Two-Directional Simultaneous Inference for High-Dimensional Models

Wei Liu\textsuperscript{a}, Huazhen Lin\textsuperscript{a}, Jin Liu\textsuperscript{b}, and Shurong Zheng\textsuperscript{c}

\textsuperscript{a}Center of Statistical Research and School of Statistics, Southwestern University of Finance and Economics, Chengdu, China; \textsuperscript{b}Centre for Quantitative Medicine, Program in Health Services \& Systems Research, Duke-NUS Medical School, Singapore; \textsuperscript{c}School of Mathematics and Statistics, Northeast Normal University, Changchun, China

ABSTRACT

This article proposes a general two-directional simultaneous inference (TOSI) framework for high-dimensional models with a manifest variable or latent variable structure, for example, high-dimensional mean models, high-dimensional sparse regression models, and high-dimensional latent factors models. TOSI performs simultaneous inference on a set of parameters from two directions, one to test whether the assumed zero parameters indeed are zeros and one to test whether exist zeros in the parameter set of nonzeros. As a result, we can better identify whether the parameters are zeros, thereby keeping the data structure fully and parsimoniously expressed. We theoretically prove that the single-split TOSI is asymptotically unbiased and the multi-split version of TOSI can control the Type I error below the presupposed significance level. Simulations are conducted to examine the performance of the proposed method in finite sample situations and two real datasets are analyzed. The results show that the TOSI method can provide more predictive and more interpretable estimators than existing methods.

1. Introduction

Over the past two decades, great progress has been made in the field of high-dimensional data, where the number of parameters can be much larger than the sample size. The most popular and powerful methods for handling high-dimensional data are regularization methods (Tibshirani 1996; Fan and Li 2001) and screening methods (Fan and Lv 2008; Ma et al. 2017), which can be used to separate the set of parameters into a set $G^0$ and its complement set $G^{\text{no}}$, which are inactive (or zero) and active (or nonzero) sets, respectively. Given these two sets, two natural problems arise: (a) whether all the elements in $G^0$ are insignificant; and (b) whether all the elements in $G^{\text{no}}$ are significant. Clearly, we can fully and parsimoniously express the data structure once we address these two problems; thus, it is important to perform statistical inference that quantifies the uncertainty associated with the two problems. We formally formulate the problems (1) and (2) as

\begin{align*}
H_{0,G^0} : \theta_j = 0, \forall j \in G^0, & \text{ versus } \quad H_{1,G^0} : \theta_j \neq 0, \exists j \in G^0, \quad (1) \\
H_{0,G^{\text{no}}} : \theta_j = 0, \forall j \in G^{\text{no}}, & \text{ versus } \quad H_{1,G^{\text{no}}} : \theta_j \neq 0, \exists j \in G^{\text{no}}, \quad (2)
\end{align*}

where $G^0$ and $G^{\text{no}}$ are the previously specified inactive and active sets, respectively, and $\{\theta_j \in \mathbb{R}^p, j = 1, \ldots, p\}$ is a set of parameters of interest with large $p$ and fixed $q$. For example, $\theta_j$ is the $j$th regression coefficient in a high-dimensional sparse regression model, $\theta_j$ is the mean of the $j$th variable in a high-dimensional mean model, and $\theta_j$ is the $j$th loading vector in a latent factor model.

Existing methods focus on assessing problem (1), while statistical inference to quantify the uncertainty associated with the identification of a group of important variables (problem (2)) is totally ignored in the literature. In fact, with the inference in the active set to eliminate the possibility of including true zeros as nonzeros in the set, we can obtain interpretable and simpler models. However, the assessment of problem (2) is more difficult than that of problem (1) since the explicit form is available for problem (1) under the null hypothesis, but not for problem (2), where each $\theta_j$ can take any value under the null hypothesis. In this article, under a general framework for high-dimensional data, we propose a two-directional simultaneous test (TOSI) to test $H_{0,G^0}$ and $H_{0,G^{\text{no}}}$, that is, whether all elements in $G^0$ are insignificant (problem (1)) and whether all elements in $G^{\text{no}}$ are significant (problem (2)).

