SUPPLEMENTARY MATERIALS

Covariate Information Matrix for Sufficient Dimension Reduction

Weixin Yao[*],
1337 Olmsted Hall, Department of Statistics,
University of California, Riverside, Riverside, CA 92521
Debmalya Nandy[†], **Bruce G. Lindsay**[‡] and **Francesca Chiaromonte**[§¶]
325 Thomas Building, Department of Statistics,
Penn State University, University Park, PA 16802

[*]W. Yao is the corresponding author and was supported by NSF grant DMS-1461677.

[†]D. Nandy and W. Yao contributed equally to this work and are jointly first-authors of this article.

[‡]During the development of this work, B. G. Lindsay passed away due to an illness. We honor the contributions of our dear friend, generous mentor and brilliant colleague.

[§]F. Chiaromonte, B. G. Lindsay and D. Nandy were supported by NSF grant DMS-1407639. F. Chiaromonte is also affiliated with the Institute of Economics and EMbeDS, Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà 33, Pisa, Italy 56127.

[¶]We gratefully acknowledge Drs. Subir Ghosh and Zhiwei Zhang of the University of California Riverside, Zejin Gao of University of Wisconsin-Madison, and Drs. Bing Li, Yanyuan Ma, Matthew Reimherr and Lingzhou Xue of Penn State, for their useful comments. We acknowledge Dr. Nikolay V. Balashov of Penn State for clarifications on the Ozone data application. We are also grateful to Drs. Yanyuan Ma (Penn State), Liping Zhu (Shanghai University) and Peng Zeng (Auburn University) for generously sharing codes respectively for the semiparametric, dMAVE and Fourier Sufficient Dimension Reduction methods. Our special thanks go to Dr. Yingcun Xia (National University of Singapore) for sharing code on an extension of the SR method still in preparation, and for his insightful responses to several questions. Ge Zhao, an advisee of Dr. Ma in the Statistics graduate program at Penn State, was instrumental in adapting FORTRAN code for our simulations and real data application. Finally, we are in debt to anonymous reviewers and editors whose comments helped us greatly improve our work.

1 Proofs of Propositions

1.1 Proposition 3.1

Proof. Let the conditional density of $Y \mid \mathbf{X}$ be $f(y|\mathbf{x}) = f(y|\boldsymbol{\beta}_1^T\mathbf{x}, \ldots, \boldsymbol{\beta}_d^T\mathbf{x})$, where the $p \times d$ matrix $\mathbf{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d)$ contains the spanning vectors for the CD, and $\mathbf{B}^T\mathbf{B} = I$. Then

$$\nabla_{\mathbf{x}} \log f(y|\mathbf{x}) = \sum_{j=1}^{d} g_j(y|\boldsymbol{\beta}_1^T\mathbf{x}, \ldots, \boldsymbol{\beta}_d^T\mathbf{x})\boldsymbol{\beta}_j$$

where

$$g_j(y|t_1, \ldots, t_d) = \frac{\partial \log f(y|t_1, \ldots, t_d)}{\partial t_j} = \frac{1}{f(y|t_1, \ldots, t_d)} \frac{\partial f(y|t_1, \ldots, t_d)}{\partial t_j}.$$

Therefore

$$\begin{aligned}
\mathbb{C}_{\mathbf{X}} &= \int f(\mathbf{x}, y) \sum_{j=1}^{d} \sum_{k=1}^{d} \left\{ g_j(y|\boldsymbol{\beta}_1^T\mathbf{x}, \ldots, \boldsymbol{\beta}_d^T\mathbf{x}) g_k(y|\boldsymbol{\beta}_1^T\mathbf{x}, \ldots, \boldsymbol{\beta}_d^T\mathbf{x}) \boldsymbol{\beta}_j \boldsymbol{\beta}_k^T \right\} d\mathbf{x}dy \\
&= \sum_{j=1}^{d} \sum_{k=1}^{d} E\left\{ g_j(y|\boldsymbol{\beta}_1^T\mathbf{x}, \ldots, \boldsymbol{\beta}_d^T\mathbf{x}) g_k(y|\boldsymbol{\beta}_1^T\mathbf{x}, \ldots, \boldsymbol{\beta}_d^T\mathbf{x}) \right\} \boldsymbol{\beta}_j \boldsymbol{\beta}_k^T \\
&= (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d)\mathbf{G}(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d)^T
\end{aligned}$$

where $\mathbf{G}$ is a $d \times d$ matrix with $(j,k)^{th}$ element $E\left\{ g_j(y|\boldsymbol{\beta}_1^T\mathbf{x}, \ldots, \boldsymbol{\beta}_d^T\mathbf{x}) g_k(y|\boldsymbol{\beta}_1^T\mathbf{x}, \ldots, \boldsymbol{\beta}_d^T\mathbf{x}) \right\}$. Thus, the rank of $\mathbb{C}_{\mathbf{X}}$ is equal to the rank of $\mathbf{G}$. The rank of $\mathbf{G}$ will be less than d if and only if $g_1(y|\boldsymbol{\beta}_1^T\mathbf{x}, \ldots, \boldsymbol{\beta}_d^T\mathbf{x}), \ldots, g_d(y|\boldsymbol{\beta}_1^T\mathbf{x}, \ldots, \boldsymbol{\beta}_d^T\mathbf{x})$ are linearly dependent, which occurs if and only if the rank of the CS is less than d. Further, if we diagonalize $\mathbf{G}$ as $\mathbf{G} = \Phi\Lambda\Phi^T$, we can rewrite $\mathbb{C}_{\mathbf{X}} = \mathbf{B}^*\Lambda\mathbf{B}^{*T}$, where the columns of $\mathbf{B}^* = \mathbf{B}\Phi$ are the eigen-vectors of $\mathbb{C}_{\mathbf{X}}$. In addition, the space spanned by $\mathbf{B}^*$ is same as the space spanned by $\mathbf{B}$, i.e. the CS. $\square$

1.2 Proposition 3.2

Proof. Let $\mathbf{W} = A\mathbf{X} + a$, and assume A is invertible. Then the density of $\mathbf{W}$ is $f_{\mathbf{W}}(\mathbf{w}) = f_{\mathbf{X}}(A^{-1}(\mathbf{w} - a))|\det A^{-1}|$, where $\det A^{-1}$ is the determinant of A^{-1}. This implies $\nabla_{\mathbf{w}} \log f_{\mathbf{W}}(\mathbf{w}) = A^{-T}\nabla_{\mathbf{x}} \log f_{\mathbf{X}}(A^{-1}(\mathbf{w} - a))$. Therefore, the DIM for the transformed variable $\mathbf{W}$ is

$$\mathbb{J}_{\mathbf{W}} = A^{-T}\mathbb{J}_{\mathbf{X}}A^{-1}. \tag{1}$$

Using Equation (1) we get $\mathbb{J}_{\mathbf{W}|Y} - \mathbb{J}_{\mathbf{W}} = A^{-T}(\mathbb{J}_{\mathbf{X}|Y} - \mathbb{J}_{\mathbf{X}})A^{-1}$. Therefore, Equation (3) in Proposition 3.4 gives $\mathbb{C}_{\mathbf{W}} = A^{-T}\mathbb{C}_{\mathbf{X}}A^{-1}$. $\square$

1.3 Proposition 3.3

Proof. Use Proposition 3.2 to write

$$\mathbb{C}_{\tilde{\mathbf{Z}}} = (\Gamma_{\mathbf{Z}}^T\Sigma_{\mathbf{X}}^{-1/2})^{-T}\mathbb{C}_{\mathbf{X}}(\Gamma_{\mathbf{Z}}^T\Sigma_{\mathbf{X}}^{-1/2})^{-1} = \Gamma_{\mathbf{Z}}^T\Sigma_{\mathbf{X}}^{1/2}\mathbb{C}_{\mathbf{X}}\Sigma_{\mathbf{X}}^{1/2}\Gamma_{\mathbf{Z}} = \Lambda_{\mathbf{Z}}.$$

Therefore,

$$\mathbb{C}_{\mathbf{X}}\Sigma_{\mathbf{X}}^{1/2}\Gamma_{\mathbf{Z}} = (\Gamma_{\mathbf{Z}}^T\Sigma_{\mathbf{X}}^{1/2})^{-1}\Lambda_{\mathbf{Z}} = \Sigma_{\mathbf{X}}^{-1/2}\Gamma_{\mathbf{Z}}\Lambda_{\mathbf{Z}};$$

so that

$$\mathbb{C}_{\mathbf{X}}\Sigma_{\mathbf{X}}(\Sigma_{\mathbf{X}}^{-1/2}\Gamma_{\mathbf{Z}}) = (\Sigma_{\mathbf{X}}^{-1/2}\Gamma_{\mathbf{Z}})\Lambda_{\mathbf{Z}}.$$

Since $\Lambda_{\mathbf{Z}}$ is diagonal, $G = \Sigma_{\mathbf{X}}^{-1/2}\Gamma_{\mathbf{Z}}$ corresponds to the right eigen-vectors of $\mathbb{C}_{\mathbf{X}}\Sigma_{\mathbf{X}}$.

$\square$

1.4 Proposition 3.4

Proof. We apply the standard definition of DIM for the joint density $f(y, \mathbf{x})$ of the vector $(Y, \mathbf{X}^T)^T \in \mathbb{R}^{p+1}$ to obtain $\mathbb{J}_{Y,\mathbf{X}}$. Next we consider the lower right $p \times p$ sub-matrix of $\mathbb{J}_{Y,\mathbf{X}}$; that is

$$\mathbb{J}_{Y,\mathbf{X}}(2,2) = \int\int [\nabla_{\mathbf{x}} \log f(y, \mathbf{x})][\nabla_{\mathbf{x}} \log f(y, \mathbf{x})]^T f(y, \mathbf{x}) dy d\mathbf{x}.$$

Since $\nabla_{\mathbf{x}} \log f(y, \mathbf{x}) = \nabla_{\mathbf{x}} \log f^{(y)}(\mathbf{x})$, we have

$$\begin{aligned} \mathbb{J}_{Y,\mathbf{X}}(2,2) &= \int\int [\nabla_{\mathbf{x}} \log f^{(y)}(\mathbf{x})][\nabla_{\mathbf{x}} \log f^{(y)}(\mathbf{x})]^T f^{(y)}(\mathbf{x}) f(y) d\mathbf{x} dy \\ &= \int \mathbb{J}_{\mathbf{X}|Y=y} f(y) dy = \mathbb{J}_{\mathbf{X}|Y}. \end{aligned} \tag{2}$$

Next, we apply the orthogonal score decomposition (see Section 3.2 in Lindsay and Yao (2012)) $\nabla_{\mathbf{x}} \log f(y, \mathbf{x}) = \nabla_{\mathbf{x}} \log f^{(y)}(\mathbf{x}) + \nabla_{\mathbf{x}} \log f(\mathbf{x})$ to show that

$$\mathbb{J}_{Y,\mathbf{X}}(2,2) = \mathbb{C}_{\mathbf{X}} + \mathbb{J}_{\mathbf{X}}. \tag{3}$$

Equations (2) and (3) lead to the statement of the proposition.

□

1.5 Proposition 4.1

Proof. Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two subspaces of $\mathbb{R}^p$, both of dimension q. Then $p-q$ is the dimension of their null spaces $\mathcal{S}_1^{\perp}$ and $\mathcal{S}_2^{\perp}$. We have

$$\begin{aligned}
R_o^2(\mathcal{S}_1, \mathcal{S}_2) &= \frac{1}{p-q}\mathrm{tr}(Q_{\mathcal{S}_1}Q_{\mathcal{S}_2}) = \frac{1}{p-q}\mathrm{tr}((I_p - P_{\mathcal{S}_1})(I_p - P_{\mathcal{S}_2})) \\
&= \frac{1}{p-q}(p-q-q+\mathrm{tr}(P_{\mathcal{S}_1}P_{\mathcal{S}_2})) = \frac{1}{p-q}(p-2q+qR^2(\mathcal{S}_1, \mathcal{S}_2)) \\
&= \frac{1}{p-q}(p-q-q(1-R^2(\mathcal{S}_1, \mathcal{S}_2))) = 1 - \frac{q}{p-q}(1-R^2(\mathcal{S}_1, \mathcal{S}_2)).
\end{aligned}$$

□

1.6 Proposition 4.2

Lemma: Let $\mathcal{S}$ be a subspace of $\mathbb{R}^p$ with dimension $d \le p$, and let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two subspaces of $\mathcal{S}$, both of dimension q. Also let $P_{\mathcal{S}_1}$ and $P_{\mathcal{S}_2}$ be two orthogonal projection matrices onto $\mathcal{S}_1$ and $\mathcal{S}_2$ respectively. If there are k $(\le q)$ common bases directions between $\mathcal{S}_1$ and $\mathcal{S}_2$, then $\mathrm{tr}(P_{\mathcal{S}_1}P_{\mathcal{S}_2}) = k$.

Proof. Without loss of generality, let us consider the standard basis $\{e_1, \ldots, e_p\}$ of $\mathbb{R}^p$ (each $e_i \in \mathbb{R}^p$ has a 1 in its i^{th} component and 0's everywhere else) and assume that the k common directions between $\mathcal{S}_1$ and $\mathcal{S}_2$ are $\{e_1, \ldots, e_k\}$. Note that $P_{\mathcal{S}_1}P_{\mathcal{S}_2}e_j = P_{\mathcal{S}_1}e_j = e_j$ for $j = 1, \ldots, k$. For $j = k+1, \ldots, p$, if $P_{\mathcal{S}_1}$ contains the direction e_j but $P_{\mathcal{S}_2}$ does not, then $P_{\mathcal{S}_1}P_{\mathcal{S}_2}e_j = 0$; if $P_{\mathcal{S}_1}$ does not contain e_j but $P_{\mathcal{S}_2}$ does, then $P_{\mathcal{S}_1}P_{\mathcal{S}_2}e_j = P_{\mathcal{S}_1}e_j = 0$; if neither $P_{\mathcal{S}_1}$ nor $P_{\mathcal{S}_2}$ contains e_j, then $P_{\mathcal{S}_1}P_{\mathcal{S}_2}e_j = 0$. Therefore, $P_{\mathcal{S}_1}P_{\mathcal{S}_2}e_j = 0$ for all $j = k+1, \ldots, p$. From the above argument, the matrix $P_{\mathcal{S}_1}P_{\mathcal{S}_2}$ has eigenvalue 1 with multiplicity k, and eigenvalue 0 with multiplicity $p-k$. Therefore, $\mathrm{tr}(P_{\mathcal{S}_1}P_{\mathcal{S}_2}) = k$. □

Proof of Proposition 4.2:

Proof. Let $p_k = \Pr[\mathcal{S}_1$ and $\mathcal{S}_2$ have exactly k directions in common$]$. Note that, using the above Lemma, $E[R^2(\mathcal{S}_1, \mathcal{S}_2)] = E[\frac{1}{q}\mathrm{tr}(P_{\mathcal{S}_1}P_{\mathcal{S}_2})] = \frac{1}{q}\sum_{k=1}^{q}\mathrm{tr}(P_{\mathcal{S}_1}P_{\mathcal{S}_2})p_k = \frac{1}{q}\sum_{k=1}^{q}kp_k$.

Out of d bases directions, q directions in $\mathcal{S}_1$ and $\mathcal{S}_2$ can be chosen in $C_1 = \binom{d}{q}\binom{d}{q}$ ways. For the favorable cases, we proceed as follows: fix one between $\mathcal{S}_1$ and $\mathcal{S}_2$, in this subspace q out of d directions can be chosen in $\binom{d}{q}$ ways. Then, for the other subspace, set aside k common directions, which can be chosen in $\binom{q}{k}$ ways. The rest of the $(q-k)$ directions can be chosen from $(d-q)$ directions in $\binom{d-q}{q-k}$ ways. Hence, the total number of favorable cases is $C_2 = \binom{d}{q}\binom{q}{k}\binom{d-q}{q-k}$. It follows that

$$p_k = \frac{C_2}{C_1} = \frac{\binom{q}{k}\binom{d-q}{q-k}}{\binom{d}{q}}, \quad k \in \{max(0, 2q-d), \ldots, q\}. \tag{4}$$

Equation (4) is an hypergeometric probability distribution with parameters d (which represents the population size) and q (which represents the number of draws without replacement, as well as the exact number of 'successes' in the population in our case). Given that $\mathrm{tr}(P_{\mathcal{S}_1}P_{\mathcal{S}_2}) = k$ with k common directions, $E[\mathrm{tr}(P_{\mathcal{S}_1}P_{\mathcal{S}_2})]$ is simply the expectation of such hypergeometric distribution, i.e.

$$E[R^2(\mathcal{S}_1, \mathcal{S}_2)] = \frac{1}{q}E[\mathrm{tr}(P_{\mathcal{S}_1}P_{\mathcal{S}_2})] = \frac{1}{q}[\frac{q^2}{d}] = \frac{q}{d}$$

which is an increasing function of q. □

The following two sections include the motivation behind the Density Information Matrix (DIM) and the Covariate Information Matrix (CIM) and has been added based on reviewers' comments.

2 Additional Background/Motivation about Density Information Matrix and Covariate Information Matrix

In this section, we will introduce a unified framework called the *Fisher Discrimination Matrix* (FDM) (Zhou, 2017), that can provide the motivation behind the density information matrix and covariate information matrix. FDM can be used to find the best directions to discriminate between two densities based on a simple eigen-analysis. Suppose we want to compare two possible densities $f(\mathbf{x})$ and $g(\mathbf{x})$, respectively, for a random variable $\mathbf{X}_{p\times 1}$. In the applications, we will consider the density f to be the true unknown density which is estimated via non-parametric methods, and the density g will be some model for the data, parametric or semi-parametric. FDM then finds the best directions of $\mathbf{X}$ that violate the model assumption of g. Different model assumptions of g result in different applications of the FDM. Alternatively, f and g can represent two distinct populations we wish to compare.

Define the *sample space score vector* $\mathbf{u}_f(\mathbf{x})$ for a density f to be $\mathbf{u}_f(\mathbf{x}) = \nabla_{\mathbf{x}} \log f(\mathbf{x})$ and the basic *discrimination score* for comparing f and g to be $\mathbf{u}_f(\mathbf{x}) - \mathbf{u}_g(\mathbf{x})$. We define the discrimination information matrix to be the matrix quadratic form in the discrimination scores, given by

$$\mathbb{D}_w(f, g) = \int (\mathbf{u}_f - \mathbf{u}_g)(\mathbf{u}_f - \mathbf{u}_g)^T w(\mathbf{x}) d\mathbf{x}, \tag{5}$$

where $w(\mathbf{x})$ is a context-specific weighting density that can be e.g. f, g, or some hybrid of f and g. We will sometimes denote this matrix as $\mathbb{D}_{\mathbf{X}} = \mathbb{D}_w(f, g)$, so as to indicate which random variable is under consideration. This matrix summarizes the local discrimination directions for separating f and g, and will be $\mathbf{0}_{p\times p}$ if and only if $f = g$, assuming w has full support on $\mathbb{R}^p$. The eigen analysis of $\mathbb{D}_w(f, g)$ can give us the linear directions that best discriminate between densities f and g. The trace of the matrix $\mathbb{D}$ provides a measure of the disagreement between f and g. When f and g are normal densities, this matrix is $\mathbb{D}_w(f, g) = \Sigma^{-1}(\mu_f - \mu_g) \cdot (\mu_f - \mu_g)^T \Sigma^{-1}$ with rank

1, and the non-null eigen-vector is the linear discriminant $\Sigma^{-1}(\mu_f - \mu_g)$. However, using more than one direction from the eigen-analysis of $\mathbb{D}_w(f, g)$, we can move beyond the traditional linear discriminant analysis to *multiple linear discriminants*.

We start by identifying the concept of *sufficiency* in this context. Let's assume, $\mathbf{X} = (\mathbf{X}_1^T, \mathbf{X}_2^T)^T$. The conventional definition of sufficiency indicates that $\mathbf{X}_1$ is sufficient for comparing f and g if the conditional densities for f and g are the same:

$$f(\mathbf{x}_2|\mathbf{x}_1) = g(\mathbf{x}_2|\mathbf{x}_1).$$

So, $\log f(\mathbf{x}) - \log g(\mathbf{x}) = \log f_1(\mathbf{x}_1) - \log g_1(\mathbf{x}_1)$, where f_1 and g_1 are marginal densities of $\mathbf{X}_1$. Hence, the discriminant function only depends on $\mathbf{X}_1$. In this case, it is obvious that

$$\mathbb{D}_w(f, g) = \begin{bmatrix} \mathbb{D}_w(f_1, g_1) & 0 \\ 0 & 0 \end{bmatrix}.$$

That is, the discrimination information matrix, via its positive information, identifies the variables that are sufficient for discriminating between f and g, as well as the set of variables in the null space that are ignorable. More generally, we will say that the zero eigen-values of $\mathbb{D}_w(f, g)$ correspond to an *ignorable subspace*, one that is irrelevant in the comparison of f and g, and its orthogonal complement space is the *sufficient* subspace.

2.1 Motivation for DIM

Friedman and Tukey (1974) developed projection pursuit to explore a high-dimensional data by examining the marginal distributions of low-dimensional linear projections. As argued later by Huber (1985) and Jones and Sibson (1987), the Gaussian distribution is the least interesting one, and therefore, the most interesting directions should be those that show the least normality. One drawback of the resulting methods has been their high computational cost, especially for high-dimensional data, due to the need to search through the large number of possible projections.

We introduce a simple way to perform projection pursuit based on the proposed discrimination matrix. Let $g = \phi(\mathbf{x})$ be the standard multivariate normal density (also called *white noise density* in Hui and Lindsay (2010)), which refers to the null hypothesis that all variables are uninteresting, and $f(\mathbf{x})$ be the true density of $\mathbf{X}$, to be estimated nonparametrically. If $w(\mathbf{x}) = f(\mathbf{x})$, Equation (5) gives

$$\mathbb{D}_f(f, g) = \int (\mathbf{u}_f - \mathbf{u}_g)(\mathbf{u}_f - \mathbf{u}_g)^T f(\mathbf{x}) d\mathbf{x} = \mathbb{J}_f - I_p,$$

where

$$\mathbb{J}_f = \mathrm{E}_f\{\mathbf{u}_f(\mathbf{X})\mathbf{u}_f(\mathbf{X})^T\},$$

where $\mathbb{J}_f$ is the so called *Density Information Matrix* for f defined in Section 2.1, and use the notation $\mathbb{J}_{\mathbf{X}}$, expressed with a random variable instead of $\mathbb{J}_f$ using f (density of $\mathbf{X}$), when it is useful for clarity regarding different variables involved.

Since g is the standard multivariate normal density, the task of projection pursuit can be considered as finding the directions that can best discriminate between g and f. Therefore, the null space of $\mathbb{D}_f(f, g)$ provides the uninteresting directions, and its orthogonal complement provides the interesting directions. In addition, the null space of $\mathbb{D}_f(f, g)$ is the same as the eigen-space of $\mathbb{J}_f$ associated with eigenvalue 1. Therefore, a simple eigen-analysis of $\mathbb{J}_f$ can be used for the projection pursuit that reveals interesting directions of the original multivariate variable $\mathbf{X}$.

Suppose $\Gamma^T \mathbb{J}_f \Gamma = \Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_p\}$. WLOG, we assume $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. Then $\mathbf{Z} = \Gamma^T \mathbf{X} = (Z_1, \ldots, Z_p)^T$ has sample space information matrix Λ (assuming $\mathbf{X}$ is standardized), and the diagonal entries of Λ measure the information in each $\mathbf{Z}$ coordinate in terms of extent of discrimination between $\mathbf{Z}$ and the normal density. We then use the uncorrelated $\mathbf{Z}$ coordinates with the greatest information in the data analysis. Equivalently, we are discarding the $\mathbf{Z}$ coordinates with the least information. More specifically, if $\lambda_{k+1} = 1$, then $\lambda_j = 1$, for $j \geq k+1$. Therefore, $Z_{k+1}, \ldots, Z_p$ are ignorable coordinates (also called *white noise coordinates* in Hui and Lindsay (2010)) in the comparison of f and the white noise density g. Therefore, we can simply use the "informative" projection coordinates $Z_1, Z_2, \ldots, Z_k$ for further data analysis. In practice, we might also just use a smaller set of projected variable $\{Z_1, \ldots, Z_m\}$, where $m < k$ (such as $m = 1$ or 2), and call them as the m *most informative projected variables*. Hui and Lindsay (2010) provided more detailed discussion of the applications of $\mathbb{J}_f$ for projection pursuit and called it a *white noise matrix*.

Lindsay and Yao (2012) discussed more applications of $\mathbb{J}_f$, such as Independent Component Analysis (ICA) models and graphical models. If $\mathbf{X}$ is generated by an ICA model (Jutten and Herault, 1991; Comon, 1994), Lindsay and Yao (2012) proved that the transformed variables $\mathbf{Z} = \Gamma^T\mathbf{X}$, given in the previous paragraph, are the independent component variables. Therefore, the proposed discrimination matrix can be also used as an alternative tool for ICA. The ICA is closely related to the famous *cocktail-party problem* and the related methods are sometimes called *blind source separation* or *blind signal separation*. See Hyvärinen and Oja (2000) for an introduction to ICA model and its applications. In addition, Lindsay and Yao (2012) also proved that the conditional independence of $\mathbf{X}$ implies the corresponding off-diagonal zero values of $\mathbb{J}_f$. Thus, the proposed discrimination matrix FDM can also be applied to graphical models.

2.2 Motivation for CIM

In a regression setting, suppose we have a univariate response variable Y, and p-dimensional covariate $\mathbf{X}_{p\times 1}$. In order to find the informative directions of $\mathbf{X}$ related to Y, let's consider the inverse regression/conditional density $f(\mathbf{x}|y)$. Under the null hypothesis that $\mathbf{X}$ is independent of

Y, $g = f(\mathbf{x})$ in (5). Then the discrimination information matrix, with $w = f(y, \mathbf{x})$, is

$$\mathbb{D}_w(f, g) = \int \{\nabla_{\mathbf{x}} \log f(\mathbf{x}|y) - \nabla_{\mathbf{x}} \log f(\mathbf{x})\} \{\nabla_{\mathbf{x}} \log f(\mathbf{x}|y) - \nabla_{\mathbf{x}} \log f(\mathbf{x})\}^T f(y, \mathbf{x}) dy d\mathbf{x}.$$

An eigen-analysis of $\mathbb{D}_w(f, g)$ would reveal the directions of $\mathbf{X}$ that can best discriminate between $f(\mathbf{x}|y)$ and $f(\mathbf{x})$, i.e. the directions of $\mathbf{X}$ that can best explain Y. The null space of $\mathbb{D}_w(f, g)$ reveals the directions of $\mathbf{X}$ that are not related to Y.

In addition, note that

$$\begin{aligned}
&\nabla_{\mathbf{x}} \log f(y, \mathbf{x}) = \nabla_{\mathbf{x}} \log f(\mathbf{x}|y), \\
&\nabla_{\mathbf{x}} \log f(y, \mathbf{x}) = \nabla_{\mathbf{x}} \log(f(y|\mathbf{x})) + \nabla_{\mathbf{x}} \log(f(\mathbf{x})).
\end{aligned}$$

Therefore,

$$\mathbb{D}_w(f, g) = \int \nabla_{\mathbf{x}} \log f(y|\mathbf{x}) \nabla_{\mathbf{x}}^T \log f(y|\mathbf{x}) f(y, \mathbf{x}) dy d\mathbf{x} = \int \mathbb{F}_{\mathbf{x}} f(\mathbf{x}) d\mathbf{x} \stackrel{\text{def}}{=} \mathbb{C}_{\mathbf{X}},$$

where

$$\begin{aligned}
\mathbb{F}_{\mathbf{x}} &= \mathrm{E}_{Y|\mathbf{x}} \left\{\nabla_{\mathbf{x}} \log f(Y|\mathbf{x}) \nabla_{\mathbf{x}}^T \log(f(Y|\mathbf{x})\right\} \\
&= \int \left[\nabla_{\mathbf{x}} \log f(y|\mathbf{x}) \nabla_{\mathbf{x}}^T \log(f(y|\mathbf{x})\right] f(y \mid \mathbf{x}) dy.
\end{aligned}$$

We can think of $\mathbb{F}_{\mathbf{x}}$ as the *conditional covariate information matrix* for $Y|\mathbf{x}$, considering $\mathbf{x}$ as "parameter" in the $f(y|\mathbf{x})$ distribution. $\mathbb{F}_{\mathbf{x}}$ tells us how much Fisher information about $\mathbf{x}$ exists when we take a single observation of Y. It is a local measure in the sense that it depends on $\mathbf{x}$. Of course, it is a bit unconventional that $\mathbb{F}_{\mathbf{x}}$ measures the Fisher information in an observed "parameter", and not in an unknown parameter. However, it does provide a natural way of assessing how sensitive the distribution of $Y|\mathbf{x}$ is to changes in $\mathbf{x}$. Henceforth, we will call $\mathbb{C}_{\mathbf{X}} = \mathrm{E}\{\mathbb{F}_{\mathbf{x}}\}$ the *(mean) Covariate Information Matrix* (CIM), which can be used to perform dimension reduction.

3 Insight on the CIM for a binary response

Consider the case of a binary response $Y \in \{y_1, y_2\}$. Here $f(\mathbf{x}) = \pi_1 f(\mathbf{x} \mid y_1) + \pi_2 f(\mathbf{x} \mid y_2)$. For $j = 1, 2$, simplifying notation, let $\pi_j(\mathbf{x}) = P(Y = y_j \mid \mathbf{x})$ and $U_j(\mathbf{x}) = \nabla_{\mathbf{x}} \log f(\mathbf{x} \mid y_j)$. With manipulations similar to those in the proof of Proposition 3.4, one can derive

$$\mathbb{C}_{\mathbf{X}} = \int (U_1(\mathbf{x}) - U_2(\mathbf{x}))(U_1(\mathbf{x}) - U_2(\mathbf{x}))^T \pi_1(\mathbf{x})\pi_2(\mathbf{x}) d\mathbf{x}. \tag{6}$$

Thinking of this as a special case of comparing two populations with densities f and g, what we are computing is a type of weighted information distance matrix for discrimination problems:

$$\mathbb{Q}_w(f, g) = \int (U_f(\mathbf{x}) - U_g(\mathbf{x}))(U_f(\mathbf{x}) - U_g(\mathbf{x}))^T w(\mathbf{x}) d\mathbf{x},$$

where w is a weighting density. Comparing Equations (5) and (6), $w(\mathbf{x}) \propto \pi_1(\mathbf{x})\pi_2(\mathbf{x})$, and since $\pi_1(\mathbf{x}) + \pi_2(\mathbf{x}) = 1$, $\pi_1(\mathbf{x})\pi_2(\mathbf{x}) \leq 1/4$, the maximum is attained at $\pi_1(\mathbf{x}) = \pi_2(\mathbf{x}) = 1/2$. Also, the weighting goes to zero as either of the posteriors approaches 1. This implies that we assign maximum weight to values of $\mathbf{x}$ which are most central to the discrimination problem. Particularly, if f and g are normal densities with means $\mu_{\mathbf{f}}$ and $\mu_{\mathbf{g}}$ and common covariance Σ, $U_f(\mathbf{x}) - U_g(\mathbf{x}) = \Sigma^{-1}(\mu_f - \mu_g)$, independent of $\mathbf{x}$. It follows that $\mathbb{Q}_w(f, g)$ is proportional to $\Sigma^{-1}(\mu_f - \mu_g)(\mu_f - \mu_g)^T\Sigma^{-1}$, which has rank 1 with the eigenvector corresponding to the sole non-zero eigenvalue proportional to $\Sigma^{-1}(\mu_f - \mu_g)$. Thus, in this particular case, the CIM identifies the *Fisher's linear discriminant direction.*

The following section has been added based on the reviewers' comments.

4 Computation of Density Information Matrix using f2 Method

In this section, we give the explicit formula of the f2 computation (Hui and Lindsay, 2010). Let the variable $\mathbf{s}$ have the density

$$f_2(\mathbf{s}) \equiv \frac{f^2(\mathbf{s})}{\int f^2(\mathbf{x})d\mathbf{x}},$$

where $f(\mathbf{x})$ is the density of $\mathbf{x}$. We proposed to estimate the density information for f by the density information for $f_2(\mathbf{s})$, denote as $\mathbf{J_s}$ or $\mathbf{J}_{f_2}$

$$\boldsymbol{J}_{f_2} = \frac{\int \bigtriangledown_{\boldsymbol{x}} f \times \bigtriangledown_{\boldsymbol{x}} f^T d\boldsymbol{x}}{\int f^2(\boldsymbol{x})d\boldsymbol{x}}.$$

Suppose a multivariate sample $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n$ are drawn from a density f. We can first estimate f by the multivariate kernel estimate

$$\hat{f}_{\mathbf{H}}(\mathbf{x}) = \sum_{i=1}^{n} \frac{1}{n|\mathbf{H}|}\phi_p(\mathbf{x} - \mathbf{x}_i\,;\, \mathbf{0}, \mathbf{H}^2),$$

where $\phi_p(\cdot\,;\,\mathbf{0}, \mathbf{H}^2)$ is the p-variate Gaussian density with mean $\mathbf{0}$ and covariance $\mathbf{H}^2$.

Here, we choose normal density as Kernel function $\boldsymbol{K}$ and use the bandwidth recommended by Bowman and Foster (1993),

$$\mathbf{H}_{opt} = (\frac{4}{p+2})^{\frac{1}{p+4}} \mathbf{\Sigma}^{\frac{1}{2}} n^{-\frac{1}{p+4}}.$$

where the unknown $\mathbf{\Sigma}$ is usually replaced by its sample estimate. Then, the estimated density information matrix has the following form

$$\hat{\boldsymbol{J}}_{f_2} = \frac{\int \bigtriangledown_{\boldsymbol{x}} \hat{f}_{\boldsymbol{H}} \times \bigtriangledown_{\boldsymbol{x}} \hat{f}_{\boldsymbol{H}}^T d\boldsymbol{x}}{\int \hat{f}_{\boldsymbol{H}}^2(\boldsymbol{x})d\boldsymbol{x}}.$$

where

$$\int \hat{f}_{\boldsymbol{H}}^2(\boldsymbol{x})d\boldsymbol{x} = \frac{1}{n^2}\sum_{i=i}^{n}\sum_{j=1}^{n}\phi_p(\boldsymbol{x}_i - \boldsymbol{x}_j; 0, 2\mathbf{H}^2),$$

$$\int \nabla_{\boldsymbol{x}}\hat{f}_{\boldsymbol{H}} \times \nabla_{\boldsymbol{x}}\hat{f}_{\boldsymbol{H}}^T d\boldsymbol{x} = \frac{1}{n^2}\sum_{i=i}^{n}\sum_{j=1}^{n}\phi_p(\boldsymbol{x}_i - \boldsymbol{x}_j; 0, 2\boldsymbol{H}^2) \times \left[\frac{\boldsymbol{H}^{-2}}{2} + \frac{\boldsymbol{H}^{-2}}{2}(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T\frac{\boldsymbol{H}^{-2}}{2}\right].$$

Based on the above formula, we can see that one of the major advantages of *f2* computation method is that it provides explicit formula for all integrations when normal kernel is used.

