$ singular with rank $pq - k$, the main results of Yoo (2015) need the following assumption: $\text{rank}(\Delta_\beta) = \text{rank}(\Delta_\beta^\top \boldsymbol{\Gamma}^- \Delta_\beta)$.

According to Yoo (2015), the assumption is mild in practice, because the true dimension d of $\mathcal{S}_{Y|X}$ often turns out to be one or two in many SDR problems. Define that $\boldsymbol{\Gamma}^-$ is a Moore-Penrose inverse such that $\boldsymbol{\Gamma}^- \boldsymbol{\Gamma} \boldsymbol{\Gamma}^- = \boldsymbol{\Gamma}^-$ and $\boldsymbol{\Gamma} \boldsymbol{\Gamma}^- \boldsymbol{\Gamma} = \boldsymbol{\Gamma}$.

Result 1. Assume that $\text{rank}(\Delta_\theta) = \text{rank}(\Delta_\theta \boldsymbol{\Gamma}^- \Delta_\theta)$. Let $\hat{\boldsymbol{\eta}}^*$ and $\boldsymbol{\gamma}^*$ be the solutions to minimizing $F_d^S(\mathbf{B}, \mathbf{C}) = \{\text{vec}(\hat{\boldsymbol{\beta}}) - \text{vec}(\mathbf{BC})\}^\top \boldsymbol{\Gamma}^- \{\text{vec}(\hat{\boldsymbol{\beta}}) - \text{vec}(\mathbf{BC})\}$. Then, $n\hat{F}_d^{S*} = nF_d^S(\hat{\boldsymbol{\eta}}^*, \boldsymbol{\gamma}^*) \sim \chi_{(p-d)(q-d)-k}^2$ asymptotically.

Result 2. Assume that $\text{rank}(\Delta_\theta) = \text{rank}(\Delta_\theta \boldsymbol{\Gamma}^- \Delta_\theta)$. Define $\mathcal{V} = \{\mathbf{V} : \mathcal{S}(\mathbf{V}) = \mathcal{S}(\boldsymbol{\Gamma})\}$. Then $\text{vec}(\hat{\boldsymbol{\eta}}^* \boldsymbol{\gamma}^*)$ has the minimum variance of $\text{vec}(\hat{\mathbf{B}}\hat{\mathbf{C}})$ from (1.2) with any inner-product matrix in \mathcal{V} .

Since it is not always guaranteed that $\mathcal{S}(\mathbf{V}) = \mathcal{S}(\boldsymbol{\Gamma})$, the asymptotic efficiency in Result 2 is more restrictive than the asymptotic χ^2 statistic in Result 1.

3. Numerical studies

3.1. Model configurations

If \mathbf{X} have multi-collinearity, $\boldsymbol{\Gamma}$ can be singular. However, if so, the estimation of $\mathcal{S}_{Y|X}$ through the OLS is not desirable. Instead, the simulation models are set up for the responses to have high-correlation, which also can induce the singularity of $\boldsymbol{\Gamma}$ just like the example given in Yoo (2015). We consider the following six models for numerical studies:

Model 1. $(X_1, \dots, X_5) \stackrel{iid}{\sim} N(0, 1) \perp\!\!\!\perp (\varepsilon_1, \dots, \varepsilon_4) \stackrel{iid}{\sim} N(0, 1)$
 $Y_1 = X_1 + \varepsilon_1; Y_2 = X_1 + \varepsilon_2; Y_3 = X_1 + \varepsilon_3; Y_4 = Y_1 + s\varepsilon_4.$

Model 2. $(X_1, \dots, X_5) \stackrel{iid}{\sim} N(0, 1) \perp\!\!\!\perp (\varepsilon_1, \dots, \varepsilon_4) \stackrel{iid}{\sim} N(0, 1)$
 $Y_1 = \exp(X_1) + \varepsilon_1; Y_2 = \exp(X_1) + \varepsilon_2; Y_3 = \exp(X_1) + \varepsilon_3; Y_4 = Y_1 + s\varepsilon_4.$

Model 3. $(X_1, \dots, X_5) \stackrel{iid}{\sim} N(0, 1) \perp\!\!\!\perp (\varepsilon_1, \dots, \varepsilon_4) \stackrel{iid}{\sim} t_5$
 $Y_1 = X_1 + \varepsilon_1; Y_2 = X_1 + \varepsilon_2; Y_3 = X_1 + \varepsilon_3; Y_4 = Y_1 + s\varepsilon_4.$

Model 4. $(X_1, \dots, X_5) \stackrel{iid}{\sim} N(0, 1) \perp\!\!\!\perp (\varepsilon_1, \dots, \varepsilon_4) \stackrel{iid}{\sim} t_5$
 $Y_1 = \exp(X_1) + \varepsilon_1; Y_2 = \exp(X_1) + \varepsilon_2; Y_3 = \exp(X_1) + \varepsilon_3; Y_4 = Y_1 + s\varepsilon_4.$

Model 5. $(X_1, \dots, X_5) \stackrel{iid}{\sim} t_5 \perp\!\!\!\perp (\varepsilon_1, \dots, \varepsilon_4) \stackrel{iid}{\sim} N(0, 1)$
 $Y_1 = X_1(X_1 + X_2) + \varepsilon_1; Y_2 = \exp(X_1 + X_2) + \varepsilon_2; Y_3 = \exp(X_2) + \varepsilon_3; Y_4 = Y_1 + s\varepsilon_4.$

