Statistics and Its Interface Volume 14 (2021) 417–430

Feature screening via Bergsma-Dassios sign correlation learning

Daojiang He*, Xinxin Hao, Kai Xu, Lei He, and Youxin Liu

Robust rank correlation screening (RRCS) procedure that is built on Kendall τ, has been suggested by Li, Peng, Zhang and Zhu (2012) as a robust alternative to the sure independence screening (SIS) method that is based on the Pearson's correlation. However, as a drawback for certain applications is that τ may be zero even if there is an association between two random variables, RRCS is not omnibus, only having an ability to detect monotonic effects. In this paper, we use the Bergsma-Dassios sign correlation (Bergsma and Dassios, 2014, τ_b^*) to introduce a new SIS procedure. We advocate using the τ_b^*-SIS for three reasons. First, as τ_b^* possesses the necessary and intuitive properties as a correlation index, the τ_b^*-SIS has a better screening ability for nonlinear effects including interactions and heterogeneity compared with the RRCS. Second, as τ_b^* is a natural extension of τ, the τ_b^*-SIS is conceptually simple, easy to implement and robust to the presence of extreme values and outliers in the observations. Third, without assuming any moment condition on the response and predictors, the τ_b^*-SIS enjoys several appealing properties, such as the sure screening property, ranking consistency property and the characteristic of minimum model size. We demonstrate the merits of the τ_b^*-SIS procedure through extensive Monte Carlo experiments and illustrate the method through a real-data example.

Keywords and phrases: Bergsma-Dassios sign correlation, Feature screening, Kendall τ, Sure screening property, Ranking consistency property, Minimum model size.

1. INTRODUCTION

With the development of modern scientific techniques, ultrahigh dimensional data sets can be conveniently obtained in diverse fields of the sciences, engineering and humanities. Much importance has been placed on the problem of variable selection which plays a significant role in data analysis.

The idea of shrinkage estimation with penalization is very important and can be applied to handle high dimensional data, with examples given of bridge regression (Frank and Friedman, 1993; Huang et al., 2008), LASSO (Tibshirani, 1996; Van de Geer, 2008), elastic-net (Zou and Hastie, 2005), adaptive LASSO (Zou, 2006), SCAD (Fan and Li, 2001; Fan and Lv, 2011; Fan and Peng, 2004) and Dantzig selector (Candes and Tao, 2007). However, these methods may not perform well when the dimension of predictor variables p is much larger than sample size n. An alternative approach that has been advocated in Fan and Lv (2008) is to first perform variable screening to reduce the number of predictors to a relative large scale that is smaller than or equal to the sample size, and then finish the final variable selection and parameter estimation simultaneously through a more sophisticated technique. Since the seminal work of Fan and Lv (2008) on sure independence screening (SIS, hereafter) which means that all truly important predictors can be selected with probability approaching one as the sample size diverges to ∞ for linear regressions with Gaussian predictors and responses, there has been a surge of interest on ultrahigh-dimensional variable screening. Fan and Song (2010) and Fan, Feng and Song (2011) proposed a feature-screening method based on ranking the maximum marginal likelihood estimates in generalized linear models. Zhu, Li, Li and Zhu (2011) proposed a sure independent ranking and screening procedure (SIRS, hereafter) to screen significant predictors in multi-index models. Li, Peng, Zhang and Zhu (2012) proposed a robust rank correlation screening method which is based on Kendall's τ (Kendall, 1938). Li, Zhong and Zhu (2012) proposed a feature screening procedure (DC-SIS, hereafter) based on the distance correlation (Székely et al., 2007; Székely and Rizzo, 2009). Shao and Zhang (2014) proposed a model-free screening procedure (MDC-SIS, hereafter) based on the martingale difference correlation. Kong, Li, Fan and Lv (2017) proposed a two-stage interaction identification method, called the interaction pursuit via distance correlation (IPDC). Pan, Wang, Xiao and Zhu (2019) proposed a generic nonparametric sure independence screening procedure, called BCor-SIS, on the basis of Ball correlation.

It is known in the literature that each screening method targets a certain aspect of dependence. Let us look more closely at the robust rank correlation screening (RRCS) method that essentially uses the same premises as ours. Since Kendall's τ has robust advantages over Pearson correlation as showed in Kendall (1962), the RRCS procedure can be considered as a robust alternative to the SIS method. However, as a drawback for certain applications is that τ may be zero even if there is an association between two random variables, RRCS is not omnibus, only having an ability

*Corresponding author.

to detect monotonic effects. Inspired by the RRCS, we use the Bergsma-Dassios sign correlation (Bergsma and Dassios, 2014, τ_b^*) to introduce a new SIS procedure. We advocate using the τ_b^*-SIS for three reasons. First, as τ_b^* possesses the necessary and intuitive properties as a correlation index, the τ_b^*-SIS has a better screening ability for nonlinear effects including interactions and heterogeneity than RRCS. Second, as τ_b^* is a natural extension of τ, the τ_b^*-SIS is conceptually simple, easy to implement and robust to the presence of extreme values and outliers in the observations. Third, without assuming any moment condition on the response and predictors, the τ_b^*-SIS enjoys several appealing properties, such as the sure screening property, ranking consistency property and the characteristic of minimum model size.

The rest of this paper is organized as follows. Section 2 presents some preliminary results for τ_b^*, develops the new feature screening approach and studies its asymptotic properties. Section 3 demonstrates the merits of the τ_b^*-SIS procedure through Monte Carlo experiments. Section 4 gives a real data example to illustrate the proposed methodology. All technical proofs are presented in the Appendix.

2. METHOD

2.1 Preliminary: Bergsma-Dassios sign correlation

The Bergsma-Dassios sign correlation τ_b^* between two random variables U and V was recently introduced by Bergsma and Dassios (2014). Apply $\{(U_i, V_i)\}_{i=1}^n, n \geq 4$ to denote i.i.d. replications of (U, V). Bergsma and Dassios (2014) defined the Bergsma-Dassios sign covariance between U and V without finite moments to be the nonnegative number τ^* given by

$$\tau^*(U,V) = \mathbb{E}a(U_1,U_2,U_3,U_4)a(V_1,V_2,V_3,V_4),$$

where $a(z_1,z_2,z_3,z_4) = sign(|z_1-z_2|+|z_3-z_4|-|z_1-z_3|-|z_2-z_4|)$. The τ_b^* between U and V without finite moments is defined as

$$\tau_b^*(U,V) = \frac{\tau^*(U,V)}{\sqrt{\tau^*(U,U)\tau^*(V,V)}}.$$

Bergsma and Dassios (2014) showed that τ_b^* has many desirable properties. A remarkable property is that $\tau_b^*(U,V) = 0$ if and only if U and V are independent. This partly motivates us to use it in a feature screening procedure. Further, Bergsma and Dassios (2014) stated that

$$\tau^*(U,V) = 4S_1 + 4S_2 - 8S_3,$$

where $S_i, i = 1,2,3$, are defined as:

$$\begin{aligned} S_1 &= \mathbb{E}\left[I(U_1,U_2<U_3,U_4)I(V_1,V_2<V_3,V_4)\right], \\ S_2 &= \mathbb{E}\left[I(U_1,U_2<U_3,U_4)I(V_1,V_2>V_3,V_4)\right], \\ S_3 &= \mathbb{E}\left[I(U_1,U_2<U_3,U_4)I(V_1,V_3<V_2,V_4)\right], \end{aligned}$$

where $I(U_1,U_2<U_3,U_4)$ is shortened for $I(U_1<U_3 \& U_1<U_4 \& U_2<U_3 \& U_2<U_4)$, and & is logical AND. We can estimate S_1, S_2 and S_3 through the usual moment estimators, which are

$$\begin{aligned} \hat{S}_1 &= \frac{1}{n^4}\sum_{i,j,l,m=1}^{n}[I(U_i,U_j<U_l,U_m)I(V_i,V_j<V_l,V_m)], \\ \hat{S}_2 &= \frac{1}{n^4}\sum_{i,j,l,m=1}^{n}[I(U_i,U_j<U_l,U_m)I(V_i,V_j>V_l,V_m)], \\ \hat{S}_3 &= \frac{1}{n^4}\sum_{i,j,l,m=1}^{n}[I(U_i,U_j<U_l,U_m)I(V_i,V_l<V_j,V_m)]. \end{aligned}$$

Consequently, a natural estimator of $\tau^*(U,V)$ is given by

$$\hat{\tau}^*(U,V) = 4\hat{S}_1 + 4\hat{S}_2 - 8\hat{S}_3.$$

Accordingly, the sample $\tau_b^*(U,V)$ between U and V can be defined by

$$\hat{\tau}_b^*(U,V) = \frac{\hat{\tau}^*(U,V)}{\sqrt{\hat{\tau}^*(U,U)\hat{\tau}^*(V,V)}}.$$

2.2 The τ_b^*-based feature screening

In this subsection, we propose a sure independence screening procedure based on the τ_b^* for ultrahigh dimensional data. As the dimensionality p extremely exceeds the sample size n in ultrahigh dimensional data, the main focus is on screening out as many unimportant predictors as possible. Let Y be the response variable with support Ψ_y, $\mathbf{x} = (X_1, X_2, \ldots, X_p)^T$ be the predictor vector, and $F(y|\mathbf{x})$ denote the conditional distribution function of Y given $\mathbf{x}$. The screening procedure using τ_b^* can also be established without a specified regression model. First the index set of the active predictors which contributes to response variable is defined as

$$D = \{k : F(y|\mathbf{x}) \text{ functionally depends on } X_k \text{ for some } y \in \Psi_y\},$$

and the index set of the inactive predictors is defined as

$$I = \{k : F(y|\mathbf{x}) \text{ does not functionally depend on } X_k \text{ for any } y \in \Psi_y\}.$$

Then we can further denote $\mathbf{x}_D = \{X_k : k \in D\}$ as active predictors and $\mathbf{x}_I = \{X_k : k \in I\}$ as inactive predictors. From this definition, we can clearly understand that the predictors $\mathbf{x}_I$ are independent of Y when the important predictors $\mathbf{x}_D$ are given. The goal of an independence ranking and screening procedure is to obtain a reduced model with a proper scale which can almost fully contain active predictors $\mathbf{x}_D$ by using an independence screening method. Thus the following method can be taken to shrink the full model $\{1,2,\ldots,p\}$ to an appropriate submodel by ranking all the predictors and picking out the important ones in that the $\mathbf{x}_I$ are equivalent to redundant predictors.