5 Full Simulation Results

Here we present the detailed results including the full tables and the figures corresponding to all the simulation models studied in the main text.

5.1 Model (1)

Table S1 presents the performance of CIM, other SDR methods and a 'Random' benchmark for Model (1). The table can also be viewed online at the following link:

http://bit.ly/Model1Results

Dimension estimation plots for CIM (using $L = 3$ and 5 slices and SNR's ≈ 10 and ≈ 2.5) based on $B = 500$ bootstrap replicates are shown in Figure S1, and their boxplot versions in Figures S2 and S3, respectively for SNR's ≈ 10 and ≈ 2.5.

5.2 Model (2)

Table S2 presents the performance of CIM, other SDR methods and a 'Random' benchmark for Model (2). The table can also be viewed online at the following link:

http://bit.ly/Model2Results

Dimension estimation plots for CIM (using $L = 3$ and 5 slices and SNR's ≈ 10 and ≈ 2.75) based on $B = 500$ bootstrap replicates are shown in Figure S4, and their boxplot versions in Figures S5 and S6, respectively for SNR's ≈ 10 and ≈ 2.75.

5.3 Model (3)

Table S3 presents the performance of CIM, other SDR methods and a 'Random' benchmark for Model (3). The table can also be viewed online at the following link:

http://bit.ly/Model3Results

Dimension estimation plots for CIM (using $L = 5$ and 8 slices and SNR's ≈ 0.06 and ≈ 0.25) based on $B = 500$ bootstrap replicates are shown in Figure S7, and their boxplot versions in Figures S8 and S9, respectively for SNR's ≈ 0.06 and ≈ 0.25.

Recall that for heteroscedastic additive error models we defined SNR, a ratio between signals in the mean and that in variance components, modulated by σ (see Section 5 of the main text). Although with an abuse of notation we simply denote this as SNR, it is different than the conventional signal-to-noise ratio. The different σ's (along with $\beta_2^T \mathbf{X}$) used in model (3) give rise to the SNR's presented in Table S3.

5.4 Model (4)

Table S4 presents the performance of CIM, other SDR methods and a 'Random' benchmark for Model (4). The table can also be viewed online at the following link:

http://bit.ly/Model4Results

Dimension estimation plots for CIM (using $L = 5$ and 8 slices and SNR's ≈ 0.06 and ≈ 0.25) based on $B = 500$ bootstrap replicates are shown in Figure S10, and their boxplot versions in Figures S11 and S12, respectively for SNR's ≈ 0.06 and ≈ 0.25. The different σ's (along with $\beta_2^T\mathbf{X}$) used in model (4) give rise to the SNR's summarized in Table S4.

5.5 Model (5)

Table S5 presents the performance of CIM, other SDR methods and a 'Random' benchmark for Model (5). The table can also be viewed online at the following link:

http://bit.ly/Model5Results

Dimension estimation plots for CIM (using $L = 5$ and 8 slices and SNR's ≈ 0.02 and ≈ 0.24) based on $B = 500$ bootstrap replicates are shown in Figure S13, and their boxplot versions in Figures S14 and S15, respectively for SNR's ≈ 0.02 and ≈ 0.24. The SNR's corresponding to different σ's are presented in Table S5.

The diagnostic plots along with their boxplot versions in Figures S1-S15 are updated based on the revised definition of the *Signal-to-Noise Ratio* (SNR).

5.6 Model (6)

Table S6 presents the performance of CIM, other SDR methods and a 'Random' benchmark for a discrete $Y \in \{0, 1, 2, 3\}$. The table can also be viewed online at the following link:

http://bit.ly/Model6Results

SNR computation

Following the definition of Signal-to-Noise Ratio (SNR) for the heteroscedastic error simulations with continuous Y, we compute the same for discrete-Y examples as well.

We define, $\beta_1^T \mathbf{x} = \gamma_1(\mathbf{x})$ and $\beta_2^T \mathbf{x} = \gamma_2(\mathbf{x})$. Firstly, we compute the conditional mean.

$$\begin{aligned}
E\left[Y|\mathbf{X}=\mathbf{x}\right] &= E\left[I(\gamma_1(\mathbf{X}) + \sigma\epsilon > 1) + 2 \cdot I(\gamma_2(\mathbf{X}) \cdot \sigma\epsilon > 1) | \mathbf{X}=\mathbf{x}\right] \\
&= P\left(\epsilon > \frac{1-\gamma_1(\mathbf{x})}{\sigma}\right) + 2 \cdot P\left(\gamma_2(\mathbf{x}) \cdot \sigma\epsilon > 1\right) \\
&= \Phi\left(-\frac{1-\gamma_1(\mathbf{x})}{\sigma}\right) + 2 \cdot P\left(\gamma_2(\mathbf{x}) \cdot \sigma\epsilon > 1\right) \\
&\text{(where } \Phi \text{ is the distribution function of standard normal)} \\
&= p_1(\mathbf{x}, \sigma) + 2 \cdot p_2(\mathbf{x}, \sigma), \text{ for notational simplicity.} \qquad (7)
\end{aligned}$$

Note that, $p_2(\mathbf{x}, \sigma)$ simplifies to:

$$\begin{aligned}
p_2(\mathbf{x}, \sigma) &= P\left(\gamma_2(\mathbf{x}) \cdot \sigma\epsilon > 1\right) \qquad (8) \\
&= \begin{cases} \Phi\left(-\frac{1}{\sigma\gamma_2(\mathbf{x})}\right) \text{ for } \gamma_2(\mathbf{x}) > 0; \\ \Phi\left(\frac{1}{\sigma\gamma_2(\mathbf{x})}\right) \text{ for } \gamma_2(\mathbf{x}) < 0; \\ 0 \text{ for } \gamma_2(\mathbf{x}) = 0. \end{cases} \\
&= \begin{cases} \Phi\left(-\frac{1}{\sigma|\gamma_2(\mathbf{x})|}\right) \text{ for } \gamma_2(\mathbf{x}) \neq 0; \\ 0 \text{ for } \gamma_2(\mathbf{x}) = 0. \end{cases} \qquad (9) \\
&= \Phi\left(-\frac{1}{\sigma|\gamma_2(\mathbf{x})|}\right) \cdot I(\gamma_2(\mathbf{x}) \neq 0). \qquad (10)
\end{aligned}$$

Next, we compute the conditional variance.

$$
\begin{aligned}
Var\left[Y|\mathbf{X}=\mathbf{x}\right] = & Var\left[I(\gamma_1(\mathbf{X})+\sigma\epsilon>1)|\mathbf{X}=\mathbf{x}\right] + 4\cdot Var\left[I(\gamma_2(\mathbf{X})\cdot\sigma\epsilon>1)|\mathbf{X}=\mathbf{x}\right] \\
& + 2\cdot Cov\left[I(\gamma_1(\mathbf{X})+\sigma\epsilon>1), I(\gamma_2(\mathbf{X})\cdot\sigma\epsilon>1)|\mathbf{X}=\mathbf{x}\right] \\
& = p_1(\mathbf{x},\sigma)(1-p_1(\mathbf{x},\sigma)) + 4\cdot p_2(\mathbf{x},\sigma)(1-p_2(\mathbf{x},\sigma)) \\
& + 2\cdot Cov\left[I(\gamma_1(\mathbf{x})+\sigma\epsilon>1), I(\gamma_2(\mathbf{x})\cdot\sigma\epsilon>1)\right]. \qquad (11)
\end{aligned}
$$

The covariance term in Equation (11) is computed as follows:

$$
\begin{aligned}
& Cov\left[I(\gamma_1(\mathbf{x})+\sigma\epsilon>1), I(\gamma_2(\mathbf{x})\cdot\sigma\epsilon>1)\right] \\
= & E\left[I(\gamma_1(\mathbf{x})+\sigma\epsilon>1)\cdot I(\gamma_2(\mathbf{x})\cdot\sigma\epsilon>1)\right] - E\left[I(\gamma_1(\mathbf{x})+\sigma\epsilon>1)\right]\cdot E\left[I(\gamma_2(\mathbf{x})\cdot\sigma\epsilon>1)\right] \\
= & p_{1,2}(\mathbf{x},\sigma) - p_1(\mathbf{x},\sigma)\cdot p_2(\mathbf{x},\sigma),
\end{aligned}
$$

where

$$
\begin{aligned}
p_{1,2}(\mathbf{x},\sigma) = & \begin{cases} P\left(\epsilon > \max\{\frac{1-\gamma_1(\mathbf{x})}{\sigma}, \frac{1}{\sigma\gamma_2(\mathbf{x})}\}\right) \text{ for } \gamma_2(\mathbf{x})>0; \\ P\left(\frac{1-\gamma_1(\mathbf{x})}{\sigma} < \epsilon < \frac{1}{\sigma\gamma_2(\mathbf{x})}\right) \text{ for } \gamma_2(\mathbf{x})<0; \\ 0 \text{ for } \gamma_2(\mathbf{x})=0. \end{cases} \\
= & \begin{cases} \Phi\left(-\max\{\frac{1-\gamma_1(\mathbf{x})}{\sigma}, \frac{1}{\sigma\gamma_2(\mathbf{x})}\}\right) \text{ for } \gamma_2(\mathbf{x})>0; \\ \max\left\{0, \Phi\left(\frac{1}{\sigma\gamma_2(\mathbf{x})}\right) - \Phi\left(\frac{1-\gamma_1(\mathbf{x})}{\sigma}\right)\right\} \text{ for } \gamma_2(\mathbf{x})<0; \\ 0 \text{ for } \gamma_2(\mathbf{x})=0. \end{cases} \\
= & \Phi\left(-\max\{\frac{1-\gamma_1(\mathbf{x})}{\sigma}, \frac{1}{\sigma\gamma_2(\mathbf{x})}\}\right)\cdot I(\gamma_2(\mathbf{x})\geq 0) \\
& + \max\left\{0, \Phi\left(\frac{1}{\sigma\gamma_2(\mathbf{x})}\right) - \Phi\left(\frac{1-\gamma_1(\mathbf{x})}{\sigma}\right)\right\}\cdot I(\gamma_2(\mathbf{x})<0). \qquad (12)
\end{aligned}
$$

We define the SNR as:

$$SNR = \frac{Var\left(E[Y|\mathbf{X}]\right)}{E\left[Var(Y|\mathbf{X})\right]}. \tag{13}$$

Note that, the probabilities $p_1(\mathbf{x},\sigma), p_2(\mathbf{x},\sigma)$ and $p_{1,2}(\mathbf{x},\sigma)$ can be estimated from the data by simply replacing $\mathbf{x}$ with $\mathbf{x}_i$. Therefore, using Equations (7) and (10), we estimate the numerator in Equation (13) with the sample variance using the data $\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$, as shown below:

$$\widehat{Var}(E[Y|\mathbf{X}]) = Var\{p_1(\mathbf{x}_i,\sigma) + 2\cdot p_2(\mathbf{x}_i,\sigma); i = 1,2,\ldots,n\}.$$

Similarly, using $p_1(\mathbf{x}_i,\sigma), p_2(\mathbf{x}_i,\sigma)$, $p_{1,2}(\mathbf{x}_i,\sigma)$, and Equations (7), (10), (11), and (12), the denominator in Equation (13) can be estimated by the sample mean of $Var(Y|\mathbf{x}_i; i = 1,2,\ldots,n)$.

We repeat the calculations 200 times, and report the average SNR as the final estimate of (13).

5.7 Model (7)

Table S7 presents the performance of CIM, other SDR methods and a 'Random' benchmark for a discrete $Y \in \{0,1,2\}$. The table can also be viewed online at the following link:

http://bit.ly/Model7Results

SNR computation

We denote, $\beta_1^T\mathbf{x} = \gamma_1(\mathbf{x})$, $\beta_2^T\mathbf{x} = \gamma_2(\mathbf{x})$, and $Y_0 = 2\gamma_1(\mathbf{x}) + 2\exp(\gamma_2(\mathbf{x}))\cdot\sigma\epsilon$.

Firstly, we compute the conditional mean.

$$
\begin{aligned}
E\left[Y|\mathbf{X}=\mathbf{x}\right] &= E\left[I(-2<Y_0<2)+2\cdot I(Y_0\geq 2)|\mathbf{X}=\mathbf{x}\right]\\
&= P\left(-2<Y_0<2\right)+2\cdot P\left(Y_0\geq 2\right)\\
&= P\left(Y_0<2\right)-P\left(Y_0\leq -2\right)+2P\left(Y_0\geq 2\right)\\
&= P\left(\epsilon<\frac{2-2\gamma_1(\mathbf{x})}{2\exp(\gamma_2(\mathbf{x}))\sigma}\right)-P\left(\epsilon\leq\frac{-2-2\gamma_1(\mathbf{x})}{2\exp(\gamma_2(\mathbf{x}))\sigma}\right)+2\cdot P\left(\epsilon>\frac{2-2\gamma_1(\mathbf{x})}{2\exp(\gamma_2(\mathbf{x}))\sigma}\right)\\
&= \Phi\left(\frac{2-2\gamma_1(\mathbf{x})}{2\exp(\gamma_2(\mathbf{x}))\sigma}\right)-\Phi\left(\frac{-2-2\gamma_1(\mathbf{x})}{2\exp(\gamma_2(\mathbf{x}))\sigma}\right)+2\cdot\Phi\left(-\frac{2-2\gamma_1(\mathbf{x})}{2\exp(\gamma_2(\mathbf{x}))\sigma}\right)\\
&\quad \text{(where } \Phi \text{ is the distribution function of standard normal)}\\
&= p_1(\mathbf{x},\sigma)+2\cdot p_2(\mathbf{x},\sigma),\ \text{for notational simplicity.}
\end{aligned}
\tag{14}
$$

Next, we compute the conditional variance.

$$
\begin{aligned}
Var\left[Y|\mathbf{X}=\mathbf{x}\right] =& Var\left[I(-2<Y_0<2)|\mathbf{X}=\mathbf{x}\right]+4\cdot Var\left[I(Y_0\geq 2)|\mathbf{X}=\mathbf{x}\right]\\
&+2\cdot Cov\left[I(-2<Y_0<2),I(Y_0\geq 2)|\mathbf{X}=\mathbf{x}\right]\\
=& p_1(\mathbf{x},\sigma)(1-p_1(\mathbf{x},\sigma))+4\cdot p_2(\mathbf{x},\sigma)(1-p_2(\mathbf{x},\sigma))\\
&+2\cdot Cov\left[I(-2<Y_0<2),I(Y_0\geq 2)|\mathbf{X}=\mathbf{x}\right].
\end{aligned}
\tag{15}
$$

The covariance term in Equation (15) is computed as follows:

$$
\begin{aligned}
&Cov\left[I(-2<Y_0<2),I(Y_0\geq 2)|\mathbf{X}=\mathbf{x}\right]\\
=&E\left[I(-2<Y_0<2)\cdot I(Y_0\geq 2)|\mathbf{X}=\mathbf{x}\right]-E\left[I(-2<Y_0<2)|\mathbf{X}=\mathbf{x}\right]\cdot E\left[I(Y_0\geq 2)|\mathbf{X}=\mathbf{x}\right]\\
=&-p_1(\mathbf{x},\sigma)\cdot p_2(\mathbf{x},\sigma),
\end{aligned}
\tag{16}
$$

where $p_1(\mathbf{x},\sigma)$ and $p_2(\mathbf{x},\sigma)$ are defined in Equation (14).

We define the SNR as:

$$SNR = \frac{Var\left(E[Y|\mathbf{X}]\right)}{E\left[Var(Y|\mathbf{X})\right]}. \tag{17}$$

Note that, the probabilities $p_1(\mathbf{x}, \sigma)$ and $p_2(\mathbf{x}, \sigma)$ can be estimated from the data by simply replacing $\mathbf{x}$ with $\mathbf{x}_i$. Therefore, using Equation (14), we estimate the numerator in Equation (17) with the sample variance using the data $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, as shown below:

$$\widehat{Var}(E[Y|\mathbf{X}]) = Var\{p_1(\mathbf{x}_i, \sigma) + 2 \cdot p_2(\mathbf{x}_i, \sigma); i = 1, 2, \ldots, n\}.$$

Similarly, using $p_1(\mathbf{x}_i, \sigma)$, $p_2(\mathbf{x}_i, \sigma)$, and Equations (14), (15), and (16), the denominator in Equation (17) can be estimated by the sample mean of $Var(Y|\mathbf{x}_i; i = 1, 2, \ldots, n)$. We repeat the calculations 200 times, and report the average SNR as the final estimate of (17).

5.8 Model (8)

Table S8 presents the performance of CIM, other SDR methods and a 'Random' benchmark for a discrete $Y \in \{0, 1, 2\}$. The table can also be viewed online at the following link:

http://bit.ly/Model8Results

SNR computation

We denote, $\beta_1^T\mathbf{x} = \gamma_1(\mathbf{x})$, $\beta_2^T\mathbf{x} = \gamma_2(\mathbf{x})$, and $Y_0 = 2\gamma_1^2(\mathbf{x}) + 2\exp(\gamma_2(\mathbf{x})) \cdot \sigma\epsilon$.

Firstly, we compute the conditional mean.

$$\begin{aligned}
E\left[Y|\mathbf{X}=\mathbf{x}\right] &= E\left[I(-2<Y_0<2)+2\cdot I(Y_0\geq 2)|\mathbf{X}=\mathbf{x}\right]\\
&= P\left(-2<Y_0<2\right)+2\cdot P\left(Y_0\geq 2\right)\\
&= P\left(Y_0<2\right)-P\left(Y_0\leq -2\right)+2P\left(Y_0\geq 2\right)\\
&= P\left(\epsilon<\frac{2-2\gamma_1^2(\mathbf{x})}{2\exp(\gamma_2(\mathbf{x}))\sigma}\right)-P\left(\epsilon\leq\frac{-2-2\gamma_1^2(\mathbf{x})}{2\exp(\gamma_2(\mathbf{x}))\sigma}\right)+2\cdot P\left(\epsilon>\frac{2-2\gamma_1^2(\mathbf{x})}{2\exp(\gamma_2(\mathbf{x}))\sigma}\right)\\
&= \Phi\left(\frac{2-2\gamma_1^2(\mathbf{x})}{2\exp(\gamma_2(\mathbf{x}))\sigma}\right)-\Phi\left(\frac{-2-2\gamma_1^2(\mathbf{x})}{2\exp(\gamma_2(\mathbf{x}))\sigma}\right)+2\cdot\Phi\left(-\frac{2-2\gamma_1^2(\mathbf{x})}{2\exp(\gamma_2(\mathbf{x}))\sigma}\right)\\
&\quad\text{(where }\Phi\text{ is the distribution function of standard normal)}\\
&= p_1(\mathbf{x},\sigma)+2\cdot p_2(\mathbf{x},\sigma),\ \text{for notational simplicity.}
\end{aligned}\tag{18}$$

Next, we compute the conditional variance.

$$\begin{aligned}
Var\left[Y|\mathbf{X}=\mathbf{x}\right] =& Var\left[I(-2<Y_0<2)|\mathbf{X}=\mathbf{x}\right]+4\cdot Var\left[I(Y_0\geq 2)|\mathbf{X}=\mathbf{x}\right]\\
&+2\cdot Cov\left[I(-2<Y_0<2),I(Y_0\geq 2)|\mathbf{X}=\mathbf{x}\right]\\
=& p_1(\mathbf{x},\sigma)(1-p_1(\mathbf{x},\sigma))+4\cdot p_2(\mathbf{x},\sigma)(1-p_2(\mathbf{x},\sigma))\\
&+2\cdot Cov\left[I(-2<Y_0<2),I(Y_0\geq 2)|\mathbf{X}=\mathbf{x}\right].
\end{aligned}\tag{19}$$

The covariance term in Equation (19) is computed as follows:

$$\begin{aligned}
&Cov\left[I(-2<Y_0<2),I(Y_0\geq 2)|\mathbf{X}=\mathbf{x}\right]\\
=&E\left[I(-2<Y_0<2)\cdot I(Y_0\geq 2)|\mathbf{X}=\mathbf{x}\right]-E\left[I(-2<Y_0<2)|\mathbf{X}=\mathbf{x}\right]\cdot E\left[I(Y_0\geq 2)|\mathbf{X}=\mathbf{x}\right]\\
=&-p_1(\mathbf{x},\sigma)\cdot p_2(\mathbf{x},\sigma),
\end{aligned}\tag{20}$$

where $p_1(\mathbf{x},\sigma)$ and $p_2(\mathbf{x},\sigma)$ are defined in Eqn. (18).

We define the SNR as:

$$SNR = \frac{Var\left(E[Y|\mathbf{X}]\right)}{E\left[Var(Y|\mathbf{X})\right]}. \tag{21}$$

Note that, the probabilities $p_1(\mathbf{x}, \sigma)$ and $p_2(\mathbf{x}, \sigma)$ can be estimated from the data by simply replacing $\mathbf{x}$ with $\mathbf{x}_i$. Therefore, using Equation (18), we estimate the numerator in Equation (21) with the sample variance using the data $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, as shown below:

$$\widehat{Var}(E[Y|\mathbf{X}]) = Var\{p_1(\mathbf{x}_i, \sigma) + 2 \cdot p_2(\mathbf{x}_i, \sigma); i = 1, 2, \ldots, n\}.$$

Similarly, using $p_1(\mathbf{x}_i, \sigma)$, $p_2(\mathbf{x}_i, \sigma)$, and Equations (18), (19), and (20), the denominator in Equation (21) can be estimated by the sample mean of $Var(Y|\mathbf{x}_i; i = 1, 2, \ldots, n)$. We repeat the calculations 200 times, and report the average SNR as the final estimate of (21).

5.9 Computation Time

Table S11 presents in detail the comparative times (in seconds) to generate $d = 2$ output directions for models (2), (5), and (8) using $n = 400$, Independent $\mathbf{X}$, and specified σ's.

5.10 Some Remarks

- Performance of CIM depends on the number of slices (see Tables S1 - S5); small values (e.g. $L = 3$) are best for homoscedastic models whereas moderate ones (e.g. $L = 5$) appear necessary for heteroscedastic models. Finer slicing (e.g. $L = 8, 10$) worsens performance, likely due to kernel density estimation deteriorating with fewer points in each slice.

- Estimation can result in a negative definite $\hat{\mathbb{C}}_{\mathbf{X}} = \hat{\mathbb{J}}_{\mathbf{X}|Y} - \hat{\mathbb{J}}_{\mathbf{X}}$, in which case, one can set the negative eigen-values to zero.

- For the sake of fair comparison, we used the same simulation datasets for the Fourier and the Semiparametric methods which are implemented in R and FORTRAN90 programming languages respectively.

- Due to the difficult task of tuning the 4 bandwidths (for the CS structural dimension $d = 2$) involved in the Semiparametric method, we did not include this method for comparison in the main text. Rather, we focused on models (1), (2), and (5), and after extensive efforts, managed to find a "clean" set of bandwidths producing reliable results for models (1) and (2), only for a subset of the scenarios considered. Please see Tables S9 and S10 (also found in the sheets named "WithSemiparam" in the online links for full results for models (1) and (2)) for the comparative performance with all other methods. In addition to tuning the 4 bandwidths, the semiparametric method requires initialization; results in Tables S9 and S10 are based on initializing the algorithm with perturbed versions of the true $\boldsymbol{\beta}$ vectors.

6 Additional Details on Real Data Applications

6.1 Wine Recognition Data

Dimension estimation plot for CIM with $L = 3$ (number of response classes) and the its boxplot versions are provided in Figure S16. The panels in Figure S17 show scatter-plots of the standardized Wine data projected onto the 2-dimensional CS estimated by (a) CIM, (b) SIR and (c) SR, as well as a random 2D projection for benchmarking.

6.2 Ozone Data

Firstly, Figure S18 shows the autocorrelation function for the response (Ozone concentration) up to lag = 25; strong autocorrelation exists, also at large lags.

Dimension estimation plots for **CIM** with varying number of slices ($L \in [3, 10]$) and the their boxplot versions are provided in Figures S19 - S24.

The trace correlations (R) among the first CIM directions, the second CIM directions and the 2D CIM estimates of the CS obtained using varying number of slices $L \in [3, 10]$ are close to 1, indicating that the estimation is stable across varying L. These results are presented in Tables S12, S13 and S14, respectively.

Fixing $\hat{d} = 2$, the trace correlations between first directions, second directions and estimated 2D subspaces obtained via CIM, and SIR, SAVE, SR (using $L = 3, 5, 8, 10$), PHD, MAVE, dMAVE, Fourier, and Semiparametric methods are provided in Table S15.

Using our proposed dimension estimation plot for SIR with $L = 3, 5, 8, 10$ (Figures S25 - S28), we get a strong indication of $\hat{d} = 1$ across all L's, confirming that SIR misses the second direction, along which the response is symmetric in mean, and heteroschedastic.

Similarly, using our proposed dimension estimation plots for SAVE (Figures S29 - S32), we get a strong indication of $\hat{d} = 1$ for $L = 8$ and 10 and of $\hat{d} = 2$ for $L = 3$ and 5.

Finally, Figure S33 displays scatterplots of the response against the first and the second CIM directions obtained using $L = 5$ and 10 slices, and also that of the first two CIM directions against each other in each case.

Table S1: Mean (SD) of trace correlation (R) in 200 repetitions for **Model (1)** using $n = 200$ and 400; SNR's $10, 5, 2.5$ and 1; $L = 5, 10$ for SIR, SAVE, SR and $2, 3, 5, 8, 10$ for CIM; *Independent* (Ind.), *Correlated* (Corr.) and *Non-linear* (NonL) **X**. Highest R in each row is **boldfaced**.

σ, Sample Size (SNR)	X	SIR (5)	SIR (10)	SAVE (5)	SAVE (10)	SR (5)	SR (10)	PHD	MAVE	dMAVE	Fourier	CIM (2)	CIM (3)	CIM (5)	CIM (8)	CIM (10)	Random
$\sigma = \mathbf{0.25}, \mathbf{n} = \mathbf{200}$ (10.91, 9.47, 9.47)	**Ind.**	0.42 (0.129)	0.419 (0.122)	0.918 (0.052)	0.873 (0.080)	0.7 (0.203)	0.572 (0.214)	0.944 (0.029)	**0.997** (0.002)	0.964 (0.060)	0.809 (0.090)	0.916 (0.056)	0.92 (0.049)	0.908 (0.066)	0.859 (0.104)	0.826 (0.1)	0.421 (0.133)
	Corr.	0.43 (0.132)	0.419 (0.144)	0.866 (0.061)	0.819 (0.079)	0.695 (0.194)	0.557 (0.207)	0.893 (0.045)	**0.995** (0.003)	0.943 (0.075)	0.777 (0.083)	0.865 (0.063)	0.874 (0.061)	0.851 (0.081)	0.816 (0.089)	0.776 (0.098)	0.417 (0.132)
	NonL	0.401 (0.132)	0.409 (0.15)	0.784 (0.097)	0.740 (0.100)	0.594 (0.197)	0.547 (0.198)	0.859 (0.085)	**0.988** (0.026)	0.959 (0.055)	0.606 (0.134)	0.787 (0.107)	0.78 (0.099)	0.761 (0.095)	0.734 (0.090)	0.714 (0.105)	0.427 (0.130)
$\sigma = \mathbf{0.25}, \mathbf{n} = \mathbf{400}$ (10.9, 9.87, 9.48)	**Ind.**	0.442 (0.138)	0.458 (0.141)	0.969 (0.014)	0.962 (0.017)	0.868 (0.173)	0.745 (0.202)	0.977 (0.009)	**0.999** (0.000)	0.995 (0.003)	0.926 (0.064)	0.966 (0.014)	0.972 (0.011)	0.973 (0.011)	0.971 (0.014)	0.966 (0.017)	0.419 (0.124)
	Corr.	0.406 (0.135)	0.408 (0.131)	0.942 (0.031)	0.933 (0.033)	0.872 (0.147)	0.694 (0.208)	0.948 (0.025)	**0.998** (0.001)	0.991 (0.006)	0.868 (0.083)	0.944 (0.031)	0.952 (0.026)	0.952 (0.030)	0.943 (0.032)	0.941 (0.034)	0.429 (0.126)
	NonL	0.418 (0.132)	0.423 (0.139)	0.841 (0.099)	0.806 (0.097)	0.687 (0.185)	0.616 (0.216)	0.902 (0.080)	**0.997** (0.003)	0.990 (0.007)	0.645 (0.125)	0.875 (0.099)	0.870 (0.096)	0.838 (0.103)	0.809 (0.107)	0.780 (0.096)	0.423 (0.130)
$\sigma = \mathbf{0.35}, \mathbf{n} = \mathbf{200}$ (5.56, 4.83, 4.83)	**Ind.**	0.434 (0.138)	0.439 (0.144)	0.911 (0.047)	0.864 (0.083)	0.667 (0.212)	0.553 (0.199)	0.947 (0.024)	**0.994** (0.003)	0.939 (0.077)	0.792 (0.094)	0.911 (0.053)	0.914 (0.054)	0.906 (0.056)	0.854 (0.092)	0.818 (0.100)	0.443 (0.118)
	Corr.	0.421 (0.139)	0.421 (0.140)	0.849 (0.076)	0.798 (0.099)	0.657 (0.197)	0.542 (0.194)	0.886 (0.055)	**0.990** (0.005)	0.914 (0.090)	0.762 (0.084)	0.854 (0.076)	0.859 (0.074)	0.836 (0.089)	0.800 (0.098)	0.745 (0.103)	0.433 (0.119)
	NonL	0.416 (0.147)	0.424 (0.158)	0.773 (0.105)	0.712 (0.111)	0.593 (0.185)	0.523 (0.193)	0.848 (0.107)	**0.982** (0.033)	0.947 (0.066)	0.609 (0.136)	0.794 (0.104)	0.777 (0.104)	0.761 (0.099)	0.728 (0.113)	0.706 (0.106)	0.421 (0.129)
$\sigma = \mathbf{0.35}, \mathbf{n} = \mathbf{400}$ (5.56, 4.83, 4.84)	**Ind.**	0.424 (0.129)	0.417 (0.132)	0.964 (0.017)	0.954 (0.023)	0.854 (0.164)	0.712 (0.209)	0.973 (0.011)	**0.998** (0.001)	0.992 (0.004)	0.922 (0.059)	0.965 (0.016)	0.970 (0.012)	0.970 (0.014)	0.964 (0.025)	0.959 (0.023)	0.404 (0.129)
	Corr.	0.426 (0.138)	0.426 (0.130)	0.935 (0.033)	0.921 (0.040)	0.846 (0.156)	0.667 (0.212)	0.944 (0.026)	**0.996** (0.002)	0.985 (0.015)	0.871 (0.076)	0.940 (0.032)	0.949 (0.023)	0.942 (0.035)	0.933 (0.036)	0.923 (0.048)	0.437 (0.131)
	NonL	0.411 (0.137)	0.412 (0.141)	0.827 (0.094)	0.796 (0.107)	0.681 (0.189)	0.601 (0.208)	0.900 (0.091)	**0.996** (0.004)	0.984 (0.023)	0.633 (0.131)	0.871 (0.093)	0.859 (0.093)	0.834 (0.095)	0.794 (0.092)	0.774 (0.095)	0.437 (0.131)
$\sigma = \mathbf{0.50}, \mathbf{n} = \mathbf{200}$ (2.73, 2.37, 2.37)	**Ind.**	0.427 (0.136)	0.448 (0.133)	0.879 (0.075)	0.815 (0.097)	0.640 (0.202)	0.554 (0.194)	0.934 (0.033)	**0.987** (0.007)	0.903 (0.094)	0.774 (0.106)	0.901 (0.064)	0.904 (0.062)	0.876 (0.079)	0.817 (0.100)	0.775 (0.104)	0.424 (0.129)
	Corr.	0.413 (0.136)	0.417 (0.133)	0.822 (0.086)	0.761 (0.101)	0.620 (0.186)	0.523 (0.172)	0.878 (0.061)	**0.978** (0.015)	0.877 (0.095)	0.725 (0.089)	0.844 (0.082)	0.833 (0.082)	0.797 (0.099)	0.739 (0.111)	0.692 (0.107)	0.437 (0.126)
	NonL	0.428 (0.137)	0.431 (0.144)	0.743 (0.112)	0.682 (0.112)	0.591 (0.176)	0.523 (0.180)	0.834 (0.097)	**0.961** (0.057)	0.921 (0.069)	0.588 (0.128)	0.77 (0.113)	0.769 (0.110)	0.744 (0.111)	0.707 (0.105)	0.688 (0.106)	0.432 (0.123)
$\sigma = \mathbf{0.50}, \mathbf{n} = \mathbf{400}$ (2.73, 2.37, 2.37)	**Ind.**	0.42 (0.133)	0.417 (0.133)	0.956 (0.022)	0.943 (0.032)	0.805 (0.176)	0.648 (0.220)	0.969 (0.013)	**0.995** (0.002)	0.982 (0.029)	0.893 (0.071)	0.959 (0.024)	0.965 (0.016)	0.961 (0.018)	0.952 (0.028)	0.939 (0.042)	0.432 (0.130)
	Corr.	0.415 (0.129)	0.412 (0.126)	0.924 (0.035)	0.903 (0.048)	0.768 (0.180)	0.586 (0.215)	0.937 (0.030)	**0.991** (0.005)	0.964 (0.043)	0.831 (0.083)	0.93 (0.035)	0.935 (0.027)	0.929 (0.031)	0.911 (0.049)	0.893 (0.058)	0.431 (0.119)
	NonL	0.416 (0.141)	0.412 (0.141)	0.797 (0.095)	0.754 (0.099)	0.686 (0.200)	0.588 (0.192)	0.890 (0.086)	**0.990** (0.006)	0.974 (0.026)	0.615 (0.142)	0.862 (0.102)	0.849 (0.094)	0.819 (0.093)	0.779 (0.092)	0.754 (0.092)	0.420 (0.132)
$\sigma = \mathbf{0.8}, \mathbf{0.75}, \mathbf{0.75}$, $\mathbf{n} = \mathbf{200}$ (1.07, 1.05, 1.05)	**Ind.**	0.423 (0.134)	0.409 (0.139)	0.820 (0.099)	0.746 (0.122)	0.580 (0.188)	0.473 (0.175)	0.896 (0.059)	**0.935** (0.068)	0.829 (0.106)	0.701 (0.120)	0.854 (0.096)	0.844 (0.096)	0.797 (0.112)	0.739 (0.122)	0.716 (0.110)	0.431 (0.130)
	Corr.	0.416 (0.132)	0.418 (0.128)	0.745 (0.100)	0.685 (0.106)	0.550 (0.193)	0.470 (0.157)	0.833 (0.072)	**0.909** (0.082)	0.786 (0.097)	0.655 (0.105)	0.784 (0.094)	0.773 (0.086)	0.728 (0.113)	0.675 (0.125)	0.631 (0.132)	0.429 (0.124)
	NonL	0.429 (0.141)	0.423 (0.141)	0.695 (0.113)	0.632 (0.111)	0.566 (0.183)	0.487 (0.172)	0.796 (0.100)	**0.892** (0.106)	0.834 (0.099)	0.551 (0.136)	0.743 (0.102)	0.720 (0.115)	0.691 (0.112)	0.662 (0.115)	0.640 (0.115)	0.430 (0.129)
$\sigma = \mathbf{0.8}, \mathbf{0.75}, \mathbf{0.75}$, $\mathbf{n} = \mathbf{400}$ (1.06, 1.05, 1.05)	**Ind.**	0.439 (0.135)	0.426 (0.143)	0.926 (0.042)	0.890 (0.066)	0.729 (0.195)	0.565 (0.210)	0.953 (0.023)	**0.981** (0.012)	0.938 (0.068)	0.829 (0.095)	0.941 (0.026)	0.939 (0.032)	0.932 (0.038)	0.907 (0.058)	0.887 (0.071)	0.443 (0.123)
	Corr.	0.420 (0.135)	0.420 (0.124)	0.885 (0.057)	0.844 (0.073)	0.676 (0.179)	0.523 (0.177)	0.915 (0.037)	**0.972** (0.020)	0.904 (0.079)	0.786 (0.098)	0.904 (0.045)	0.902 (0.048)	0.886 (0.061)	0.847 (0.081)	0.816 (0.090)	0.428 (0.125)
	NonL	0.435 (0.138)	0.438 (0.139)	0.774 (0.105)	0.720 (0.103)	0.661 (0.197)	0.566 (0.193)	0.868 (0.094)	**0.961** (0.059)	0.952 (0.034)	0.592 (0.134)	0.832 (0.103)	0.824 (0.093)	0.791 (0.103)	0.747 (0.115)	0.730 (0.105)	0.412 (0.131)

Table S2: Mean (SD) of trace correlation (R) in 200 repetitions for **Model (2)** using $n = 200$ and 400; SNR's $10, 4.7, 2.75$ and 1; $L = 5, 10$ for SIR, SAVE, SR and $2, 3, 5, 8, 10$ for CIM; *Independent* (Ind.), *Correlated* (Corr.) and *Non-linear* (NonL) **X**. Highest R in each row is **boldfaced**.