Model 6. $(X_1, \dots, X_5) \stackrel{iid}{\sim} t_5 \perp\!\!\!\perp (\varepsilon_1, \dots, \varepsilon_4) \stackrel{iid}{\sim} t_5$

$$Y_1 = X_1(X_1 + X_2) + \varepsilon_1; Y_2 = \exp(X_1 + X_2) + \varepsilon_2; Y_3 = \exp(X_2) + \varepsilon_3; Y_4 = Y_1 + s\varepsilon_4.$$

In all simulation models, two responses of Y_1 and Y_4 have high-correlation depending on values of $s = 0.01, 0.1, 0.2, 0.3, 0.4, 0.5$. The structural dimension d for Models 1–4 is equal to one, whose central subspace is spanned by the column of $(1, 0, 0, 0, 0)^T$, equivalently X_1 . However, for Models 5 and 6, $\mathcal{S}_{Y|X}$ is spanned by the two column of $((1, 0, 0, 0, 0)^T, (0, 1, 0, 0, 0)^T)$, equivalently (X_1, X_2) , so d is equal to two. The predictors for Models 1–4 are independently generated from $N(0, 1)$, while those for Models 5–6 are randomly sampled from t distribution with five degrees of freedom (t_5). For Models 1, 2, and 5 and Models 3, 4, and 6, the random errors are independently generated from $N(0, 1)$ and t_5 , respectively. Models 2 and 4 have non-linear mean structures, and Models 5 and 6 have a non-linear relationship between predictors and non-linear means together.

The reason to use t distribution is because of the need to investigate the asymptotic behaviors of Yoo (2015) under thick-tailed distributions, which are more prone to outliers than $N(0, 1)$. Sample sizes n are changed as $n = 25, 50, 100, 200$, and each simulation model is iterated 500 times.

The singularity k varies from 0 to $(p - d) - 1$ with increment by 1. The case of $k = 0$ is equivalent to the optimal SDR proposed by Yoo and Cook (2007). In the test of $H_0 : d = 0$, the values of k can be $0, 1, 2, 3, \dots, (p - 1)$, while those are $0, 1, 2, \dots, (p - 2)$ for testing $H_0 : d = 1$. Reasoning of not using the values of k larger than $(p - d) - 1$ is as follows.

In the simulation, if Y_1 and Y_4 are highly correlated, it would be ideal that $\boldsymbol{\eta}$ is estimated after removing either Y_1 or Y_4 or after replacing Y_1 and Y_4 by its linear combination such as $Y_{\text{new}} = (Y_1 + Y_4)/2$. The dimension is then reduced from four to three, and the degrees of freedom for the dimension test of $H_0 : d = m$ is $(p - m)((r - 1) - m)$. The differences in degrees of freedom between before and after the removal is $(p - m)$. Therefore, the use of the singularity k larger than $(p - m) - 1$ intrinsically induces the reduction of data, before starting the dimension reduction. So, such cases are ruled out. It is recommended to use the guidelines to select the range of k in practice.

3.2. Results

To summarize a basis estimation, the averages of $|r|$ of a square root of R^2 from the OLS regression of $X_1|\hat{\boldsymbol{\eta}}^T \mathbf{X}$ for Models 1–4 and the averages of $|r_i|$ from $X_i|\hat{\boldsymbol{\eta}}^T \mathbf{X}$, $i = 1, 2$, for Models 5 and 6 are computed. If the true basis is well-estimated, the averages of $|r|$ should be close to one.

To measure how well the true dimension d is estimated, for Models 1–4 and Models 5–6, the percentages of the correct decisions that $\hat{d} = 1$ and $\hat{d} = 2$ and the percentages of the decisions that $\hat{d} \geq 2$ and $\hat{d} \geq 3$, which represents the observed levels, with level 5% are computed, respectively. If the true dimension is well-estimated, the former and latter percentages should be close to 95% and 5%, respectively.

First, we discuss the basis estimation results. Varying s with fixing $n = 50$, the averages are reported in Figure 1. For Models 5 and 6, the averages of $|r_1|$ alone are reported because the averages of $|r_2|$ are about the same as those of $|r|$ for Models 1–4. According to Figure 1, for all simulation models, there are no notable differences varying the values of s ; however, Yoo and Cook (2007) show surprisingly poor performances. The minimum averages of $|r|$ is over 0.8 and better with $n = 50$. Summarizing these observations, Yoo (2015) is not a practical cause of concern for the basis estimation.

Figures 2 and 3 summarized the dimension estimation for Model 1, which are the percentages of the decision that $\hat{d} = 1$ and the observed level, respectively. The figures represent the characteristic behaviors observed in all the other simulation models with $d = 1$. Figures 4 and 5 present the summary

Table 1: Dimension test in Minneapolis school data in Section 3.3

	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$
$H_0 : d = 0$	0.000	0.000	0.000	0.000	0.000
$H_0 : d = 1$	0.027	0.061	0.294	0.279	N/A
$H_0 : d = 2$	0.534	N/A	N/A	N/A	N/A

of the dimension estimation for Models 5 and 6 with $s = 0.01, 0.2, 0.5$, because the results for $s = 0.1, 0.3, 0.4$ are similar. Figures 2–5 indicate that, for any values of k under consideration, Yoo (2015) shows better performance in both the correct decision percentages and the observed level than the Yoo-Cook method. Especially, with a smaller sample size such as $n = 25$ and $n = 50$, there exist notable differences between Yoo (2015) and the Yoo-Cook method in Models 1–4. The small sample behaviors are better with a larger value of k . In addition, by the comparison between Model 5 and Model 6 in Figures 4 and 5, the dimension estimation is impacted by the distribution of predictors rather than that of errors, when a nonlinear relation exists. Figures 6 and 7 zoom Figures 2–5 at fixing $n = 50$ for Models 1–6. According to the figures, the dimension estimation seems independent of on the size of s , rather than the choices of k . In addition the distribution of predictors provide a greater effect to the accuracy in the dimension estimation than those of the random errors. According to the observation, Yoo (2015) can improve the Yoo-Cook method in the dimension estimation, and $k = 1$ is recommended as the default value.