Existing methods for assessing problem (1) can be roughly divided into two categories: $p$-value adjustment methods (PAMs) and simultaneous inference methods (SIMs). PAMs are proposed for testing a single parameter (TSP) in high-dimensional sparse regression models (Zhang and Zhang 2014; De Geer et al. 2014). Specifically, by performing TSP on $H_0 : \theta_j = 0$, versus $H_1 : \theta_j \neq 0$ for each parameter $\theta_j$ in $G^0$, we can obtain a set of $p$-values. For these nominal $p$-values, the PAMs

CONTACT Hua-Zhen Lin linhz@swufe.edu.cn Center of Statistical Research and School of Statistics, Southwestern University of Finance and Economics, Chengdu, China.

Supplementary materials for this article are available online. Please go to www.tandfonline.com/UBES.

© 2023 American Statistical Association
for problem (1) are proposed to control the family-wise error rate (FWER) (Holm 1979) or false discovery rate (Benjamini and Yekutieli 2001). Recently, the sample-splitting techniques are commonly used under PAM framework. For example, Rinaldo, Wasserman, and Gsell (2019) and Barber and Candès (2019) used single-split method for the post-selection inference so that the uncertainty from the estimation and variable selection can be ignored. To achieve multi-adjustment of each p-value of regression coefficients, Wasserman and Roeder (2009) proposed a single-split method, Meinshausen, Meier, and Bühlmann (2009) proposed a multi-split method, and Mandonzo and Bühlmann (2016) proposed a hierarchical version of the multi-split method in Meinshausen, Meier, and Bühlmann (2009). However, the PAMs lack power due to the strictness of FWER methods. Several SIMs are developed to improve the power. The max-type test statistic is a typical SIM. Jiang (2004) proposed a max-type test for a high-dimensional correlation matrix with a restriction of \( p \leq n \). For a set of mean parameters \( \mu_j \) in a high-dimensional mean model with \( p = o(\exp(n^c)) (0 < c < 1) \), Chernozhukov, Chetverikov, and Kato (2013) and Lou and Wu (2017) proposed a multiplier bootstrap method to conduct simultaneous testing based on \( \max |\hat{\beta}_j - \mu_j| \). Zhang and Cheng (2017) and Dezeure, Bühlmann, and Zhang (2017) considered \( \max |\hat{\beta}_j - \beta| \) to make simultaneous inference for high-dimensional sparse regression models under homogenous and heterogeneous errors, respectively. To enhance the power of the max-type test, Zhang and Cheng (2017) also proposed a single-split method that divides a sample into two parts, called a three-step procedure, performing variable selection using the first part and simultaneous inference using the second part. Because only part of the data are used for inference using the single-split method, the testing power is also restricted. Since the max-type statistic used in the above works does not have a known asymptotic distribution, the bootstrap method is often used to determine the critical value of the test statistic (Chernozhukov, Chetverikov, and Kato 2013; Zhang and Cheng 2017; Dezeure, Bühlmann, and Zhang 2017).

However, these max-type methods suffer from limitations. First, the max-type test can only quantify the statistical uncertainty associated with the identification of a group of insignificant features, that is, problem (1). Second, they are proposed to make simultaneous inference about problem (1) for specific models, for example, high-dimensional sparse regression models or high dimensional mean model, the corresponding test and inference differ by case. Third, bootstrap methods are computationally intensive and may fail when the assumption of independent observations does not hold (Chernozhukov, Chetverikov, and Kato 2013; Zhang and Cheng 2017), such as in latent models where the latent factors are estimated and, hence, correlated with each other.

To overcome the aforementioned problems, we propose a general framework based on sample splitting. Our contributions are as follows.

**Generality:** Existing inference methods focus on such as high-dimensional mean or variance models and high-dimensional sparse regression analysis, in which the corresponding test and inference may differ case by case. In the article, we provide a generalized framework for two-directional inference for the models mentioned above, as well as models which have not been considered in the literature, such as latent variable models.