σ, Sample Size (SNR)	X	SIR (5)	SIR (10)	SAVE (5)	SAVE (10)	SR (5)	SR (10)	PHD	MAVE	dMAVE	Fourier	CIM (2)	CIM (3)	CIM (5)	CIM (8)	CIM (10)	Random
$\sigma = \mathbf{0.55}, \mathbf{n} = \mathbf{200}$ (9.93, 9.93, 9.97)	**Ind.**	0.722 (0.044)	0.725 (0.050)	0.802 (0.091)	0.710 (0.046)	0.860 (0.132)	0.812 (0.118)	0.729 (0.048)	**0.995** (0.002)	0.990 (0.004)	0.959 (0.022)	0.953 (0.018)	0.962 (0.016)	0.956 (0.021)	0.926 (0.046)	0.879 (0.078)	0.429 (0.126)
	Corr.	0.766 (0.073)	0.775 (0.087)	0.765 (0.089)	0.679 (0.068)	0.901 (0.113)	0.859 (0.119)	0.720 (0.043)	**0.991** (0.004)	0.982 (0.008)	0.927 (0.038)	0.924 (0.032)	0.939 (0.029)	0.930 (0.038)	0.890 (0.065)	0.843 (0.091)	0.424 (0.136)
	NonL	0.788 (0.062)	0.814 (0.062)	0.692 (0.042)	0.663 (0.061)	0.763 (0.079)	0.756 (0.069)	0.776 (0.105)	**0.987** (0.029)	0.947 (0.072)	0.820 (0.055)	0.792 (0.103)	0.767 (0.097)	0.753 (0.090)	0.731 (0.072)	0.722 (0.072)	0.437 (0.129)
$\sigma = \mathbf{0.55}, \mathbf{n} = \mathbf{400}$ (9.9, 9.92, 9.93)	**Ind.**	0.733 (0.048)	0.739 (0.055)	0.964 (0.029)	0.891 (0.080)	0.943 (0.108)	0.891 (0.129)	0.733 (0.042)	**0.998** (0.001)	0.995 (0.002)	0.984 (0.007)	0.979 (0.008)	0.984 (0.006)	0.984 (0.006)	0.982 (0.007)	0.980 (0.008)	0.437 (0.128)
	Corr.	0.818 (0.077)	0.826 (0.083)	0.926 (0.056)	0.816 (0.093)	0.980 (0.046)	0.960 (0.082)	0.731 (0.045)	**0.996** (0.002)	0.992 (0.003)	0.966 (0.015)	0.968 (0.014)	0.976 (0.011)	0.977 (0.010)	0.973 (0.013)	0.969 (0.013)	0.439 (0.116)
	NonL	0.837 (0.053)	0.865 (0.049)	0.701 (0.029)	0.691 (0.027)	0.780 (0.089)	0.789 (0.087)	0.789 (0.118)	**0.996** (0.002)	0.986 (0.022)	0.847 (0.044)	0.811 (0.101)	0.794 (0.101)	0.785 (0.096)	0.759 (0.088)	0.745 (0.082)	0.423 (0.115)
$\sigma = \mathbf{0.8}, \mathbf{n} = \mathbf{200}$ (4.69, 4.69, 4.71)	**Ind.**	0.720 (0.049)	0.725 (0.055)	0.758 (0.076)	0.703 (0.044)	0.833 (0.128)	0.781 (0.112)	0.725 (0.046)	**0.988** (0.005)	0.979 (0.009)	0.948 (0.028)	0.942 (0.028)	0.948 (0.021)	0.935 (0.031)	0.891 (0.067)	0.840 (0.088)	0.429 (0.130)
	Corr.	0.743 (0.067)	0.752 (0.078)	0.718 (0.060)	0.668 (0.069)	0.854 (0.123)	0.814 (0.113)	0.716 (0.051)	**0.981** (0.010)	0.965 (0.019)	0.918 (0.034)	0.906 (0.045)	0.922 (0.038)	0.894 (0.060)	0.833 (0.088)	0.772 (0.100)	0.422 (0.124)
	NonL	0.773 (0.069)	0.800 (0.069)	0.685 (0.041)	0.648 (0.073)	0.741 (0.072)	0.746 (0.074)	0.764 (0.103)	**0.968** (0.055)	0.892 (0.102)	0.799 (0.062)	0.799 (0.103)	0.768 (0.092)	0.746 (0.080)	0.721 (0.066)	0.708 (0.066)	0.426 (0.124)
$\sigma = \mathbf{0.8}, \mathbf{n} = \mathbf{400}$ (4.68, 4.69, 4.69)	**Ind.**	0.731 (0.045)	0.734 (0.051)	0.911 (0.070)	0.791 (0.081)	0.923 (0.115)	0.880 (0.124)	0.736 (0.047)	**0.995** (0.002)	0.990 (0.004)	0.980 (0.009)	0.974 (0.009)	0.979 (0.008)	0.979 (0.009)	0.975 (0.011)	0.971 (0.012)	0.435 (0.134)
	Corr.	0.799 (0.076)	0.806 (0.087)	0.843 (0.087)	0.748 (0.070)	0.965 (0.061)	0.931 (0.096)	0.736 (0.050)	**0.992** (0.003)	0.984 (0.007)	0.960 (0.018)	0.959 (0.016)	0.966 (0.014)	0.968 (0.013)	0.963 (0.015)	0.954 (0.030)	0.428 (0.136)
	NonL	0.821 (0.056)	0.850 (0.052)	0.696 (0.019)	0.687 (0.029)	0.767 (0.086)	0.774 (0.085)	0.798 (0.118)	**0.991** (0.005)	0.968 (0.044)	0.826 (0.048)	0.800 (0.103)	0.777 (0.097)	0.748 (0.084)	0.733 (0.068)	0.718 (0.056)	0.422 (0.125)
$\sigma = \mathbf{1.05}, \mathbf{n} = \mathbf{200}$ (2.72, 2.72, 2.74)	**Ind.**	0.709 (0.043)	0.71 (0.046)	0.725 (0.058)	0.694 (0.037)	0.790 (0.120)	0.760 (0.110)	0.724 (0.049)	**0.979** (0.010)	0.962 (0.026)	0.935 (0.040)	0.924 (0.037)	0.929 (0.037)	0.905 (0.057)	0.837 (0.088)	0.780 (0.101)	0.420 (0.120)
	Corr.	0.720 (0.063)	0.735 (0.074)	0.693 (0.058)	0.655 (0.083)	0.827 (0.116)	0.772 (0.107)	0.714 (0.049)	**0.967** (0.019)	0.938 (0.031)	0.901 (0.043)	0.877 (0.065)	0.883 (0.063)	0.855 (0.084)	0.783 (0.098)	0.733 (0.107)	0.436 (0.132)
	NonL	0.752 (0.070)	0.775 (0.075)	0.674 (0.047)	0.642 (0.081)	0.728 (0.070)	0.720 (0.066)	0.767 (0.103)	**0.932** (0.080)	0.848 (0.107)	0.769 (0.069)	0.772 (0.095)	0.766 (0.095)	0.726 (0.072)	0.701 (0.059)	0.697 (0.062)	0.423 (0.130)
$\sigma = \mathbf{1.05}, \mathbf{n} = \mathbf{400}$ (2.72, 2.72, 2.72)	**Ind.**	0.726 (0.047)	0.730 (0.054)	0.825 (0.088)	0.744 (0.058)	0.884 (0.128)	0.841 (0.123)	0.735 (0.050)	**0.990** (0.004)	0.981 (0.008)	0.975 (0.010)	0.966 (0.014)	0.972 (0.011)	0.971 (0.014)	0.964 (0.016)	0.959 (0.018)	0.425 (0.130)
	Corr.	0.791 (0.080)	0.795 (0.085)	0.769 (0.077)	0.719 (0.048)	0.944 (0.078)	0.906 (0.103)	0.729 (0.046)	**0.985** (0.007)	0.972 (0.012)	0.952 (0.021)	0.950 (0.022)	0.955 (0.020)	0.953 (0.019)	0.941 (0.031)	0.928 (0.045)	0.404 (0.128)
	NonL	0.809 (0.058)	0.828 (0.062)	0.696 (0.024)	0.686 (0.030)	0.763 (0.092)	0.763 (0.080)	0.791 (0.115)	**0.981** (0.029)	0.937 (0.071)	0.802 (0.055)	0.786 (0.098)	0.775 (0.096)	0.751 (0.085)	0.727 (0.063)	0.718 (0.063)	0.426 (0.133)
$\sigma = \mathbf{1.75}, \mathbf{n} = \mathbf{200}$ (0.98, 0.98, 0.99)	**Ind.**	0.691 (0.057)	0.682 (0.060)	0.695 (0.041)	0.674 (0.065)	0.733 (0.106)	0.678 (0.102)	0.723 (0.054)	**0.921** (0.045)	0.843 (0.101)	0.875 (0.082)	0.853 (0.071)	0.838 (0.077)	0.797 (0.085)	0.722 (0.099)	0.668 (0.116)	0.415 (0.129)
	Corr.	0.670 (0.076)	0.668 (0.081)	0.671 (0.063)	0.633 (0.091)	0.717 (0.109)	0.665 (0.113)	0.708 (0.053)	**0.890** (0.059)	0.781 (0.109)	0.829 (0.080)	0.797 (0.085)	0.784 (0.089)	0.742 (0.091)	0.675 (0.096)	0.608 (0.120)	0.437 (0.127)
	NonL	0.693 (0.074)	0.697 (0.085)	0.656 (0.055)	0.620 (0.084)	0.692 (0.075)	0.659 (0.092)	0.714 (0.102)	**0.813** (0.114)	0.726 (0.114)	0.708 (0.070)	0.752 (0.084)	0.726 (0.080)	0.693 (0.072)	0.672 (0.068)	0.658 (0.079)	0.434 (0.119)
$\sigma = \mathbf{1.75}, \mathbf{n} = \mathbf{400}$ (0.98, 0.98, 0.98)	**Ind.**	0.712 (0.047)	0.710 (0.048)	0.742 (0.057)	0.716 (0.039)	0.822 (0.126)	0.767 (0.121)	0.725 (0.044)	**0.963** (0.021)	0.944 (0.030)	0.950 (0.027)	0.933 (0.031)	0.939 (0.026)	0.931 (0.037)	0.907 (0.048)	0.886 (0.064)	0.418 (0.124)
	Corr.	0.724 (0.069)	0.734 (0.073)	0.719 (0.052)	0.700 (0.044)	0.839 (0.107)	0.772 (0.113)	0.723 (0.051)	**0.946** (0.030)	0.909 (0.056)	0.918 (0.036)	0.903 (0.052)	0.905 (0.054)	0.886 (0.063)	0.849 (0.077)	0.822 (0.082)	0.437 (0.122)
	NonL	0.750 (0.066)	0.769 (0.074)	0.678 (0.034)	0.662 (0.052)	0.742 (0.083)	0.713 (0.071)	0.767 (0.102)	**0.907** (0.090)	0.830 (0.111)	0.734 (0.056)	0.764 (0.093)	0.754 (0.086)	0.732 (0.076)	0.713 (0.066)	0.700 (0.057)	0.442 (0.129)

Table S3: Mean (SD) of trace correlation (R) in 200 repetitions for **Model (3)** using $n = 200$ and 400; $\sigma = 0.32, 0.45, 0.64, 1$ and 4; $L = 5, 10$ for SIR, SAVE, SR and $2, 3, 5, 8, 10$ for CIM; *Independent* (Ind.), *Correlated* (Corr.) and *Non-linear* (NonL) $\mathbf{X}$. SNR values within parentheses correspond to Ind., Corr., and NonL $\mathbf{X}$ respectively. Highest R in each row is **boldfaced**.

σ, Sample Size (SNR)	X	SIR (5)	SIR (10)	SAVE (5)	SAVE (10)	SR (5)	SR (10)	PHD	MAVE	dMAVE	Fourier	CIM (2)	CIM (3)	CIM (5)	CIM (8)	CIM (10)	Random
$\sigma = \mathbf{0.32}, \mathbf{n} = \mathbf{200}$ (9.87, 9.9, 38.69)	**Ind.**	0.738 (0.039)	0.74 (0.039)	0.742 (0.042)	0.732 (0.072)	0.751 (0.053)	0.748 (0.049)	0.584 (0.111)	**0.863** (0.077)	0.772 (0.068)	0.748 (0.049)	0.739 (0.046)	0.744 (0.048)	0.751 (0.047)	0.764 (0.061)	0.763 (0.060)	0.426 (0.123)
	Corr.	0.711 (0.019)	0.716 (0.024)	0.707 (0.020)	0.680 (0.066)	0.720 (0.034)	0.718 (0.023)	0.419 (0.106)	**0.790** (0.063)	0.731 (0.041)	0.714 (0.023)	0.689 (0.030)	0.711 (0.030)	0.716 (0.024)	0.722 (0.035)	0.718 (0.033)	0.430 (0.130)
	NonL	0.658 (0.047)	0.688 (0.033)	0.519 (0.069)	0.497 (0.046)	0.711 (0.027)	0.717 (0.033)	0.415 (0.126)	**0.796** (0.065)	0.715 (0.023)	0.678 (0.022)	0.667 (0.019)	0.683 (0.014)	0.693 (0.011)	0.696 (0.010)	0.697 (0.010)	0.436 (0.127)
$\sigma = \mathbf{0.32}, \mathbf{n} = \mathbf{400}$ (9.81, 9.81, 38.5)	**Ind.**	0.743 (0.042)	0.741 (0.042)	0.750 (0.051)	0.753 (0.050)	0.753 (0.057)	0.743 (0.045)	0.595 (0.095)	**0.881** (0.072)	0.804 (0.090)	0.751 (0.048)	0.751 (0.053)	0.756 (0.054)	0.762 (0.061)	0.771 (0.063)	0.771 (0.065)	0.432 (0.124)
	Corr.	0.714 (0.016)	0.719 (0.021)	0.720 (0.022)	0.722 (0.024)	0.720 (0.023)	0.721 (0.019)	0.417 (0.127)	**0.823** (0.062)	0.753 (0.060)	0.721 (0.028)	0.709 (0.027)	0.726 (0.035)	0.730 (0.032)	0.732 (0.033)	0.731 (0.033)	0.436 (0.133)
	NonL	0.681 (0.027)	0.699 (0.023)	0.522 (0.042)	0.525 (0.051)	0.713 (0.023)	0.721 (0.031)	0.432 (0.116)	**0.828** (0.063)	0.713 (0.017)	0.689 (0.013)	0.672 (0.015)	0.686 (0.010)	0.695 (0.007)	0.699 (0.006)	0.700 (0.006)	0.436 (0.126)
$\sigma = \mathbf{0.45}, \mathbf{n} = \mathbf{200}$ (4.99, 5, 19.56)	**Ind.**	0.736 (0.042)	0.741 (0.047)	0.750 (0.052)	0.712 (0.089)	0.747 (0.051)	0.746 (0.046)	0.584 (0.104)	**0.862** (0.079)	0.800 (0.091)	0.749 (0.052)	0.740 (0.052)	0.748 (0.050)	0.758 (0.058)	0.766 (0.067)	0.760 (0.060)	0.425 (0.122)
	Corr.	0.709 (0.025)	0.708 (0.022)	0.704 (0.035)	0.617 (0.111)	0.711 (0.020)	0.715 (0.026)	0.414 (0.114)	**0.790** (0.061)	0.752 (0.064)	0.715 (0.029)	0.688 (0.037)	0.709 (0.030)	0.716 (0.036)	0.720 (0.037)	0.710 (0.032)	0.451 (0.122)
	NonL	0.660 (0.039)	0.687 (0.037)	0.518 (0.067)	0.502 (0.057)	0.707 (0.026)	0.712 (0.034)	0.429 (0.110)	**0.801** (0.066)	0.726 (0.042)	0.676 (0.020)	0.665 (0.022)	0.679 (0.018)	0.686 (0.015)	0.690 (0.103)	0.689 (0.016)	0.441 (0.130)
$\sigma = \mathbf{0.45}, \mathbf{n} = \mathbf{400}$ (4.96, 4.96, 19.47)	**Ind.**	0.744 (0.042)	0.744 (0.047)	0.774 (0.066)	0.772 (0.067)	0.764 (0.061)	0.748 (0.050)	0.586 (0.111)	**0.881** (0.071)	0.875 (0.109)	0.769 (0.062)	0.748 (0.053)	0.764 (0.062)	0.788 (0.071)	0.784 (0.073)	0.786 (0.073)	0.412 (0.135)
	Corr.	0.716 (0.022)	0.717 (0.020)	0.727 (0.036)	0.728 (0.040)	0.725 (0.037)	0.721 (0.029)	0.434 (0.116)	**0.829** (0.064)	0.816 (0.102)	0.724 (0.033)	0.706 (0.027)	0.726 (0.035)	0.737 (0.042)	0.738 (0.042)	0.739 (0.044)	0.425 (0.123)
	NonL	0.678 (0.027)	0.697 (0.020)	0.522 (0.047)	0.522 (0.050)	0.714 (0.026)	0.723 (0.034)	0.432 (0.111)	**0.836** (0.059)	0.725 (0.034)	0.688 (0.013)	0.667 (0.016)	0.682 (0.012)	0.689 (0.010)	0.693 (0.009)	0.694 (0.008)	0.436 (0.128)
$\sigma = \mathbf{0.64}, \mathbf{n} = \mathbf{200}$ (2.47, 2.47, 9.67)	**Ind.**	0.735 (0.041)	0.738 (0.045)	0.760 (0.068)	0.619 (0.137)	0.750 (0.057)	0.750 (0.056)	0.585 (0.103)	**0.864** (0.073)	0.853 (0.106)	0.772 (0.062)	0.737 (0.050)	0.760 (0.063)	0.769 (0.064)	0.763 (0.062)	0.767 (0.070)	0.427 (0.124)
	Corr.	0.698 (0.025)	0.702 (0.024)	0.684 (0.054)	0.470 (0.155)	0.709 (0.034)	0.708 (0.026)	0.407 (0.117)	0.796 (0.068)	**0.801** (0.090)	0.709 (0.036)	0.673 (0.034)	0.703 (0.040)	0.715 (0.039)	0.707 (0.045)	0.699 (0.038)	0.421 (0.131)
	NonL	0.656 (0.049)	0.685 (0.035)	0.510 (0.065)	0.487 (0.048)	0.701 (0.032)	0.704 (0.032)	0.424 (0.105)	**0.790** (0.070)	0.754 (0.076)	0.669 (0.024)	0.658 (0.025)	0.670 (0.022)	0.677 (0.018)	0.683 (0.019)	0.679 (0.021)	0.421 (0.124)
$\sigma = \mathbf{0.64}, \mathbf{n} = \mathbf{400}$ (2.45, 2.45, 9.63)	**Ind.**	0.739 (0.038)	0.739 (0.040)	0.803 (0.079)	0.812 (0.081)	0.773 (0.083)	0.764 (0.069)	0.588 (0.111)	0.888 (0.065)	**0.960** (0.067)	0.792 (0.076)	0.753 (0.056)	0.796 (0.076)	0.816 (0.080)	0.819 (0.079)	0.810 (0.079)	0.418 (0.132)
	Corr.	0.708 (0.019)	0.711 (0.020)	0.743 (0.057)	0.739 (0.052)	0.731 (0.053)	0.723 (0.043)	0.433 (0.119)	0.815 (0.059)	**0.928** (0.075)	0.738 (0.048)	0.707 (0.035)	0.740 (0.054)	0.749 (0.053)	0.748 (0.053)	0.747 (0.051)	0.434 (0.127)
	NonL	0.676 (0.032)	0.696 (0.020)	0.512 (0.050)	0.511 (0.043)	0.713 (0.035)	0.719 (0.041)	0.462 (0.113)	**0.833** (0.063)	0.771 (0.081)	0.687 (0.014)	0.665 (0.020)	0.674 (0.014)	0.683 (0.013)	0.686 (0.012)	0.688 (0.013)	0.431 (0.123)
$\sigma = \mathbf{1}, \mathbf{n} = \mathbf{200}$ (1.01, 1.01, 3.96)	**Ind.**	0.730 (0.045)	0.734 (0.045)	0.734 (0.104)	0.555 (0.141)	0.752 (0.070)	0.744 (0.057)	0.564 (0.101)	0.845 (0.078)	**0.932** (0.086)	0.799 (0.081)	0.731 (0.050)	0.772 (0.074)	0.781 (0.078)	0.758 (0.065)	0.764 (0.069)	0.432 (0.123)
	Corr.	0.681 (0.029)	0.684 (0.032)	0.615 (0.108)	0.427 (0.124)	0.706 (0.048)	0.712 (0.051)	0.416 (0.128)	0.766 (0.078)	**0.881** (0.082)	0.725 (0.054)	0.661 (0.042)	0.705 (0.057)	0.714 (0.053)	0.695 (0.058)	0.681 (0.059)	0.436 (0.121)
	NonL	0.641 (0.046)	0.668 (0.041)	0.511 (0.063)	0.490 (0.058)	0.686 (0.050)	0.692 (0.055)	0.421 (0.118)	0.747 (0.069)	**0.822** (0.097)	0.664 (0.029)	0.648 (0.034)	0.654 (0.032)	0.663 (0.029)	0.664 (0.032)	0.661 (0.031)	0.439 (0.125)
$\sigma = \mathbf{1}, \mathbf{n} = \mathbf{400}$ (1, 1, 3.94)	**Ind.**	0.739 (0.043)	0.737 (0.042)	0.869 (0.082)	0.861 (0.086)	0.805 (0.104)	0.771 (0.082)	0.587 (0.112)	0.885 (0.066)	**0.988** (0.029)	0.850 (0.087)	0.754 (0.059)	0.854 (0.085)	0.876 (0.079)	0.874 (0.075)	0.854 (0.082)	0.405 (0.126)
	Corr.	0.701 (0.026)	0.704 (0.029)	0.770 (0.069)	0.749 (0.093)	0.743 (0.084)	0.741 (0.076)	0.421 (0.123)	0.790 (0.072)	**0.966** (0.026)	0.778 (0.069)	0.697 (0.037)	0.768 (0.067)	0.792 (0.072)	0.782 (0.068)	0.774 (0.065)	0.434 (0.125)
	NonL	0.674 (0.033)	0.694 (0.029)	0.517 (0.060)	0.517 (0.048)	0.738 (0.075)	0.741 (0.077)	0.437 (0.118)	0.802 (0.066)	**0.898** (0.081)	0.679 (0.017)	0.655 (0.025)	0.662 (0.023)	0.668 (0.021)	0.672 (0.019)	0.673 (0.021)	0.431 (0.119)
$\sigma = \mathbf{4}, \mathbf{n} = \mathbf{200}$ (0.06, 0.06, 0.25)	**Ind.**	0.672 (0.067)	0.660 (0.071)	0.680 (0.071)	0.665 (0.077)	0.715 (0.128)	0.660 (0.126)	0.581 (0.111)	0.708 (0.102)	**0.932** (0.046)	0.831 (0.090)	0.650 (0.084)	0.793 (0.100)	0.776 (0.096)	0.713 (0.114)	0.676 (0.116)	0.430 (0.124)
	Corr.	0.538 (0.093)	0.515 (0.092)	0.555 (0.092)	0.53 (0.098)	0.607 (0.139)	0.536 (0.156)	0.408 (0.109)	0.546 (0.117)	**0.837** (0.069)	0.668 (0.109)	0.498 (0.107)	0.636 (0.107)	0.622 (0.111)	0.551 (0.131)	0.504 (0.128)	0.431 (0.110)
	NonL	0.524 (0.091)	0.518 (0.093)	0.540 (0.104)	0.518 (0.107)	0.602 (0.141)	0.547 (0.133)	0.440 (0.118)	0.581 (0.107)	**0.846** (0.068)	0.588 (0.079)	0.568 (0.100)	0.600 (0.109)	0.602 (0.122)	0.586 (0.108)	0.559 (0.122)	0.437 (0.127)
$\sigma = \mathbf{4}, \mathbf{n} = \mathbf{400}$ (0.06, 0.06, 0.25)	**Ind.**	0.708 (0.054)	0.707 (0.056)	0.753 (0.067)	0.733 (0.059)	0.838 (0.136)	0.793 (0.129)	0.589 (0.105)	0.747 (0.109)	**0.978** (0.011)	0.929 (0.043)	0.712 (0.070)	0.925 (0.046)	0.921 (0.052)	0.900 (0.060)	0.873 (0.075)	0.430 (0.134)
	Corr.	0.604 (0.065)	0.605 (0.069)	0.670 (0.053)	0.647 (0.053)	0.768 (0.156)	0.681 (0.146)	0.436 (0.123)	0.583 (0.115)	**0.924** (0.030)	0.818 (0.074)	0.593 (0.072)	0.808 (0.069)	0.818 (0.062)	0.782 (0.077)	0.751 (0.082)	0.423 (0.130)
	NonL	0.583 (0.071)	0.595 (0.073)	0.588 (0.109)	0.588 (0.108)	0.755 (0.156)	0.690 (0.139)	0.441 (0.117)	0.632 (0.111)	**0.927** (0.032)	0.635 (0.055)	0.609 (0.075)	0.668 (0.109)	0.699 (0.110)	0.691 (0.104)	0.678 (0.095)	0.440 (0.131)

Table S4: Mean (SD) of trace correlation (R) in 200 repetitions for **Model (4)** using $n = 200$ and 400; $\sigma = 0.32, 0.45, 0.64, 1$ and 4; $L = 5, 10$ for SIR, SAVE, SR and $2, 3, 5, 8, 10$ for CIM; *Independent* (Ind.), *Correlated* (Corr.) and *Non-linear* (NonL) **X**. SNR values within parentheses correspond to Ind., Corr., and NonL **X** respectively. Highest R in each row is **boldfaced**.

σ, Sample Size (SNR)	X	SIR (5)	SIR (10)	SAVE (5)	SAVE (10)	SR (5)	SR (10)	PHD	MAVE	dMAVE	Fourier	CIM (2)	CIM (3)	CIM (5)	CIM (8)	CIM (10)	Random
$\sigma = \mathbf{0.32}, \mathbf{n} = \mathbf{200}$ (9.96, 10.06, 38.75)	**Ind.**	0.741 (0.042)	0.741 (0.043)	0.744 (0.049)	0.735 (0.045)	0.744 (0.044)	0.744 (0.049)	0.568 (0.108)	**0.855** (0.077)	0.764 (0.066)	0.744 (0.044)	0.734 (0.044)	0.749 (0.054)	0.754 (0.054)	0.751 (0.048)	0.755 (0.053)	0.424 (0.135)
	Corr.	0.713 (0.024)	0.717 (0.024)	0.709 (0.024)	0.674 (0.064)	0.716 (0.028)	0.719 (0.026)	0.423 (0.109)	**0.791** (0.065)	0.729 (0.039)	0.715 (0.028)	0.688 (0.032)	0.710 (0.026)	0.716 (0.025)	0.719 (0.031)	0.719 (0.034)	0.432 (0.133)
	NonL	0.669 (0.047)	0.695 (0.038)	0.513 (0.066)	0.500 (0.051)	0.712 (0.023)	0.718 (0.033)	0.420 (0.119)	**0.799** (0.070)	0.724 (0.032)	0.679 (0.022)	0.669 (0.019)	0.686 (0.016)	0.695 (0.012)	0.699 (0.010)	0.700 (0.012)	0.437 (0.117)
$\sigma = \mathbf{0.32}, \mathbf{n} = \mathbf{400}$ (9.91, 10.02, 38.6)	**Ind.**	0.740 (0.042)	0.742 (0.041)	0.751 (0.051)	0.752 (0.051)	0.745 (0.049)	0.745 (0.047)	0.592 (0.104)	**0.891** (0.072)	0.790 (0.085)	0.748 (0.047)	0.752 (0.054)	0.756 (0.054)	0.759 (0.056)	0.764 (0.060)	0.764 (0.058)	0.434 (0.132)
	Corr.	0.716 (0.020)	0.718 (0.019)	0.722 (0.027)	0.721 (0.026)	0.721 (0.025)	0.720 (0.024)	0.426 (0.123)	**0.818** (0.064)	0.754 (0.062)	0.722 (0.028)	0.711 (0.029)	0.721 (0.028)	0.730 (0.034)	0.733 (0.034)	0.734 (0.035)	0.444 (0.135)
	NonL	0.695 (0.033)	0.710 (0.026)	0.522 (0.049)	0.523 (0.053)	0.719 (0.023)	0.731 (0.035)	0.466 (0.112)	**0.827** (0.065)	0.725 (0.029)	0.690 (0.011)	0.674 (0.013)	0.688 (0.010)	0.697 (0.007)	0.700 (0.006)	0.701 (0.006)	0.425 (0.132)
$\sigma = \mathbf{0.45}, \mathbf{n} = \mathbf{200}$ (5.04, 5.09, 19.59)	**Ind.**	0.738 (0.040)	0.74 (0.043)	0.747 (0.052)	0.689 (0.113)	0.745 (0.049)	0.739 (0.044)	0.569 (0.111)	**0.859** (0.075)	0.796 (0.088)	0.757 (0.057)	0.737 (0.045)	0.746 (0.050)	0.755 (0.057)	0.755 (0.057)	0.751 (0.055)	0.446 (0.123)
	Corr.	0.712 (0.029)	0.710 (0.022)	0.706 (0.035)	0.610 (0.126)	0.717 (0.033)	0.715 (0.027)	0.408 (0.116)	**0.789** (0.065)	0.744 (0.063)	0.719 (0.032)	0.687 (0.035)	0.707 (0.028)	0.716 (0.031)	0.714 (0.032)	0.713 (0.034)	0.437 (0.123)
	NonL	0.663 (0.043)	0.691 (0.038)	0.512 (0.065)	0.494 (0.048)	0.706 (0.022)	0.716 (0.034)	0.431 (0.128)	**0.799** (0.066)	0.723 (0.039)	0.675 (0.021)	0.663 (0.025)	0.676 (0.017)	0.685 (0.016)	0.690 (0.015)	0.689 (0.019)	0.416 (0.127)
$\sigma = \mathbf{0.45}, \mathbf{n} = \mathbf{400}$ (5.01, 5.07, 19.52)	**Ind.**	0.743 (0.042)	0.747 (0.045)	0.766 (0.062)	0.767 (0.060)	0.750 (0.054)	0.759 (0.059)	0.595 (0.097)	**0.888** (0.072)	0.870 (0.106)	0.763 (0.059)	0.750 (0.050)	0.766 (0.062)	0.785 (0.073)	0.774 (0.067)	0.783 (0.068)	0.421 (0.127)
	Corr.	0.712 (0.015)	0.716 (0.018)	0.729 (0.038)	0.728 (0.036)	0.727 (0.037)	0.728 (0.036)	0.433 (0.118)	0.824 (0.064)	**0.826** (0.098)	0.726 (0.033)	0.704 (0.026)	0.727 (0.040)	0.738 (0.045)	0.741 (0.042)	0.747 (0.048)	0.434 (0.129)
	NonL	0.686 (0.034)	0.706 (0.027)	0.522 (0.047)	0.519 (0.057)	0.716 (0.023)	0.729 (0.039)	0.457 (0.117)	**0.829** (0.060)	0.735 (0.047)	0.689 (0.013)	0.670 (0.015)	0.683 (0.011)	0.691 (0.009)	0.695 (0.008)	0.696 (0.008)	0.432 (0.120)
$\sigma = \mathbf{0.64}, \mathbf{n} = \mathbf{200}$ (2.49, 2.52, 9.69)	**Ind.**	0.737 (0.040)	0.737 (0.040)	0.749 (0.062)	0.624 (0.137)	0.754 (0.063)	0.750 (0.058)	0.581 (0.104)	0.865 (0.075)	**0.874** (0.100)	0.763 (0.059)	0.738 (0.050)	0.755 (0.057)	0.765 (0.067)	0.763 (0.059)	0.761 (0.069)	0.446 (0.120)
	Corr.	0.698 (0.027)	0.698 (0.022)	0.685 (0.051)	0.480 (0.157)	0.707 (0.033)	0.707 (0.031)	0.415 (0.113)	0.782 (0.067)	**0.792** (0.091)	0.714 (0.037)	0.677 (0.031)	0.700 (0.034)	0.712 (0.038)	0.712 (0.043)	0.699 (0.042)	0.441 (0.130)
	NonL	0.656 (0.048)	0.680 (0.037)	0.512 (0.065)	0.489 (0.049)	0.702 (0.033)	0.709 (0.043)	0.426 (0.112)	**0.782** (0.065)	0.752 (0.070)	0.674 (0.022)	0.657 (0.029)	0.668 (0.025)	0.678 (0.019)	0.679 (0.023)	0.678 (0.025)	0.418 (0.130)
$\sigma = \mathbf{0.64}, \mathbf{n} = \mathbf{400}$ (2.48, 2.50, 9.65)	**Ind.**	0.740 (0.040)	0.742 (0.044)	0.800 (0.079)	0.807 (0.082)	0.763 (0.071)	0.771 (0.073)	0.579 (0.112)	0.875 (0.072)	**0.957** (0.074)	0.792 (0.076)	0.752 (0.054)	0.788 (0.073)	0.818 (0.082)	0.811 (0.080)	0.821 (0.080)	0.440 (0.132)
	Corr.	0.712 (0.025)	0.714 (0.023)	0.737 (0.045)	0.736 (0.056)	0.735 (0.057)	0.729 (0.047)	0.437 (0.115)	0.817 (0.061)	**0.928** (0.070)	0.740 (0.046)	0.707 (0.032)	0.739 (0.049)	0.749 (0.052)	0.749 (0.052)	0.747 (0.051)	0.422 (0.126)
	NonL	0.689 (0.036)	0.707 (0.033)	0.519 (0.049)	0.517 (0.041)	0.722 (0.042)	0.732 (0.055)	0.466 (0.110)	**0.833** (0.063)	0.782 (0.081)	0.687 (0.013)	0.665 (0.019)	0.676 (0.013)	0.683 (0.012)	0.687 (0.011)	0.688 (0.010)	0.431 (0.142)
$\sigma = \mathbf{1}, \mathbf{n} = \mathbf{200}$ (1.02, 1.03, 3.97)	**Ind.**	0.732 (0.043)	0.734 (0.041)	0.755 (0.094)	0.570 (0.121)	0.754 (0.075)	0.752 (0.068)	0.573 (0.110)	0.863 (0.076)	**0.941** (0.077)	0.808 (0.080)	0.732 (0.051)	0.780 (0.078)	0.798 (0.079)	0.778 (0.076)	0.765 (0.076)	0.431 (0.135)
	Corr.	0.684 (0.031)	0.686 (0.030)	0.607 (0.128)	0.404 (0.121)	0.701 (0.047)	0.707 (0.051)	0.420 (0.100)	0.749 (0.070)	**0.869** (0.094)	0.731 (0.061)	0.662 (0.046)	0.700 (0.054)	0.708 (0.057)	0.693 (0.052)	0.669 (0.058)	0.434 (0.125)
	NonL	0.652 (0.044)	0.674 (0.038)	0.510 (0.081)	0.492 (0.068)	0.696 (0.059)	0.690 (0.051)	0.440 (0.117)	0.756 (0.075)	**0.824** (0.098)	0.665 (0.028)	0.651 (0.037)	0.656 (0.035)	0.664 (0.027)	0.663 (0.032)	0.658 (0.039)	0.428 (0.129)
$\sigma = \mathbf{1}, \mathbf{n} = \mathbf{400}$ (1.02, 1.03, 3.95)	**Ind.**	0.735 (0.039)	0.738 (0.046)	0.864 (0.081)	0.860 (0.087)	0.780 (0.091)	0.769 (0.080)	0.585 (0.110)	0.889 (0.069)	**0.991** (0.005)	0.864 (0.082)	0.752 (0.058)	0.839 (0.086)	0.872 (0.080)	0.865 (0.079)	0.852 (0.087)	0.436 (0.130)
	Corr.	0.700 (0.023)	0.701 (0.025)	0.770 (0.072)	0.737 (0.088)	0.750 (0.086)	0.736 (0.067)	0.438 (0.106)	0.795 (0.069)	**0.966** (0.030)	0.773 (0.067)	0.695 (0.035)	0.766 (0.070)	0.789 (0.075)	0.784 (0.068)	0.773 (0.067)	0.416 (0.126)
	NonL	0.674 (0.033)	0.692 (0.028)	0.520 (0.048)	0.516 (0.053)	0.737 (0.070)	0.735 (0.067)	0.450 (0.114)	0.803 (0.068)	**0.897** (0.083)	0.681 (0.017)	0.655 (0.030)	0.662 (0.023)	0.668 (0.022)	0.673 (0.019)	0.674 (0.020)	0.435 (0.126)
$\sigma = \mathbf{4}, \mathbf{n} = \mathbf{200}$ (0.06, 0.06, 0.25)	**Ind.**	0.662 (0.069)	0.666 (0.078)	0.685 (0.081)	0.668 (0.077)	0.723 (0.121)	0.676 (0.117)	0.584 (0.099)	0.705 (0.124)	**0.932** (0.057)	0.840 (0.081)	0.648 (0.085)	0.796 (0.104)	0.791 (0.098)	0.734 (0.110)	0.690 (0.118)	0.438 (0.136)
	Corr.	0.540 (0.087)	0.532 (0.091)	0.570 (0.088)	0.547 (0.087)	0.603 (0.157)	0.558 (0.153)	0.406 (0.113)	0.549 (0.110)	**0.845** (0.065)	0.670 (0.102)	0.507 (0.097)	0.647 (0.109)	0.634 (0.108)	0.588 (0.113)	0.517 (0.124)	0.419 (0.122)
	NonL	0.515 (0.096)	0.525 (0.096)	0.533 (0.106)	0.511 (0.111)	0.617 (0.136)	0.560 (0.142)	0.364 (0.123)	0.586 (0.118)	**0.848** (0.057)	0.589 (0.076)	0.571 (0.095)	0.600 (0.109)	0.618 (0.105)	0.579 (0.113)	0.573 (0.118)	0.424 (0.129)
$\sigma = \mathbf{4}, \mathbf{n} = \mathbf{400}$ (0.06, 0.06, 0.25)	**Ind.**	0.704 (0.055)	0.709 (0.061)	0.760 (0.071)	0.733 (0.058)	0.837 (0.131)	0.773 (0.130)	0.579 (0.104)	0.747 (0.096)	**0.978** (0.011)	0.928 (0.042)	0.712 (0.058)	0.929 (0.042)	0.926 (0.049)	0.897 (0.069)	0.885 (0.071)	0.412 (0.117)
	Corr.	0.611 (0.059)	0.610 (0.060)	0.661 (0.059)	0.644 (0.052)	0.763 (0.152)	0.697 (0.143)	0.421 (0.116)	0.590 (0.101)	**0.925** (0.032)	0.808 (0.072)	0.588 (0.068)	0.812 (0.070)	0.808 (0.069)	0.778 (0.082)	0.752 (0.096)	0.430 (0.128)
	NonL	0.584 (0.073)	0.591 (0.070)	0.594 (0.092)	0.595 (0.098)	0.756 (0.154)	0.671 (0.152)	0.447 (0.130)	0.628 (0.107)	**0.923** (0.034)	0.633 (0.048)	0.606 (0.071)	0.676 (0.100)	0.708 (0.099)	0.689 (0.088)	0.678 (0.083)	0.431 (0.122)