Summarizing the numerical studies, Yoo (2015) can have clear advantages over the existing Yoo-Cook method.

3.3. Real data example: Minneapolis school data

For the illustration purpose of Yoo (2015), a dataset to measure the performance of students in Minneapolis schools ($n = 63$) are analyzed (Yoo, 2009). The dataset contains the percentages P_{\bullet} of students in a school scoring above (A) and below (B) average on standardized fourth and sixth grade reading comprehension tests, which are used as the four dimensional responses, $Y = (P_{A4}, P_{B4}, P_{A6}, P_{B6})^T$. Subtracting either pair of grade specific percentages from 100 gives the percentage of students scoring about average on the test. The same five predictors used in Yoo (2009) are considered: (1) the pupil teacher ratio, (2) the square roots of the percentage of children receiving aid to families with dependent children, (3) the percentage of children not living with both biological parents (B), (4) the percentage of adults in the school area who completed high school, and (5) the percentage of persons in the area below the federal poverty level. The transformation is done to induce the required condition to estimate $S_{Y|X}$.

The methods by Yoo (2015) and Yoo and Cook (2007), equivalently Yoo (2015) with $k = 0$, are applied to estimate the structural dimension. Table 1 reports the p -values used to test $H_0 : d = m$, $m = 0, 1, 2$. Table 1 indicates that, with level 5%, the Yoo-Cook method determines that $\hat{d} = 2$, while Yoo (2015) with $k = 1, 2, 3$ concludes that $\hat{d} = 1$. The sample correlation coefficients between P_{A4} and P_{B4} and between P_{A6} and P_{B6} are -0.7699 and -0.8331 , respectively. This indicates that P_{A4} and P_{B4} and P_{A6} and P_{B6} have strong linear relationships. Also, according to numerical studies, with smaller samples, Yoo (2015) with non-zero k is preferable to $k = 0$. Combining this, it would be more desirable to decide that $\hat{d} = 1$.

4. Conclusion

Optimal SDR is a paradigm to estimate the central subspace by minimizing the quadratic objective functions with known initial estimates. Optimality means χ^2 dimension test and asymptotic efficiency.

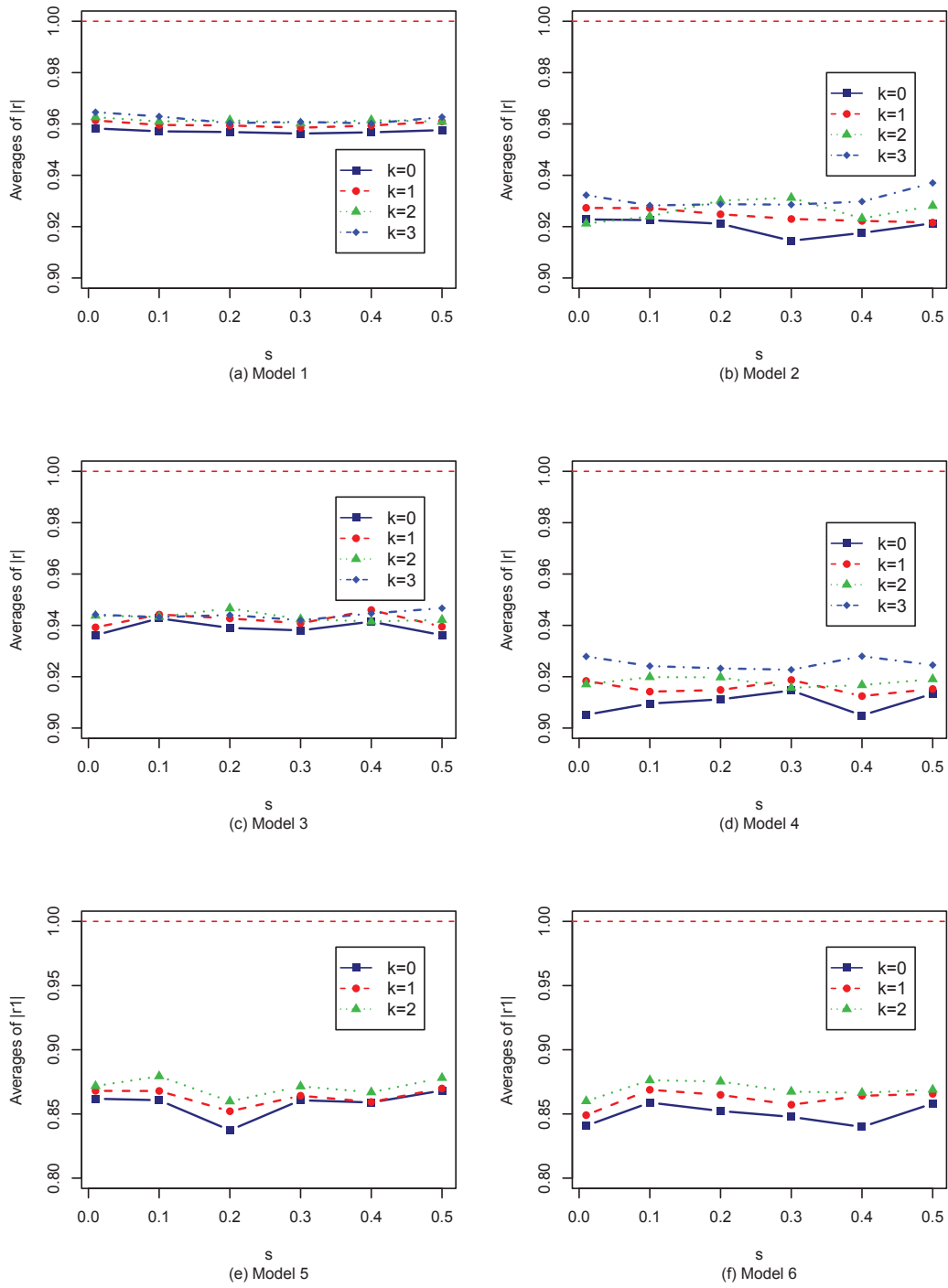


Figure 1: Averages of $|r|$ s varying s with $n = 50$.