**Interpretability:** By better identifying the sets of zeros and nonzeros, we can explicitly explore the latent structure in data, thereby achieving interpretability. Furthermore, when we identify the sets of zeros \( G(\lambda) \) and nonzeros \( G^{\text{no}}(\lambda) \) based on extra samples for any given tuning parameter \( \lambda \), the TOSI method can choose a tuning parameter \( \lambda \) for which \( H_{0,G(\lambda)} \) is accepted and \( H_{0,G^{\text{no}}(\lambda)} \) is rejected if such \( \lambda \) exists, that is, the TOSI method can select \( \lambda \) so that the resulting sets of zeros \( G(\lambda) \) and nonzeros \( G^{\text{no}}(\lambda) \) are statistically insignificant and significant, respectively. Hence, the \( \lambda \) selected by TOSI is meaningful to identify important and unimportant variables. This is also observed in our simulation studies and two motivating data of liquor sales and criminal data. For example, in the high-dimensional sparse regression model of Experiment 1 in Section 6, the LASSO with \( \lambda \) chosen using cross-validation exactly selects important variables at a frequency (CS) of 10% on average, but the LASSO with \( \lambda \) chosen using TOSI can achieve a frequency of 95% on average (Table 2). In our motivating data of a liquor sales data where \( \lambda = 0.1866 \) selected using 10-fold cross-validation identifies nine important among 249 variables, while \( \lambda = 0.3172 \) selected using TOSI further identifies three unimportant variables from the nine variables, as presented in Table 4. Further checking via existing testing methods for \( H_{0,G(\lambda)} \) with \( \lambda = 0.3172 \) show that the 243 variables indeed were unimportant. Thus, the analysis results obtained via TOSI were more interpretable than those from CV LASSO. Similarly, for the criminal data, \( \lambda = 0.1788 \) selected using 10-fold cross-validation identifies nine important variables and 91 unimportant variables, while \( \lambda = 0.1828 \) selected using TOSI further identified three unimportant variables from the nine variables, as presented in Table 5.

**Computation:** Denote \( |G| \) and \( |G^{\text{no}}| \) as the size of \( G \) and \( G^{\text{no}} \), respectively. Using the two-stage test, we convert the simultaneous test for \( H_{0,G} \) and \( H_{0,G^{\text{no}}} \) into \( |G| \) and \( |G^{\text{no}}| \) TSPs on \( H_0 : \theta_j = 0 \), versus \( H_1 : \theta_j \neq 0 \) for \( j \in G \) and \( G^{\text{no}} \), respectively. Any TSP method can be used. Hence, the computation and programming are very simple.

**Asymptotic theory:** We establish the validity of the TOSI methods of two versions, including a single-split version and a multi-split version. We prove that the single-split TOSI is asymptotically unbiased and the multi-split version of TOSI can control the Type I error below the prespecified significance level. The study in this article is the first attempt to discuss a two-directional simultaneous test.

The rest of this article is organized as follows. In Section 2, we present the general TOSI framework. The theoretical properties are investigated in Section 3. We introduce the inference for sparse latent factor models and the applications for selection of penalty parameters in Sections 4 and 5, respectively. The performance of the proposed testing procedure is evaluated via simulation studies in Section 6. In Section 7, we apply the TOSI method to analyze two real datasets with a sparse linear regression model. A brief discussion about further research along this direction is provided in Section 8. Technical proofs are relegated to the supplementary materials. In addition, we implement our
2. General Framework

2.1. TOSI Inference with $L = 1$

Let $L$ denote the sample splitting times. We start with $L = 1$. Consider $\{z_i, i = 1, \ldots, n\}$ iid samples from a population $Z \sim \mathcal{N}(\mu, \sigma^2)$, where $\sigma$ can exceed $n$. We are interested in testing a set of parameters $\{\theta_j \in \mathbb{R}, j = 1, \ldots, p\}$ with fixed integer $q$. Denote $G^0 \subseteq [p]$ as any subset of interest, where $[p] = \{1, \ldots, p\}$. In practice, $G^0$ can be the indices of parameters that are penalized to zeros by using an extra sample independent of $z_i$. We consider hypotheses (1) and (2). Throughout this article, we allow the size $|G^0|$ and $|G^{op}|$ to grow as fast as $n$ which can be the exponential order of $n$.

Suppose there exists an estimator $\hat{\theta}_i$ of $\theta_j$ satisfying $\sqrt{n}(\hat{\theta}_j - \theta_j) \overset{d}{\rightarrow} N(0, \Sigma_j)$, where $\Sigma_j$ can be consistently estimated by $\hat{\Sigma}_j$. We randomly split the data $\{z_i\}_{i=1}^{n}$ into two parts, $D_1$ and $D_2$. Without loss of generality, we take $|D_1| = |D_2| = n/2 = \hat{n}$ by assuming $n$ to be even. To test the null hypothesis in problem (1) that all $\theta_j$s with $j \in G^0$ are zeros, we propose a two-stage maximum (ToMax) test as below.