Table S5: Mean (SD) of trace correlation (R) in 200 repetitions for **Model (5)** using $n = 200$ and 400; $\sigma = 0.01, 0.03, 0.04, 0.06$ and 0.2; $L = 5, 10$ for SIR, SAVE, SR and $2, 3, 5, 8, 10$ for CIM; *Independent* (Ind.), *Correlated* (Corr.) and *Non-linear* (NonL) **X**. SNR values within parentheses correspond to Ind., Corr., and NonL **X** respectively. Highest R in each row is **boldfaced**.

σ, Sample Size (SNR)	X	SIR (5)	SIR (10)	SAVE (5)	SAVE (10)	SR (5)	SR (10)	PHD	MAVE	dMAVE	Fourier	CIM (2)	CIM (3)	CIM (5)	CIM (8)	CIM (10)	Random
$\sigma = \mathbf{0.01}, \mathbf{n} = \mathbf{200}$ (8.49, 10.24, 12.94)	**Ind.**	0.424 (0.131)	0.434 (0.144)	0.760 (0.064)	0.809 (0.081)	0.696 (0.151)	0.651 (0.172)	0.765 (0.068)	**0.881** (0.079)	0.823 (0.103)	0.745 (0.057)	0.744 (0.050)	0.751 (0.054)	0.766 (0.067)	0.778 (0.072)	0.771 (0.065)	0.422 (0.119)
	Corr.	0.348 (0.120)	0.353 (0.119)	0.713 (0.049)	0.727 (0.067)	0.671 (0.152)	0.625 (0.188)	0.720 (0.049)	**0.847** (0.074)	0.765 (0.072)	0.713 (0.047)	0.719 (0.037)	0.728 (0.043)	0.734 (0.049)	0.736 (0.051)	0.735 (0.056)	0.449 (0.131)
	NonL	0.695 (0.052)	0.701 (0.049)	0.656 (0.071)	0.560 (0.109)	0.719 (0.065)	0.723 (0.075)	0.601 (0.080)	**0.837** (0.077)	0.799 (0.090)	0.696 (0.020)	0.704 (0.020)	0.707 (0.025)	0.711 (0.022)	0.713 (0.032)	0.710 (0.029)	0.446 (0.127)
$\sigma = \mathbf{0.01}, \mathbf{n} = \mathbf{400}$ (8.31, 9.98, 12.65)	**Ind.**	0.420 (0.129)	0.431 (0.153)	0.789 (0.078)	0.882 (0.079)	0.743 (0.074)	0.742 (0.127)	0.766 (0.066)	0.903 (0.071)	**0.933** (0.089)	0.764 (0.060)	0.748 (0.054)	0.758 (0.055)	0.802 (0.080)	0.861 (0.086)	0.865 (0.082)	0.429 (0.124)
	Corr.	0.346 (0.119)	0.355 (0.120)	0.734 (0.046)	0.784 (0.076)	0.724 (0.079)	0.720 (0.100)	0.739 (0.048)	**0.872** (0.070)	0.823 (0.098)	0.726 (0.036)	0.730 (0.040)	0.732 (0.040)	0.749 (0.048)	0.783 (0.072)	0.800 (0.078)	0.430 (0.124)
	NonL	0.710 (0.043)	0.715 (0.049)	0.700 (0.019)	0.678 (0.053)	0.744 (0.070)	0.776 (0.097)	0.632 (0.061)	**0.881** (0.074)	0.838 (0.105)	0.699 (0.011)	0.710 (0.013)	0.709 (0.009)	0.714 (0.015)	0.715 (0.016)	0.715 (0.018)	0.433 (0.112)
$\sigma = \mathbf{0.03}, \mathbf{n} = \mathbf{200}$ (0.94, 1.14, 1.44)	**Ind.**	0.423 (0.133)	0.427 (0.147)	0.839 (0.092)	0.855 (0.083)	0.704 (0.157)	0.635 (0.176)	0.794 (0.070)	0.888 (0.072)	**0.973** (0.049)	0.811 (0.081)	0.736 (0.047)	0.789 (0.081)	0.853 (0.089)	0.831 (0.089)	0.793 (0.087)	0.437 (0.122)
	Corr.	0.360 (0.120)	0.362 (0.125)	0.757 (0.077)	0.771 (0.082)	0.649 (0.173)	0.582 (0.194)	0.734 (0.066)	0.838 (0.076)	**0.918** (0.085)	0.736 (0.069)	0.711 (0.043)	0.733 (0.058)	0.777 (0.079)	0.752 (0.078)	0.737 (0.076)	0.435 (0.130)
	NonL	0.687 (0.058)	0.689 (0.058)	0.646 (0.091)	0.592 (0.116)	0.751 (0.099)	0.733 (0.089)	0.602 (0.089)	0.841 (0.078)	**0.957** (0.039)	0.700 (0.030)	0.699 (0.044)	0.708 (0.039)	0.739 (0.063)	0.719 (0.049)	0.717 (0.052)	0.432 (0.127)
$\sigma = \mathbf{0.03}, \mathbf{n} = \mathbf{400}$ (0.92, 1.11, 1.41)	**Ind.**	0.415 (0.144)	0.428 (0.148)	0.946 (0.041)	0.957 (0.025)	0.804 (0.122)	0.719 (0.180)	0.815 (0.065)	0.913 (0.066)	**0.994** (0.003)	0.873 (0.089)	0.747 (0.057)	0.873 (0.082)	0.954 (0.040)	0.956 (0.034)	0.945 (0.043)	0.425 (0.139)
	Corr.	0.344 (0.128)	0.344 (0.129)	0.843 (0.083)	0.891 (0.054)	0.766 (0.102)	0.709 (0.159)	0.763 (0.063)	0.876 (0.060)	**0.982** (0.014)	0.779 (0.079)	0.726 (0.045)	0.773 (0.075)	0.863 (0.079)	0.887 (0.065)	0.874 (0.068)	0.421 (0.123)
	NonL	0.706 (0.046)	0.713 (0.056)	0.700 (0.060)	0.691 (0.071)	0.811 (0.122)	0.789 (0.109)	0.626 (0.079)	0.866 (0.069)	**0.984** (0.020)	0.704 (0.023)	0.709 (0.027)	0.716 (0.036)	0.763 (0.080)	0.742 (0.066)	0.733 (0.055)	0.445 (0.122)
$\sigma = \mathbf{0.04}, \mathbf{n} = \mathbf{200}$ (0.53, 0.64, 0.81)	**Ind.**	0.420 (0.133)	0.411 (0.142)	0.891 (0.069)	0.890 (0.071)	0.706 (0.161)	0.584 (0.182)	0.804 (0.072)	0.881 (0.073)	**0.985** (0.013)	0.841 (0.082)	0.733 (0.048)	0.833 (0.086)	0.892 (0.073)	0.857 (0.089)	0.819 (0.089)	0.430 (0.126)
	Corr.	0.347 (0.123)	0.352 (0.129)	0.780 (0.085)	0.785 (0.089)	0.652 (0.171)	0.541 (0.195)	0.729 (0.073)	0.822 (0.081)	**0.958** (0.041)	0.758 (0.076)	0.705 (0.045)	0.745 (0.068)	0.798 (0.083)	0.777 (0.084)	0.741 (0.080)	0.444 (0.123)
	NonL	0.681 (0.054)	0.686 (0.057)	0.644 (0.091)	0.601 (0.123)	0.743 (0.098)	0.736 (0.094)	0.590 (0.101)	0.828 (0.083)	**0.962** (0.030)	0.701 (0.036)	0.696 (0.037)	0.712 (0.050)	0.751 (0.076)	0.736 (0.063)	0.727 (0.063)	0.437 (0.136)
$\sigma = \mathbf{0.04}, \mathbf{n} = \mathbf{400}$ (0.52, 0.62, 0.79)	**Ind.**	0.419 (0.144)	0.435 (0.133)	0.961 (0.028)	0.965 (0.020)	0.813 (0.129)	0.747 (0.167)	0.829 (0.067)	0.913 (0.061)	**0.995** (0.002)	0.911 (0.067)	0.750 (0.058)	0.915 (0.065)	0.968 (0.022)	0.963 (0.024)	0.960 (0.025)	0.425 (0.133)
	Corr.	0.352 (0.116)	0.364 (0.118)	0.890 (0.056)	0.912 (0.039)	0.782 (0.125)	0.713 (0.169)	0.767 (0.066)	0.862 (0.067)	**0.986** (0.007)	0.832 (0.084)	0.732 (0.049)	0.831 (0.077)	0.906 (0.052)	0.912 (0.048)	0.898 (0.057)	0.441 (0.117)
	NonL	0.704 (0.047)	0.705 (0.049)	0.704 (0.062)	0.698 (0.066)	0.821 (0.123)	0.797 (0.121)	0.607 (0.103)	0.849 (0.077)	**0.987** (0.006)	0.702 (0.020)	0.712 (0.031)	0.722 (0.043)	0.785 (0.094)	0.760 (0.077)	0.748 (0.066)	0.428 (0.130)
$\sigma = \mathbf{0.06}, \mathbf{n} = \mathbf{200}$ (0.24, 0.28, 0.36)	**Ind.**	0.424 (0.133)	0.417 (0.138)	0.912 (0.063)	0.895 (0.065)	0.685 (0.179)	0.603 (0.174)	0.787 (0.084)	0.843 (0.105)	**0.987** (0.007)	0.853 (0.086)	0.733 (0.048)	0.862 (0.086)	0.908 (0.064)	0.869 (0.078)	0.815 (0.100)	0.439 (0.129)
	Corr.	0.367 (0.112)	0.358 (0.126)	0.824 (0.075)	0.802 (0.077)	0.667 (0.157)	0.548 (0.185)	0.709 (0.071)	0.773 (0.104)	**0.969** (0.019)	0.761 (0.088)	0.697 (0.047)	0.768 (0.082)	0.839 (0.075)	0.782 (0.092)	0.737 (0.105)	0.427 (0.131)
	NonL	0.672 (0.060)	0.671 (0.068)	0.670 (0.104)	0.644 (0.111)	0.752 (0.105)	0.726 (0.108)	0.567 (0.118)	0.779 (0.107)	**0.966** (0.019)	0.703 (0.044)	0.691 (0.049)	0.739 (0.075)	0.775 (0.095)	0.742 (0.076)	0.723 (0.071)	0.419 (0.128)
$\sigma = \mathbf{0.06}, \mathbf{n} = \mathbf{400}$ (0.23, 0.28, 0.35)	**Ind.**	0.437 (0.130)	0.428 (0.136)	0.972 (0.013)	0.968 (0.013)	0.860 (0.143)	0.713 (0.207)	0.824 (0.066)	0.868 (0.094)	**0.995** (0.002)	0.940 (0.050)	0.748 (0.056)	0.952 (0.041)	0.976 (0.013)	0.970 (0.019)	0.965 (0.022)	0.421 (0.117)
	Corr.	0.331 (0.123)	0.342 (0.123)	0.927 (0.031)	0.922 (0.033)	0.783 (0.161)	0.661 (0.211)	0.752 (0.069)	0.802 (0.103)	**0.988** (0.006)	0.871 (0.068)	0.725 (0.050)	0.881 (0.057)	0.937 (0.033)	0.927 (0.041)	0.922 (0.041)	0.430 (0.109)
	NonL	0.701 (0.055)	0.706 (0.059)	0.733 (0.069)	0.727 (0.063)	0.844 (0.130)	0.792 (0.118)	0.611 (0.100)	0.821 (0.095)	**0.987** (0.006)	0.703 (0.028)	0.711 (0.038)	0.777 (0.093)	0.833 (0.099)	0.799 (0.096)	0.786 (0.089)	0.421 (0.128)
$\sigma = \mathbf{0.20}, \mathbf{n} = \mathbf{200}$ (0.02, 0.03, 0.03)	**Ind.**	0.427 (0.130)	0.422 (0.137)	0.882 (0.073)	0.840 (0.083)	0.642 (0.189)	0.562 (0.192)	0.645 (0.099)	0.675 (0.086)	**0.893** (0.099)	0.741 (0.073)	0.689 (0.065)	0.867 (0.082)	0.870 (0.078)	0.809 (0.090)	0.759 (0.096)	0.422 (0.134)
	Corr.	0.351 (0.119)	0.342 (0.128)	0.804 (0.075)	0.755 (0.088)	0.603 (0.199)	0.521 (0.185)	0.543 (0.108)	0.617 (0.101)	**0.863** (0.100)	0.670 (0.068)	0.626 (0.075)	0.791 (0.080)	0.796 (0.089)	0.736 (0.094)	0.678 (0.104)	0.432 (0.135)
	NonL	0.595 (0.091)	0.599 (0.091)	0.728 (0.088)	0.701 (0.085)	0.730 (0.136)	0.683 (0.135)	0.560 (0.113)	0.649 (0.100)	**0.904** (0.060)	0.657 (0.075)	0.626 (0.089)	0.785 (0.100)	0.794 (0.094)	0.760 (0.100)	0.731 (0.105)	0.441 (0.127)
$\sigma = \mathbf{0.20}, \mathbf{n} = \mathbf{400}$ (0.02, 0.02, 0.03)	**Ind.**	0.415 (0.134)	0.411 (0.139)	0.957 (0.030)	0.943 (0.040)	0.796 (0.134)	0.675 (0.187)	0.700 (0.088)	0.698 (0.084)	**0.980** (0.041)	0.798 (0.081)	0.724 (0.055)	0.958 (0.023)	0.957 (0.034)	0.948 (0.035)	0.931 (0.051)	0.440 (0.140)
	Corr.	0.355 (0.132)	0.355 (0.133)	0.918 (0.035)	0.900 (0.044)	0.783 (0.156)	0.645 (0.205)	0.589 (0.111)	0.626 (0.085)	**0.974** (0.017)	0.740 (0.081)	0.683 (0.052)	0.913 (0.035)	0.927 (0.029)	0.910 (0.043)	0.897 (0.048)	0.426 (0.130)
	NonL	0.648 (0.068)	0.656 (0.065)	0.787 (0.087)	0.757 (0.074)	0.867 (0.118)	0.810 (0.130)	0.573 (0.110)	0.665 (0.103)	**0.952** (0.035)	0.690 (0.047)	0.674 (0.069)	0.880 (0.081)	0.893 (0.070)	0.887 (0.078)	0.874 (0.081)	0.418 (0.123)

Table S6: Mean (SD) of trace correlation (R) in 200 repetitions for **Model (6)** using $n = 200 - 400$, $\sigma = 3$ and 6 (SNRs 0.12 and 0.06 respectively) and *Independent* $\mathbf{X}$. Slice-based methods, namely SIR, SAVE, SR, and CIM naturally use $L = 4$. Highest R in each row is **boldfaced**.

	SIR	**SAVE**	**SR**	**PHD**	**MAVE**	**dMAVE**	**Fourier**	**CIM**	**Random**
					$\sigma = \mathbf{3}$, **SNR $\approx$ 0.12**				
n=200	**0.926** (0.026)	0.567 (0.140)	0.911 (0.045)	0.493 (0.138)	0.583 (0.137)	0.799 (0.087)	0.906 (0.026)	0.855 (0.077)	0.430 (0.130)
n=300	**0.950** (0.017)	0.682 (0.094)	0.948 (0.022)	0.526 (0.138)	0.621 (0.138)	0.835 (0.091)	0.941 (0.023)	0.932 (0.043)	0.424 (0.130)
n=400	**0.961** (0.015)	0.739 (0.093)	0.958 (0.028)	0.545 (0.130)	0.641 (0.123)	0.879 (0.078)	0.955 (0.020)	0.957 (0.019)	0.433 (0.124)
					$\sigma = \mathbf{6}$, **SNR $\approx$ 0.06**				
n=200	**0.866** (0.056)	0.645 (0.095)	0.828 (0.069)	0.448 (0.137)	0.519 (0.142)	0.752 (0.067)	0.813 (0.075)	0.778 (0.075)	0.451 (0.122)
n=300	**0.901** (0.049)	0.720 (0.063)	0.867 (0.068)	0.469 (0.132)	0.554 (0.130)	0.772 (0.064)	0.862 (0.072)	0.839 (0.079)	0.449 (0.127)
n=400	**0.926** (0.036)	0.743 (0.061)	0.888 (0.064)	0.490 (0.134)	0.529 (0.132)	0.786 (0.071)	0.885 (0.069)	0.888 (0.069)	0.436 (0.133)

Table S7: Mean (SD) of trace correlation (R) in 200 repetitions for **Model (7)** using $n = 200 - 400$, $\sigma = 2, 3,$ and 4 (SNRs $0.12, 0.08,$ and 0.06 respectively) and *Independent* $\mathbf{X}$. Slice-based methods, namely SIR, SAVE, SR, and CIM naturally use $L = 3$. Highest R in each row is **boldfaced**.

	SIR	**SAVE**	**SR**	**PHD**	**MAVE**	**dMAVE**	**Fourier**	**CIM**	**Random**
					$\sigma = \mathbf{2}, \mathbf{SNR} \approx \mathbf{0.12}$				
$\mathbf{n = 200}$	0.893 (0.041)	0.730 (0.107)	**0.908** (0.040)	0.568 (0.142)	0.676 (0.087)	0.851 (0.085)	0.898 (0.039)	0.891 (0.054)	0.416 (0.126)
$\mathbf{n = 300}$	0.927 (0.027)	0.826 (0.088)	**0.943** (0.022)	0.623 (0.153)	0.719 (0.076)	0.921 (0.046)	0.931 (0.025)	0.940 (0.025)	0.420 (0.128)
$\mathbf{n = 400}$	0.944 (0.021)	0.892 (0.065)	**0.961** (0.014)	0.710 (0.127)	0.734 (0.081)	0.952 (0.022)	0.949 (0.019)	**0.961** (0.014)	0.436 (0.127)
					$\sigma = \mathbf{3}, \mathbf{SNR} \approx \mathbf{0.08}$				
$\mathbf{n = 200}$	0.868 (0.047)	0.644 (0.134)	0.862 (0.054)	0.541 (0.157)	0.631 (0.097)	0.775 (0.106)	**0.872** (0.043)	0.842 (0.074)	0.430 (0.126)
$\mathbf{n = 300}$	0.909 (0.035)	0.752 (0.114)	0.908 (0.043)	0.607 (0.150)	0.670 (0.088)	0.841 (0.092)	**0.913** (0.034)	0.910 (0.048)	0.434 (0.128)
$\mathbf{n = 400}$	0.927 (0.024)	0.814 (0.101)	0.937 (0.026)	0.671 (0.135)	0.693 (0.086)	0.905 (0.065)	0.932 (0.023)	**0.943** (0.025)	0.427 (0.126)
					$\sigma = \mathbf{4}, \mathbf{SNR} \approx \mathbf{0.06}$				
$\mathbf{n = 200}$	0.846 (0.053)	0.583 (0.138)	0.820 (0.077)	0.528 (0.143)	0.590 (0.107)	0.697 (0.117)	**0.849** (0.051)	0.789 (0.096)	0.432 (0.124)
$\mathbf{n = 300}$	0.884 (0.044)	0.682 (0.132)	0.874 (0.062)	0.591 (0.156)	0.634 (0.095)	0.781 (0.094)	**0.889** (0.044)	0.872 (0.066)	0.429 (0.133)
$\mathbf{n = 400}$	0.910 (0.036)	0.744 (0.122)	0.908 (0.043)	0.621 (0.139)	0.661 (0.110)	0.823 (0.103)	**0.914** (0.034)	0.908 (0.054)	0.426 (0.129)

Table S8: Mean (SD) of trace correlation (R) in 200 repetitions for **Model (8)** using $n = 200 - 400$, $\sigma = 3, 4$, and 5 (SNRs $0.07, 0.06$, and 0.05 respectively) and *Independent* $\mathbf{X}$. Slice-based methods, namely SIR, SAVE, SR, and CIM naturally use $L = 3$. Highest R in each row is **boldfaced**.

	SIR	**SAVE**	**SR**	**PHD**	**MAVE**	**dMAVE**	**Fourier**	**CIM**	**Random**
					$\sigma = \mathbf{3}$, **SNR $\approx$ 0.07**				
n = 200	0.699 (0.055)	0.643 (0.101)	0.726 (0.105)	0.619 (0.108)	0.576 (0.152)	0.766 (0.121)	0.736 (0.081)	**0.823** (0.090)	0.439 (0.129)
n = 300	0.704 (0.042)	0.720 (0.091)	0.749 (0.112)	0.667 (0.059)	0.653 (0.134)	0.844 (0.103)	0.776 (0.085)	**0.897** (0.054)	0.430 (0.130)
n = 400	0.718 (0.049)	0.762 (0.095)	0.786 (0.125)	0.681 (0.048)	0.681 (0.119)	0.894 (0.088)	0.821 (0.093)	**0.924** (0.045)	0.429 (0.122)
					$\sigma = \mathbf{4}$, **SNR $\approx$ 0.06**				
n = 200	0.686 (0.047)	0.594 (0.130)	0.686 (0.088)	0.598 (0.095)	0.541 (0.146)	0.660 (0.145)	0.710 (0.070)	**0.776** (0.102)	0.426 (0.134)
n = 300	0.711 (0.056)	0.673 (0.105)	0.742 (0.110)	0.623 (0.109)	0.594 (0.133)	0.759 (0.127)	0.757 (0.085)	**0.849** (0.090)	0.416 (0.128)
n = 400	0.718 (0.045)	0.723 (0.100)	0.760 (0.103)	0.658 (0.066)	0.633 (0.136)	0.819 (0.113)	0.783 (0.085)	**0.896** (0.065)	0.432 (0.119)
					$\sigma = \mathbf{5}$, **SNR $\approx$ 0.05**				
n = 200	0.689 (0.056)	0.556 (0.133)	0.676 (0.075)	0.542 (0.128)	0.520 (0.144)	0.599 (0.148)	0.702 (0.069)	**0.727** (0.096)	0.440 (0.121)
n = 300	0.701 (0.050)	0.643 (0.112)	0.710 (0.088)	0.604 (0.110)	0.551 (0.145)	0.693 (0.135)	0.735 (0.075)	**0.799** (0.097)	0.422 (0.117)
n = 400	0.713 (0.047)	0.685 (0.095)	0.736 (0.090)	0.622 (0.078)	0.579 (0.140)	0.741 (0.132)	0.767 (0.082)	**0.856** (0.076)	0.436 (0.126)

Table S9: Mean (SD) of trace correlation (R) in 200 repetitions for **Model (1)** for the subset of scenarios where **Semiparamteric** method produces reliable outputs for a chosen set of bandwidths. Highest R in each row is **boldfaced**.

σ, Sample Size (SNR)	X	SIR (5)	SIR (10)	SAVE (5)	SAVE (10)	SR (5)	SR (10)	PHD	MAVE	dMAVE	Fourier	Semiparametric	CIM (2)	CIM (3)	CIM (5)	CIM (8)	CIM (10)	Random
$\sigma = \mathbf{0.25}, \mathbf{n} = \mathbf{200}$	**Ind.**	0.42 (0.129)	0.419 (0.122)	0.918 (0.052)	0.873 (0.080)	0.7 (0.203)	0.572 (0.214)	0.944 (0.029)	**0.997** (0.002)	0.964 (0.060)	0.809 (0.090)	0.994 (0.003)	0.916 (0.056)	0.92 (0.049)	0.908 (0.066)	0.859 (0.104)	0.826 (0.1)	0.421 (0.133)
(10.91, 9.47)	**Corr.**	0.43 (0.132)	0.419 (0.144)	0.866 (0.061)	0.819 (0.079)	0.695 (0.194)	0.557 (0.207)	0.893 (0.045)	**0.995** (0.003)	0.943 (0.075)	0.777 (0.083)	0.990 (0.005)	0.865 (0.063)	0.874 (0.061)	0.851 (0.081)	0.816 (0.089)	0.776 (0.098)	0.417 (0.132)
$\sigma = \mathbf{0.25}, \mathbf{n} = \mathbf{400}$	**Ind.**	0.442 (0.138)	0.458 (0.141)	0.969 (0.014)	0.962 (0.017)	0.868 (0.173)	0.745 (0.202)	0.977 (0.009)	**0.999** (0.000)	0.995 (0.003)	0.926 (0.064)	0.998 (0.001)	0.966 (0.014)	0.972 (0.011)	0.973 (0.011)	0.971 (0.014)	0.966 (0.017)	0.419 (0.124)
(10.9, 9.87)	**Corr.**	0.406 (0.135)	0.408 (0.131)	0.942 (0.031)	0.933 (0.033)	0.872 (0.147)	0.694 (0.208)	0.948 (0.025)	**0.998** (0.001)	0.991 (0.006)	0.868 (0.083)	0.996 (0.002)	0.944 (0.031)	0.952 (0.026)	0.952 (0.030)	0.943 (0.032)	0.941 (0.034)	0.429 (0.126)
$\sigma = \mathbf{0.35}, \mathbf{n} = \mathbf{200}$	**Ind.**	0.434 (0.138)	0.439 (0.144)	0.911 (0.047)	0.864 (0.083)	0.667 (0.212)	0.553 (0.199)	0.947 (0.024)	**0.994** (0.003)	0.939 (0.077)	0.792 (0.094)	0.991 (0.004)	0.911 (0.053)	0.914 (0.054)	0.906 (0.056)	0.854 (0.092)	0.818 (0.100)	0.443 (0.118)
(5.56, 4.83)	**Corr.**	0.421 (0.139)	0.421 (0.140)	0.849 (0.076)	0.798 (0.099)	0.657 (0.197)	0.542 (0.194)	0.886 (0.055)	**0.990** (0.005)	0.914 (0.090)	0.762 (0.084)	0.985 (0.008)	0.854 (0.076)	0.859 (0.074)	0.836 (0.089)	0.800 (0.098)	0.745 (0.103)	0.433 (0.119)
$\sigma = \mathbf{0.35}, \mathbf{n} = \mathbf{400}$	**Ind.**	0.424 (0.129)	0.417 (0.132)	0.964 (0.017)	0.954 (0.023)	0.854 (0.164)	0.712 (0.209)	0.973 (0.011)	**0.998** (0.001)	0.992 (0.004)	0.922 (0.059)	0.997 (0.001)	0.965 (0.016)	0.970 (0.012)	0.970 (0.014)	0.964 (0.025)	0.959 (0.023)	0.404 (0.129)
(5.56, 4.83)	**Corr.**	0.426 (0.138)	0.426 (0.130)	0.935 (0.033)	0.921 (0.040)	0.846 (0.156)	0.667 (0.212)	0.944 (0.026)	**0.996** (0.002)	0.985 (0.015)	0.871 (0.076)	0.994 (0.003)	0.940 (0.032)	0.949 (0.023)	0.942 (0.035)	0.933 (0.036)	0.923 (0.048)	0.437 (0.131)
$\sigma = \mathbf{0.50}, \mathbf{n} = \mathbf{400}$	**Ind.**	0.42 (0.133)	0.417 (0.133)	0.956 (0.022)	0.943 (0.032)	0.805 (0.176)	0.648 (0.220)	0.969 (0.013)	**0.995** (0.002)	0.982 (0.029)	0.893 (0.071)	0.994 (0.002)	0.959 (0.024)	0.965 (0.016)	0.961 (0.018)	0.952 (0.028)	0.939 (0.042)	0.432 (0.130)
(2.73, 2.37)	**Corr.**	0.415 (0.129)	0.412 (0.126)	0.924 (0.035)	0.903 (0.048)	0.768 (0.180)	0.586 (0.215)	0.937 (0.030)	**0.991** (0.005)	0.964 (0.043)	0.831 (0.083)	0.989 (0.005)	0.93 (0.035)	0.935 (0.027)	0.929 (0.031)	0.911 (0.049)	0.893 (0.058)	0.431 (0.119)

Table S10: Mean (SD) of trace correlation (R) in 200 repetitions for **Model (2)** for the subset of scenarios where **Semiparamteric** method produces reliable outputs for a chosen set of bandwidths. Highest R in each row is **boldfaced**.