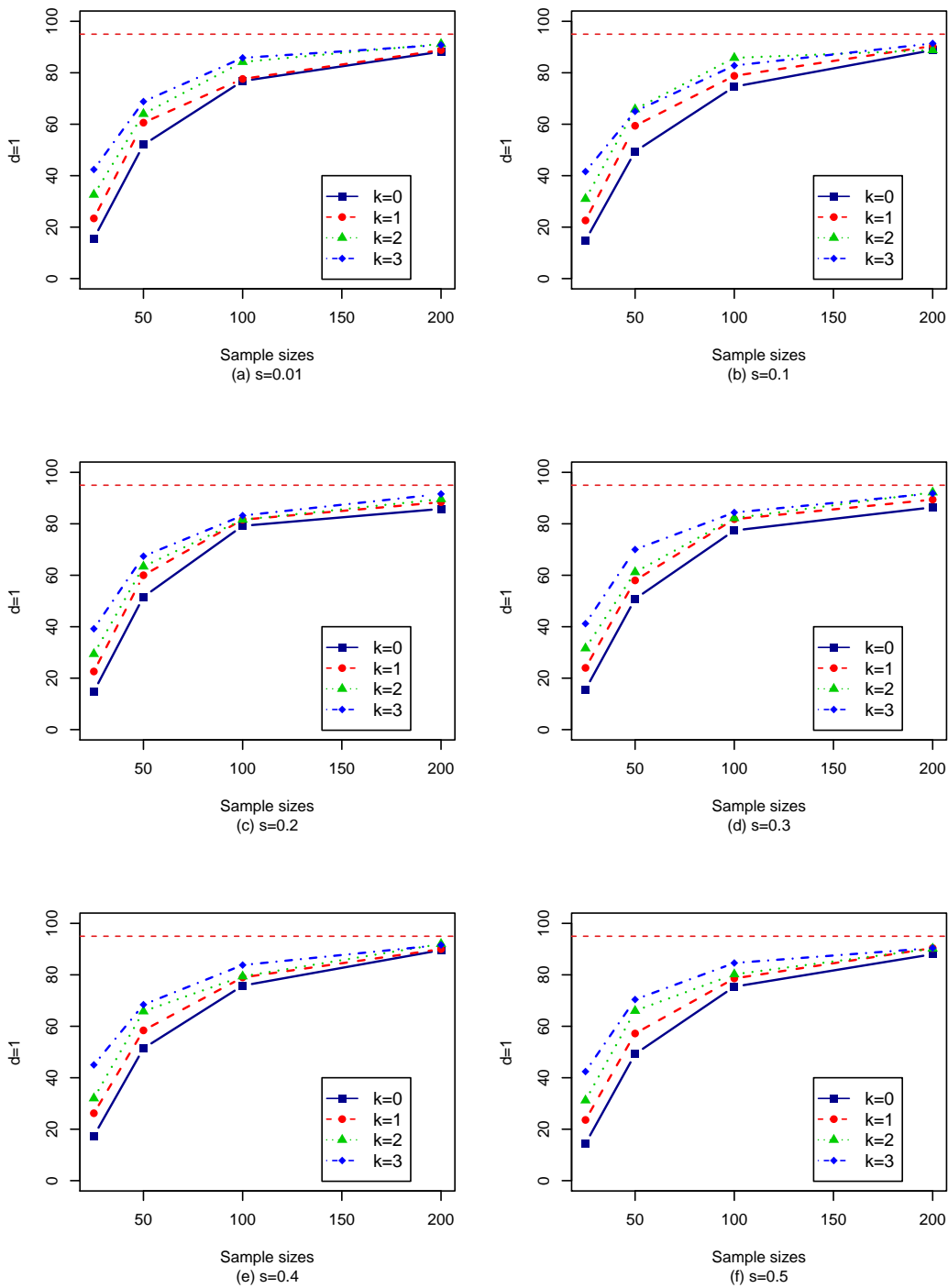


Figure 2: Model 1: percentages of the decision that $\hat{d} = 1$; red dashed-line, 95% reference.