Based upon a random sample $\{\mathbf{x}_i, Y_i\}$, $i = 1, 2, \ldots, n$, we write

$$\omega_k = \tau_b^*(X_k, Y), \quad \hat{\omega}_k = \hat{\tau}_b^*(X_k, Y),$$

for $k = 1, 2, \ldots, p$. Then ω_k is applied as a marginal utility to rank the predictors at the population level and correspondingly an appropriate $\hat{\omega}_k$ is used as a criterion to select important predictors in sample. Note that $\omega_k = 0$ if and only if the independence between X_k and Y holds. Then we can naturally apply ω_k as an effective measure to distinguish the active predictor subset from the inactive predictor subset because $\omega_k > 0$ for $k \in D$ and $\omega_k = 0$ for $k \in I$. It also implies that the τ_b^*-based variable screening is model-free in that there is no need to specify a regression model in its definition, and able to identify both linear and nonlinear relationships between the response and predictors. Thus it leads to the following definition which is used for choosing a submodel

$$\hat{D}^* = \{k : \hat{\omega}_k \geq cn^{-\kappa},\ 1 \leq k \leq p\},$$

where c and κ are prespecified threshold values that will be explained later in Condition (C1). We refer this procedure to the τ_b^*-based sure independence screening (τ_b^*-SIS for short).

In the following, we study the theoretical properties of the τ_b^*-SIS. When we refer to independence screening procedure, we understand that the most crucial property we have to guarantee is the sure screening property. That is, all active predictors can be embraced in the reduced model with pretty high probability. Thus, we first establish the sure screening property for τ_b^*-SIS with assuming an essential condition:

(C1) The minimum τ_b^* of active predictors satisfies

$$\min_{k \in D} \omega_k \geq 2cn^{-\kappa},$$

for some constants $c > 0$ and $0 \leq \kappa < 1/2$.

This condition puts forward an idea that the minimum true signal must possess a lower threshold and the order of the signal can ensure that it vanishes to zero along with the sample size n approaching to infinity.

Theorem 1 (Sure Screening Property)**.** *Under the condition (C1), there exists a positive constant $c_1 > 0$, such that*

$$P\left(\max_{1 \leq k \leq p} |\hat{\omega}_k - \omega_k| \geq cn^{-\kappa}\right) \leq O\left(p \exp\left\{-c_1 n^{1-2\kappa}\right\}\right),$$

and

$$P\left(D \subseteq \hat{D}^*\right) \geq 1 - O\left(s_n \exp\left\{-c_1 n^{1-2\kappa}\right\}\right),$$

where s_n is the cardinality of D.

In the next we explore another property for independence screening, namely ranking consistency property (Zhu et al., 2011). By the definition of D and I, we recognize that the conditional distribution of Y given $\mathbf{x}_D$ is independent of $\mathbf{x}_I$. So we expect Y depends more on $\mathbf{x}_D$ than on $\mathbf{x}_I$ which means the τ_b^* index has ability to sort active and inactive predictors well in the population level. That is the following condition we additionally assume:

(C2)

$$\liminf_{p \to \infty} \left\{\min_{k \in D} \omega_k - \max_{k \in I} \omega_k\right\} \geq c_2,$$

where c_2 is a positive constant.

Then we can obtain the ranking consistency property for the τ_b^*-SIS procedure.

Theorem 2 (Ranking Consistency Property)**.** *Suppose that the condition (C2) holds in addition to the conditions of Theorem 1, then*

$$\liminf_{n \to \infty} \left\{\min_{k \in D} \hat{\omega}_k - \max_{k \in I} \hat{\omega}_k\right\} > 0$$

almost surely.

This theorem implies that the sample $\tau_b^*(X_k, Y)$ values of inactive predictors are always ranked behind those of active ones in high probability.

We can also obtain the following theorem which characterizes the size of the reduced model after screening.

Theorem 3 (Minimum Model Size)**.** *Under the conditions of Theorem 1, there exists a positive constant c_4, such that*

$$P\left\{\left|\hat{D}^*\right| \leq O\left(n^{\kappa} \sum_{k=1}^{p} |\omega_k|\right)\right\} \geq 1 - O\left(p \exp\left\{-c_4 n^{1-2\kappa}\right\}\right).$$

3. SIMULATION STUDY

In this Section, we assess the finite sample performance of the proposed τ_b^*-SIS via Monte Carlo simulation. For comparison purpose, five existing screening methods are included here: SIS, SIRS, DC-SIS, MDC-SIS and RRCS. The following three criteria are considered to evaluate the performance:

(1) S: the minimum model size that includes all active predictors. We report the 5%, 25%, 50%, 75% and 95% quantiles of S out of 500 replications.

(2) P_s: the proportion that an individual active predictor is selected for a given model size d in 500 replications.

(3) P_a: the proportion that all active predictors are selected for a given model size d in 500 replications.

The minimum model size S is a measure of the ability of a screening procedure to include all the active predictors. The smaller the S is, the better the screening method is. The effects of P_s and P_a are similar, and the ability of a screening procedure can also be measured through them. The sure screening property ensures that P_s and P_a will both tend to be one when the model size d is as large as it can be. So we consider a screening method competitive if the measured S value approaches the actual number of all active

predictors and the P_s and P_a values approach one. In the following, we choose d to be $d_1 = \lfloor n/\log n \rfloor$, $d_2 = 2\lfloor n/\log n \rfloor$ and $d_3 = 3\lfloor n/\log n \rfloor$, respectively.

Example 1. We modify the simple linear model in Fan and Lv (2008) as follows:

$$M_1: \quad Y = 5X_1 + 5X_2 + 5X_3 + \varepsilon.$$

The difference is that the error term $\varepsilon \sim t(1)$ rather than the standard normal distribution $N(0,1)$. And we generate samples of $(X_1, X_2, \ldots, X_p)$ with size n from a multivariate normal distribution with mean zero and covariance matrix $\mathbf{\Sigma} = (\sigma_{ij})_{p\times p}$. In this example, we take $\sigma_{ij} = 0.5^{|i-j|}$, $n = 20, 50, 100$ and $p = 100, 1000, 2000$.

Table 1 presents the estimated results of P_s and P_a. Overall, high dimensionality, for fixed sample sizes n, results in worse performance for all the six methods. It is readily seen that the procedures τ_b^*-SIS and RRCS outperform others in an obvious way. And RRCS always performs the best, while SIRS always performs the worst for this linear model. SIS and MDC-SIS deliver a similar performance, and DC-SIS outperforms these two methods except the cases of $n = 20$. The values of P_s and P_a for both τ_b^*-SIS and RRCS approach one as p and n increase, which indicates that they are both robust against heavy-tailed distribution due to Kendall's τ's good properties.

Table 2 shows the 5%, 25%, 50%, 75% and 95% quantiles of the minimum model size S. Note that when $n = 20, 50, 100$, $\lfloor n/\log(n) \rfloor = 6, 12, 21$, respectively. It can be seen that, in the case 3 (i.e., $n = 20, p = 1000$), all the procedures result in poor performance due to the small sample size. In other cases, τ_b^*-SIS performs almost as good as RRCS. For the competitors, the estimated values S of DC-SIS are always smaller than that of MDC-SIS, SIS and SIRS.

Example 2. Consider the following nonlinear model:

$$M_2: \quad Y = 5X_1^2 + 5X_2 + 5X_3^2 + \exp\left(X_1^2 + X_2^2\right) + \varepsilon.$$

Here the predictors $\mathbf{x} = (X_1, X_2, \ldots, X_p)^T$ are also generated from a multivariate normal distribution with mean zero and covariance matrix $\mathbf{\Sigma} = (\sigma_{ij})_{p\times p}$, where $\sigma_{ij} = \rho^{|i-j|}$, and $\rho = 0.5, 0.9$, respectively. The noise ε is distributed identically to that in Example 1. And $n = 50, 100$, and $p = 1000, 2000$, respectively.

Table 3 depicts the simulation results for P_s and P_a. It occurs to us that τ_b^*-SIS always performs the best, followed by DC-SIS, MDC-SIS, SIS, RRCS and SIRS in turn. It should be pointed out that, the performance of RRCS is significantly inferior to τ_b^*-SIS compared with Example 1, and SIRS fails completely in selecting any active predictor in this nonlinear model. These results demonstrate that the τ_b^*-based screening procedure is able to identify both linear and nonlinear relationships between response and predictors.

Table 4 presents the 5%, 25%, 50%, 75% and 95% quantiles of the minimum model size S in the case of $\rho = 0.9$. Clearly, τ_b^*-SIS outperforms all the five competitors substantially. And the resulting model based on τ_b^*-SIS can almost cover all active predictors across three different model sizes.

Example 3. Let us consider the following model:

$$M_3: \quad \log(Y) = 2X_1^2 + 5X_1X_2 + 2X_3 + \varepsilon,$$

which contains a logarithm transformation function and an interaction term X_1X_2. All the settings are the same as that in Example 2, except that we fix $\sigma_{ij} = 0.9^{|i-j|}$ for simplicity.

The simulation results are listed in Tables 5 and 6, from which it can be seen that τ_b^*-SIS shows outstanding performance in terms of the minimum modes size and the proportions of including active predictor variables. RRCS and SIRS still perform poor and the other three methods give a similar and lackluster presentation. These results support the assertion of the sure screening property of the procedure τ_b^*-SIS in which the condition of X_k and Y to be bounded is not necessary.

Example 4. In the previous three examples, the predictors are all generated from multivariate normal distributions. In this example, we consider the predictors are generated from the following mixed normal distribution:

$$\lambda\mathcal{N}(\mu_1, \mathbf{\Sigma}_1) + (1-\lambda)\mathcal{N}(\mu_2, \mathbf{\Sigma}_2),$$

where $\lambda = 0.4$, $\mu_1 = \mathbf{0}$, $\mu_2 = \mathbf{1}_p$, $\mathbf{\Sigma}_1 = (\sigma_{ij})_{p\times p}$ with σ_{ij} being 1 for $i = j$, and 0.9 otherwise, and $\mathbf{\Sigma}_2 = (\gamma_{ij})_{p\times p}$ with $\gamma_{ij} = 0.9^{|i-j|}$.

Table 7 and 8 illustrate the simulation results for P_s and P_a. An interesting observation is that all the proportions have declined to some extent. Specifically, RRCS and τ_b^*-SIS remain a robust performance, DC-SIS, MDC-SIS and SIS deliver a relatively worse performance and SIRS continues to behave poor for the three models. Compared with the observations in Example 1, RRCS also slightly outperforms τ_b^*-SIS. While for the other two models, τ_b^*-SIS behaves slightly better than RRCS. As a whole, τ_b^*-SIS performs quite satisfactory in all the considered cases.