σ, Sample Size (SNR)	X	SIR (5)	SIR (10)	SAVE (5)	SAVE (10)	SR (5)	SR (10)	PHD	MAVE	dMAVE	Fourier	Semiparametric	CIM (2)	CIM (3)	CIM (5)	CIM (8)	CIM (10)	Random
$\sigma = \mathbf{0.55}, \mathbf{n} = \mathbf{200}$	**Ind.**	0.722 (0.044)	0.725 (0.050)	0.802 (0.091)	0.710 (0.046)	0.860 (0.132)	0.812 (0.118)	0.729 (0.048)	**0.995** (0.002)	0.990 (0.004)	0.959 (0.022)	0.993 (0.003)	0.953 (0.018)	0.962 (0.016)	0.956 (0.021)	0.926 (0.046)	0.879 (0.078)	0.429 (0.126)
(9.93, 9.93)	**Corr.**	0.766 (0.073)	0.775 (0.087)	0.765 (0.089)	0.679 (0.068)	0.901 (0.113)	0.859 (0.119)	0.720 (0.043)	**0.991** (0.004)	0.982 (0.008)	0.927 (0.038)	0.988 (0.005)	0.924 (0.032)	0.939 (0.029)	0.930 (0.038)	0.890 (0.065)	0.843 (0.091)	0.424 (0.136)
$\sigma = \mathbf{0.55}, \mathbf{n} = \mathbf{400}$	**Ind.**	0.733 (0.048)	0.739 (0.055)	0.964 (0.029)	0.891 (0.080)	0.943 (0.108)	0.891 (0.129)	0.733 (0.042)	**0.998** (0.001)	0.995 (0.002)	0.984 (0.007)	0.997 (0.001)	0.979 (0.008)	0.984 (0.006)	0.984 (0.006)	0.982 (0.007)	0.980 (0.008)	0.437 (0.128)
(9.9, 9.92)	**Corr.**	0.818 (0.077)	0.826 (0.083)	0.926 (0.056)	0.816 (0.093)	0.980 (0.046)	0.960 (0.082)	0.731 (0.045)	**0.996** (0.002)	0.992 (0.003)	0.966 (0.015)	0.995 (0.002)	0.968 (0.014)	0.976 (0.011)	0.977 (0.010)	0.973 (0.013)	0.969 (0.013)	0.439 (0.116)
$\sigma = \mathbf{0.8}, \mathbf{n} = \mathbf{200}$	**Ind.**	0.720 (0.049)	0.725 (0.055)	0.758 (0.076)	0.703 (0.044)	0.833 (0.128)	0.781 (0.112)	0.725 (0.046)	**0.988** (0.005)	0.979 (0.009)	0.948 (0.028)	0.986 (0.006)	0.942 (0.028)	0.948 (0.021)	0.935 (0.031)	0.891 (0.067)	0.840 (0.088)	0.429 (0.130)
(4.69, 4.69)	**Corr.**	0.743 (0.067)	0.752 (0.078)	0.718 (0.060)	0.668 (0.069)	0.854 (0.123)	0.814 (0.113)	0.716 (0.051)	**0.981** (0.010)	0.965 (0.019)	0.918 (0.034)	0.976 (0.012)	0.906 (0.045)	0.922 (0.038)	0.894 (0.060)	0.833 (0.088)	0.772 (0.100)	0.422 (0.124)
$\sigma = \mathbf{0.8}, \mathbf{n} = \mathbf{400}$	**Ind.**	0.731 (0.045)	0.734 (0.051)	0.911 (0.070)	0.791 (0.081)	0.923 (0.115)	0.880 (0.124)	0.736 (0.047)	**0.995** (0.002)	0.990 (0.004)	0.980 (0.009)	0.994 (0.003)	0.974 (0.009)	0.979 (0.008)	0.979 (0.009)	0.975 (0.011)	0.971 (0.012)	0.435 (0.134)
(4.68, 4.69)	**Corr.**	0.799 (0.076)	0.806 (0.087)	0.843 (0.087)	0.748 (0.070)	0.965 (0.061)	0.931 (0.096)	0.736 (0.050)	**0.992** (0.003)	0.984 (0.007)	0.960 (0.018)	0.990 (0.004)	0.959 (0.016)	0.966 (0.014)	0.968 (0.013)	0.963 (0.015)	0.954 (0.030)	0.428 (0.136)
$\sigma = \mathbf{1.05}, \mathbf{n} = \mathbf{400}$	**Ind.**	0.726 (0.047)	0.730 (0.054)	0.825 (0.088)	0.744 (0.058)	0.884 (0.128)	0.841 (0.123)	0.735 (0.050)	**0.990** (0.004)	0.981 (0.008)	0.975 (0.010)	0.989 (0.004)	0.966 (0.014)	0.972 (0.011)	0.971 (0.014)	0.964 (0.016)	0.959 (0.018)	0.425 (0.130)
(2.72, 2.72)	**Corr.**	0.791 (0.080)	0.795 (0.085)	0.769 (0.077)	0.719 (0.048)	0.944 (0.078)	0.906 (0.103)	0.729 (0.046)	**0.985** (0.007)	0.972 (0.012)	0.952 (0.021)	0.983 (0.008)	0.950 (0.022)	0.955 (0.020)	0.953 (0.019)	0.941 (0.031)	0.928 (0.045)	0.404 (0.128)

Table S11: Computation time in seconds (average of 200 runs; SD in parentheses) to generate $d = 2$ output directions for Models (2), (5), and (8) with $n = 400$, Independent $\mathbf{X}$, and specified σ. Number of slices $L = 3$ naturally for Model (8) with discrete Y.

Model (2): Homoscedastic, Continuous $Y(\sigma = 0.55)$

SIR (5)	SIR (10)	SAVE (5)	SAVE (10)	SR (5)	SR (10)	PHD	MAVE	dMAVE	Fourier	Semiparametric	CIM(2)	CIM(3)	CIM(5)	CIM(8)	CIM(10)
5e-04	4e-04	6e-04	7e-04	0.8499	0.8787	0.0026	1.1625	12.1979	0.0127	1.1711	0.6527	0.5737	0.5253	0.4892	0.4809
(3e-04)	(2e-04)	(2e-04)	(2e-04)	(0.0433)	(0.0593)	(4e-04)	(0.0944)	(0.9517)	(0.00889)	(0.0779)	(0.0501)	(0.0287)	(0.0351)	(0.0257)	(0.0304)

Model (5): Heteroscedastic, Continuous $Y(\sigma = 0.20)$

SIR (5)	SIR (10)	SAVE (5)	SAVE (10)	SR (5)	SR (10)	PHD	MAVE	dMAVE	Fourier	Semiparametric	CIM(2)	CIM(3)	CIM(5)	CIM(8)	CIM(10)
5e-04	4e-04	6e-04	7e-04	0.8358	0.8660	0.0026	2.0753	12.8684	0.0131	6.7422	0.6582	0.5830	0.5290	0.4958	0.4856
(4e-04)	(1e-04)	(2e-04)	(4e-04)	(0.0141)	(0.0137)	(3e-04)	(0.0534)	(2.0811)	(0.00817)	(1.8080)	(0.0167)	(0.0175)	(0.0096)	(0.0095)	(0.0079)

Model (8): Heteroscedastic, Discrete $Y(\sigma = 4)$

SIR	SAVE	SR	PHD	MAVE	dMAVE	Fourier	Semiparametric	CIM
6e-04	8e-04	0.8202	0.0026	2.0446	17.5507	0.0132	XX	0.5906
(5e-04)	(3e-04)	(0.0149)	(5e-04)	(0.0229)	(0.1781)	(0.0088)		(0.0084)

Table S12: Trace Correlation (R) between the **first CIM directions** obtained using varying number of slices $L \in [3, 10]$ for **Ozone Data**.

	CIM (5)	**CIM (6)**	**CIM (7)**	**CIM (8)**	**CIM (10)**
CIM (3)	0.9855	0.9967	0.9967	0.9979	0.9898
CIM (5)		0.9950	0.9882	0.9918	0.9980
CIM (6)			0.9967	0.9980	0.9969
CIM (7)				0.9978	0.9888
CIM (8)					0.9930

Table S13: Trace Correlation (R) between the **second CIM directions** obtained using varying number of slices $L \in [3, 10]$ for **Ozone Data**.

	CIM (5)	**CIM (6)**	**CIM (7)**	**CIM (8)**	**CIM (10)**
CIM (3)	0.9761	0.9817	0.9821	0.9865	0.9471
CIM (5)		0.9910	0.9799	0.9835	0.9834
CIM (6)			0.9877	0.9896	0.9762
CIM (7)				0.9938	0.9786
CIM (8)					0.9775

Table S14: Trace Correlation (R) between the estimated **2D Central Subspaces** obtained by **CIM** using varying number of slices $L \in [3, 10]$ for **Ozone Data**.

CIM 2D CS	**CIM (5)**	**CIM (6)**	**CIM (7)**	**CIM (8)**	**CIM (10)**
CIM (3)	0.9933	0.9916	0.9895	0.9929	0.9774
CIM (5)		0.9973	0.9947	0.995	0.9912
CIM (6)			0.9936	0.9943	0.9886
CIM (7)				0.9961	0.9909
CIM (8)					0.9896

Table S15: Fixing $d = 2$ for **Ozone Data**, trace correlation (R) between the subspaces spanned by the first directions (DIR 1), the second directions (DIR 2), and the estimated 2D Subspaces (2D) obtained via **CIM**, and **SIR**, **SAVE**, **SR** (using $L = 3, 5, 8, 10$ slices), **PHD**, **MAVE**, **dMAVE**, **Fourier**, and **Semiparametric** methods. Highest R in each row is **boldfaced**.

		SIR (L)	**SAVE (L)**	**SR (L)**	**PHD**	**MAVE**	**dMAVE**	**Fourier**	**Semiparametric**
CIM (3)	**DIR 1**	0.9386	**0.9702**	0.9121	0.7713	0.9029	0.877	0.9358	0.5237
	DIR2	0.4607	**0.946**	0.8526	0.2759	0.3844	0.472	0.8876	0.5966
	2D	0.7838	**0.9847**	0.9362	0.6744	0.7429	0.7872	0.9472	0.8008
CIM (5)	**DIR 1**	0.9711	**0.9973**	0.9665	0.8434	0.9533	0.9378	0.9738	0.4989
	DIR2	0.3167	**0.9431**	0.7344	0.1278	0.2875	0.3894	0.8971	0.4846
	2D	0.7437	**0.9718**	0.8858	0.6644	0.7242	0.7605	0.9438	0.7719
CIM (8)	**DIR 1**	0.939	**0.9813**	0.9318	0.7996	0.9246	0.8989	0.9556	0.535
	DIR2	0.0873	**0.9163**	0.5615	0.2077	0.3198	0.411	0.8805	0.5286
	2D	0.7094	**0.9658**	0.814	0.6756	0.7337	0.7732	0.9459	0.7848
CIM (10)	**DIR 1**	0.9558	**0.9772**	0.965	0.8326	0.9401	0.927	0.9626	0.4702
	DIR2	0.6042	0.8761	0.7181	0.0766	0.2519	0.3286	**0.8831**	0.4254
	2D	0.8196	0.9283	0.8656	0.6719	0.7141	0.7421	**0.9353**	0.748

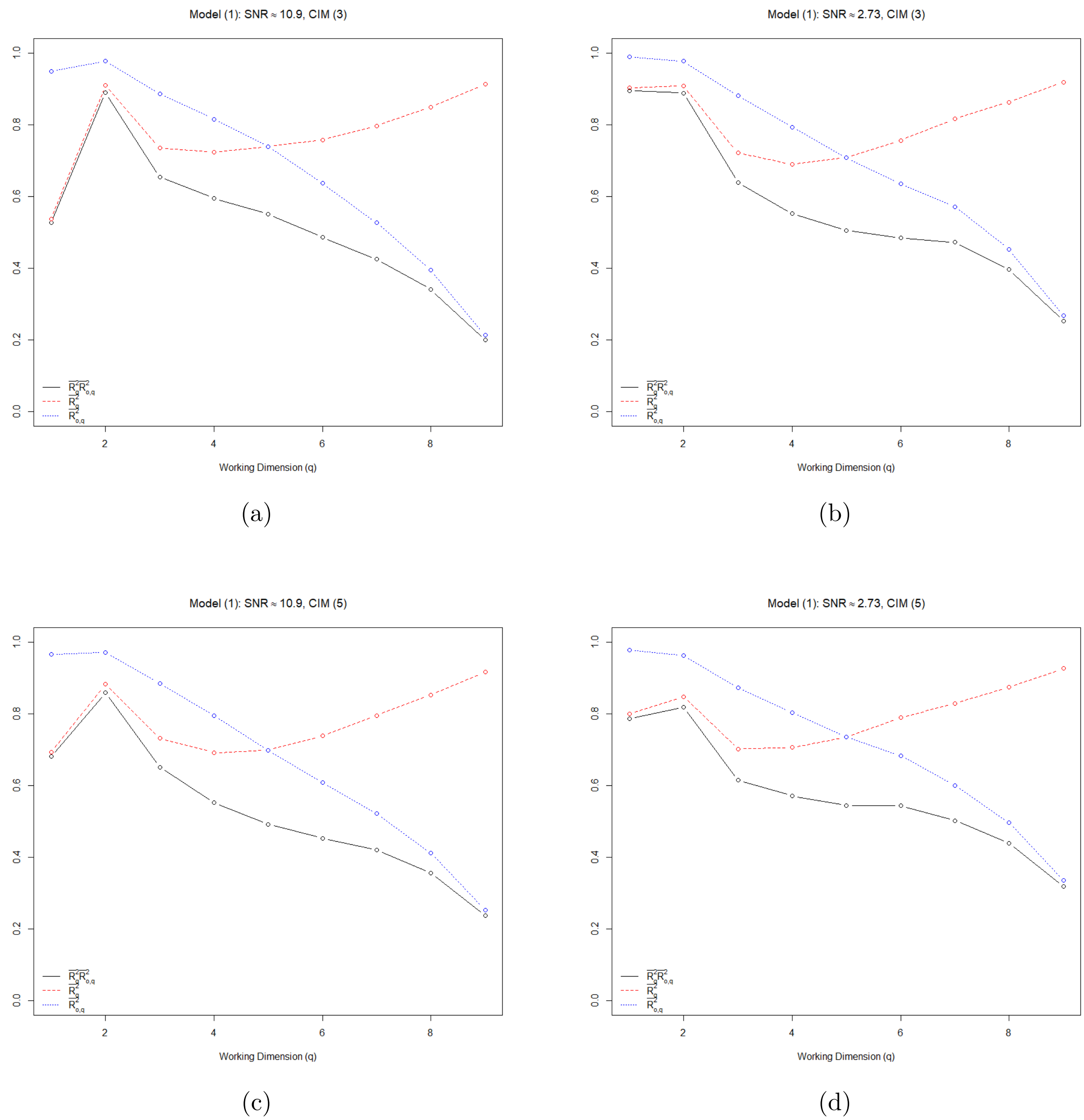


Figure S1: Dimension estimation plots (500 replicates) for **CIM** in **Model (1)** using *Independent* $\mathbf{X}$, $n = 400$, SNRs ≈ 10.9 ($\sigma = 0.25$) and ≈ 2.73 ($\sigma = 0.5$), and $L = 3$ slices (panels (a) and (b)) and 5 slices (panels (c) and (d)).

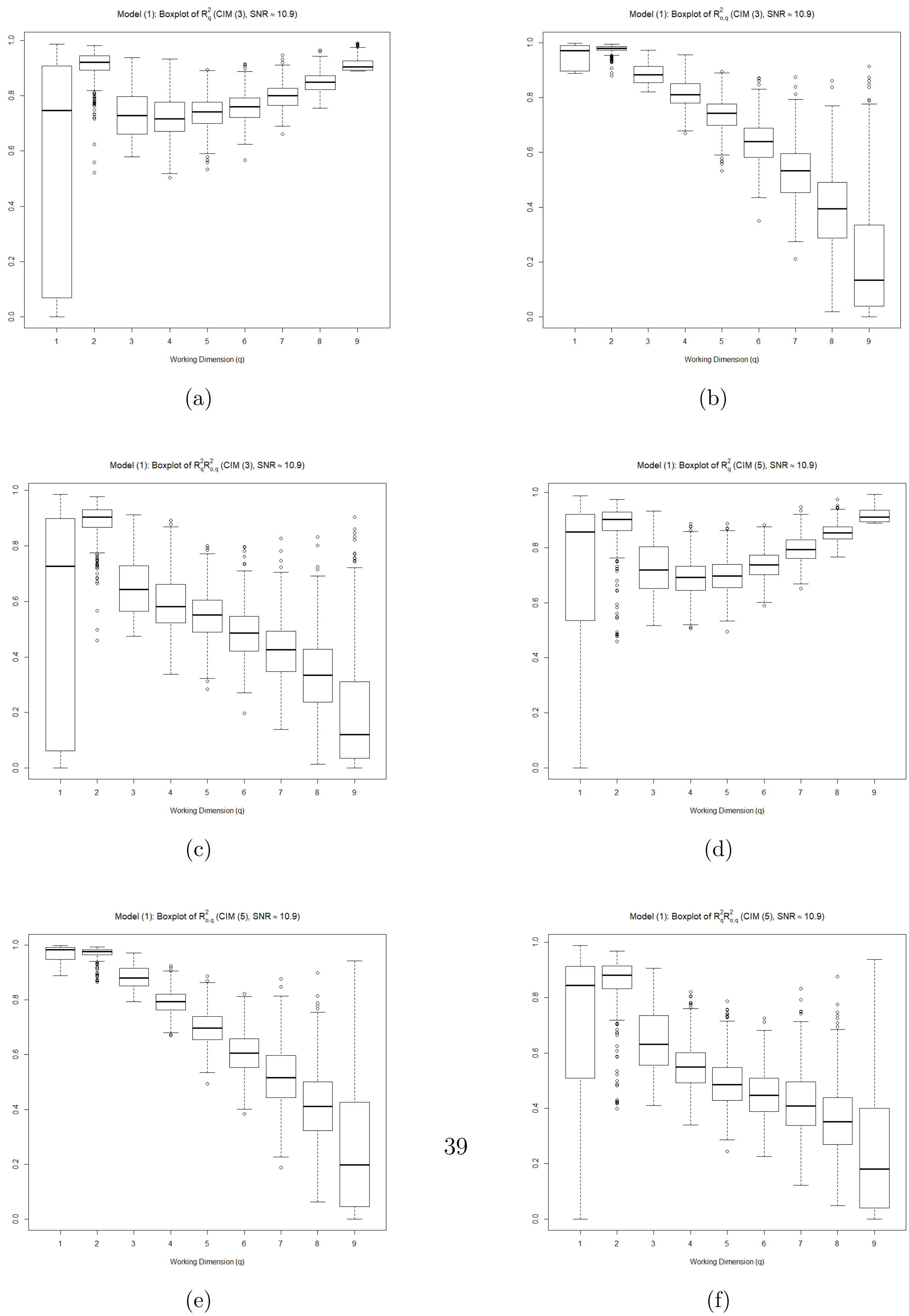
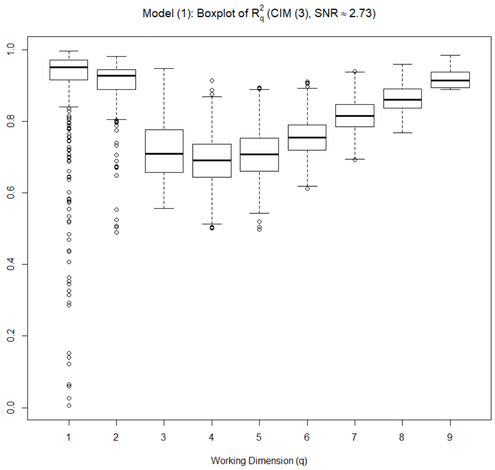
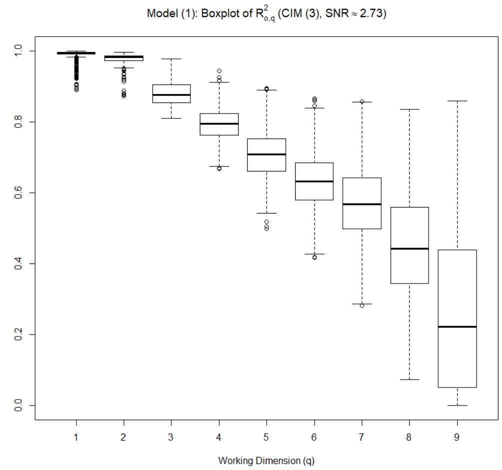


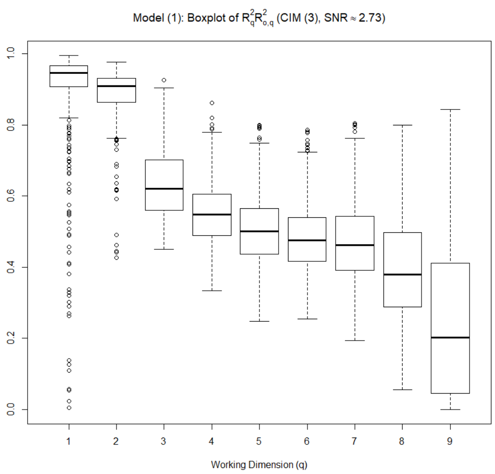
Figure S2: Boxplot versions of dimension estimation plots (500 replicates) for **CIM** in **Model (1)** using *Independent* $\mathbf{X}$, $n = 400$, SNR ≈ 10.9 ($\sigma = 0.25$), and $L = 3$ slices (panels (a)-(c)) and 5 slices (panels (d)-(f)).



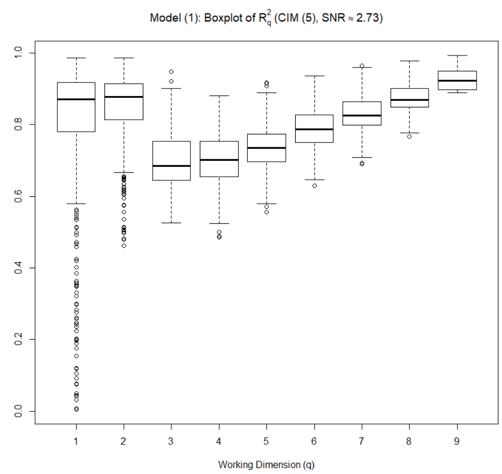
(a)



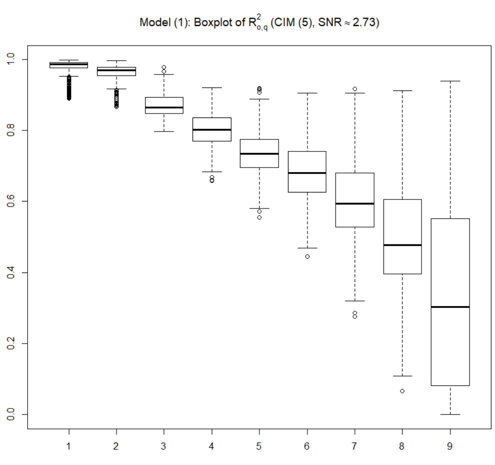
(b)



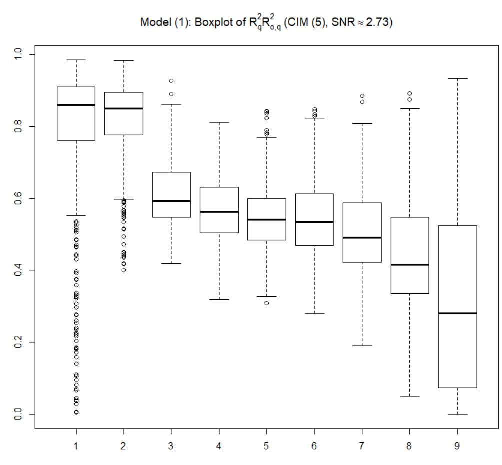
(c)



(d)



(e)



(f)

Figure S3: Boxplot versions of dimension estimation plots (500 replicates) for **CIM** in **Model (1)** using *Independent* $\mathbf{X}$, $n = 400$, SNR ≈ 2.73 ($\sigma = 0.5$), and $L = 3$ slices (panels (a)-(c)) and 5 slices (panels (d)-(f)).

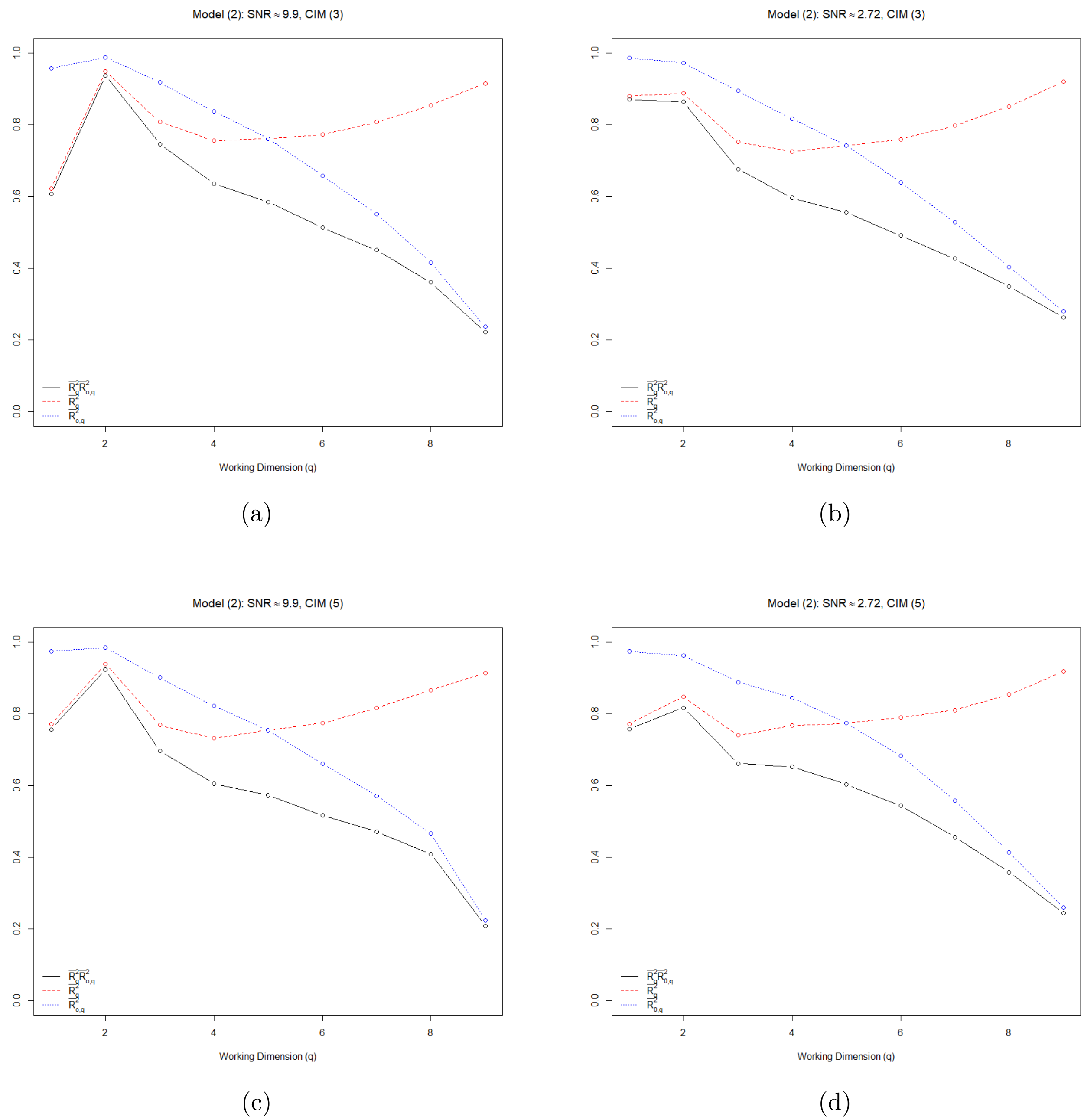


Figure S4: Dimension estimation plots (500 replicates) for **CIM** in **Model (2)** using *Independent* $\mathbf{X}$, $n = 400$, SNRs ≈ 9.9 ($\sigma = 0.55$) and ≈ 2.72 ($\sigma = 1.05$), and $L = 3$ slices (panels (a) and (b)) and 5 slices (panels (c) and (d)).

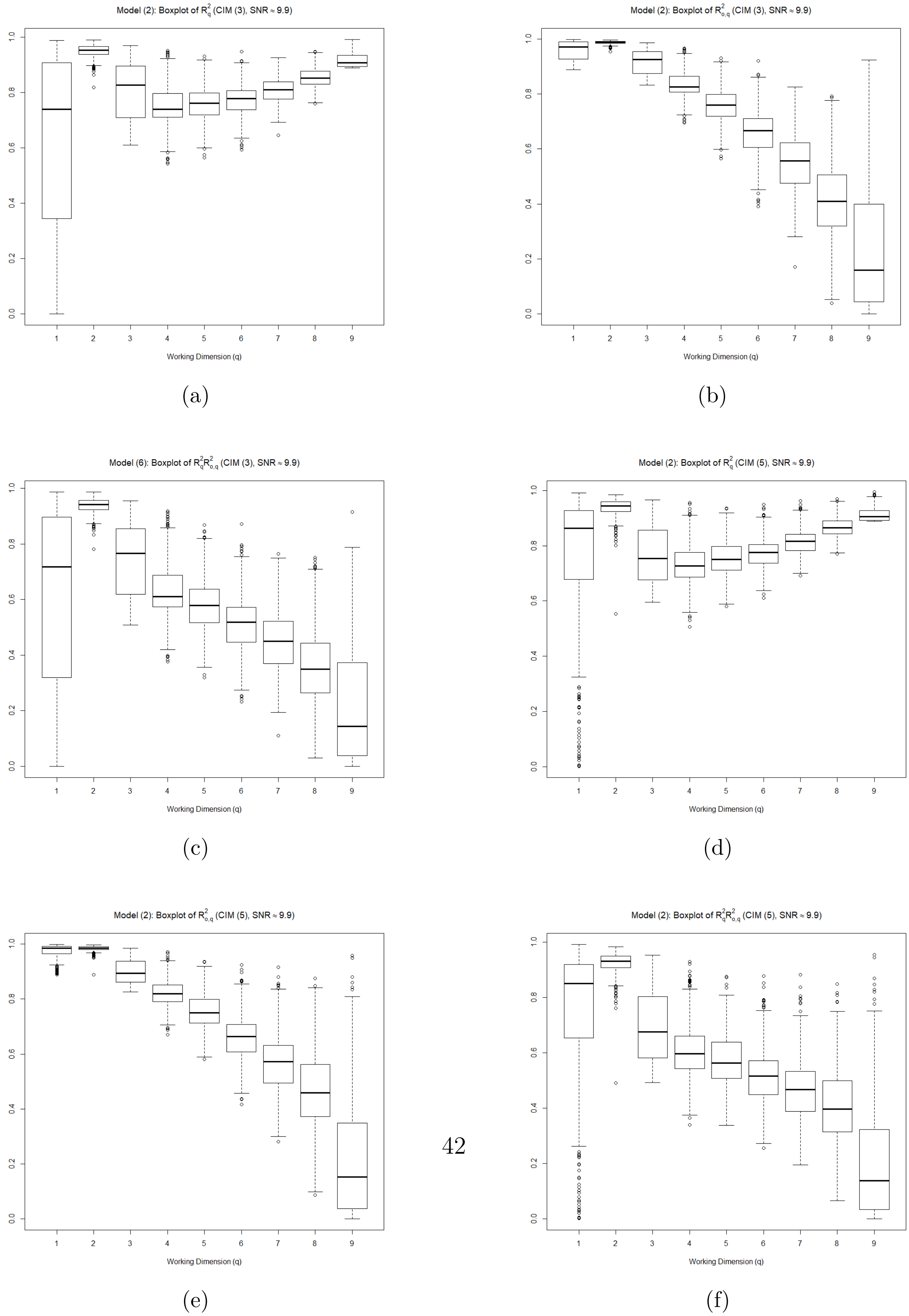
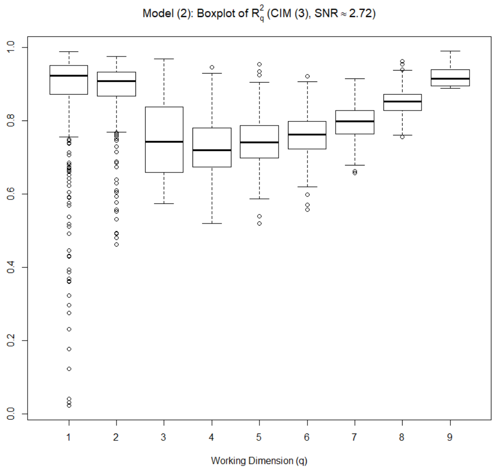
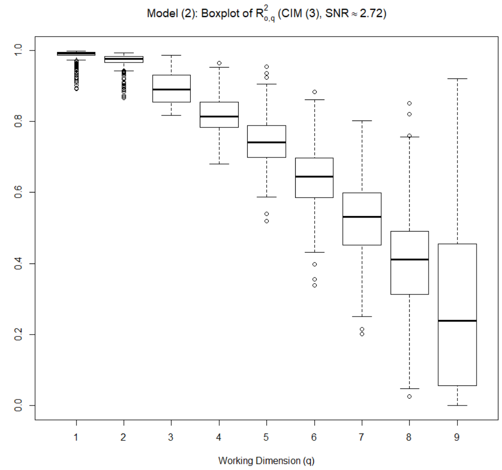


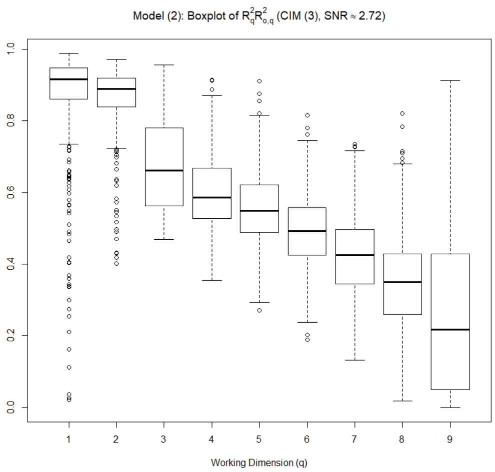
Figure S5: Boxplot versions of dimension estimation plots (500 replicates) for **CIM** in **Model (2)** using *Independent* $\mathbf{X}$, $n = 400$, SNR ≈ 9.9 ($\sigma = 0.55$), and $L = 3$ slices (panels (a)-(c)) and 5 slices (panels (d)-(f)).