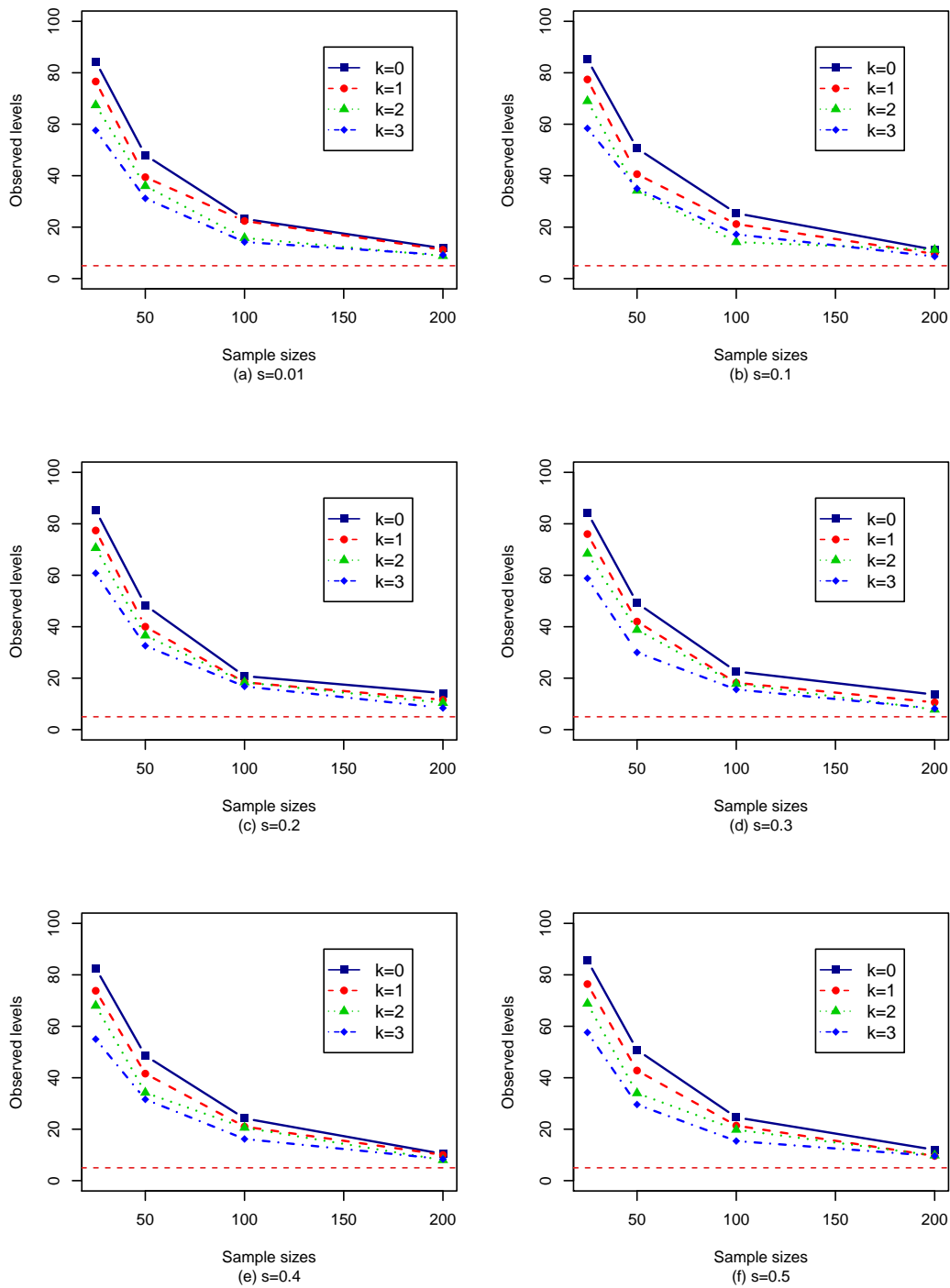


Figure 3: Model 1: observed levels; red dashed-line, 5% reference.