Example 5. Here, we discuss two procedures in recent: BCor-SIS (Pan, Wang, Xiao and Zhu (2019)) and IPDC (Kong, Li, Fan and Lv (2017)). In the above simulations, we have discussed models that are designed for good performance of τ_b^*-SIS. For a fair comparison, here we consider the following four models from Pan, Wang, Xiao and Zhu (2019):

$$\begin{aligned}
M_4: &\quad Y = X_1 + 1.25X_2 + 0.75X_8 - 2.4X_{16} + \varepsilon,\\
M_5: &\quad Y = 3X_1X_5 + 2X_{10} + 3X_{15} + \varepsilon,\\
M_6: &\quad Y = 3X_1^2 + 5X_2 + 5X_8 - 8X_{16} + \varepsilon,\\
M_7: &\quad Y = 2I(\omega > 0)\omega + 1\\
&\qquad \text{with } \omega = 5X_1^2 - 5X_2^2 + 3X_8 + 2X_{16} + \varepsilon,
\end{aligned}$$

where $\varepsilon \sim N(0,1)$.

Table 1. The proportions of P_s and P_a in Example 1

	DC-SIS				MDC-SIS				SIS				SIRS				τ_b^*-SIS				RRCS			
	P_s			P_a	P_s			P_a	P_s			P_a	P_s			P_a	P_s			P_a	P_s			P_a
Size	X_1	X_2	X_3	All	X_1	X_2	X_3	All	X_1	X_2	X_3	All	X_1	X_2	X_3	All	X_1	X_2	X_3	All	X_1	X_2	X_3	All
Case 1: p=100 and n=20																								
d_1	0.784	0.922	0.784	0.614	0.782	0.850	0.814	0.680	0.800	0.850	0.810	0.706	0.236	0.360	0.272	0.044	0.810	0.960	0.830	0.656	0.872	0.982	0.876	0.764
d_2	0.872	0.958	0.886	0.780	0.860	0.880	0.860	0.782	0.860	0.890	0.870	0.792	0.392	0.510	0.410	0.150	0.920	0.980	0.910	0.834	0.950	0.992	0.952	0.900
d_3	0.910	0.960	0.926	0.848	0.880	0.910	0.890	0.814	0.880	0.920	0.900	0.832	0.492	0.646	0.510	0.228	0.930	0.980	0.930	0.858	0.962	0.992	0.978	0.934
Case 2: p=100 and n=50																								
d_1	0.984	0.990	0.988	0.978	0.852	0.864	0.872	0.798	0.864	0.882	0.884	0.816	0.762	0.896	0.768	0.568	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
d_2	0.992	0.998	0.994	0.988	0.888	0.902	0.916	0.850	0.902	0.918	0.916	0.860	0.914	0.980	0.910	0.832	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
d_3	0.998	1.000	0.996	0.994	0.918	0.926	0.932	0.876	0.926	0.926	0.994	0.882	0.976	0.998	0.980	0.956	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Case 3: p=1000 and n=20																								
d_1	0.498	0.690	0.508	0.248	0.574	0.714	0.536	0.358	0.592	0.722	0.570	0.404	0.086	0.084	0.060	0.006	0.560	0.808	0.564	0.272	0.658	0.858	0.630	0.382
d_2	0.624	0.764	0.606	0.372	0.646	0.756	0.638	0.488	0.672	0.768	0.658	0.524	0.130	0.134	0.106	0.014	0.680	0.866	0.676	0.424	0.756	0.924	0.752	0.556
d_3	0.682	0.798	0.666	0.460	0.688	0.780	0.684	0.560	0.718	0.784	0.698	0.592	0.156	0.182	0.142	0.018	0.738	0.904	0.734	0.518	0.822	0.936	0.798	0.638
Case 4: p=1000 and n=50																								
d_1	0.934	0.966	0.936	0.910	0.782	0.822	0.796	0.724	0.772	0.814	0.796	0.722	0.396	0.534	0.392	0.150	0.992	0.998	0.992	0.982	0.996	1.000	0.992	0.988
d_2	0.952	0.972	0.954	0.934	0.806	0.848	0.808	0.756	0.798	0.840	0.812	0.754	0.542	0.650	0.536	0.278	0.996	1.000	0.996	0.992	0.996	1.000	0.994	0.990
d_3	0.960	0.974	0.958	0.940	0.816	0.858	0.818	0.768	0.808	0.860	0.824	0.766	0.642	0.734	0.628	0.368	0.996	1.000	0.998	0.994	0.996	1.000	0.994	0.990
Case 5: p=1000 and n=100																								
d_1	0.976	0.980	0.970	0.968	0.782	0.830	0.800	0.746	0.784	0.820	0.802	0.754	0.854	0.950	0.866	0.740	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
d_2	0.982	0.984	0.980	0.978	0.808	0.838	0.816	0.762	0.810	0.840	0.828	0.780	0.914	0.978	0.936	0.854	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
d_3	0.984	0.990	0.980	0.980	0.818	0.850	0.828	0.770	0.830	0.858	0.832	0.784	0.954	0.992	0.962	0.916	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Case 6: p=2000 and n=50																								
d_1	0.936	0.960	0.930	0.900	0.798	0.858	0.786	0.734	0.800	0.852	0.784	0.738	0.284	0.424	0.258	0.056	0.994	1.000	0.982	0.976	1.000	1.000	0.990	0.990
d_2	0.948	0.964	0.948	0.920	0.834	0.870	0.800	0.770	0.832	0.868	0.812	0.778	0.392	0.544	0.360	0.136	1.000	1.000	0.994	0.994	1.000	1.000	0.992	0.992
d_3	0.960	0.966	0.958	0.936	0.846	0.878	0.816	0.786	0.840	0.878	0.828	0.794	0.468	0.602	0.434	0.198	1.000	1.000	0.996	0.996	1.000	1.000	0.996	0.996
Case 7: p=2000 and n=100																								
d_1	0.968	0.974	0.962	0.958	0.762	0.810	0.774	0.700	0.764	0.810	0.778	0.710	0.750	0.906	0.798	0.582	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
d_2	0.976	0.980	0.968	0.962	0.780	0.824	0.802	0.728	0.788	0.828	0.802	0.744	0.830	0.948	0.876	0.720	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
d_3	0.978	0.986	0.972	0.966	0.812	0.840	0.810	0.748	0.810	0.842	0.814	0.756	0.886	0.968	0.914	0.796	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 2. The 5%, 25%, 50%, 75%, and 95% quantiles of the minimum model size S out of 500 replications in Example 1

DC-SIS					MDC-SIS					SIS					SIRS					τ_b^*-SIS					RRCS				
5%	25%	50%	75%	95%	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
Case 1: p=100 and n=20																													
3	3	5	11	47	3	3	4	10	79	3	3	4	8	81	7	20	35	53	77	3	3	4	9	67	3	3	4	6	21
Case 2: p=100 and n=50																													
3	3	3	3	4	3	3	3	6	80	3	3	3	5	79	3	5	10	20	34	3	3	3	3	3	3	3	3	3	3
Case 3: p=1000 and n=20																													
3	7	22	96	565	3	4	13	101	801	3	4	11	85	749	40	170	342	531	794	3	6	18	53	683	3	4	10	31	202
Case 4: p=1000 and n=50																													
3	3	3	3	59	3	3	3	21	850	3	3	3	20	784	5	22	67	145	326	3	3	3	3	5	3	3	3	3	4
Case 5: p=1000 and n=100																													
3	3	3	3	3	3	3	3	22	870	3	3	3	20	835	3	4	8	23	87	3	3	3	3	3	3	3	3	3	3
Case 6: p=2000 and n=50																													
3	3	3	4	72	3	3	3	18	1466	3	3	3	16	1386	11	49	144	301	611	3	3	3	3	7	3	3	3	3	5
Case 7: p=2000 and n=100																													
3	3	3	3	6	3	3	3	66	1528	3	3	3	56	1639	3	6	16	49	166	3	3	3	3	3	3	3	3	3	3