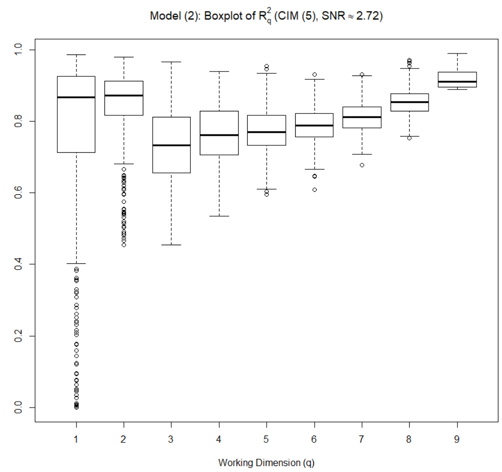
(a)



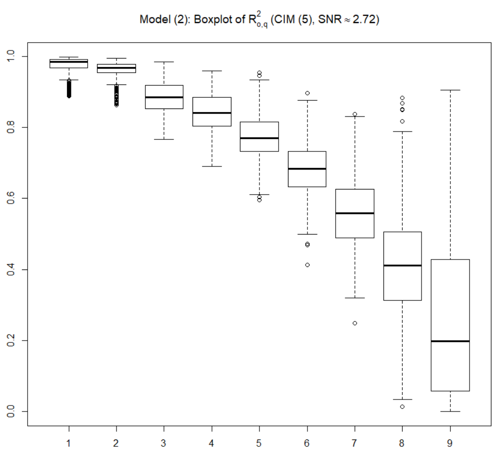
(b)



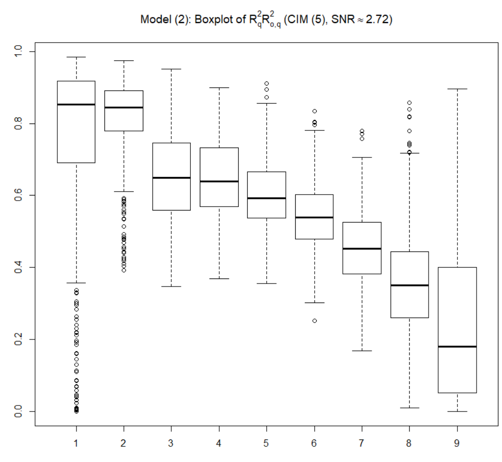
(c)



(d)

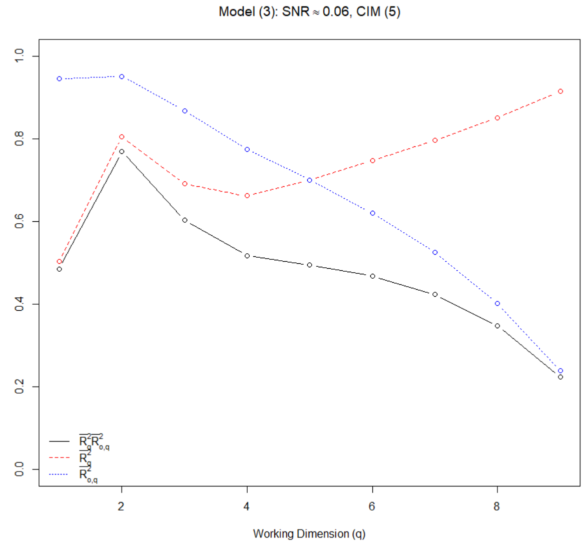


(e)

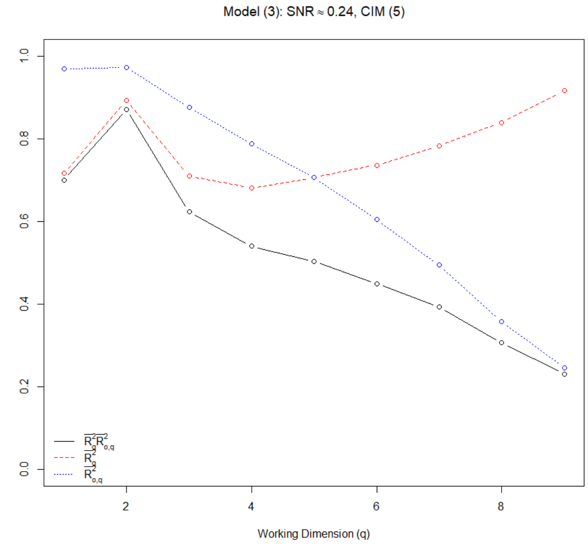


(f)

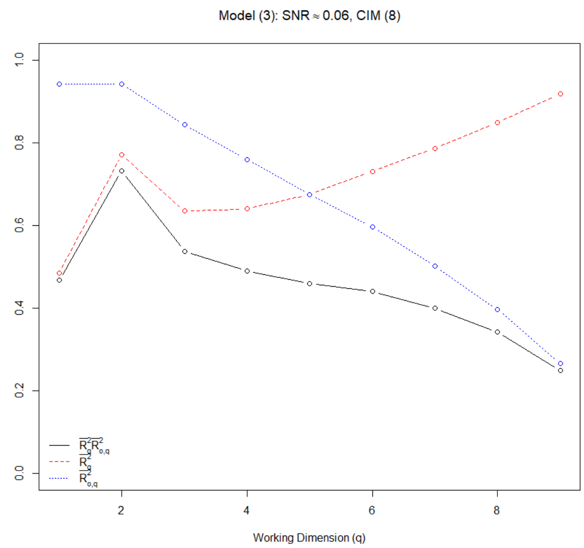
Figure S6: Boxplot versions of dimension estimation plots (500 replicates) for **CIM** in **Model (2)** using *Independent* $\mathbf{X}$, $n = 400$, SNR ≈ 2.72 ($\sigma = 1.05$), and $L = 3$ slices (panels (a)-(c)) and 5 slices (panels (d)-(f)).



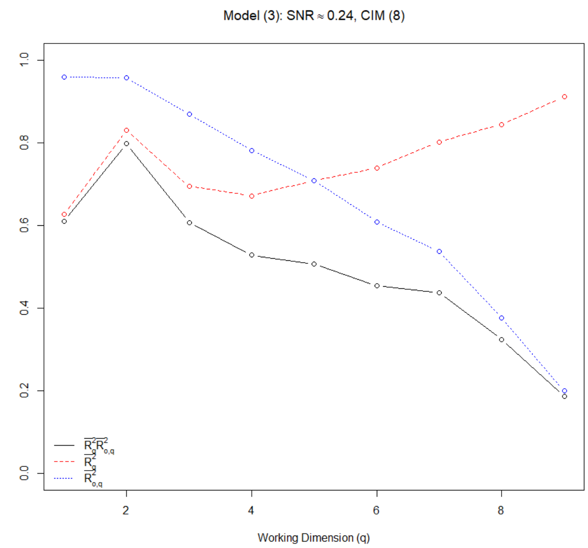
(a)



(b)

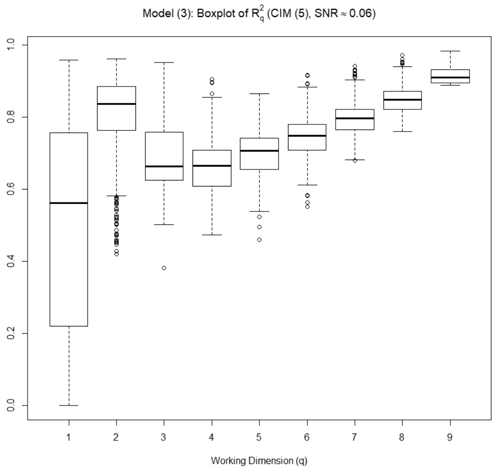


(c)

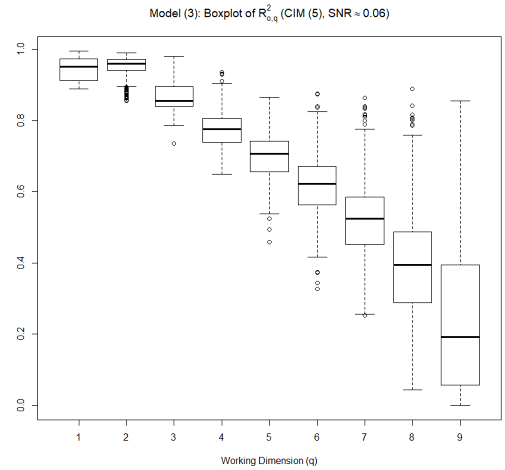


(d)

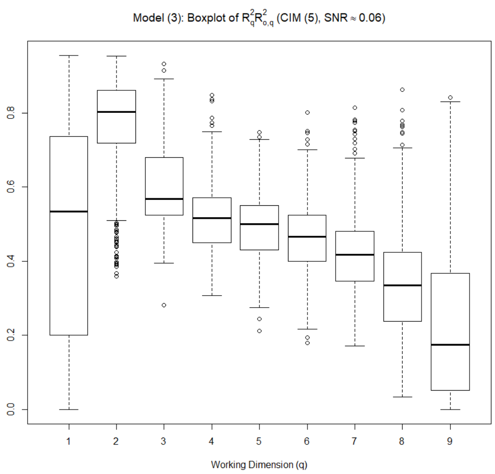
Figure S7: Dimension estimation plots (500 replicates) for **CIM** in **Model (3)** using *Independent* $\mathbf{X}$, $n = 400$, SNR's ≈ 0.06 ($\sigma = 4$) and ≈ 0.24 ($\sigma = 2.03$), and $L = 5$ slices (panels (a) and (b)) and 8 slices (panels (c) and (d)).



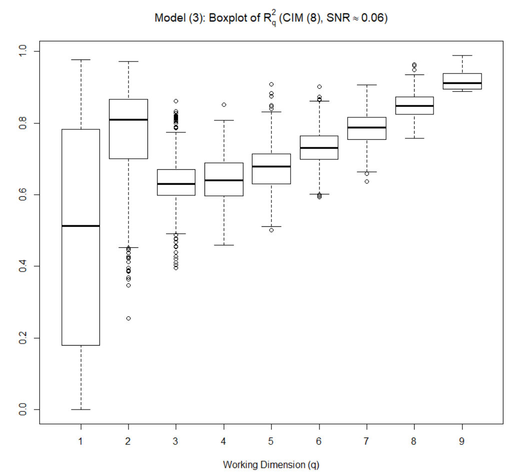
(a)



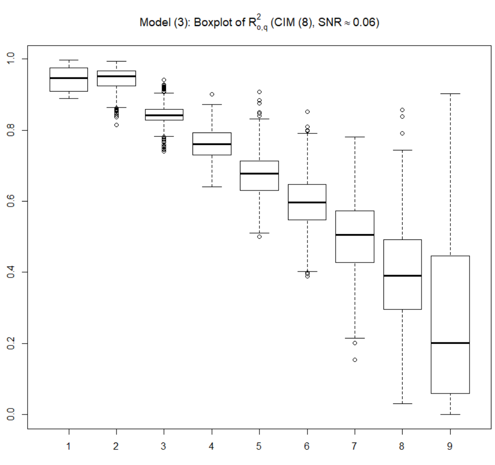
(b)



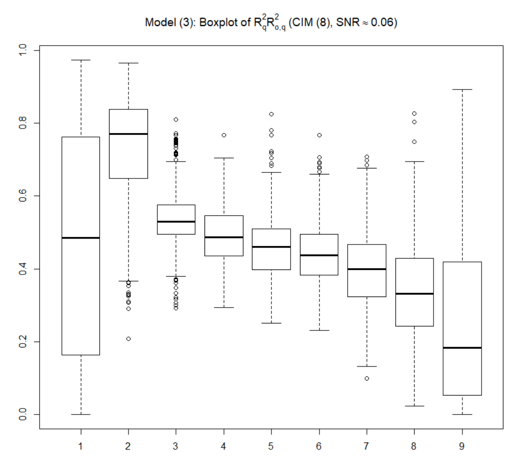
(c)



(d)



(e)



(f)

Figure S8: Boxplot versions of dimension estimation plots (500 replicates) for **CIM** in **Model (3)** using *Independent* $\mathbf{X}$, $n = 400$, SNR ≈ 0.06 ($\sigma = 4$), and $L = 5$ slices (panels (a)-(c)) and 8 slices (panels (d)-(f)).

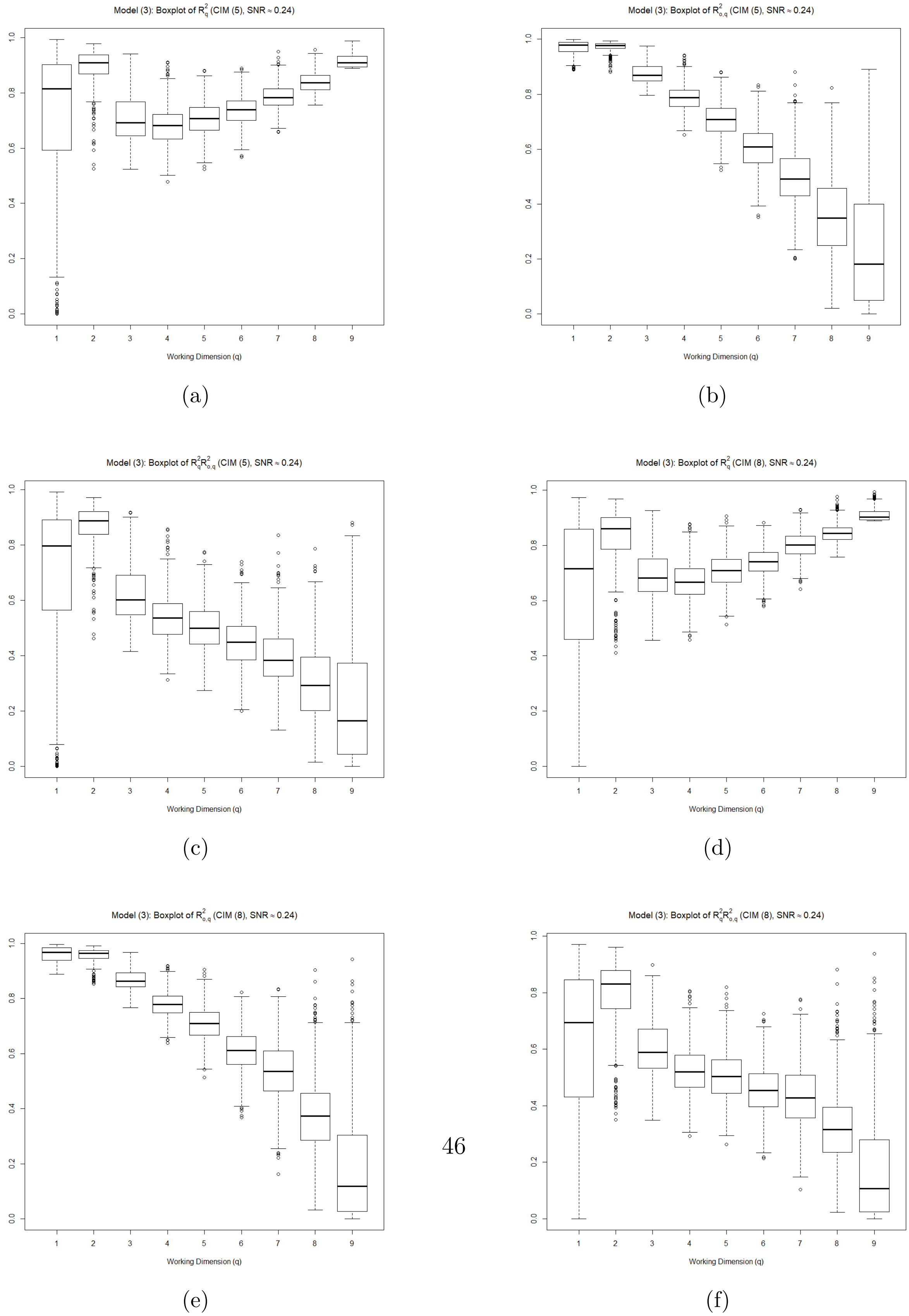
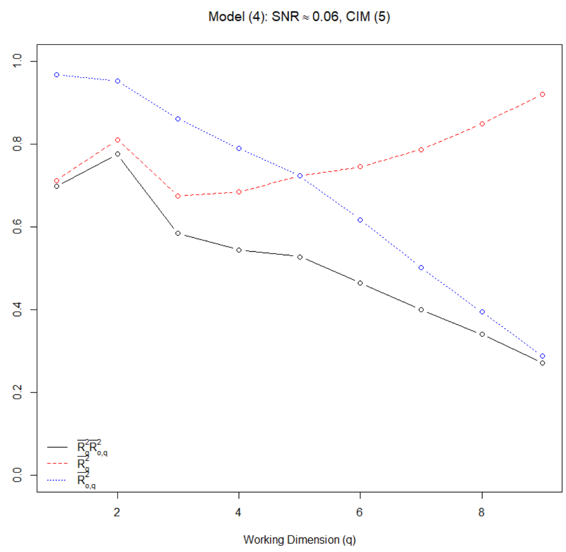
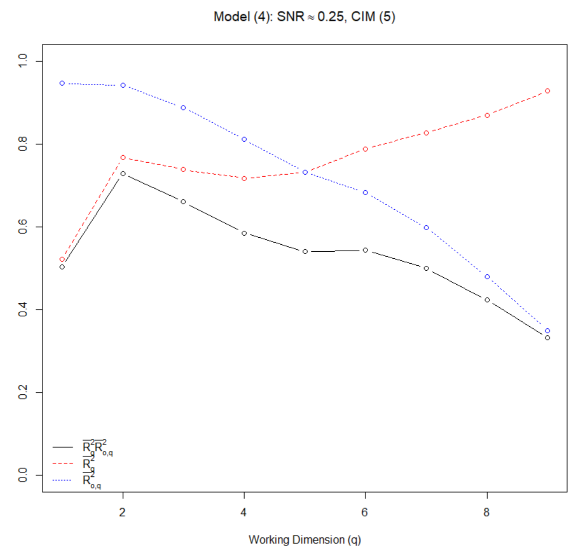


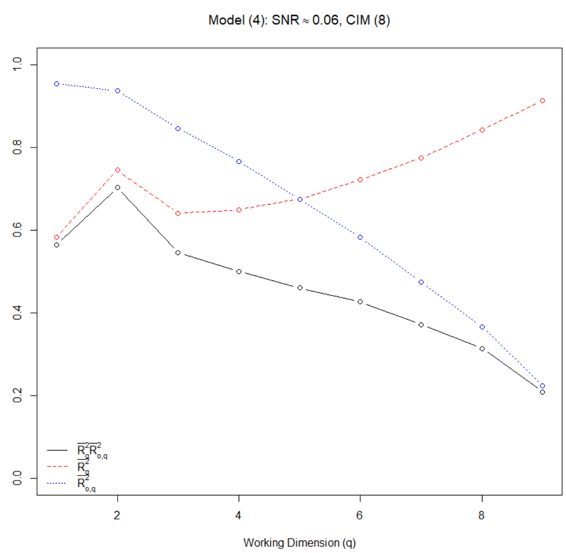
Figure S9: Boxplot versions of dimension estimation plots (500 replicates) for **CIM** in **Model (3)** using *Independent* $\mathbf{X}$, $n = 400$, SNR ≈ 0.24 ($\sigma = 2.03$), and $L = 5$ slices (panels (a)-(c)) and 8 slices (panels (d)-(f)).



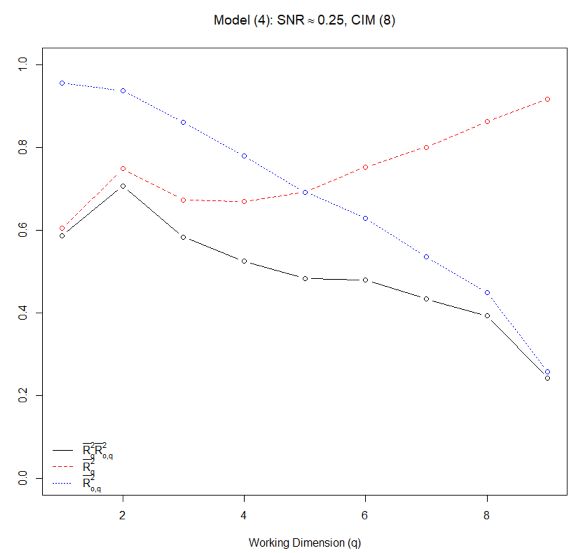
(a)



(b)

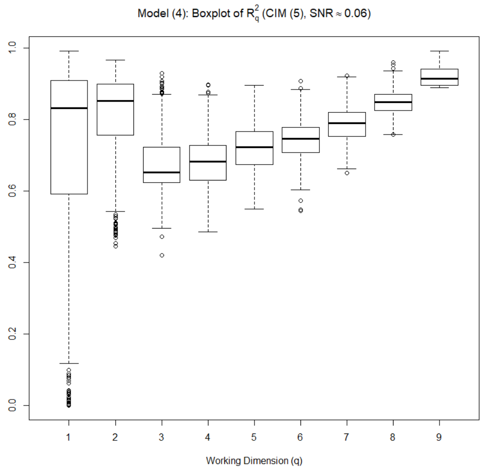


(c)

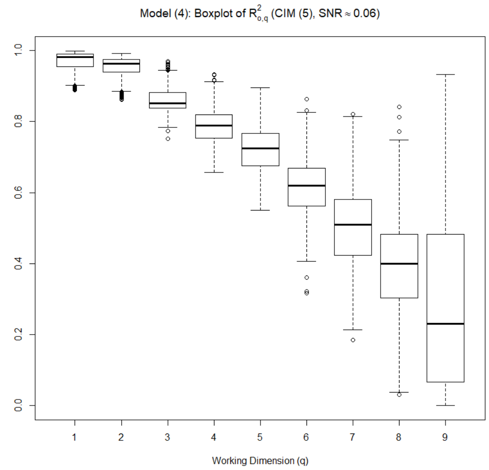


(d)

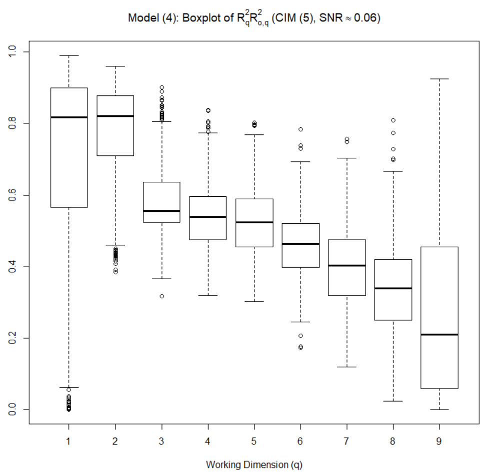
Figure S10: Dimension estimation plots (500 replicates) for **CIM** in **Model (4)** using *Independent* $\mathbf{X}$, $n = 400$, SNR's ≈ 0.06 ($\sigma = 4$) and ≈ 0.25 ($\sigma = 2.03$), and $L = 5$ slices (panels (a) and (b)) and 8 slices (panels (c) and (d)).



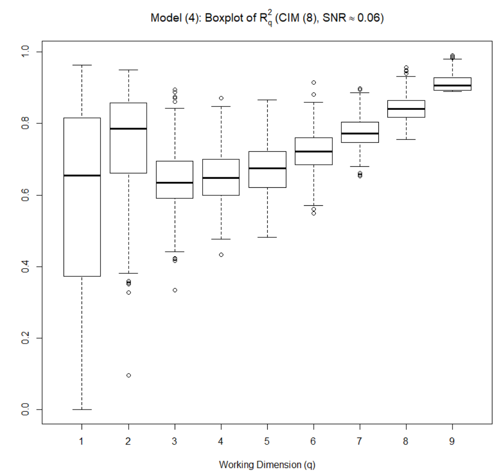
(a)



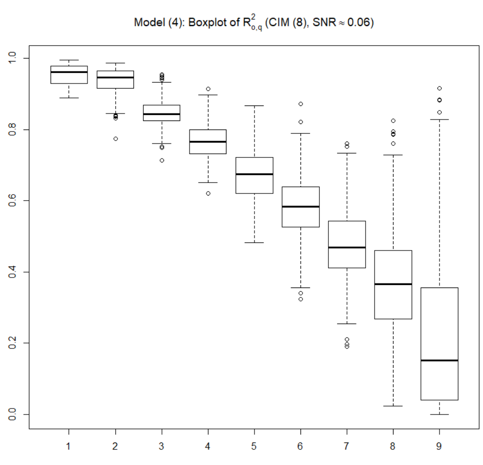
(b)



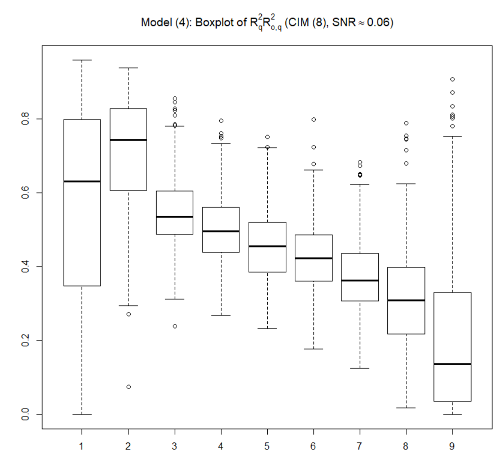
(c)



(d)



(e)



(f)

Figure S11: Boxplot versions of dimension estimation plots (500 replicates) for **CIM** in **Model (4)** using *Independent* $\mathbf{X}$, $n = 400$, SNR ≈ 0.06 ($\sigma = 4$), and $L = 5$ slices (panels (a)-(c)) and 8 slices (panels (d)-(f)).

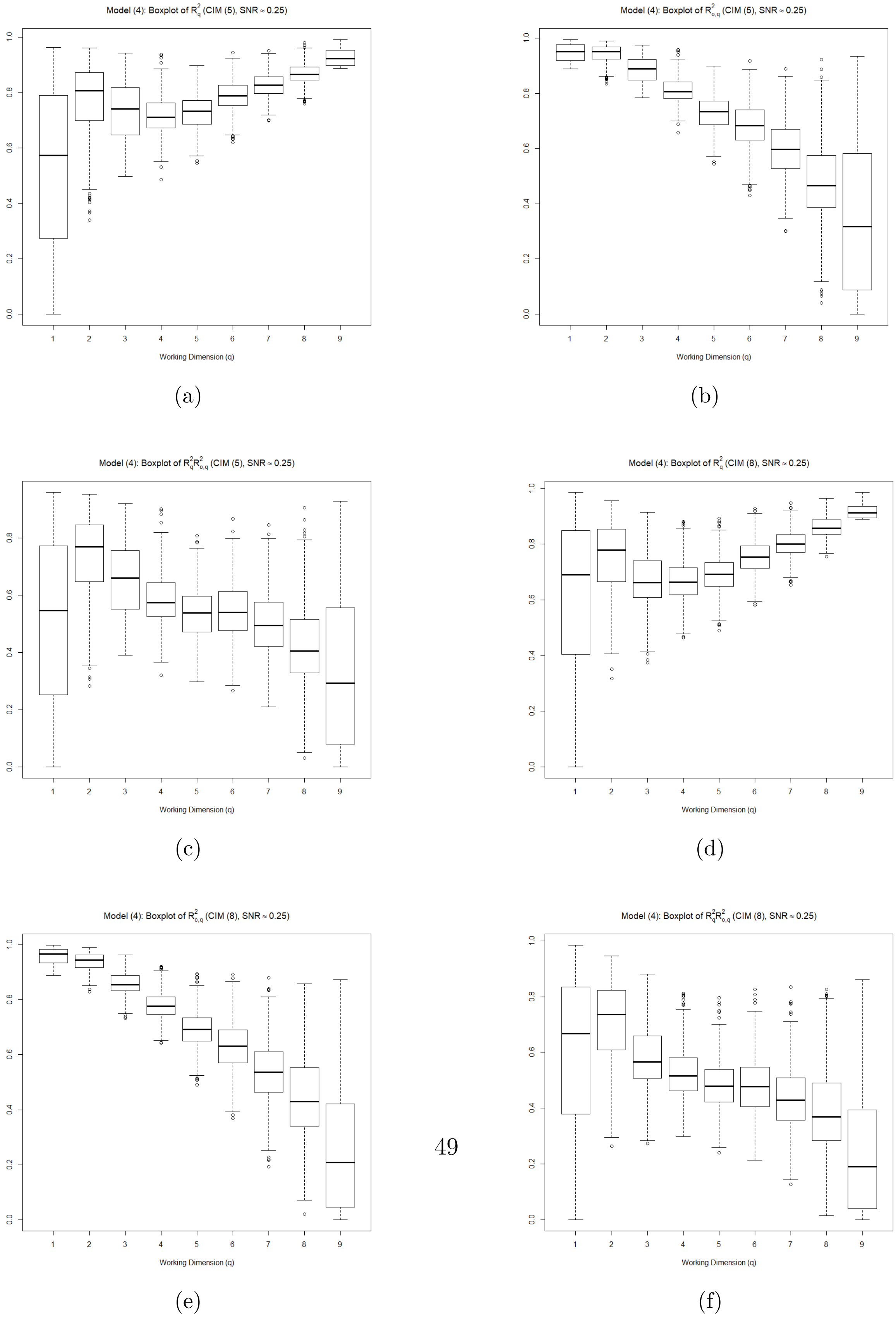
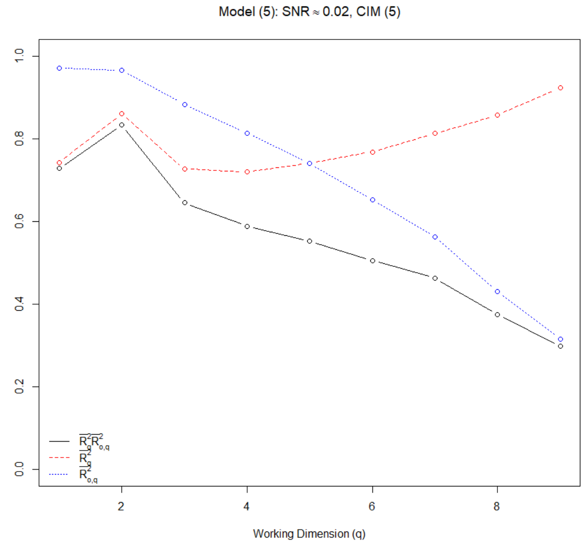
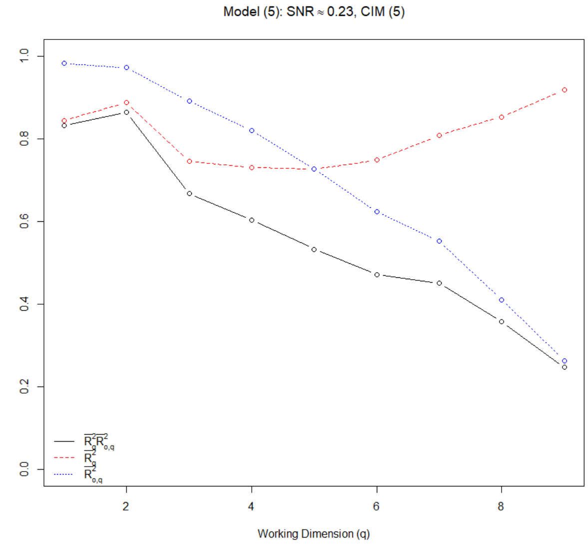


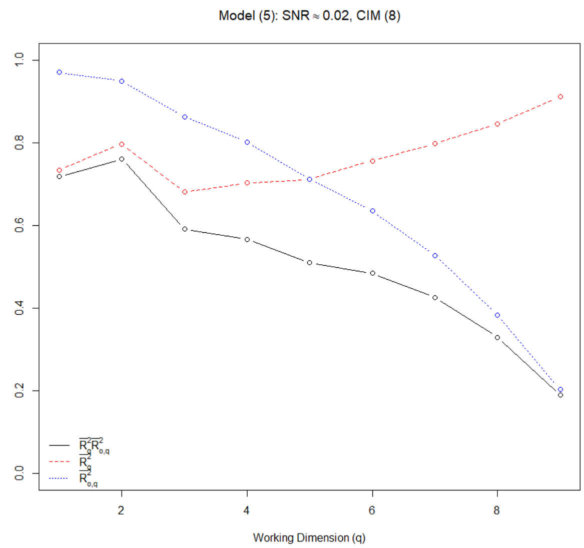
Figure S12: Boxplot versions of dimension estimation plots (500 replicates) for **CIM** in **Model (4)** using *Independent* $\mathbf{X}$, $n = 400$, SNR ≈ 0.25 ($\sigma = 2.03$), and $L = 5$ slices (panels (a)-(c)) and 8 slices (panels (d)-(f)).



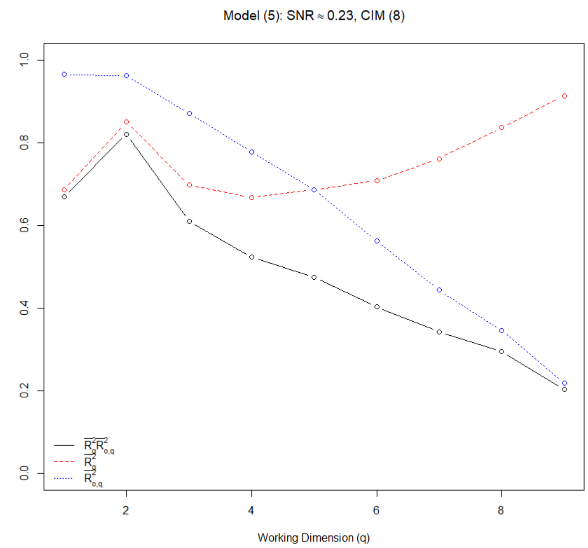
(a)



(b)

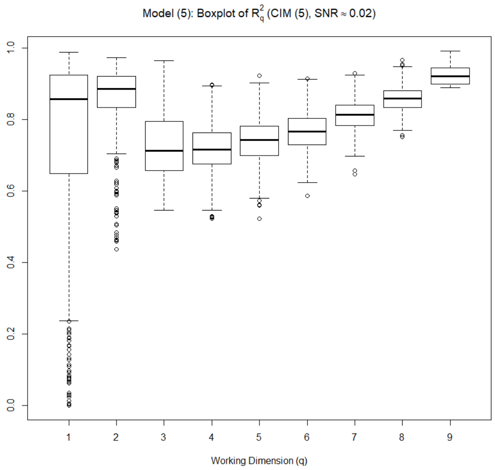


(c)

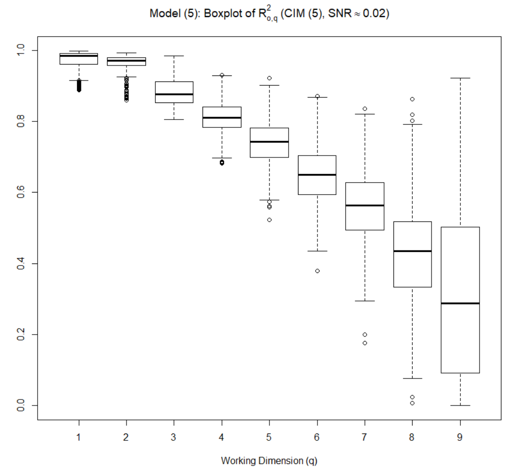


(d)

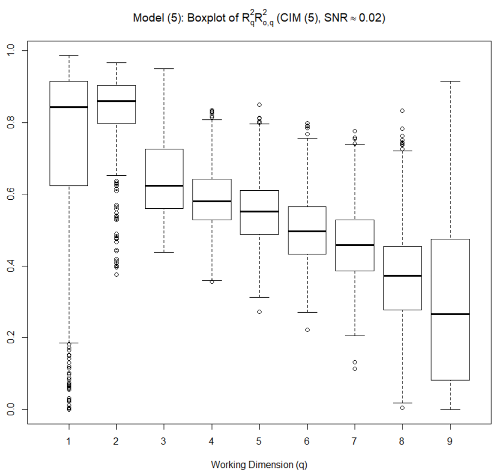
Figure S13: Dimension estimation plots (500 replicates) for **CIM** in **Model (5)** using *Independent* $\mathbf{X}$, $n = 400$, SNR's ≈ 0.02 ($\sigma = 0.2$) and ≈ 0.23 ($\sigma = 0.06$), and $L = 5$ slices (panels (a) and (b)) and 8 slices (panels (c) and (d)).