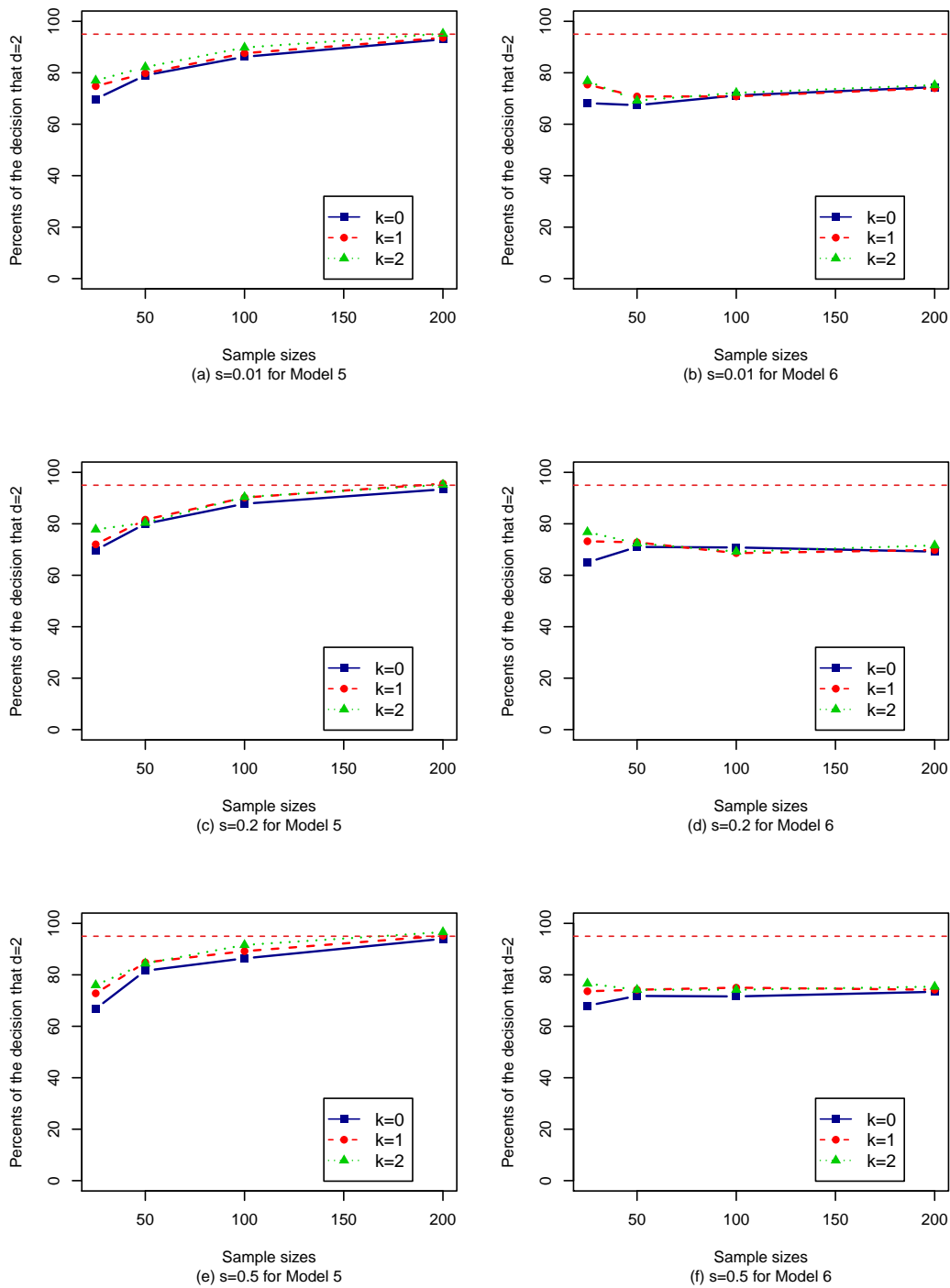


Figure 4: Models 5 and 6: percentages of the decision that $\hat{d} = 2$; red dashed-line, 95% reference.

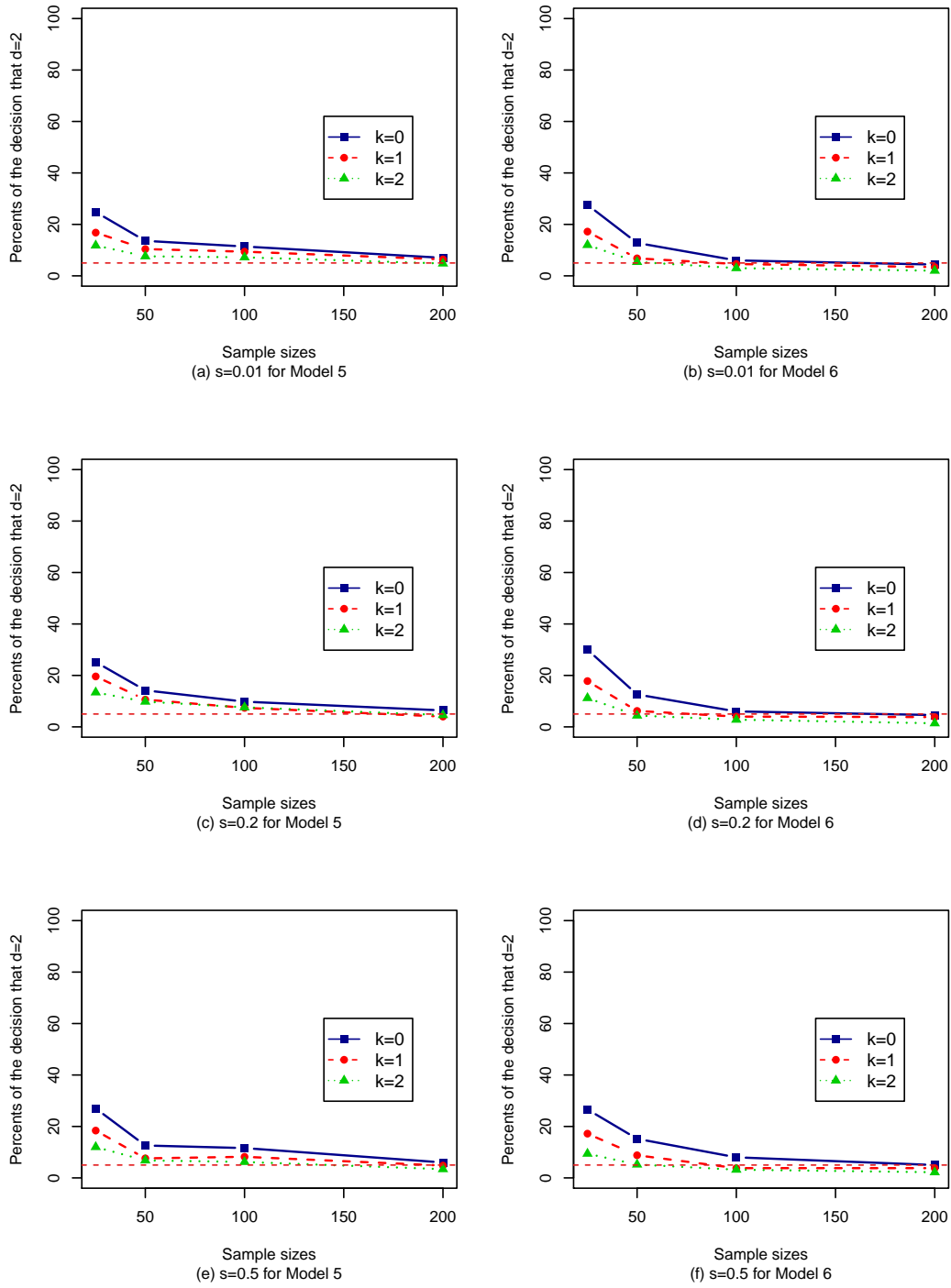


Figure 5: Models 5 and 6: observed levels; red dashed-line, 5% reference.