Table 3. The proportions of P_s and P_a in Example 2

		DC-SIS				MDC-SIS				SIS				SIRS				τ_b^*-SIS				RRCS			
		P_s			P_a	P_s			P_a	P_s			P_a	P_s			P_a	P_s			P_a	P_s			P_a
ρ	Size	X_1	X_2	X_3	All	X_1	X_2	X_3	All	X_1	X_2	X_3	All	X_1	X_2	X_3	All	X_1	X_2	X_3	All	X_1	X_2	X_3	All
p=1000 and n=50																									
0.5	d_1	0.488	0.476	0.072	0.028	0.392	0.396	0.058	0.012	0.338	0.342	0.052	0.008	0.002	0.002	0.002	0.000	0.382	0.376	0.084	0.018	0.052	0.194	0.022	0.000
	d_2	0.622	0.624	0.116	0.062	0.526	0.520	0.102	0.046	0.440	0.438	0.102	0.030	0.004	0.002	0.002	0.000	0.592	0.564	0.154	0.066	0.064	0.248	0.038	0.006
	d_3	0.708	0.700	0.142	0.078	0.612	0.602	0.138	0.070	0.512	0.506	0.122	0.046	0.008	0.006	0.004	0.000	0.692	0.690	0.222	0.124	0.098	0.312	0.046	0.006
0.9	d_1	0.664	0.668	0.436	0.328	0.538	0.556	0.352	0.224	0.444	0.486	0.322	0.194	0.000	0.000	0.000	0.000	0.918	0.946	0.690	0.654	0.146	0.252	0.130	0.100
	d_2	0.826	0.864	0.616	0.540	0.700	0.718	0.526	0.400	0.592	0.626	0.442	0.340	0.002	0.000	0.000	0.000	0.982	0.984	0.852	0.840	0.202	0.322	0.180	0.144
	d_3	0.914	0.948	0.712	0.660	0.806	0.838	0.630	0.540	0.670	0.692	0.538	0.452	0.004	0.000	0.000	0.000	0.998	0.998	0.924	0.922	0.232	0.354	0.224	0.176
p=1000 and n=100																									
0.5	d_1	0.786	0.780	0.140	0.084	0.654	0.664	0.114	0.060	0.510	0.498	0.110	0.054	0.002	0.008	0.010	0.000	0.998	0.982	0.422	0.412	0.084	0.536	0.044	0.008
	d_2	0.868	0.870	0.202	0.158	0.766	0.768	0.176	0.112	0.622	0.606	0.162	0.092	0.002	0.018	0.024	0.000	0.998	0.998	0.634	0.632	0.106	0.646	0.080	0.014
	d_3	0.902	0.912	0.258	0.206	0.812	0.814	0.228	0.148	0.660	0.664	0.212	0.132	0.008	0.052	0.040	0.000	1.000	0.998	0.766	0.764	0.148	0.702	0.124	0.032
0.9	d_1	0.932	0.948	0.684	0.654	0.828	0.822	0.574	0.512	0.660	0.668	0.470	0.404	0.000	0.000	0.002	0.000	1.000	1.000	0.998	0.998	0.298	0.530	0.288	0.234
	d_2	0.986	0.992	0.852	0.842	0.944	0.942	0.750	0.722	0.774	0.774	0.638	0.598	0.002	0.004	0.002	0.000	1.000	1.000	1.000	1.000	0.376	0.592	0.364	0.312
	d_3	0.996	0.996	0.930	0.930	0.970	0.970	0.858	0.838	0.810	0.808	0.728	0.688	0.004	0.012	0.010	0.000	1.000	1.000	1.000	1.000	0.424	0.634	0.412	0.352
p=2000 and n=50																									
0.5	d_1	0.404	0.416	0.042	0.006	0.298	0.292	0.036	0.004	0.254	0.228	0.024	0.000	0.000	0.000	0.000	0.000	0.250	0.398	0.036	0.012	0.016	0.216	0.012	0.002
	d_2	0.520	0.512	0.070	0.014	0.432	0.436	0.058	0.010	0.352	0.352	0.058	0.006	0.000	0.002	0.000	0.000	0.458	0.516	0.080	0.042	0.030	0.270	0.020	0.002
	d_3	0.608	0.590	0.088	0.036	0.496	0.502	0.088	0.032	0.410	0.424	0.080	0.024	0.000	0.004	0.004	0.000	0.558	0.598	0.106	0.058	0.050	0.312	0.026	0.002
0.9	d_1	0.534	0.573	0.324	0.198	0.386	0.388	0.238	0.126	0.328	0.330	0.196	0.106	0.000	0.000	0.000	0.000	0.888	0.918	0.528	0.494	0.118	0.232	0.090	0.064
	d_2	0.722	0.718	0.470	0.352	0.544	0.546	0.378	0.240	0.448	0.446	0.318	0.186	0.000	0.000	0.000	0.000	0.962	0.964	0.754	0.736	0.148	0.284	0.130	0.090
	d_3	0.834	0.824	0.560	0.478	0.674	0.646	0.470	0.336	0.566	0.550	0.410	0.302	0.000	0.000	0.000	0.000	0.984	0.986	0.832	0.822	0.178	0.322	0.164	0.126
p=2000 and n=100																									
0.5	d_1	0.680	0.666	0.092	0.036	0.526	0.522	0.086	0.020	0.436	0.432	0.082	0.020	0.000	0.008	0.004	0.000	0.954	0.934	0.268	0.242	0.074	0.490	0.034	0.010
	d_2	0.782	0.762	0.154	0.082	0.676	0.646	0.134	0.056	0.546	0.536	0.126	0.048	0.000	0.010	0.012	0.000	0.984	0.980	0.454	0.446	0.116	0.574	0.050	0.020
	d_3	0.818	0.824	0.190	0.122	0.734	0.718	0.174	0.090	0.606	0.588	0.174	0.094	0.000	0.022	0.014	0.000	0.996	0.986	0.594	0.586	0.148	0.614	0.066	0.024
0.9	d_1	0.828	0.866	0.614	0.530	0.698	0.714	0.484	0.374	0.588	0.610	0.400	0.322	0.000	0.000	0.000	0.000	1.000	1.000	0.998	0.998	0.212	0.458	0.214	0.164
	d_2	0.946	0.964	0.738	0.706	0.836	0.872	0.634	0.558	0.698	0.718	0.528	0.462	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.300	0.512	0.278	0.232
	d_3	0.988	0.990	0.806	0.798	0.916	0.934	0.722	0.678	0.762	0.776	0.606	0.552	0.002	0.000	0.004	0.000	1.000	1.000	1.000	1.000	0.328	0.548	0.336	0.268

Table 4. The 5%, 25%, 50%, 75%, and 95% quantiles of the minimum model size S out of 500 replications in Example 2

DC-SIS					MDC-SIS					SIS					SIRS					τ_b^*-SIS					RRCS				
5%	25%	50%	75%	95%	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
Case 1: p=1000 and n=50																													
4	9	22	50	130	4	14	34	69	178	5	17	43	138	834	340	618	772	911	988	3	4	7	17	47	6	86	392	731	955
Case 2: p=1000 and n=100																													
3	6	14	30	71	3	9	21	48	112	4	11	30	96	753	125	279	498	675	975	3	3	3	3	4	4	26	193	620	919
Case 3: p=2000 and n=50																													
4	17	39	97	281	6	26	63	141	449	8	30	80	270	1470	688	1188	1514	1783	1972	3	5	13	26	82	11	155	819	1527	1927
Case 4: p=2000 and n=100																													
3	7	19	49	151	3	11	34	79	258	4	17	50	157	1475	592	984	1329	1630	1918	3	3	3	3	6	3	56	399	1192	1880

Table 5. The proportions of P_s and P_a in Example 3

	DC-SIS				MDC-SIS				SIS				SIRS				τ_b^*-SIS				RRCS			
	P_s			P_a	P_s			P_a	P_s			P_a	P_s			P_a	P_s			P_a	P_s			P_a
Size	X_1	X_2	X_3	All	X_1	X_2	X_3	All	X_1	X_2	X_3	All	X_1	X_2	X_3	All	X_1	X_2	X_3	All	X_1	X_2	X_3	All
p=1000 and n=50																								
d_1	0.342	0.280	0.162	0.110	0.322	0.294	0.170	0.108	0.350	0.316	0.206	0.132	0.000	0.002	0.006	0.000	0.852	0.728	0.478	0.436	0.134	0.192	0.254	0.106
d_2	0.448	0.404	0.286	0.212	0.452	0.416	0.302	0.214	0.480	0.442	0.298	0.222	0.000	0.002	0.010	0.000	0.942	0.870	0.602	0.570	0.194	0.254	0.332	0.162
d_3	0.542	0.518	0.380	0.300	0.530	0.500	0.376	0.298	0.552	0.520	0.394	0.320	0.000	0.004	0.018	0.000	0.970	0.922	0.672	0.654	0.238	0.304	0.382	0.208
p=1000 and n=100																								
d_1	0.396	0.378	0.282	0.220	0.386	0.372	0.290	0.226	0.438	0.424	0.334	0.276	0.004	0.010	0.018	0.000	0.998	0.998	0.948	0.948	0.186	0.316	0.458	0.174
d_2	0.474	0.472	0.370	0.338	0.462	0.458	0.364	0.320	0.508	0.494	0.418	0.382	0.006	0.026	0.040	0.002	1.000	0.998	0.978	0.978	0.246	0.400	0.556	0.230
d_3	0.502	0.498	0.428	0.404	0.496	0.500	0.408	0.384	0.526	0.526	0.458	0.426	0.014	0.042	0.078	0.002	1.000	1.000	0.990	0.990	0.280	0.450	0.604	0.258
p=2000 and n=50																								
d_1	0.216	0.204	0.142	0.072	0.188	0.192	0.118	0.048	0.230	0.226	0.146	0.078	0.000	0.000	0.000	0.000	0.758	0.596	0.334	0.282	0.106	0.150	0.184	0.086
d_2	0.342	0.316	0.206	0.144	0.308	0.276	0.200	0.124	0.348	0.320	0.232	0.166	0.000	0.000	0.004	0.000	0.890	0.762	0.486	0.448	0.130	0.196	0.256	0.100
d_3	0.428	0.406	0.264	0.196	0.412	0.356	0.252	0.182	0.436	0.390	0.294	0.216	0.002	0.002	0.008	0.000	0.934	0.850	0.578	0.544	0.152	0.222	0.304	0.118
p=2000 and n=100																								
d_1	0.340	0.288	0.198	0.146	0.304	0.280	0.194	0.132	0.348	0.320	0.232	0.170	0.000	0.008	0.004	0.000	1.000	0.998	0.900	0.900	0.192	0.308	0.464	0.180
d_2	0.416	0.386	0.286	0.238	0.400	0.386	0.272	0.226	0.434	0.412	0.302	0.260	0.000	0.010	0.012	0.000	1.000	1.000	0.966	0.966	0.236	0.388	0.534	0.224
d_3	0.462	0.428	0.336	0.296	0.450	0.424	0.330	0.284	0.480	0.452	0.368	0.322	0.000	0.010	0.020	0.000	1.000	1.000	0.978	0.978	0.266	0.434	0.588	0.254

Table 6. The 5%, 25%, 50%, 75%, and 95% quantiles of the minimum model size S out of 500 replications in Example 3

DC-SIS					MDC-SIS					SIS					SIRS					τ_b^*-SIS					RRCS				
5%	25%	50%	75%	95%	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%	5%	25%	50%	75%	95%
Case 1: p=1000 and n=50																													
6	29	82	387	933	7	28	85	366	945	6	27	80	416	922	288	551	723	896	989	3	6	18	52	486	5	66	397	773	966
Case 2: p=1000 and n=100																													
5	23	127	612	949	5	25	141	654	961	3	25	141	654	961	154	403	628	829	976	3	3	3	7	23	4	56	275	687	951
Case 3: p=2000 and n=50																													
9	45	156	690	1827	13	56	173	710	1838	8	46	153	727	1841	482	1056	1433	1781	1960	3	11	31	93	927	6	147	673	1485	1898
Case 4: p=2000 and n=100																													
8	44	250	1249	1951	8	49	259	1310	1924	3	39	245	1226	1924	407	939	1321	1681	1950	3	3	4	9	33	4	61	370	1176	1886