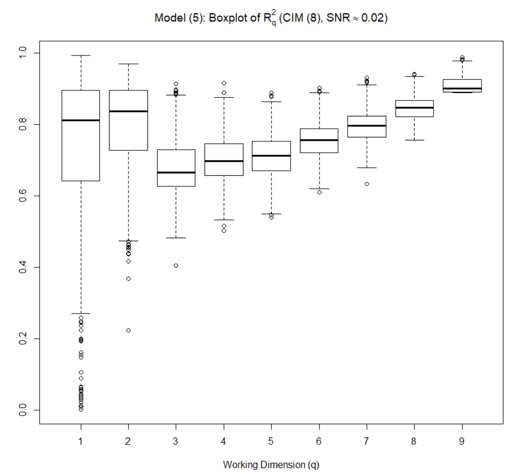
(a)



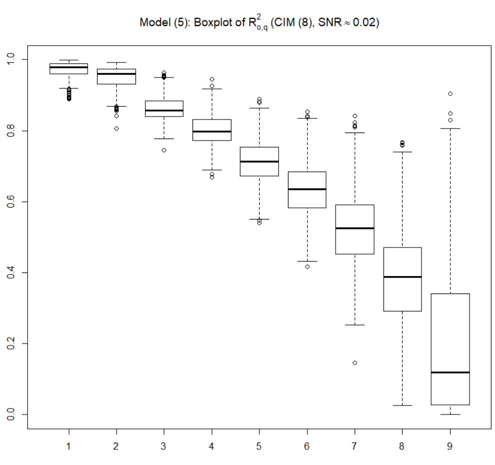
(b)



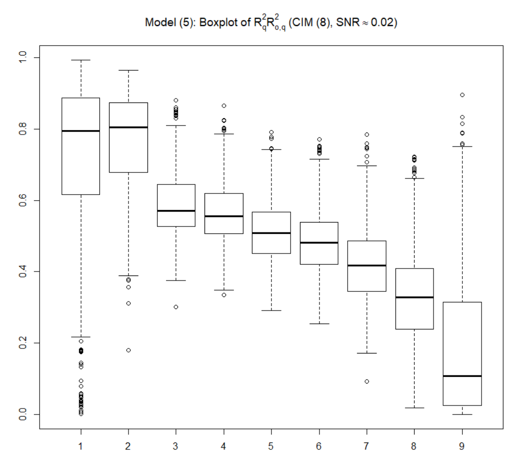
(c)



(d)

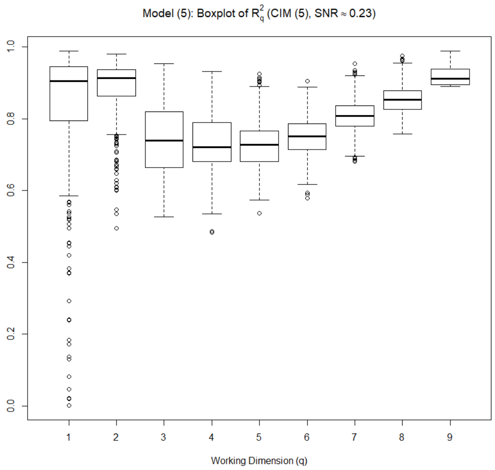


(e)

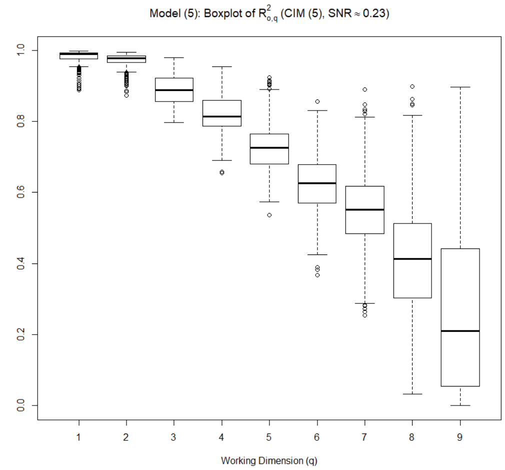


(f)

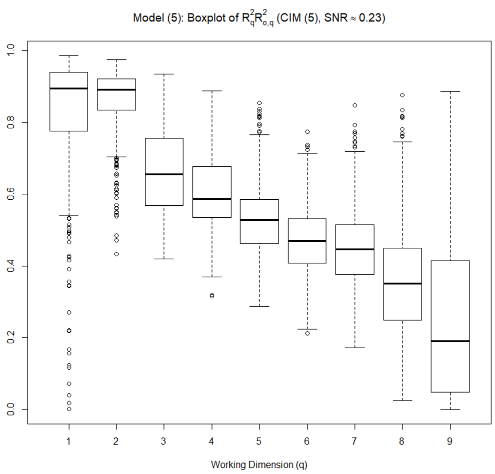
Figure S14: Boxplot versions of dimension estimation plots (500 replicates) for **CIM** in **Model (5)** using *Independent* $\mathbf{X}$, $n = 400$, SNR ≈ 0.02 ($\sigma = 0.2$), and $L = 5$ slices (panels (a)-(c)) and 8 slices (panels (d)-(f)).



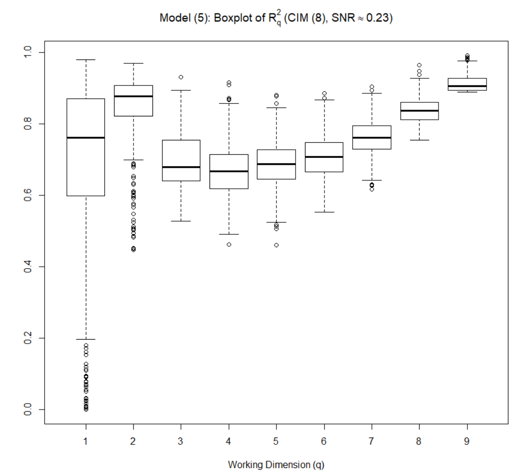
(a)



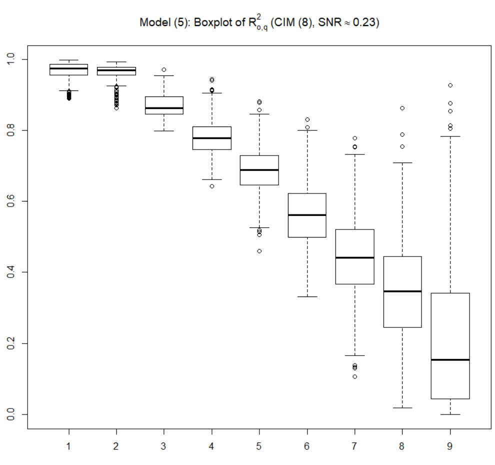
(b)



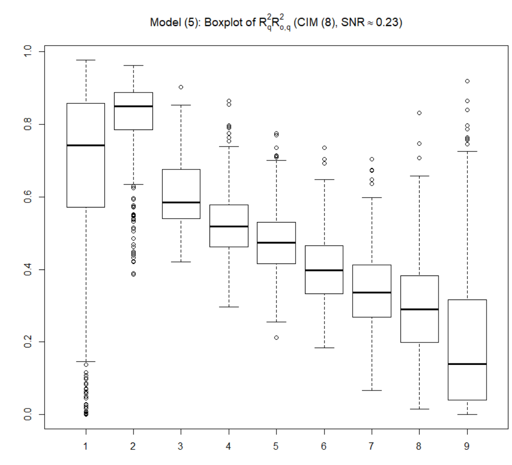
(c)



(d)

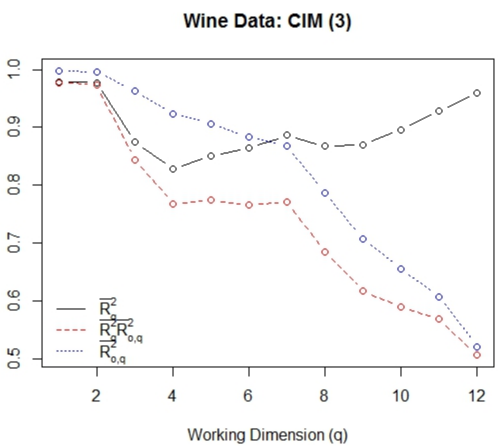


(e)

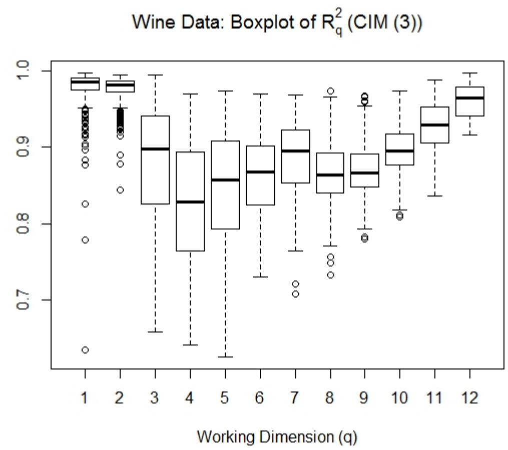


(f)

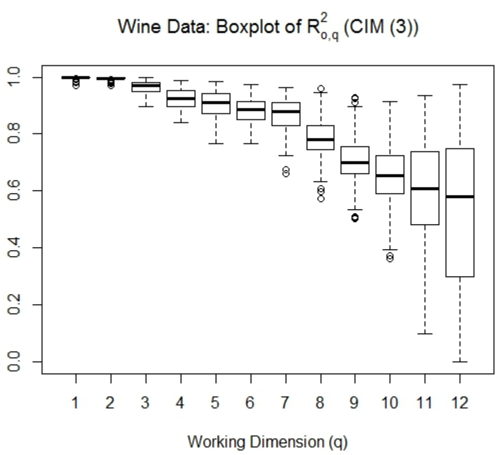
Figure S15: Boxplot versions of dimension estimation plots (500 replicates) for **CIM** in **Model (5)** using *Independent* $\mathbf{X}$, $n = 400$, SNR ≈ 0.23 ($\sigma = 0.06$), and $L = 5$ slices (panels (a)-(c)) and 8 slices (panels (d)-(f)).



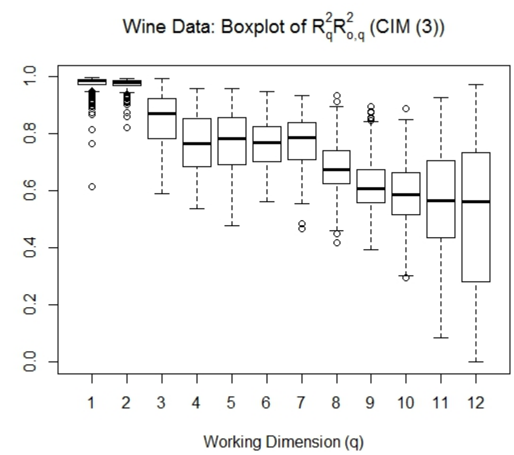
(a)



(b)

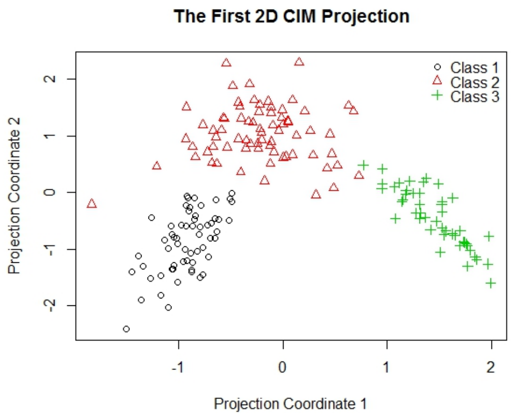


(c)

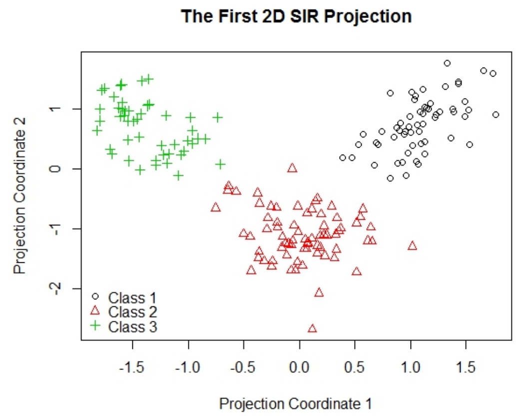


(d)

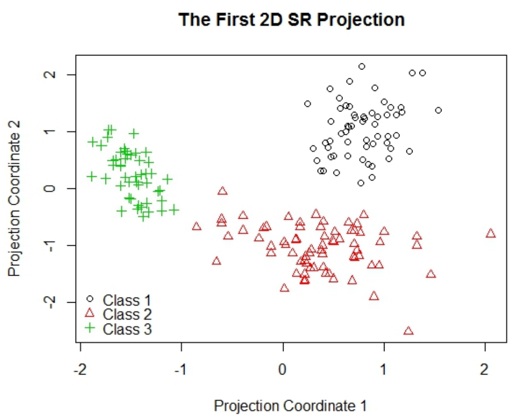
Figure S16: Panel (a): Dimension estimation plot (500 replicates) for **CIM** for **Wine Recognition Data** with $L = 3$ slices (for 3 response classes). Panels (b) - (d): Boxplot versions of (a).



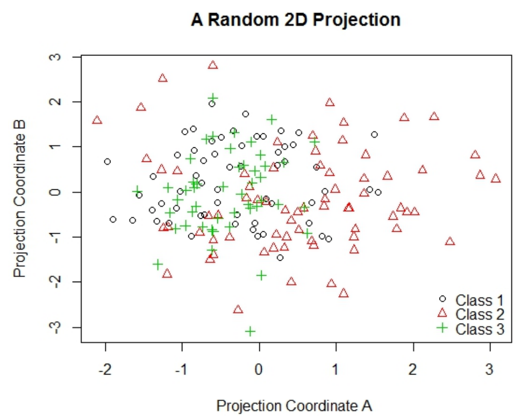
(a)



(b)



(c)



(d)

Figure S17: 2D projections of the **Wine Recognition Data** on the CS estimated via (a) **CIM**, (b) **SIR**, and (c) **SR**. (d): a **Random** 2D projection.

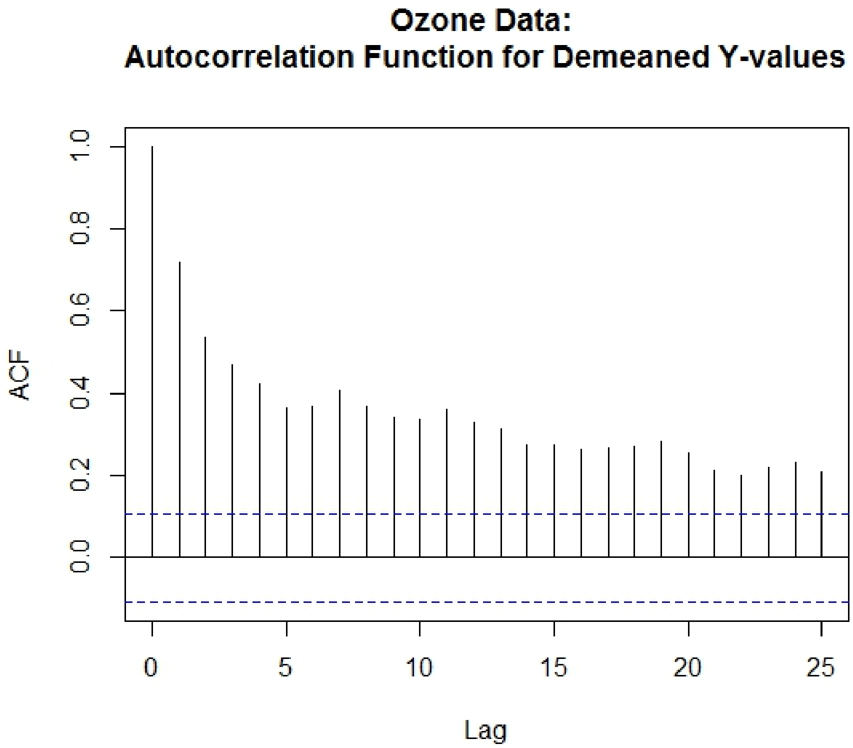
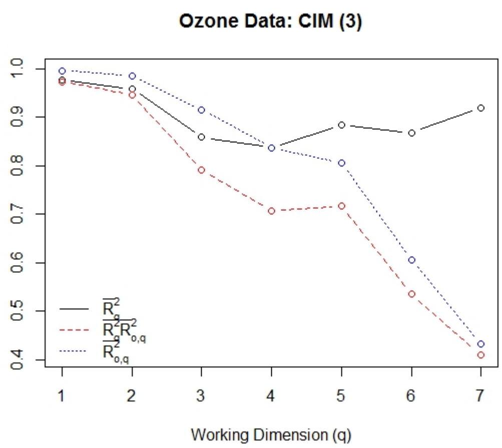
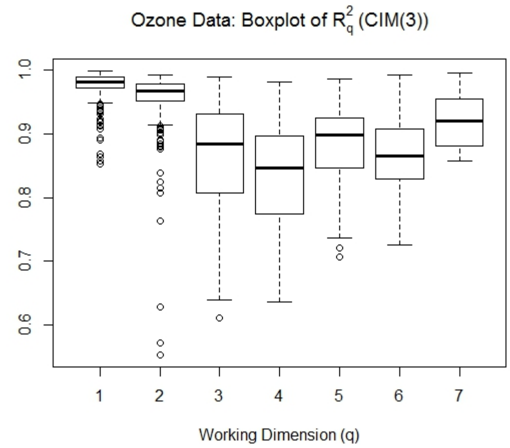


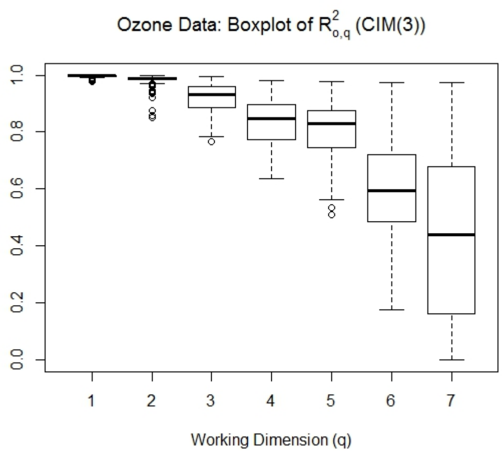
Figure S18: Autocorrelation function for the (de-meaned) Ozone concentration response in the **Ozone Data**.



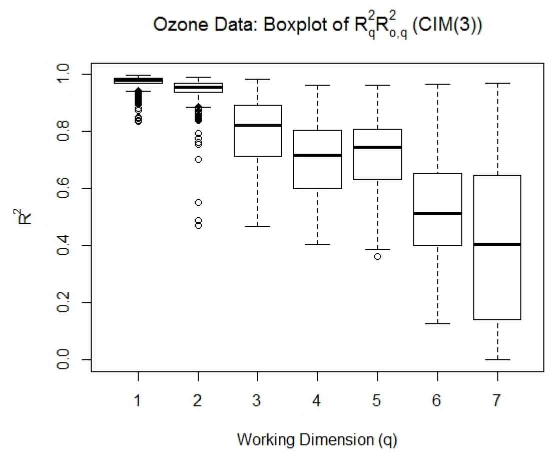
(a)



(b)

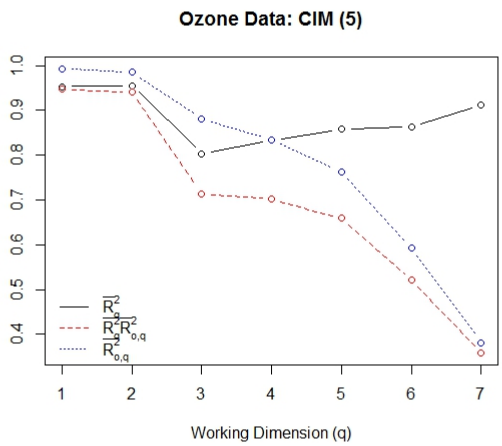


(c)

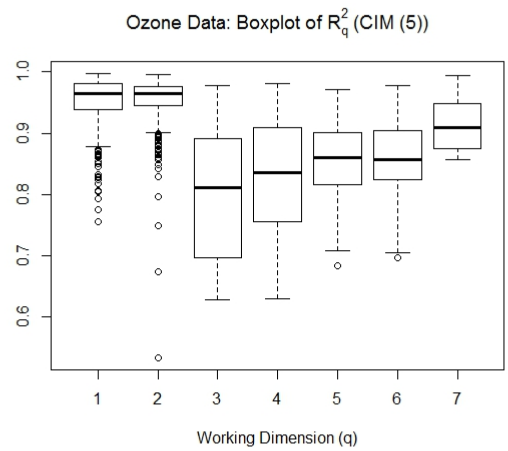


(d)

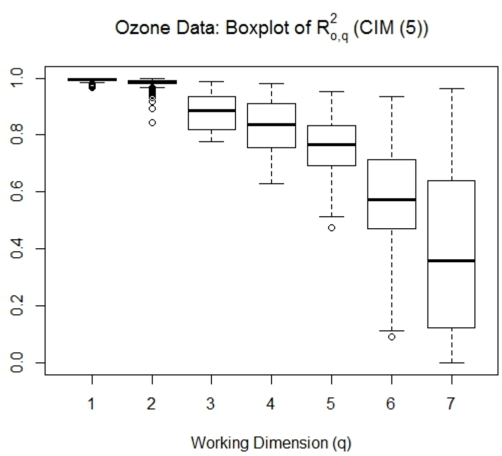
Figure S19: Panel (a): Dimension estimation plot (500 replicates) for **CIM** for **Ozone Data** with $L = 3$ slices. Panels (b) - (d): Boxplot versions of (a).



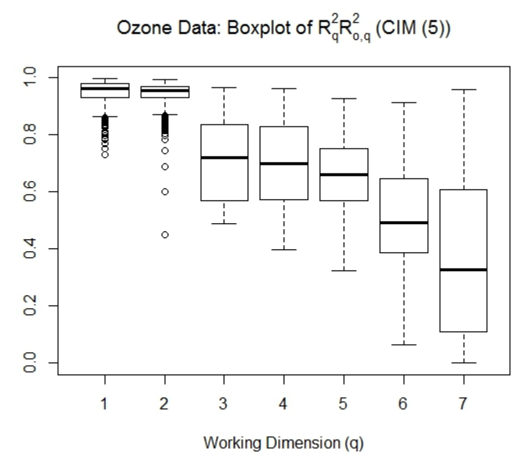
(a)



(b)



(c)



(d)

Figure S20: Panel (a): Dimension estimation plot (500 replicates) for **CIM** for **Ozone Data** with $L = 5$ slices. Panels (b) - (d): Boxplot versions of (a).

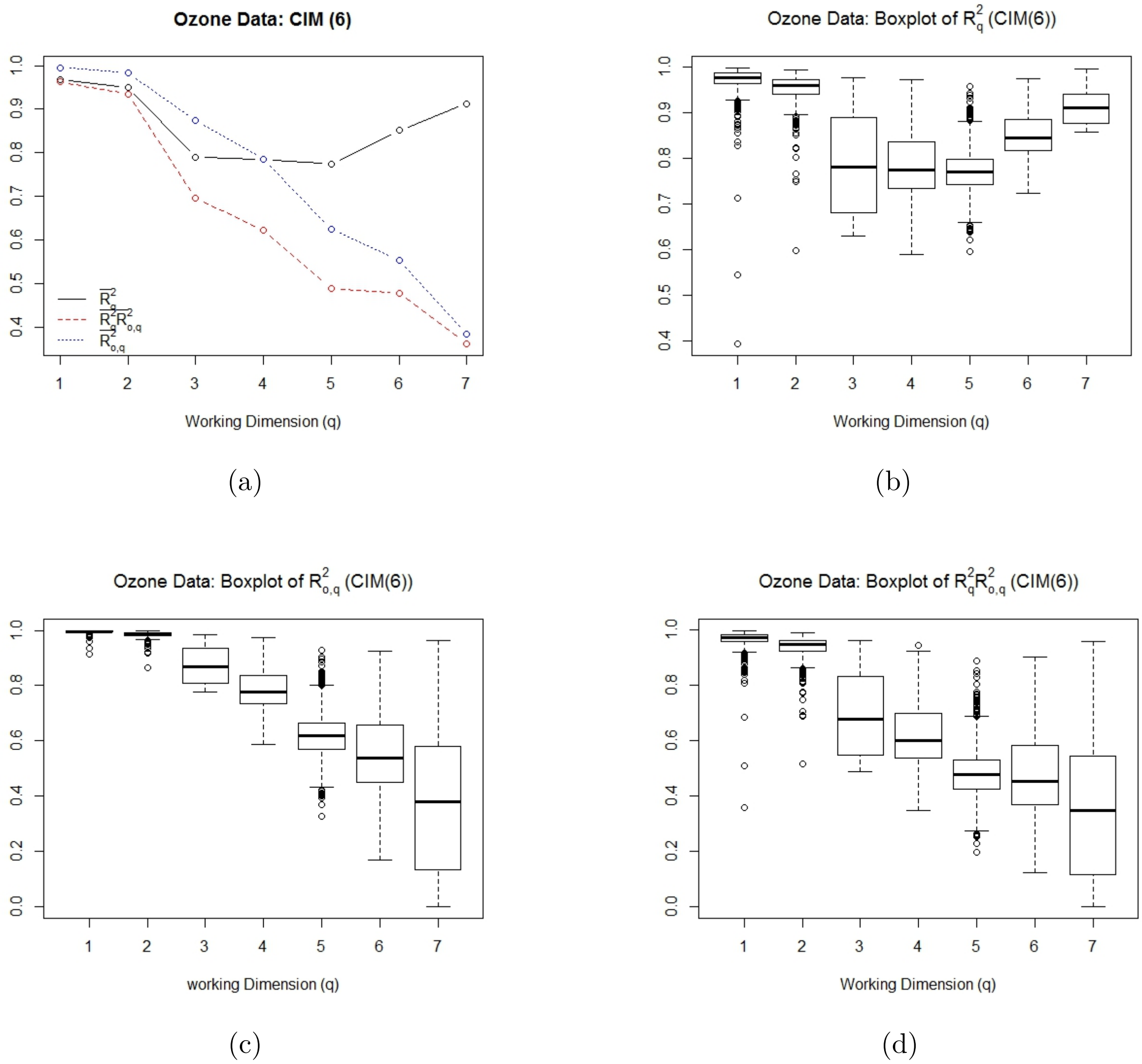
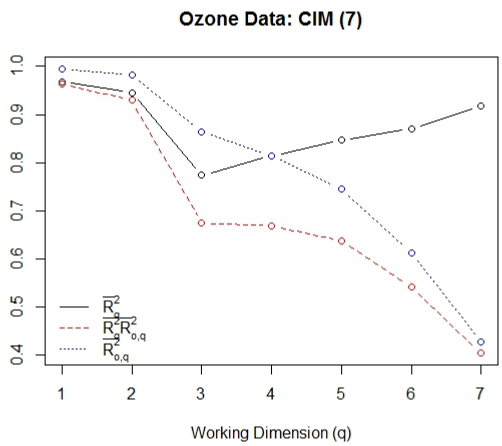
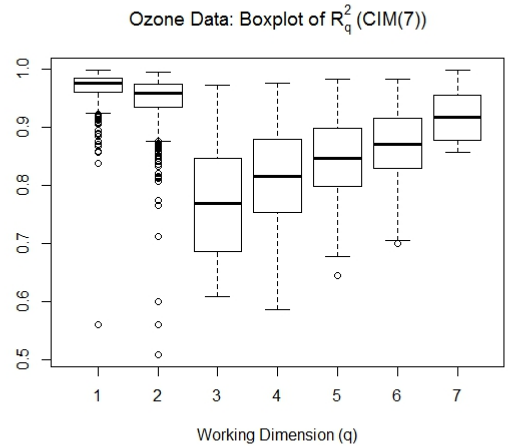


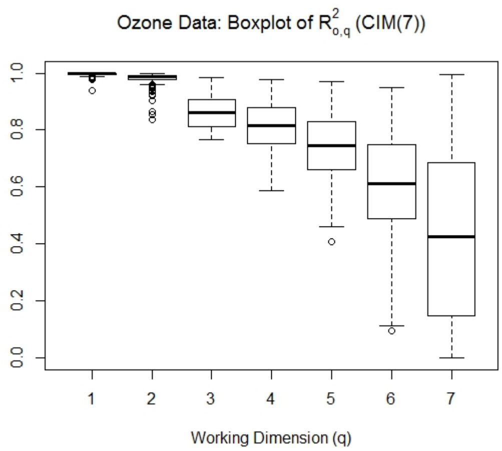
Figure S21: Panel (a): Dimension estimation plot (500 replicates) for **CIM** for **Ozone Data** with $L = 6$ slices. Panels (b) - (d): Boxplot versions of (a).



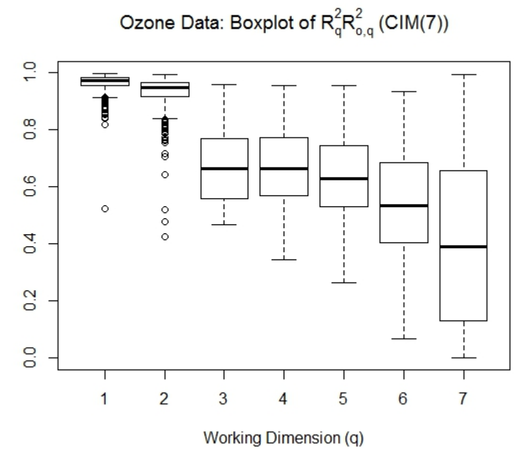
(a)



(b)

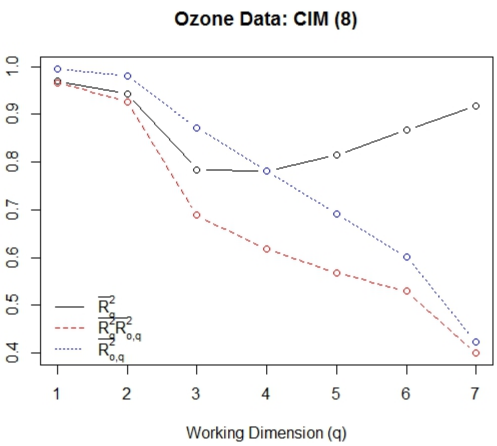


(c)

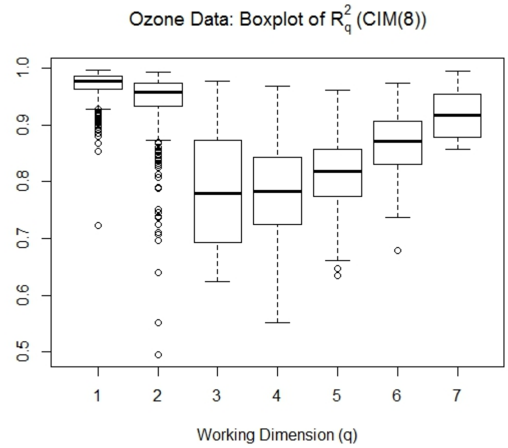


(d)

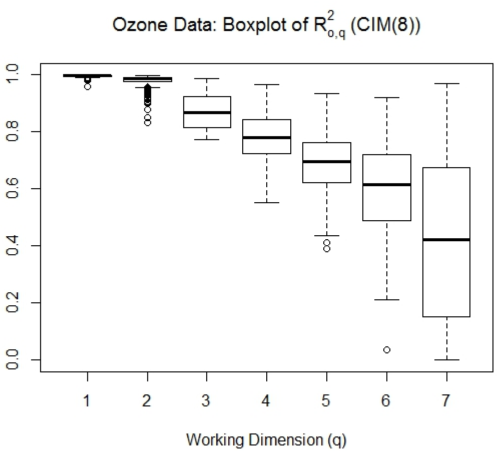
Figure S22: Panel (a): Dimension estimation plot (500 replicates) for **CIM** for **Ozone Data** with $L = 7$ slices. Panels (b) - (d): Boxplot versions of (a).



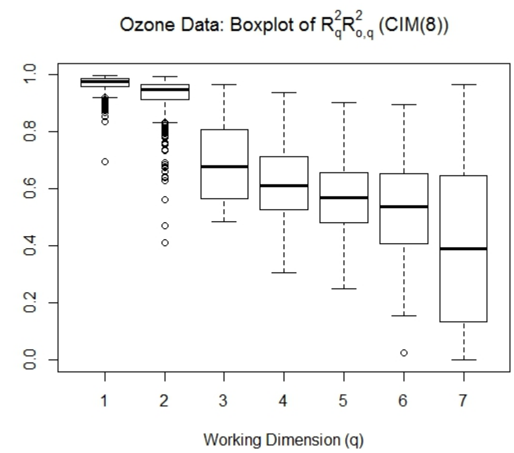
(a)



(b)

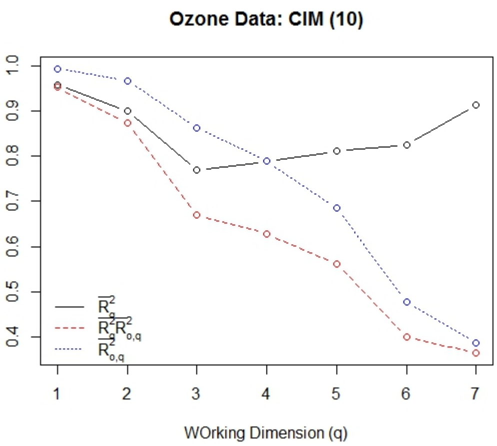


(c)

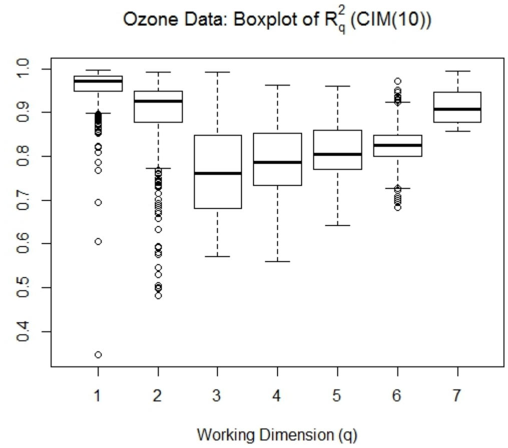


(d)

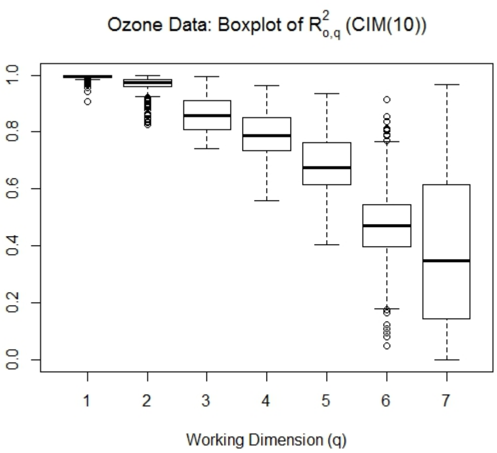
Figure S23: Panel (a): Dimension estimation plot (500 replicates) for **CIM** for **Ozone Data** with $L = 8$ slices. Panels (b) - (d): Boxplot versions of (a).