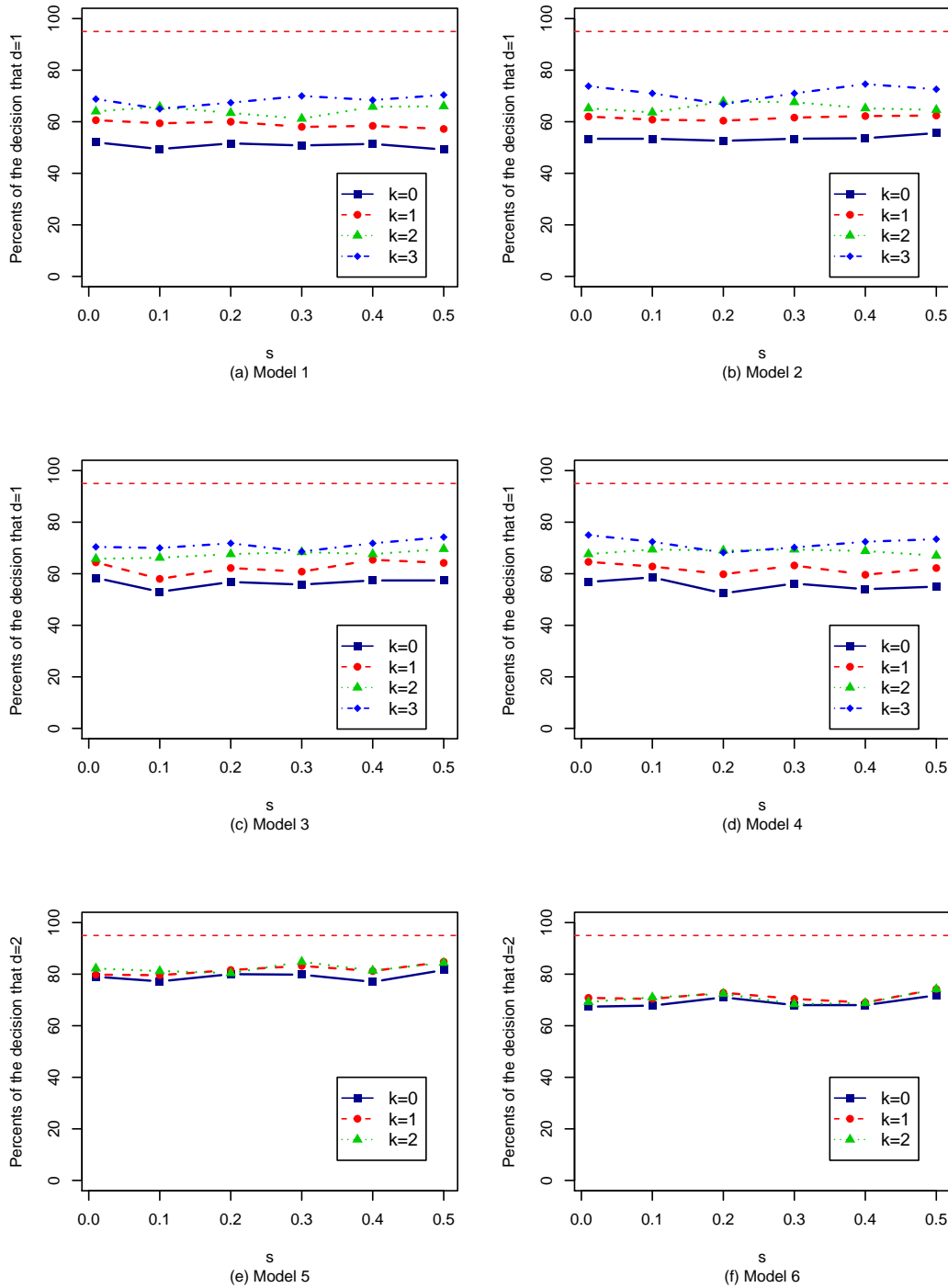


Figure 6: Percentages of the decision that $\hat{d} = d$ varying s with $n = 50$; red dashed-line, 95% reference.