Table 7. The proportions of P_a in Example 4

Pa	DC-SIS			MDC-SIS			SIS			SIRS			τ_b^*-SIS			RRCS		
	Size			Size			Size			Size			Size			Size		
Model	d_1	d_2	d_3	d_1	d_2	d_3	d_1	d_2	d_3	d_1	d_2	d_3	d_1	d_2	d_3	d_1	d_2	d_3
p=1000 and n=50																		
M_1	0.584	0.592	0.608	0.534	0.546	0.566	0.530	0.552	0.576	0.000	0.000	0.002	0.612	0.626	0.642	0.618	0.642	0.674
M_2	0.152	0.272	0.350	0.122	0.228	0.302	0.108	0.222	0.292	0.000	0.000	0.000	0.612	0.626	0.648	0.606	0.606	0.606
M_3	0.066	0.126	0.184	0.052	0.124	0.186	0.074	0.160	0.214	0.000	0.000	0.000	0.596	0.598	0.600	0.596	0.596	0.598
p=1000 and n=100																		
M_1	0.566	0.588	0.622	0.538	0.588	0.604	0.532	0.582	0.610	0.000	0.000	0.000	0.764	0.852	0.888	0.810	0.892	0.920
M_2	0.376	0.462	0.500	0.348	0.456	0.496	0.338	0.456	0.496	0.000	0.000	0.000	0.668	0.734	0.776	0.574	0.574	0.576
M_3	0.148	0.252	0.324	0.136	0.238	0.310	0.178	0.276	0.354	0.000	0.000	0.000	0.632	0.650	0.668	0.608	0.608	0.612
p=2000 and n=50																		
M_1	0.614	0.614	0.618	0.548	0.560	0.576	0.548	0.564	0.576	0.000	0.000	0.000	0.642	0.648	0.650	0.642	0.658	0.670
M_2	0.116	0.198	0.266	0.086	0.142	0.180	0.060	0.120	0.166	0.000	0.000	0.000	0.638	0.646	0.652	0.634	0.634	0.634
M_3	0.026	0.056	0.092	0.020	0.050	0.078	0.052	0.088	0.108	0.000	0.000	0.000	0.596	0.596	0.598	0.596	0.596	0.596
p=2000 and n=100																		
M_1	0.572	0.584	0.594	0.518	0.548	0.564	0.514	0.536	0.562	0.000	0.000	0.000	0.656	0.736	0.780	0.696	0.784	0.820
M_2	0.264	0.358	0.416	0.222	0.316	0.382	0.192	0.296	0.388	0.000	0.000	0.000	0.636	0.698	0.728	0.578	0.578	0.578
M_3	0.056	0.120	0.164	0.052	0.112	0.164	0.092	0.160	0.214	0.000	0.000	0.000	0.586	0.596	0.602	0.586	0.586	0.586

Table 8. The proportions of P_s in Example 4

Ps		DC-SIS			MDC-SIS			SIS			SIRS			τ_b^*-SIS			RRCS		
		size			size			size			size			size			size		
Model		d_1	d_2	d_3	d_1	d_2	d_3	d_1	d_2	d_3	d_1	d_2	d_3	d_1	d_2	d_3	d_1	d_2	d_3
p=1000 and n=50																			
M_1	X_1	0.626	0.662	0.694	0.598	0.626	0.660	0.594	0.628	0.658	0.004	0.020	0.022	0.700	0.746	0.778	0.722	0.770	0.804
	X_2	0.636	0.664	0.690	0.618	0.650	0.680	0.618	0.666	0.694	0.008	0.010	0.020	0.722	0.754	0.784	0.738	0.784	0.816
	X_3	0.650	0.684	0.718	0.624	0.652	0.674	0.624	0.660	0.688	0.018	0.032	0.048	0.730	0.770	0.802	0.758	0.814	0.848
M_2	X_1	0.366	0.500	0.570	0.314	0.438	0.532	0.290	0.436	0.522	0.002	0.002	0.002	0.794	0.828	0.860	0.622	0.626	0.632
	X_2	0.382	0.522	0.578	0.360	0.478	0.540	0.320	0.446	0.528	0.006	0.006	0.008	0.768	0.804	0.830	0.696	0.740	0.764
	X_3	0.224	0.338	0.386	0.202	0.298	0.360	0.186	0.290	0.344	0.008	0.010	0.014	0.656	0.678	0.708	0.618	0.624	0.624
M_3	X_1	0.170	0.302	0.348	0.162	0.286	0.350	0.196	0.310	0.394	0.006	0.006	0.010	0.782	0.820	0.840	0.612	0.614	0.620
	X_2	0.162	0.258	0.320	0.164	0.252	0.312	0.180	0.280	0.360	0.000	0.004	0.010	0.640	0.660	0.686	0.602	0.608	0.616
	X_3	0.134	0.216	0.274	0.128	0.210	0.258	0.152	0.248	0.294	0.008	0.018	0.026	0.622	0.628	0.640	0.634	0.658	0.678
p=1000 and n=100																			
M_1	X_1	0.700	0.758	0.796	0.630	0.686	0.716	0.640	0.684	0.718	0.034	0.048	0.066	0.894	0.940	0.964	0.926	0.964	0.974
	X_2	0.708	0.744	0.788	0.632	0.694	0.704	0.636	0.690	0.718	0.028	0.046	0.062	0.912	0.946	0.962	0.934	0.952	0.970
	X_3	0.674	0.732	0.764	0.638	0.672	0.692	0.630	0.670	0.690	0.030	0.050	0.066	0.880	0.932	0.944	0.898	0.958	0.966
M_2	X_1	0.570	0.622	0.648	0.534	0.608	0.638	0.526	0.582	0.618	0.000	0.002	0.004	0.950	0.962	0.976	0.588	0.594	0.604
	X_2	0.558	0.632	0.644	0.528	0.604	0.628	0.508	0.586	0.616	0.016	0.034	0.048	0.926	0.952	0.962	0.828	0.870	0.904
	X_3	0.424	0.496	0.534	0.420	0.490	0.528	0.406	0.486	0.522	0.010	0.014	0.028	0.706	0.772	0.812	0.578	0.590	0.598
M_3	X_1	0.330	0.398	0.460	0.320	0.408	0.456	0.356	0.442	0.486	0.002	0.004	0.012	0.966	0.978	0.992	0.618	0.624	0.632
	X_2	0.296	0.396	0.440	0.288	0.380	0.438	0.334	0.420	0.482	0.006	0.016	0.020	0.746	0.786	0.820	0.618	0.634	0.652
	X_3	0.200	0.306	0.370	0.198	0.300	0.370	0.230	0.338	0.398	0.024	0.046	0.070	0.682	0.720	0.736	0.750	0.806	0.826
p=2000 and n=50																			
M_1	X_1	0.654	0.674	0.682	0.596	0.620	0.636	0.606	0.632	0.648	0.006	0.016	0.018	0.712	0.736	0.746	0.726	0.750	0.774
	X_2	0.644	0.660	0.674	0.610	0.634	0.652	0.604	0.632	0.654	0.010	0.012	0.020	0.708	0.732	0.756	0.728	0.770	0.794
	X_3	0.650	0.660	0.666	0.616	0.634	0.662	0.624	0.642	0.660	0.014	0.016	0.020	0.704	0.742	0.756	0.728	0.768	0.790
M_2	X_1	0.286	0.416	0.488	0.228	0.340	0.416	0.186	0.314	0.388	0.000	0.000	0.002	0.778	0.822	0.846	0.644	0.646	0.652
	X_2	0.306	0.410	0.476	0.264	0.374	0.418	0.226	0.350	0.416	0.004	0.004	0.010	0.730	0.768	0.790	0.718	0.746	0.758
	X_3	0.186	0.274	0.350	0.164	0.234	0.298	0.144	0.216	0.296	0.000	0.000	0.004	0.658	0.674	0.690	0.634	0.636	0.636
M_3	X_1	0.112	0.178	0.236	0.106	0.172	0.232	0.140	0.208	0.256	0.000	0.000	0.002	0.766	0.806	0.828	0.614	0.624	0.630
	X_2	0.106	0.158	0.218	0.096	0.150	0.220	0.126	0.194	0.248	0.002	0.004	0.008	0.628	0.658	0.674	0.600	0.604	0.612
	X_3	0.074	0.126	0.172	0.066	0.118	0.154	0.100	0.154	0.186	0.004	0.012	0.016	0.608	0.616	0.626	0.634	0.652	0.668
p=2000 and n=100																			
M_1	X_1	0.650	0.698	0.726	0.612	0.652	0.672	0.606	0.638	0.662	0.008	0.016	0.032	0.794	0.854	0.898	0.830	0.894	0.918
	X_2	0.652	0.694	0.732	0.622	0.650	0.682	0.608	0.652	0.692	0.018	0.026	0.044	0.824	0.880	0.902	0.860	0.914	0.926
	X_3	0.682	0.700	0.744	0.610	0.660	0.688	0.618	0.664	0.698	0.014	0.034	0.040	0.860	0.894	0.912	0.888	0.918	0.938
M_2	X_1	0.470	0.552	0.592	0.428	0.526	0.574	0.426	0.520	0.578	0.000	0.000	0.000	0.916	0.944	0.956	0.582	0.588	0.596
	X_2	0.480	0.562	0.588	0.422	0.524	0.576	0.408	0.512	0.564	0.010	0.020	0.026	0.880	0.918	0.932	0.780	0.810	0.832
	X_3	0.330	0.414	0.458	0.300	0.392	0.430	0.268	0.372	0.426	0.004	0.008	0.014	0.692	0.748	0.770	0.580	0.584	0.588
M_3	X_1	0.184	0.282	0.338	0.158	0.270	0.324	0.210	0.324	0.386	0.002	0.002	0.002	0.926	0.954	0.966	0.588	0.596	0.602
	X_2	0.156	0.234	0.304	0.138	0.220	0.272	0.192	0.272	0.340	0.000	0.006	0.012	0.668	0.710	0.734	0.590	0.592	0.596
	X_3	0.120	0.184	0.234	0.112	0.184	0.230	0.160	0.228	0.276	0.008	0.014	0.032	0.626	0.652	0.668	0.676	0.720	0.740