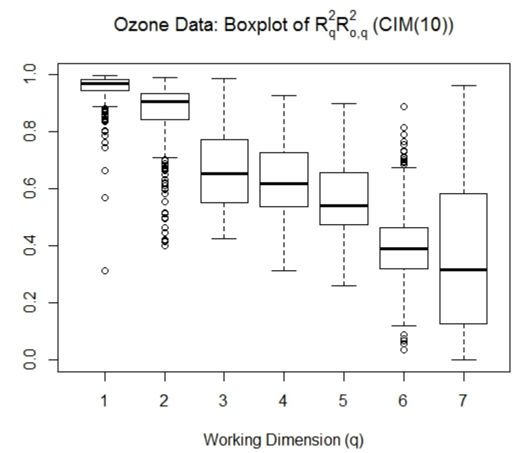
(a)



(b)

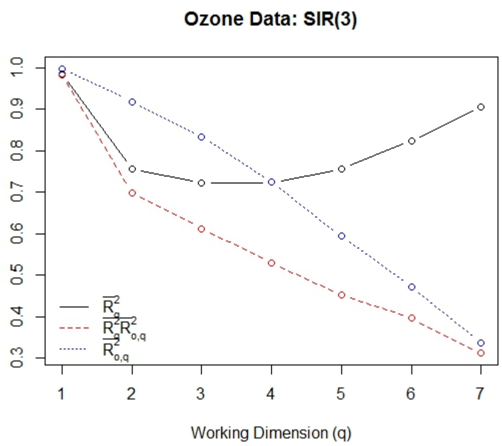


(c)

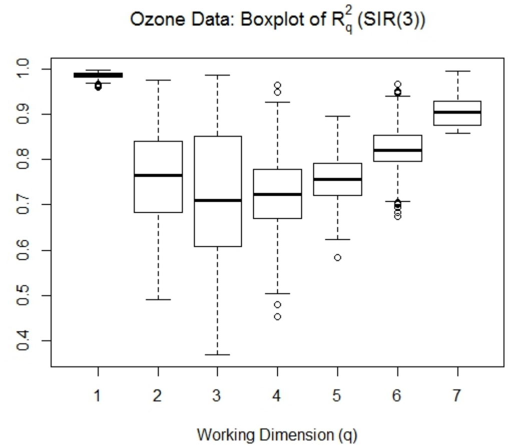


(d)

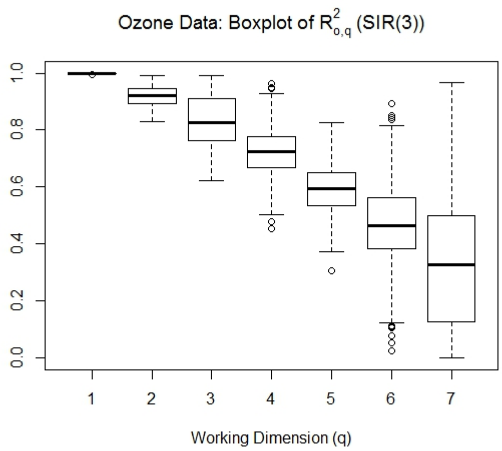
Figure S24: Panel (a): Dimension estimation plot (500 replicates) for **CIM** for **Ozone Data** with $L = 10$ slices. Panels (b) - (d): Boxplot versions of (a).



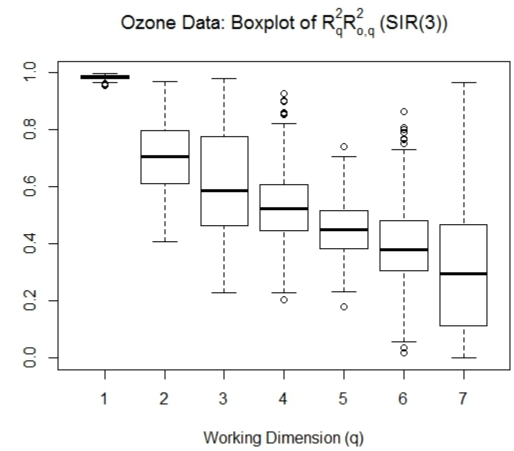
(a)



(b)

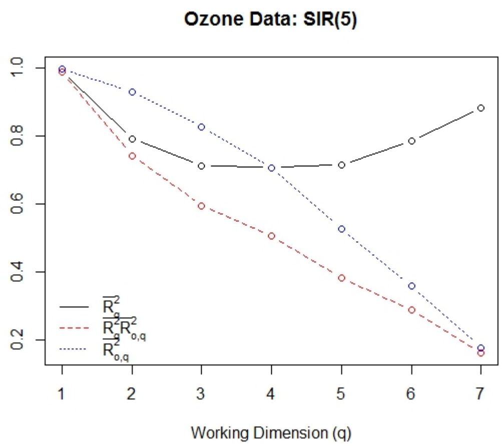


(c)

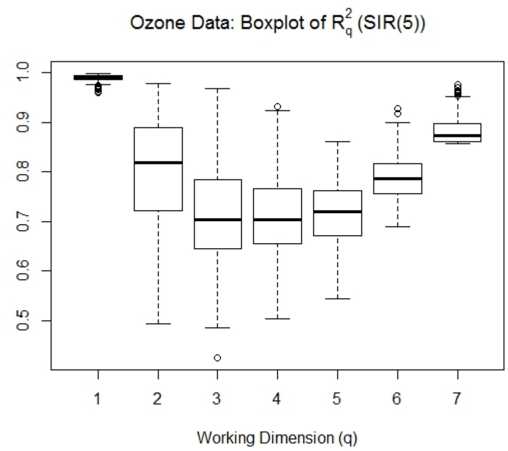


(d)

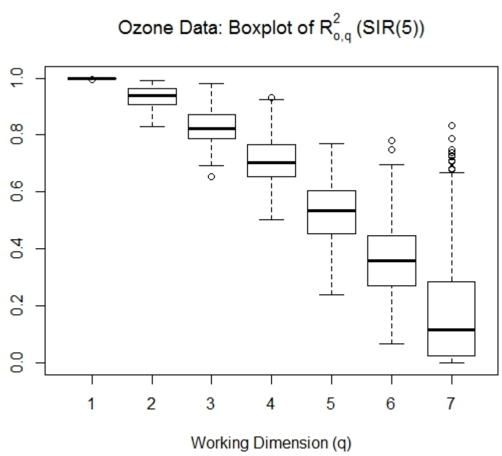
Figure S25: Panel (a): Dimension estimation plot (500 replicates) for **SIR** for **Ozone Data** with $L = 3$ slices. Panels (b) - (d): Boxplot versions of (a).



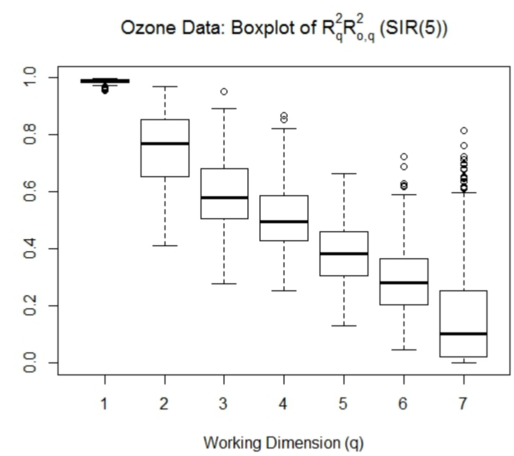
(a)



(b)



(c)



(d)

Figure S26: Panel (a): Dimension estimation plot (500 replicates) for **SIR** for **Ozone Data** with $L = 5$ slices. Panels (b) - (d): Boxplot versions of (a).

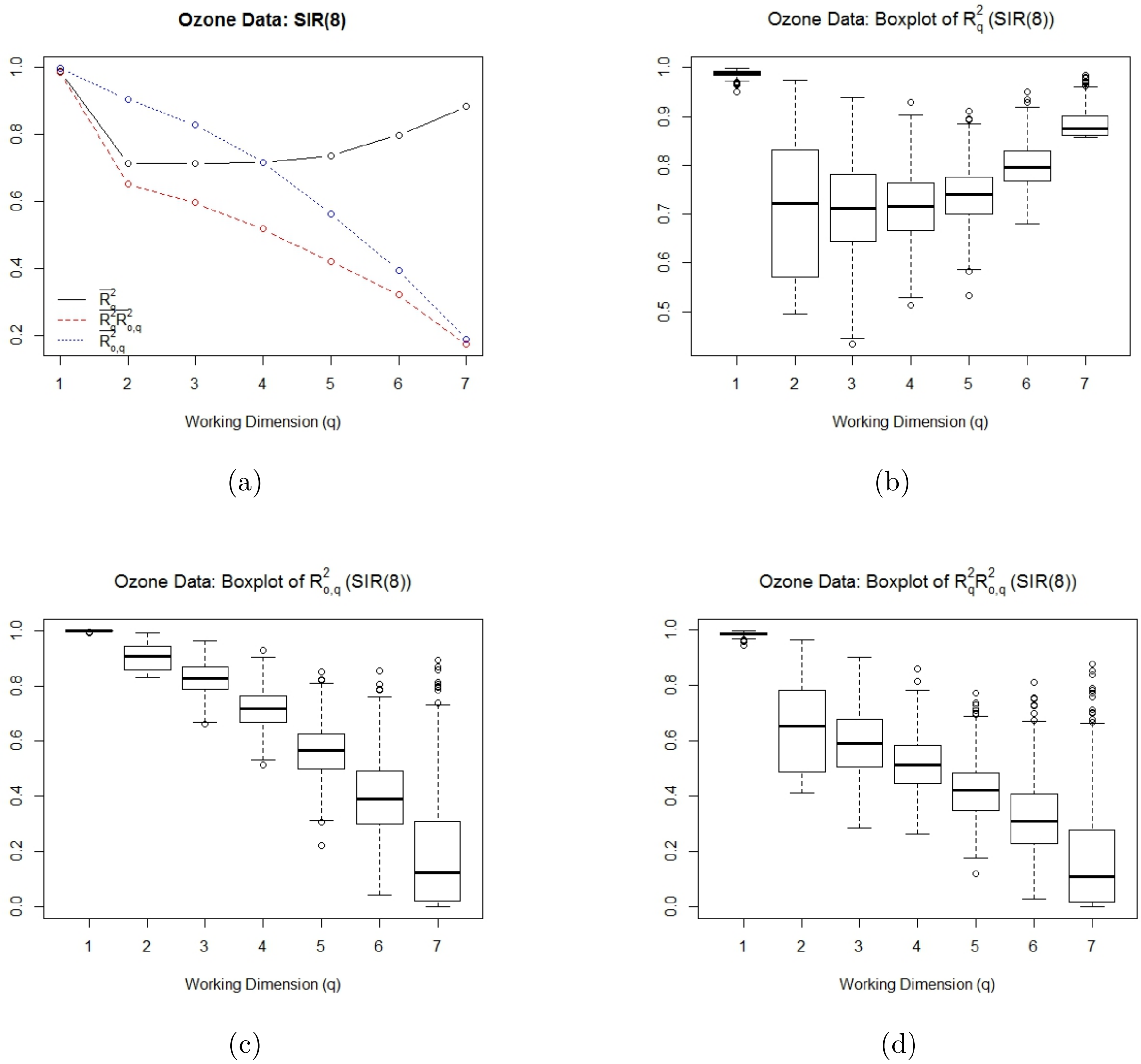


Figure S27: Panel (a): Dimension estimation plot (500 replicates) for **SIR** for **Ozone Data** with $L = 8$ slices. Panels (b) - (d): Boxplot versions of (a).

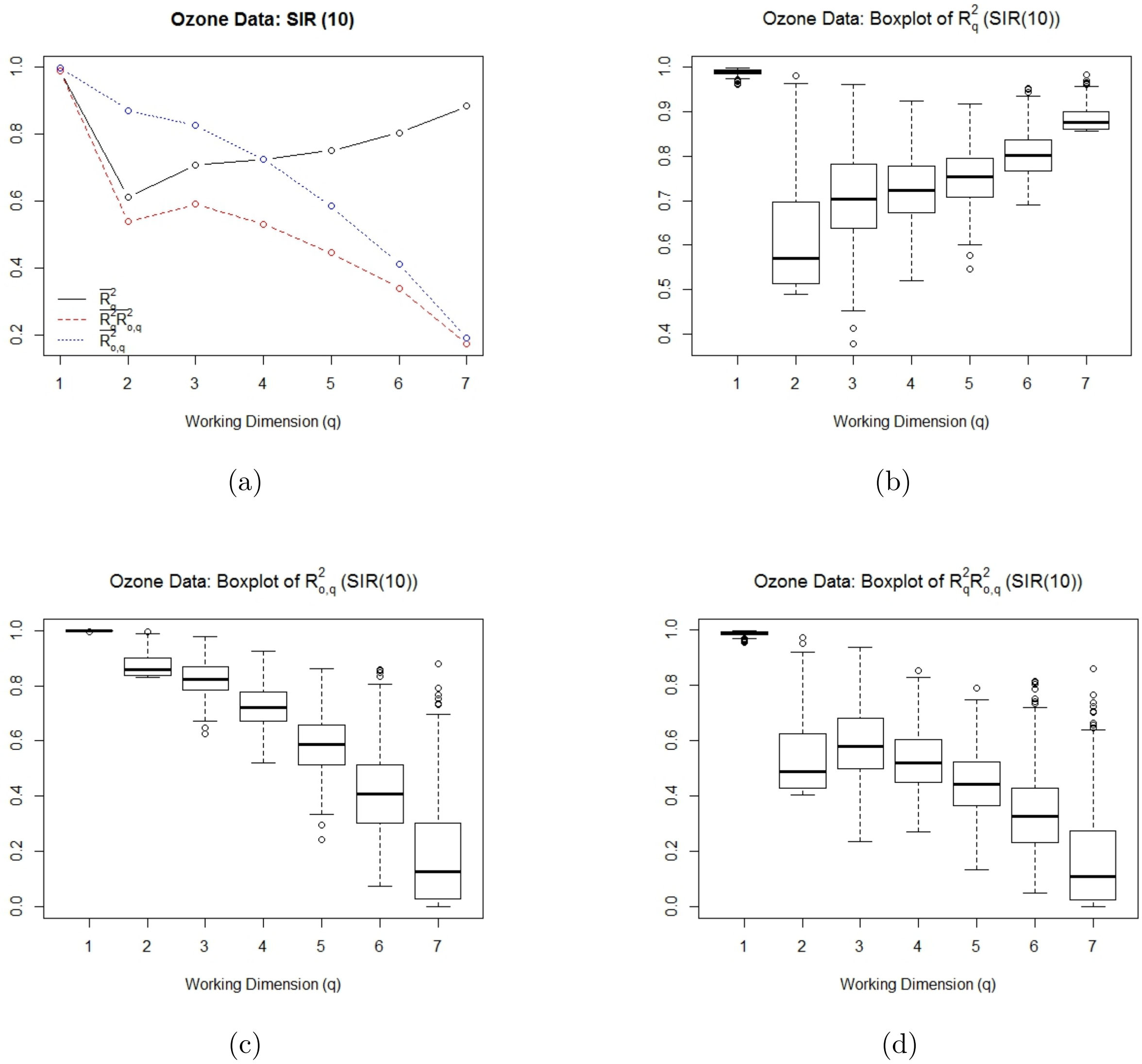


Figure S28: Panel (a): Dimension estimation plot (500 replicates) for **SIR** for **Ozone Data** with $L = 10$ slices. Panels (b) - (d): Boxplot versions of (a).

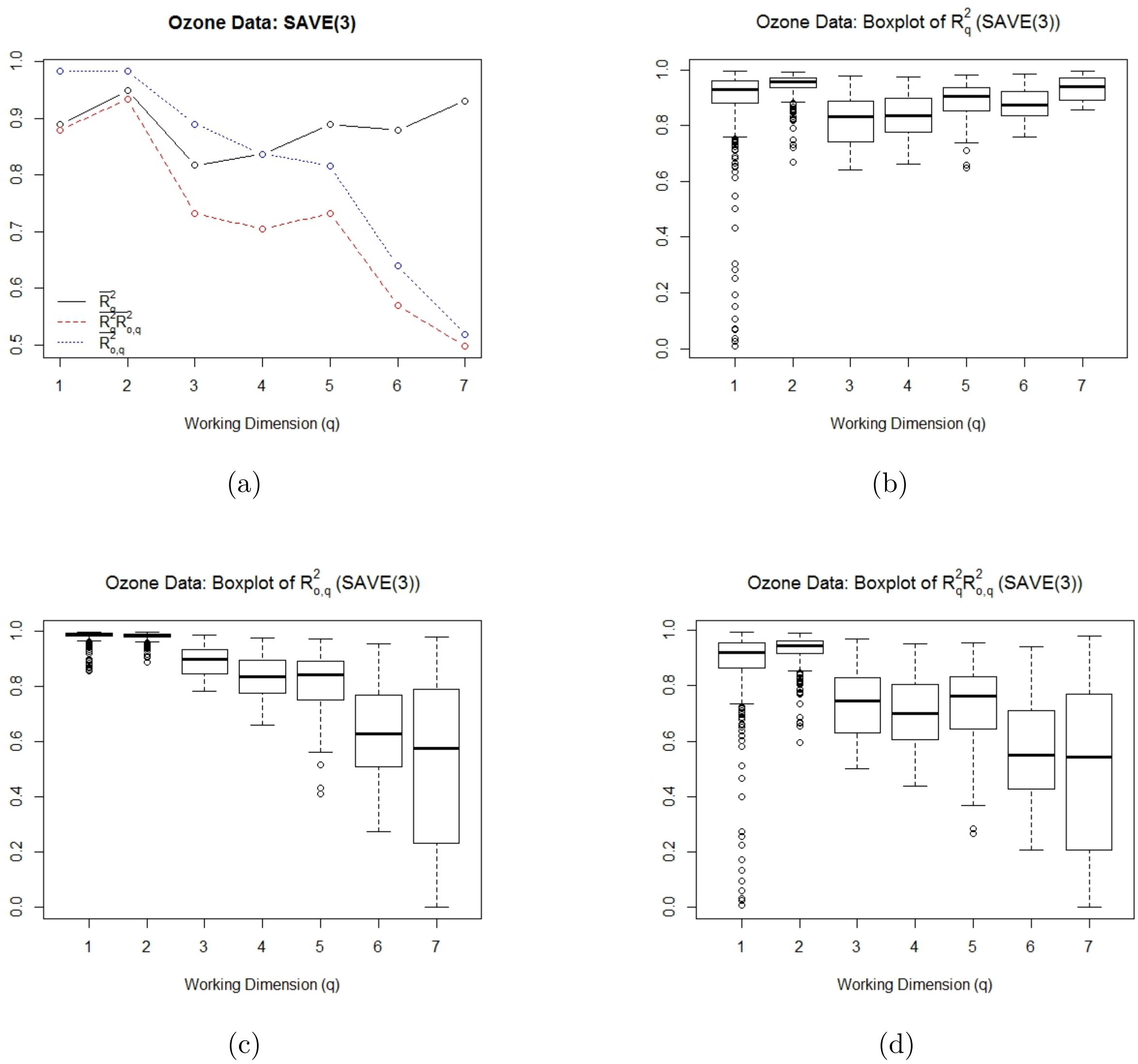
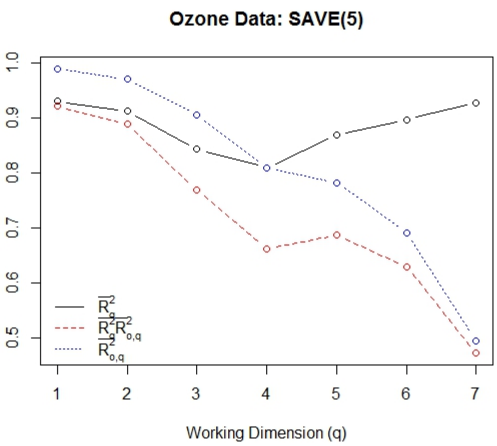
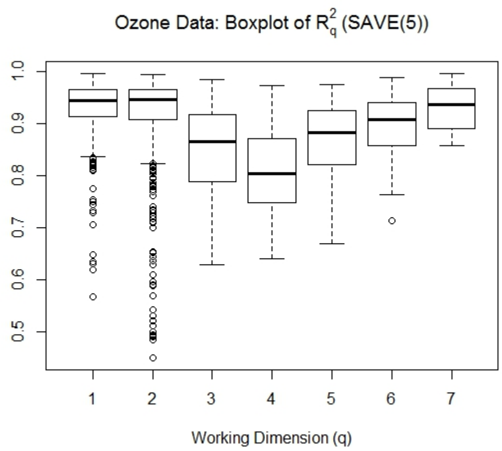


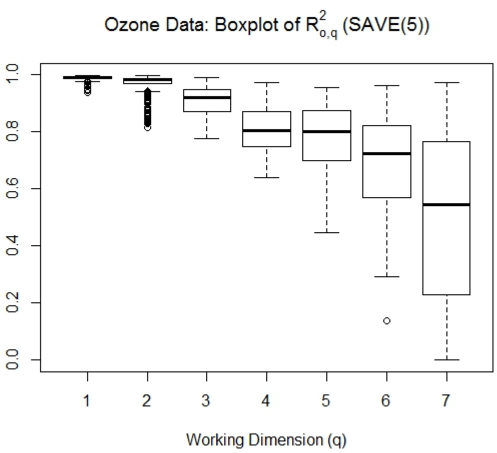
Figure S29: Panel (a): Dimension estimation plot (500 replicates) for **SAVE** for **Ozone Data** with $L = 3$ slices. Panels (b) - (d): Boxplot versions of (a).



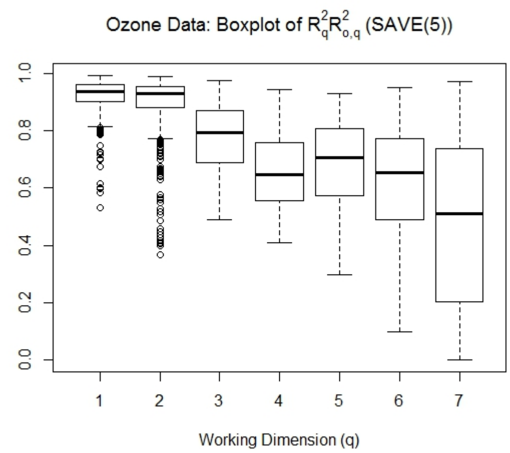
(a)



(b)

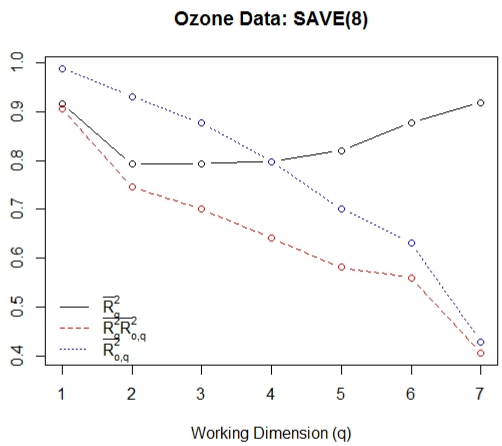


(c)

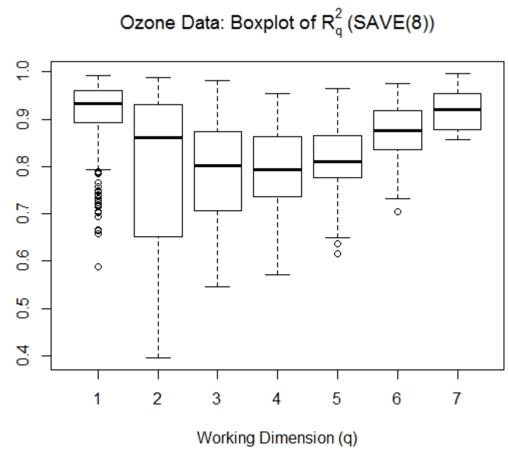


(d)

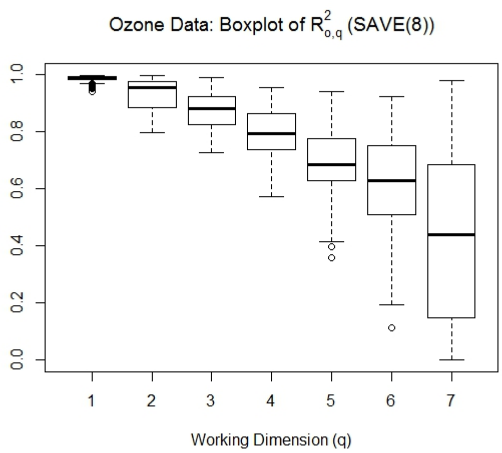
Figure S30: Panel (a): Dimension estimation plot (500 replicates) for **SAVE** for **Ozone Data** with $L = 5$ slices. Panels (b) - (d): Boxplot versions of (a).



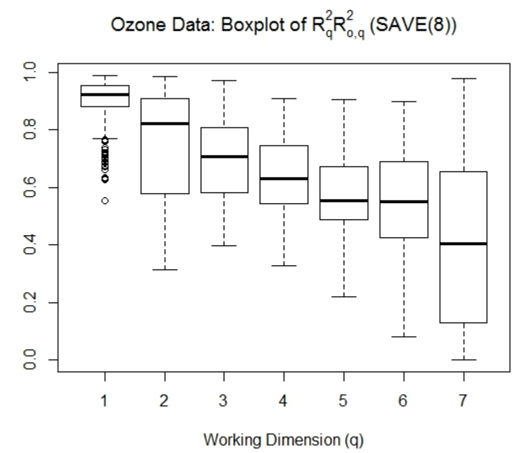
(a)



(b)

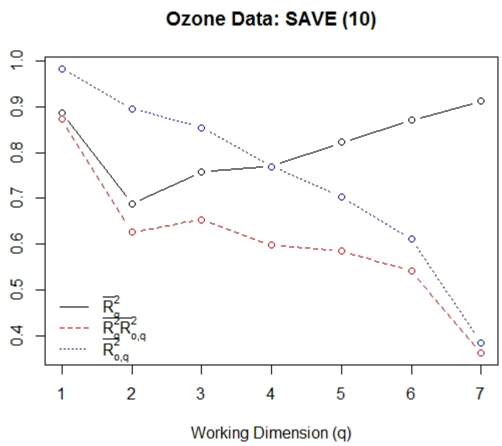


(c)

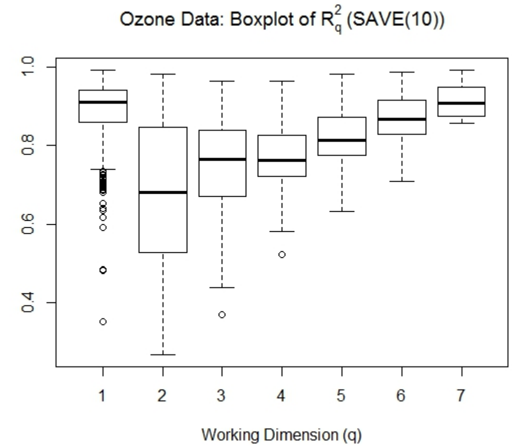


(d)

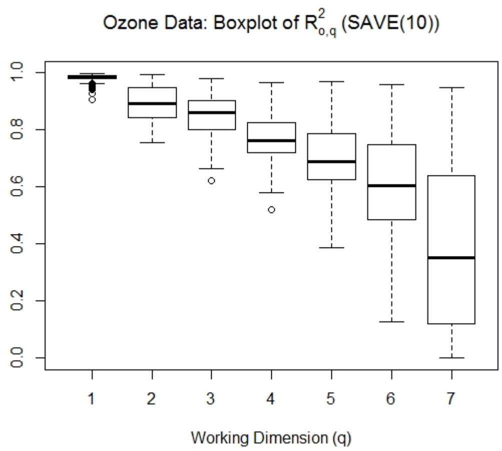
Figure S31: Panel (a): Dimension estimation plot (500 replicates) for **SAVE** for **Ozone Data** with $L = 8$ slices. Panels (b) - (d): Boxplot versions of (a).



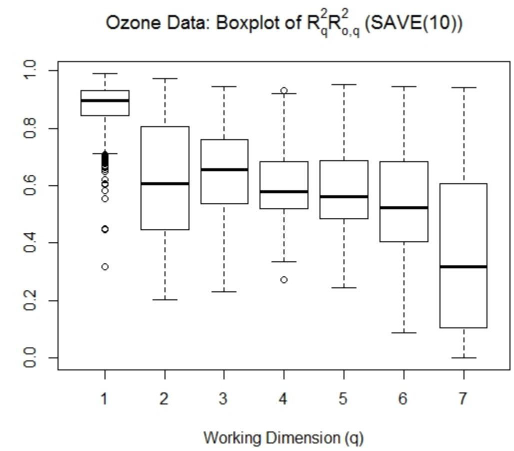
(a)



(b)

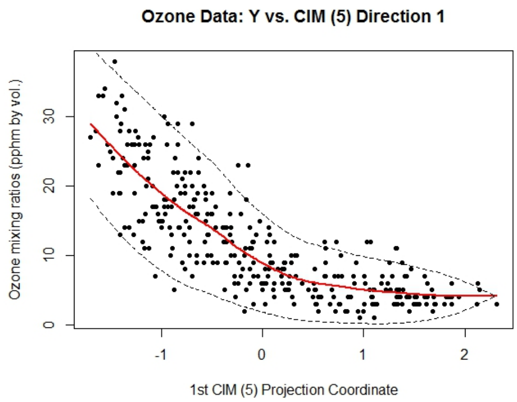


(c)

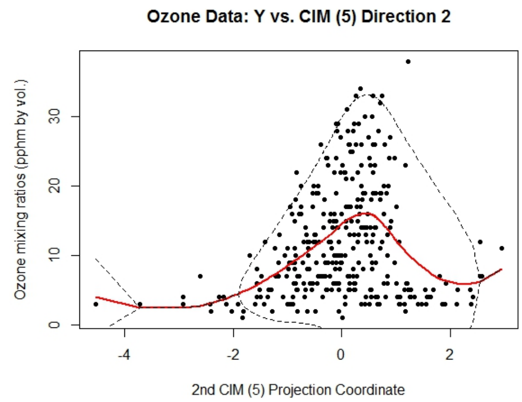


(d)

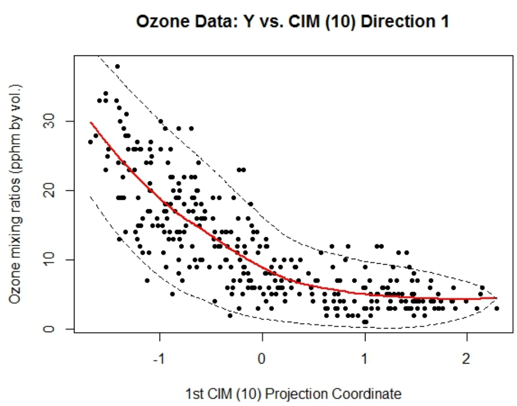
Figure S32: Panel (a): Dimension estimation plot (500 replicates) for **SAVE** for **Ozone Data** with $L = 10$ slices. Panels (b) - (d): Boxplot versions of (a).



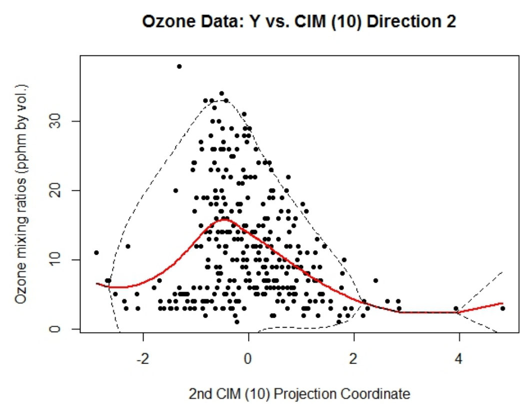
(a)



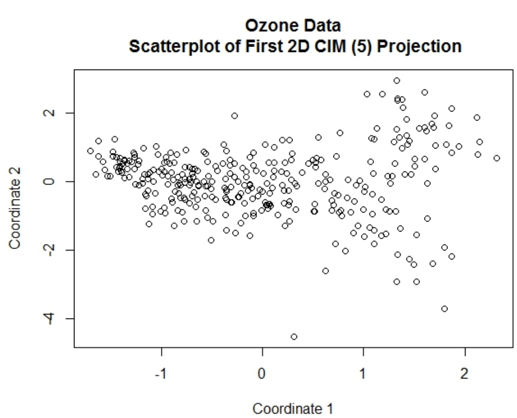
(b)



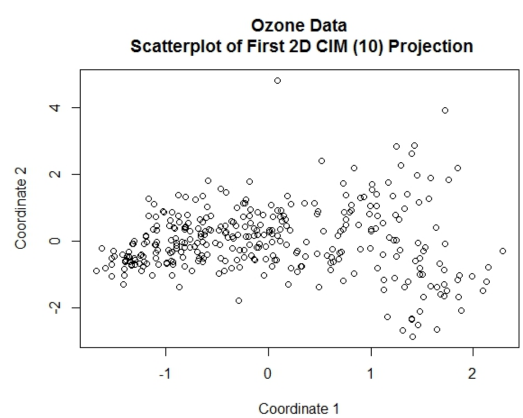
(c)



(d)



(e)



(f)

Figure S33: Panels (a) - (b): Scatterplots of Ozone concentration (mixing ratios in pphm by vol.) against the first and the second CIM directions respectively using $L = 5$ for **Ozone Data**. Solid lines are LOESS fits; dashed lines around them represent 95% prediction bands obtained using the CRAN package 'msir'. Panels (c) - (d): Similar plots using $L = 10$. Panels (e) - (f): Scatterplots of the first and the second CIM directions using $L = 5$ and $L = 10$ respectively.

References

Bowman, A. and Foster, P. (1993). Adaptive smoothing and density-based tests of multivariate normality. *Journal of the American Statistical Association*, 88(422):529–537.

Comon, P. (1994). Independent component analysis, a new concept? *Signal processing*, 36(3):287–314.

Friedman, J. H. and Tukey, J. W. (1974). A projection pursuit algorithm for exploratory data analysis.

Huber, P. J. (1985). Projection pursuit. *The annals of Statistics*, pages 435–475.

Hui, G. and Lindsay, B. G. (2010). Projection pursuit via white noise matrices. *Sankhya B*, 72(2):123–153.

Hyvärinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural networks*, 13(4):411–430.

Jones, M. C. and Sibson, R. (1987). What is projection pursuit? *Journal of the Royal Statistical Society. Series A (General)*, pages 1–37.

Jutten, C. and Herault, J. (1991). Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture. *Signal processing*, 24(1):1–10.

Lindsay, B. G. and Yao, W. (2012). Fisher information matrix: A tool for dimension reduction, projection pursuit, independent component analysis, and more. *Canadian Journal of Statistics*, 40(4):712–730.

Zhou, X. (2017). *Robust Mixed-effects Segmented Regression Models and Independent Component Analysis*. PhD thesis, Department of Statistics, University of California, Riverside.