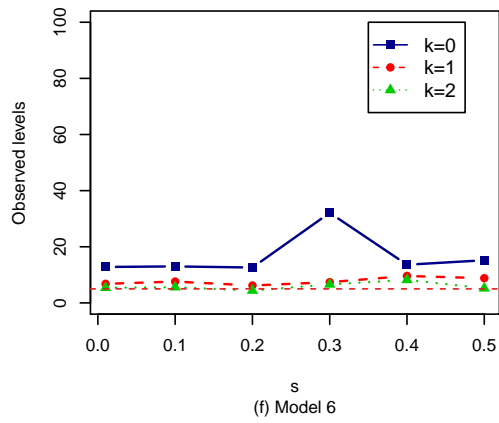
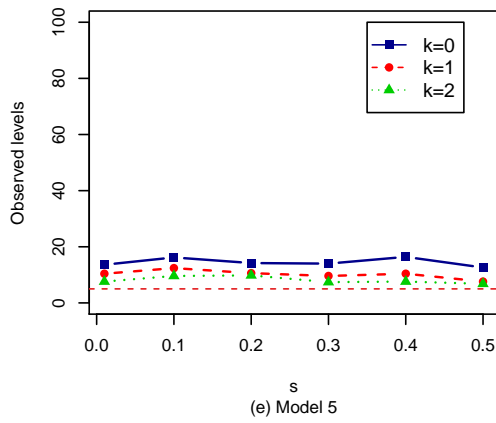
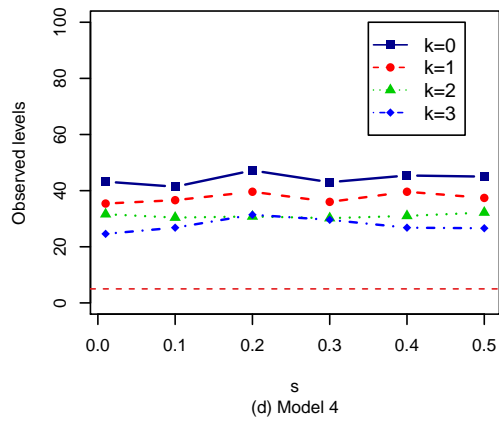
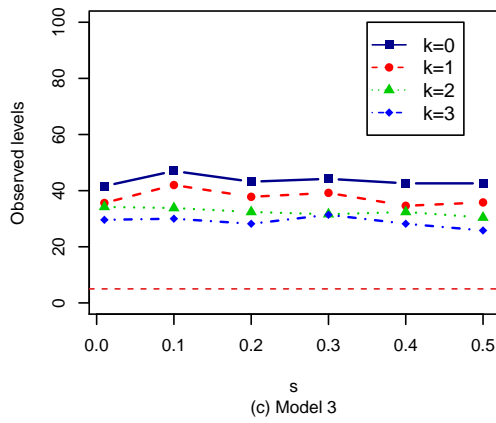
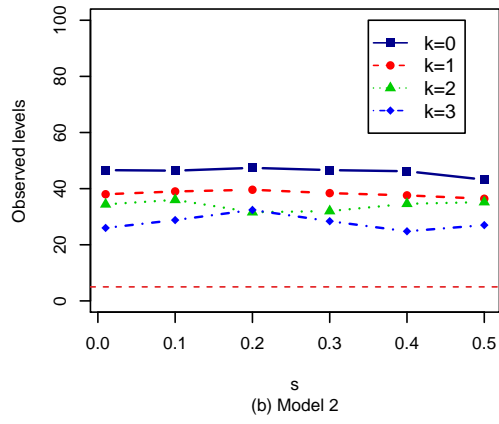
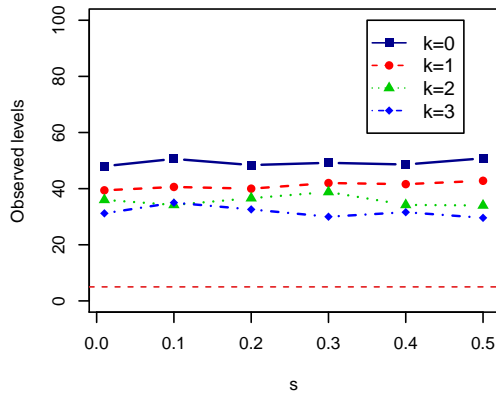


Figure 7: Observed levels varying s with $n = 50$; red dashed-line, 5% reference.

According to Yoo and Cook (2007), the optimal sufficient dimension approach shows better estimation results using the OLS to estimate $S_{Y|X}$. A possible problem arises, if the inner-product matrix in the objective function is singular. Yoo (2015) derives theoretical results for the optimal SDR with a singular inner-product matrix. However, Yoo (2015) lacks intensive numerical studies and a real data example. Therefore, various simulated models and a real data analysis should be investigated to show the practical usefulness of Yoo (2015).

In this paper, intensive numerical studies are done to show the potential advantages of Yoo (2015) over existing Yoo and Cook (2007). Especially, in the dimension estimation, Yoo (2015) shows better small sample results than Yoo and Cook (2007). A real data example also shows that the overestimation of Yoo and Cook (2007) due to small sample sizes and high correlation between responses can be prevented by using Yoo (2015). This confirms the practical usefulness of Yoo (2015).

Acknowledgments

For the corresponding author Jae Keun Yoo, this work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korean Ministry of Education (NRF-2014R1A2A1A11049389 and 2009-0093827). For Da-Hae Gwak and Min-Sun Kim, this work was supported by the BK21 Plus Project through the National Research Foundation of Korea (NRF) funded by the Korean Ministry of Education (22A20130011003).

References

- Chiaromonte F and Martinelli J (2002). Dimension reduction strategies for analyzing global gene expression data with a response, *Mathematical Biosciences*, **176**, 123–144.
- Cook RD (2003). Dimension reduction and graphical exploration in regression including survival analysis, *Statistics in Medicine*, **22**, 1399–1413.
- Cook RD and Ni L (2005). Sufficient dimension reduction via inverse regression: a minimum discrepancy approach, *Journal of the American Statistical Association*, **100**, 410–428.
- Cook RD and Zhang X (2014). Fused estimators of the central subspace in sufficient dimension reduction, *Journal of the American Statistical Association*, **109**, 815–827.
- Li L (2006). Survival prediction of diffuse large-B-cell lymphoma based on both clinical and gene expression information, *Bioinformatics*, **22**, 466–471.
- Shapiro A (1986). Asymptotic theory of overparameterized structural models, *Journal of the American Statistical Association*, **81**, 142–149.
- Yoo JK (2009). Iterative optimal sufficient dimension reduction for the conditional mean in multivariate regression, *Journal of Data Science*, **7**, 267–276.
- Yoo JK (2015). A theoretical note on optimal sufficient dimension reduction with singularity, *Statistics and Probability Letters*, **99**, 109–113.
- Yoo JK (2016a). Tutorial: dimension reduction in regression with a notion of sufficiency, *Communications for Statistical Applications and Methods*, **23**, 93–103.
- Yoo JK (2016b). Tutorial: Methodologies for sufficient dimension reduction in regression, *Communications for Statistical Applications and Methods*, **23**, 105–117.
- Yoo JK and Cook RD (2007). Optimal sufficient dimension reduction for the conditional mean in multivariate regression, *Biometrika*, **94**, 231–242.