Table 9. The proportions of P_a and P_s in Example 5

	τ_b^*-SIS					IPDC					BCor-SIS				
	P_s				P_a	P_s				P_a	P_s				P_a
	$M_4: Y = X_1 + 1.25X_2 + 0.75X_8 - 2.4X_{16} + \varepsilon$														
Size	X_1	X_2	X_8	X_{16}	All	X_1	X_2	X_8	X_{16}	All	X_1	X_2	X_8	X_{16}	All
d_1	1.000	1.000	0.682	1.000	0.682	0.948	0.982	0.090	0.974	0.082	1.000	1.000	0.432	1.000	0.432
d_2	1.000	1.000	0.814	1.000	0.814	0.978	0.994	0.154	0.988	0.150	1.000	1.000	0.574	1.000	0.574
d_3	1.000	1.000	0.866	1.000	0.866	0.984	0.994	0.194	0.994	0.190	1.000	1.000	0.640	1.000	0.640
$M_5: Y = 3X_1X_5 + 2X_{10} + 2X_{15} + \varepsilon$															
	X_1	X_5	X_{10}	X_{15}	All	X_1	X_5	X_{10}	X_{15}	All	X_1	X_5	X_{10}	X_{15}	All
d_1	0.906	1.000	1.000	1.000	0.906	1.000	1.000	1.000	1.000	1.000	0.996	1.000	1.000	1.000	0.996
d_2	0.988	1.000	1.000	1.000	0.988	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
d_3	0.992	1.000	1.000	1.000	0.992	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$M_6: Y = 3X_1^2 + 5X_2 + 5X_8 - 8X_{16} + \varepsilon$															
	X_1	X_2	X_8	X_{16}	All	X_1	X_2	X_8	X_{16}	All	X_1	X_2	X_8	X_{16}	All
d_1	1.000	1.000	1.000	1.000	1.000	0.975	0.990	0.520	1.000	0.510	1.000	1.000	0.995	1.000	0.995
d_2	1.000	1.000	1.000	1.000	1.000	0.975	1.000	0.650	1.000	0.645	1.000	1.000	1.000	1.000	1.000
d_3	1.000	1.000	1.000	1.000	1.000	0.985	1.000	0.705	1.000	0.700	1.000	1.000	1.000	1.000	1.000
$M_7: Y = 2I(\omega > 0)\omega + 1$ *with* $\omega = 5X_1^2 - 5X_2^2 + 3X_8 + 2X_{16} + \varepsilon$															
	X_1	X_2	X_8	X_{16}	All	X_1	X_2	X_8	X_{16}	All	X_1	X_2	X_8	X_{16}	All
d_1	0.994	0.720	1.000	0.874	0.614	1.000	0.398	0.108	0.042	0.000	1.000	0.996	0.998	0.738	0.732
d_2	0.998	0.866	1.000	0.922	0.796	1.000	0.576	0.168	0.086	0.010	1.000	0.998	1.000	0.828	0.826
d_3	1.000	0.942	1.000	0.940	0.884	1.000	0.696	0.200	0.112	0.020	1.000	1.000	1.000	0.874	0.874

Table 10. Top 5 genes identified by 6 methods and the adjusted R^2 and deviance explained of additive models for the microarray data

	X_{k1}	X_{k2}	X_{k3}	X_{k4}	X_{k5}	R^2	Deviance explained
DC-SIS	Msa.2134.0	Msa.2877.0	Msa.26025.0	Msa.5583.0	Msa.1590.0	91.9%	93.8%
MDC-SIS	Msa.2877.0	Msa.2134.0	Msa.741.0	Msa.1166.0	Msa.1590.0	93.3%	94.9%
SIS	Msa.2134.0	Msa.5794.0	Msa.15442.0	Msa.5727.0	Msa.7019.0	63.3%	72.3%
SIRS	Msa.2412.0	Msa.1007.0	Msa.2037.0	Msa.116.0	Msa.16991.0	60.2%	67.9%
τ_b^*-SIS	Msa.2877.0	Msa.964.0	Msa.2134.0	Msa.741.0	Msa.3041.0	93.8%	95.2%
RRCS	Msa.7019.0	Msa.5794.0	Msa.1166.0	Msa.2877.0	Msa.15442.0	85.2%	88.4%

The parameter settings are respectively and basically identical with these in Example (1.a) and (2.a) in Pan, Wang, Xiao and Zhu (2019). We generate $(X_1, X_2, \ldots, X_p)$ from a multivariate normal distribution with mean zero and covariance matrix $\mathbf{\Sigma} = (\sigma_{ij})_{p\times p}$. In M_4, $\sigma_{ij} = 0.8^{|i-j|}$, $n = 150$ and $p = 1000$. In M_5, $\sigma_{ij} = 0.9^{|i-j|}$, $n = 200$ and $p = 2000$. In M_6, $\sigma_{ij} = 0.8^{|i-j|}$, $n = 200$ and $p = 2000$. In M_7, $\sigma_{ij} = 0.5^{|i-j|}$, $n = 200$ and $p = 2000$.

From results of M_4 in Table 9 we can clearly see that the point of P_a in three methods is the selection of X_8 and all of them can successfully select predictors X_1, X_2 and X_{16}. In this view, τ_b^*-SIS performs better. Data in M_5 reveal that IPDC and BCor-SIS are suitable to the linear interaction model and τ_b^*-SIS is also competitive. For M_6 and M_7, τ_b^*-SIS and BCor-SIS outperform IPDC in general.

4. REAL DATA ANALYSIS

In this Section, we apply the proposed method to the cardiomyopathy microarray dataset, which was also analyzed by Segal, Dahlquist and Conklin (2003), Hall and Miller (2009), Li, Zhong and Zhu (2012), among others. We aim to identify a set of influential genes for overexpression of a G protein-coupled receptor (Ro1) in mice, which is beneficial to understand types of human heart disease. The dataset includes the information collected from $n = 30$ mice. The Ro1 expression level is the response Y, and other $p = 6319$ gene expression levels are the predictors X_k's.

First, we rank the predictors through all the six screening approaches and choose the top 5 genes which are listed in Table 10. To assess the performance of these 6 procedures, we fit the following additive model:

$$Y = \ell_{k1}(X_{k1}) + \ell_{k2}(X_{k2}) + \ell_{k3}(X_{k3}) + \ell_{k4}(X_{k4}) + \ell_{k5}(X_{k5}) + \varepsilon_k, \ k = 1, 2, 3, 4, 5, 6,$$

where k corresponds to different method, respectively. We use the R "mgcv" package to fit the unknown link function ℓ_{ki} and gain the results of adjusted R^2 and deviance explained for the models. As we know, the larger the val-

ues are, the better the procedure performs. Thus, we can conclude that τ_b^*-SIS performs well in identifying a set of influential genes.

APPENDIX A. SOME LEMMAS

The following two lemmas are quoted from Lemma 5.6.1.A and Theorem 5.6.1.A in Serfling (2009), which play an important role in the proof of Theorem 1.

Lemma 1. *Let $\mu = \mathbb{E}(Y)$. If $P(a \le Y \le b) = 1$, then*

$$\mathbb{E}\left[\exp\left\{s(Y-\mu)\right\}\right] \le \exp\left\{s^2(b-a)^2/8\right\}$$

holds for any $s > 0$.

Lemma 2. *Let $h(Y_1, \ldots, Y_m)$ be a kernel of the U–statistic U_n, and $\theta = \mathbb{E}\{h(Y_1, \ldots, Y_m)\}$. If $a \le h(Y_1, \ldots, Y_m) \le b$, then for any $t > 0$ and $n \ge m$,*

$$P\left(U_n - \theta \ge t\right) \le \exp\left\{-2\lfloor n/m \rfloor t^2/(b-a)^2\right\},$$

where $\lfloor n/m \rfloor$ denotes the integer part of n/m.

Owing to the symmetry of U–statistics, it follows from Lemma 2 that

$$P\left(|U_n - \theta| \ge t\right) \le 2\exp\left\{-2\lfloor n/m \rfloor t^2/(b-a)^2\right\}.$$

APPENDIX B. PROOF OF THEOREM 1

Some notations are first introduced before giving the proof. Let $\{X_k', Y'\}, \{X_k'', Y''\}, \{X_k''', Y'''\}$ be independent copies of $\{X_k, Y\}$, and define

$$\begin{aligned}
S_{k1} &= \mathbb{E}\left[I(X_k, X_k' < X_k'', X_k''')I(Y, Y' < Y'', Y''')\right], \\
S_{k2} &= \mathbb{E}\left[I(X_k, X_k' < X_k'', X_k''')I(Y, Y' > Y'', Y''')\right], \\
S_{k3} &= \mathbb{E}\left[I(X_k, X_k' < X_k'', X_k''')I(Y, Y' < Y'', Y''')\right],
\end{aligned}$$

and their sample counterparts

$$\begin{aligned}
\hat{S}_{k1} &= \frac{1}{n^4}\sum_{i,j,l,m=1}^{n}\left[I(X_{ik}, X_{jk} < X_{lk}, X_{mk})I(Y_i, Y_j < Y_l, Y_m)\right], \\
\hat{S}_{k2} &= \frac{1}{n^4}\sum_{i,j,l,m=1}^{n}\left[I(X_{ik}, X_{jk} < X_{lk}, X_{mk})I(Y_i, Y_j > Y_l, Y_m)\right], \\
\hat{S}_{k3} &= \frac{1}{n^4}\sum_{i,j,l,m=1}^{n}\left[I(X_{ik}, X_{jk} < X_{lk}, X_{mk})I(Y_i, Y_l < Y_j, Y_m)\right].
\end{aligned}$$

By the definitions of τ^* and $\hat{\tau}^*$, it follows that

$$\begin{aligned}
\tau^*(X_k, Y) &= 4S_{k1} + 4S_{k2} - 8S_{k3}, \\
\hat{\tau}^*(X_k, Y) &= 4\hat{S}_{k1} + 4\hat{S}_{k2} - 8\hat{S}_{k3}.
\end{aligned}$$

We aim to show the uniform consistency of the denominator and the numerator of $\hat{\omega}_k$ under regularity conditions, respectively. And we only deal with the numerator in the following in that the denominator has a similar form compared with the numerator. Throughout the proof, the notations C and c are generic constants, which may take different values at each appearance.

We first deal with $\hat{S}_{k1}$. Define

$$\begin{aligned}
\hat{S}_{k1}^* &= \frac{1}{\binom{n}{4}}\sum_{1\le i<j<l<m\le n}\left[I(X_{ik}, X_{jk} < X_{lk}, X_{mk})I(Y_i, Y_j < Y_l, Y_m)\right] \\
&= \frac{1}{n(n-1)(n-2)(n-3)} \\
&\quad \times \sum_{i\ne j\ne l\ne m}\left[I(X_{ik}, X_{jk} < X_{lk}, X_{mk})I(Y_i, Y_j < Y_l, Y_m)\right],
\end{aligned}$$

where the second equality follows from the lemma in Serfling (2009, Sec. 5.7.3). Clearly, $\hat{S}_{k1}^*$ is a U statistic and the following relation holds:

$$\hat{S}_{k1} = \frac{(n-1)(n-2)(n-3)}{n^3}\hat{S}_{k1}^*.$$

We shall establish the uniform consistency of $\hat{S}_{k1}^*$. By the Cauchy-Schwartz inequality,

$$\begin{aligned}
S_{k1} &= \mathbb{E}\left[I(X_{ik}, X_{jk} < X_{lk}, X_{mk})I(Y_i, Y_j < Y_l, Y_m)\right] \\
&\le \left\{\mathbb{E}I(X_{ik}, X_{jk} < X_{lk}, X_{mk})\mathbb{E}I(Y_i, Y_j < Y_l, Y_m)\right\}^{\frac{1}{2}} \\
&\le 1,
\end{aligned}$$

which implies that S_{k1} is bounded. For any given $\varepsilon > 0$, take n large enough such that $S_{k1}/n < \varepsilon$. Then, it can be shown that

(B.1)

$$\begin{aligned}
&P\left(\left|\hat{S}_{k1} - S_{k1}\right| \ge 7\varepsilon\right) \\
&= P\left(\left|\frac{(n-1)(n-2)(n-3)}{n^3}\hat{S}_{k1}^* - \frac{(n-1)(n-2)(n-3)}{n^3}S_{k1}\right.\right. \\
&\qquad \left.\left. - \frac{6n^2-11n+6}{n^3}S_{k1}\right| \ge 7\varepsilon\right) \\
&\le P\left(\left|\frac{(n-1)(n-2)(n-3)}{n^3}(\hat{S}_{k1}^* - S_{k1})\right|\right. \\
&\qquad \left. \ge 7\varepsilon - \frac{6n^2-11n+6}{n^3}S_{k1}\right) \\
&\le P\left(\left|\hat{S}_{k1}^* - S_{k1}\right| \ge \varepsilon\right).
\end{aligned}$$

To establish the uniform consistency of $\hat{S}_{k1}$, it thus suffices to show the uniform consistency of $\hat{S}_{k1}^*$.

Let

$$\begin{aligned}
&h_1(X_{ik}, Y_i; X_{jk}, Y_j; X_{lk}, Y_l; X_{mk}, Y_m) \\
&= I(X_{ik}, X_{jk} < X_{lk}, X_{mk})I(Y_i, Y_j < Y_l, Y_m)
\end{aligned}$$

be the kernel of the U statistic $\hat{S}_{k1}^*$. We decompose the kernel function h_1 into two parts:

$$h_1 = h_1 I(h_1 \le M) + h_1 I(h_1 > M),$$

where M will be specified in the following. The U statistic can now be written as follows:

$$\begin{aligned}\hat{S}^*_{k1} &= \frac{1}{n(n-1)(n-2)(n-3)} \\ &\quad \times \sum_{i\neq j\neq l\neq m} h_1(X_{ik},Y_i;X_{jk},Y_j;X_{lk},Y_l;X_{mk},Y_m) \\ &\quad \times I\{h_1(X_{ik},Y_i;X_{jk},Y_j;X_{lk},Y_l;X_{mk},Y_m)\le M\} \\ &\quad + \frac{1}{n(n-1)(n-2)(n-3)} \\ &\quad \times \sum_{i\neq j\neq l\neq m} h_1(X_{ik},Y_i;X_{jk},Y_j;X_{lk},Y_l;X_{mk},Y_m) \\ &\quad \times I\{h_1(X_{ik},Y_i;X_{jk},Y_j;X_{lk},Y_l;X_{mk},Y_m)> M\} \\ &:= \hat{S}^*_{k1,1}+\hat{S}^*_{k1,2}.\end{aligned}$$

Accordingly, we decompose S_{k1} into two parts:

$$\begin{aligned}S_{k1} &= \mathbb{E}\left[h_1(X_{ik},Y_i;X_{jk},Y_j;X_{lk},Y_l;X_{mk},Y_m)\right. \\ &\qquad \left.\times I\{h_1(X_{ik},Y_i;X_{jk},Y_j;X_{lk},Y_l;X_{mk},Y_m)\le M\}\right] \\ &\quad + \mathbb{E}\left[h_1(X_{ik},Y_i;X_{jk},Y_j;X_{lk},Y_l;X_{mk},Y_m)\right. \\ &\qquad \left.\times I\{h_1(X_{ik},Y_i;X_{jk},Y_j;X_{lk},Y_l;X_{mk},Y_m)> M\}\right] \\ &:= S_{k1,1}+S_{k1,2}.\end{aligned}$$

Clearly, $\hat{S}^*_{k1,1}$ and $\hat{S}^*_{k1,2}$ are unbiased estimators of $S_{k1,1}$ and $S_{k1,2}$, respectively.

We prove the consistency of $\hat{S}^*_{k1,1}$ first. By Markov's inequality, for any $t>0$, we have

$$\begin{aligned}&P\left(\hat{S}^*_{k1,1}-S_{k1,1}\ge\varepsilon\right) \\ &\le \exp(-t\varepsilon)\exp\left(-tS_{k1,1}\right)\mathbb{E}\left\{\exp\left(t\hat{S}^*_{k1,1}\right)\right\}.\end{aligned}$$

Note that any U statistic can be represented as an average of iid random variables by Serfling (2009, Sec. 5.1.6). That is,

$$\hat{S}^*_{k1,1}=(n!)^{-1}\sum_{n!}\Omega(X_{1k},Y_1;\ldots;X_{nk},Y_n),$$

where $\sum_{n!}$ denotes the summation over all possible permutations of $(1,\ldots,n)$, and each $\Omega(X_{1k},Y_1;\ \ldots;X_{nk},Y_n)$ is an average of $m=\lfloor n/4\rfloor$ iid random variables (i.e., $\Omega=m^{-1}\sum_r h_1^{(r)}I\{h_1^{(r)}\le M\}$). Since the exponential function is convex, it follows from Jensen's inequality that, for $0<t\le 2s_0$,

$$\begin{aligned}&\mathbb{E}\left\{\exp\left(t\hat{S}^*_{k1,1}\right)\right\} \\ &= \mathbb{E}\left[\exp\left\{t(n!)^{-1}\sum_{n!}\Omega(X_{1k},Y_1;\ldots;X_{nk},Y_n)\right\}\right] \\ &\le (n!)^{-1}\sum_{n!}\mathbb{E}\left[\exp\{t\Omega(X_{1k},Y_1;\ldots;X_{nk},Y_n)\}\right] \\ &= \mathbb{E}^m\left[\exp\left(m^{-1}th_1^{(r)}I\left\{h_1^{(r)}\le M\right\}\right)\right].\end{aligned}$$

The above inequality, together with Lemma 1, yields that

$$\begin{aligned}&P\left(\hat{S}^*_{k1,1}-S_{k1,1}\ge\varepsilon\right) \\ &\le \exp(-t\varepsilon)\mathbb{E}^m\left\{\exp\left(m^{-1}t\left[h_1^{(r)}I\left\{h_1^{(r)}\le M\right\}-S_{k1,1}\right]\right)\right\} \\ &\le \exp\{-t\varepsilon+M^2t^2/(8m)\}.\end{aligned}$$

By choosing $t=4\varepsilon m/M^2$, we have

$$P\left(\hat{S}^*_{k1,1}-S_{k1,1}\ge\varepsilon\right)\le\exp\left\{-2\varepsilon^2m/M^2\right\}.$$

Therefore, by the symmetry of U statistics, we can obtain that

$$P\left(\left|\hat{S}^*_{k1,1}-S_{k1,1}\right|\ge\varepsilon\right)\le 2\exp\left\{-2\varepsilon^2m/M^2\right\}. \tag{B.2}$$

Then we show the consistency of $\hat{S}^*_{k1,2}$. By Cauchy-Schwarz and Markov's inequalities, for any $s'>0$, we have

$$\begin{aligned}S^2_{k1,2} &\le \mathbb{E}\left\{h_1^2(X_{ik},Y_i;X_{jk},Y_j;X_{lk},Y_l;X_{mk},Y_m)\right\} \\ &\quad \times P\{h_1(X_{ik},Y_i;X_{jk},Y_j;X_{lk},Y_l;X_{mk},Y_m)>M\} \\ &\le \mathbb{E}\{h_1^2(X_{ik},Y_i;X_{jk},Y_j;X_{lk},Y_l;X_{mk},Y_m)\} \\ &\quad \times \mathbb{E}[\exp\{s'h_1(X_{ik},Y_i;X_{jk},Y_j;X_{lk},Y_l;X_{mk},Y_m)\}] \\ &\quad \cdot \exp\{-s'M\}.\end{aligned}$$

If we choose $M=cn^\gamma$, where $0<\gamma<1/2-\kappa$, then $S_{k1,2}\le\varepsilon/2$ for sufficiently large n. Thus,

$$P\left(\left|\hat{S}^*_{k1,2}-S_{k1,2}\right|>\varepsilon\right)\le P\left(\left|\hat{S}^*_{k1,2}\right|>\varepsilon/2\right).$$

Note that for any $s>0$, there holds that

$$\begin{aligned}P\left(\left|\hat{S}^*_{k1,2}\right|>\varepsilon/2\right) &\le \exp\{-s\varepsilon/2\}\mathbb{E}\left\{\exp\left(s\left|\hat{S}^*_{k1,2}\right|\right)\right\} \\ &\le C\exp(-s\varepsilon/2).\end{aligned}$$

It follows that

$$P\left(\left|\hat{S}^*_{k1,2}-S_{k1,2}\right|>\varepsilon\right)\le C\exp(-s\varepsilon/2). \tag{B.3}$$

Recall that $M=cn^\gamma$, by (B.1), (B.2) and (B.3), we have

$$\begin{aligned}P\left(\left|\hat{S}_{k1}-S_{k1}\right|\ge 14\varepsilon\right) &\le 2\exp(-\varepsilon^2n^{1-2\gamma}/2) \\ &\quad + C\exp(-s\varepsilon/2).\end{aligned} \tag{B.4}$$

The uniform consistency of $\hat{S}_{k2}$ and $\hat{S}_{k3}$ can be established similarly, that is

$$\begin{aligned}P\left(\left|\hat{S}_{k2}-S_{k2}\right|\ge 14\varepsilon\right) &\le 2\exp(-\varepsilon^2n^{1-2\gamma}/2) \\ &\quad + C\exp(-s\varepsilon/2),\end{aligned} \tag{B.5}$$

$$\begin{aligned}P\left(\left|\hat{S}_{k3}-S_{k3}\right|\ge 14\varepsilon\right) &\le 2\exp(-\varepsilon^2n^{1-2\gamma}/2) \\ &\quad + C\exp(-s\varepsilon/2).\end{aligned} \tag{B.6}$$

We thus have

$$
\begin{aligned}
&P\left\{\left|(4\hat{S}_{k1}+4\hat{S}_{k2}-8\hat{S}_{k3})-(4S_{k1}+4S_{k2}-8S_{k3})\right|\geq\varepsilon\right\}\\
&\leq P\left(\left|\hat{S}_{k1}-S_{k1}\right|\geq\varepsilon/16\right)+P\left(\left|\hat{S}_{k2}-S_{k2}\right|\geq\varepsilon/16\right)\\
&\quad+P\left(\left|\hat{S}_{k3}-S_{k3}\right|\geq\varepsilon/16\right)\\
&=O\left(\exp\left\{-c_1\varepsilon^2n^{1-2\gamma}\right\}\right),
\end{aligned}
$$

where $c_1>0$ is a constant. Till now, the convergence rate of the numerator of $\hat{\omega}_k$ is obtained.

Following similar arguments, we can also obtain the convergence rate of the denominator. In addition, it can be shown that the convergence rate of $\hat{\omega}_k$ is the same as that of its numerator. Now let $\varepsilon=cn^{-\kappa}$, where κ is such that $0<\kappa+\gamma<1/2$, then it follows that

$$
\begin{aligned}
P\Big(\max_{1\leq k\leq p}|\hat{\omega}_k-\omega_k|\geq cn^{-\kappa}\Big)&\leq p\max_{1\leq k\leq p}P\left(|\hat{\omega}_k-\omega_k|\geq cn^{-\kappa}\right)\\
&\leq O\left(p\exp\{-c_1n^{1-2(\kappa+\gamma)}\}\right)\\
&\leq O\left(p\exp\{-c_1n^{1-2\kappa}\}\right).
\end{aligned}
$$

Thus, the first part of Theorem 1 is proved.

Finally, we show the second part of Theorem 1. If $D\nsubseteq\hat{D}^*$, then there exists some $k_0\in D$ such that $\hat{\omega}_{k_0}<cn^{-\kappa}$. It follows from the condition (C1) that $|\hat{\omega}_{k_0}-\omega_{k_0}|>cn^{-\kappa}$. Thus,

$$
P\left(D\nsubseteq\hat{D}^*\right)\leq P\left(\max_{k\in D}|\hat{\omega}_k-\omega_k|\geq cn^{-\kappa}\right).
$$

Therefore,

$$
\begin{aligned}
P\left(D\subseteq\hat{D}^*\right)&\geq 1-P\left(\max_{k\in D}|\hat{\omega}_k-\omega_k|\geq cn^{-\kappa}\right)\\
&\geq 1-O\left(s_n\exp\{-c_1n^{1-2\kappa}\}\right),
\end{aligned}
$$

where s_n is the cardinality of D. The proof of Theorem 1 is completed.

APPENDIX C. PROOF OF THEOREM 2

The proof of Theorem 2 follows from that of Theorem 2.2 in Cui, Li and Zhong (2015). In fact,

$$
\begin{aligned}
&P\Big\{\min_{k\in D}\hat{\omega}_k-\max_{k\in I}\hat{\omega}_k<c_2/2\Big\}\\
&\leq P\Big\{\Big(\min_{k\in D}\hat{\omega}_k-\max_{k\in I}\hat{\omega}_k\Big)-\Big(\min_{k\in D}\omega_k-\max_{k\in I}\omega_k\Big)<-c_2/2\Big\}\\
&\leq P\Big\{\Big|\Big(\min_{k\in D}\hat{\omega}_k-\max_{k\in I}\hat{\omega}_k\Big)-\Big(\min_{k\in D}\omega_k-\max_{k\in I}\omega_k\Big)\Big|>c_2/2\Big\}\\
&\leq P\Big\{2\max_{1\leq k\leq p}|\hat{\omega}_k-\omega_k|>c_2/2\Big\}\\
&\leq O\{p\exp(-c_3n)\}.
\end{aligned}
$$

Note that $\log(p)/n=o(1)$ implies that $p\leq\exp\{(c_3/2)n\}$ for sufficiently large n. Thus, for some sufficiently large n_0,

$$
\sum_{n=n_0}^{\infty}p\exp(-c_3n)\leq\exp\left\{(c_3/2)n-c_3n\right\}\leq\sum_{n=n_0}^{\infty}n^{-2}<+\infty.
$$

Therefore, by the Borel–Contelli lemma, we have

$$
\liminf_{n\to\infty}\left\{\min_{k\in D}\hat{\omega}_k-\max_{k\in I}\hat{\omega}_k\right\}\geq c_2/2>0,\ a.s..
$$

APPENDIX D. PROOF OF THEOREM 3

For any $c>0$, the number of $\{k_1:|\omega_{k_1}|>(c/2)n^{-\kappa}\}$ is bounded by $O\left(n^{\kappa}\sum_{k=1}^{p}|\omega_k|\right)$. Then on the set $\mathcal{B}_n=\{\max_{1\leq k\leq p}|\hat{\omega}_k-\omega_k|\leq(c/2)n^{-\kappa}\}$, the number of $\{k_2:|\hat{\omega}_{k_2}|>cn^{-\kappa}\}$ can not exceed the number of $\{k_1:|\omega_{k_1}|>(c/2)n^{-\kappa}\}$. Therefore,

$$
P\left\{\left|\hat{D}^*\right|\leq O\left(n^{\kappa}\sum_{k=1}^{p}|\omega_k|\right)\right\}\geq P\left(\mathcal{B}_n\right),
$$

which, together with Theorem 1, yields Theorem 3.

ACKNOWLEDGEMENT

The authors would like to thank the Editor-in-Chief and the two anonymous Reviewers for the careful review and the valuable comments which have greatly improved the paper. Daojiang He's work is supported by the National Natural Science Foundation of China (Grant No. 11201005), the Humanities and Social Sciences Foundation of Ministry of Education, China (Grant No. 17YJC910003) and the Natural Science Foundation of Anhui Province (Grant No. 2008085MA08); Kai Xu's work is supported by the National Natural Science Foundation of China (Grant No. 11901006) and the Natural Science Foundation of Anhui Province (Grant No. 1908085QA06); Lei He's work is supported by the Natural Science Foundation of Anhui Province (Grant No. 2008085QA15).

Received 20 April 2020

REFERENCES

Bergsma, W. and Dassios, A. (2014), 'A consistent test of independence based on a sign covariance related to kendall's tau', *Bernoulli* **20**(2), 1006–1028. MR3178526

Candes, E. and Tao, T. (2007), 'The dantzig selector: Statistical estimation when p is much larger than n', *The Annals of Statistics* **35**(6), 2313–2351. MR2382644

Cui, H., Li, R. and Zhong, W. (2015), 'Model-free feature screening for ultrahigh dimensional discriminant analysis', *Journal of the American Statistical Association* **110**(510), 630–641. MR3367253

Fan, J., Feng, Y. and Song, R. (2011), 'Nonparametric independence screening in sparse ultra-high-dimensional additive models', *Journal of the American Statistical Association* **106**(494), 544–557. MR2847969

Fan, J. and Li, R. (2001), 'Variable selection via nonconcave penalized likelihood and its oracle properties', *Journal of the American Statistical Association* **96**(456), 1348–1360. MR1946581

Fan, J. and Lv, J. (2008), 'Sure independence screening for ultrahigh dimensional feature space', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**(5), 849–911. MR2530322

Fan, J. and Lv, J. (2011), 'Nonconcave penalized likelihood with np-dimensionality', *IEEE Transactions on Information Theory* **57**(8), 5467–5484. MR2849368

Fan, J. and Peng, H. (2004), 'Nonconcave penalized likelihood with a diverging number of parameters', *The Annals of Statistics* **32**(3), 928–961. MR2065194

Fan, J. and Song, R. (2010), 'Sure independence screening in generalized linear models with np-dimensionality', *The Annals of Statistics* **38**(6), 3567–3604. MR2766861

Frank, L. E. and Friedman, J. H. (1993), 'A statistical view of some chemometrics regression tools', *Technometrics* **35**(2), 109–135.

Hall, P. and Miller, H. (2009), 'Using generalized correlation to effect variable selection in very high dimensional problems', *Journal of Computational and Graphical Statistics – J. Comput. Graph. Stat.* **18**. MR2751640

Huang, J., Horowitz, J. L. and Ma, S. (2008), 'Asymptotic properties of bridge estimators in sparse high-dimensional regression models', *The Annals of Statistics* **36**(2), 587–613. MR2396808

Kendall, M. G. (1938), 'A new measure of rank correlation', *Biometrika* **30**(1/2), 81–93. MR0138175

Kendall, M. G. (1962), *Rank correlation methods*, London Griffin.

Kong, Y., Li, D., Fan, Y. and Lv, J. (2017), 'Interaction pursuit in high-dimensional multi-response regression via distance correlation', *Annals of Statistics* **45**(2), 897–922. MR3650404

Li, G., Peng, H., Zhang, J. and Zhu, L. (2012), 'Robust rank correlation based screening', *The Annals of Statistics* **40**(3), 1846–1877. MR3015046

Li, R., Zhong, W. and Zhu, L. (2012), 'Feature screening via distance correlation learning', *Journal of the American Statistical Association* **107**(499), 1129–1139. MR3010900

Pan, W., Wang, X., Xiao, W. and Zhu, H. (2019), 'A generic sure independence screening procedure', *Journal of the American statistical association* **114**(526), 928–937. MR3963192

Segal, M., Dahlquist, K. and Conklin, B. (2003), 'Regression approaches for microarray data analysis', *Journal of computational biology: a journal of computational molecular cell biology* **10**, 961–80.

Serfling, R. J. (2009), *Approximation theorems of mathematical statistics*, John Wiley & Sons. MR0595165

Shao, X. and Zhang, J. (2014), 'Martingale difference correlation and its use in high-dimensional variable screening', *Journal of the American Statistical Association* **109**(507), 1302–1318. MR3265698

Székely, G. J. and Rizzo, M. L. (2009), 'Brownian distance covariance', *The Annals of Applied Statistics* **3**(4), 1236–1265. MR2752127

Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007), 'Measuring and testing dependence by correlation of distances', *The Annals of Statistics* **35**(6), 2769–2794. MR2382665

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **58**(1), 267–288. MR1379242

Van de Geer, S. A. (2008), 'High-dimensional generalized linear models and the lasso', *The Annals of Statistics* **36**(2), 614–645. MR2396809

Zhu, L.-P., Li, L., Li, R. and Zhu, L.-X. (2011), 'Model-free feature screening for ultrahigh-dimensional data', *Journal of the American Statistical Association* **106**(496), 1464–1475. MR2896849

Zou, H. (2006), 'The adaptive lasso and its oracle properties', *Journal of the American Statistical Association* **101**(476), 1418–1429. MR2279469

Zou, H. and Hastie, T. (2005), 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(2), 301–320. MR2137327

Daojiang He
Department of Statistics
Anhui Normal University
Wuhu 241002
China
E-mail address: djheahnu@163.com

Xinxin Hao
Department of Statistics
Anhui Normal University
Wuhu 241002
China
E-mail address: 19965385868@126.com

Kai Xu
Department of Statistics
Anhui Normal University
Wuhu 241002
China
E-mail address: tjxxukai@163.com

Lei He
Department of Statistics
Anhui Normal University
Wuhu 241002
China
E-mail address: lhstat@163.com

Youxin Liu
School of Science
Nanjing University of Science and Technology
Nanjing 210094
China
E-mail address: yxliu@whit.edu.cn