Stage I: Use $D_1$ to obtain estimator $\hat{\theta}_j^{(1)}$ of $\theta_j$ and estimator $\hat{\Sigma}_{1j}$ of $\Sigma_j$, then find $j_{\max} \in G^0$ such that $\|\hat{\Sigma}_{1j_{\max}}^{-1/2}\hat{\theta}_j^{(1)}\| \geq \|\hat{\Sigma}_{1j}^{-1/2}\theta_j^{(1)}\|$ for any $j \in G^0$.

Stage II: Use $D_2$ to obtain estimator $\hat{\theta}_j^{(2)}$ of $\theta_j$, estimator $\hat{\Sigma}_{2\max}$ of $\Sigma_j$ and calculate the p-value as $p_{\max} = P(\hat{\Sigma}_{2\max}^{1/2} \hat{\theta}_j^{(2)} > \chi^2_{L, q}(q))$, where $\hat{\Sigma}_{2\max}^{1/2} = \hat{n}\hat{\theta}_j^{(2)}\hat{\Sigma}_{2\max}^{-1}\hat{\theta}_j^{(2)}$.

Note that $j_{\max}$ is a random variable determined by sample $D_1$. Intuitively, in Stage I, we select an index with the most extreme statistics in group $G^0$, and we subsequently conduct hypothesis testing for this index at Stage II.

Remark 1. If the null hypothesis that all parameters in $G^0$ are zeros holds, then ToMax is equally likely to choose any index as $j_{\max}$. Thus, the FWER can be controlled at the prespecified level. If the null hypothesis is not true and we denote $G_1^0 = \{j \in G^0 : \theta_j \neq 0\}$ and $G_0^0 = \{j \in G^0 : \theta_j = 0\}$, then we have $\|\hat{\Sigma}_{1j}^{-1/2}\theta_j^{(1)}\| \overset{p}{\rightarrow} \|\Sigma_j^{-1/2}\theta_j^{(1)}\| = O(1)$ for $j \in G_1^0$ and $\|\hat{\Sigma}_{1j_{\max}}^{-1/2}\theta_j^{(1)}\| = o_p(1)$ for $j \in G_0^0$. Under some conditions, we prove that the ToMax test is asymptotically unbiased.

Remark 2. The two stages of ToMax convert the simultaneous test into $|G^0| + 1$ TSP tests for $\theta_j = 0$, $j \in G^0$, where Stage I is equivalent to conducting $|G^0|$ TSP tests for each $j \in G^0$ because $\|\hat{\Sigma}_{1j}^{-1/2}\theta_j^{(1)}\|$ is an equivalent expression of the p-value for $\theta_j = 0$, and Stage II conducts a TSP test for the selected index $j_{\max}$. Any TSP method can be used in Stages I and II.

To test the null hypothesis in problem (2) that there exists $j \in G^{op}$ being zero, we propose a two-stage minimum (ToMin) test as below.

Stage I: Use $D_1$ to find $j_{\min} \in G^{op}$ such that $\|\hat{\Sigma}_{1j_{\min}}^{-1/2}\theta_j^{(1)}\| \leq \|\hat{\Sigma}_{1j}^{-1/2}\theta_j^{(1)}\|$ for any $j \in G^{op}$, where $\theta_j^{(1)}$ and $\hat{\Sigma}_{1j}$ are obtained based on $D_1$.

Stage II: Use $D_2$ to estimate $\hat{\theta}_{j_{\min}}^{(2)}$ and calculate the p-value as $p_{\min} = P(\hat{\Sigma}_{2\min}^{1/2} \hat{\theta}_{j_{\min}}^{(2)} > \chi^2_{L, q}(q))$, where $\hat{\Sigma}_{2\min}^{1/2} = \hat{n}\hat{\theta}_{j_{\min}}^{(2)}\hat{\Sigma}_{2\min}^{-1}\hat{\theta}_{j_{\min}}^{(2)}$, and $\hat{\Sigma}_{2\min}$ is obtained based on $D_2$.

Remark 3. In the case that the null hypothesis in problem (2) is true, that is, there exists a parameter in $G^{op}$ that is zero and denoting $G_1^{op} = \{j \in G^{op} : \theta_j \neq 0\}$ and $G_0^{op} = \{j \in G^{op} : \theta_j = 0\}$, we have $\|\hat{\Sigma}_{1j}^{-1/2}\theta_j^{(1)}\| \overset{p}{\rightarrow} \|\Sigma_j^{-1/2}\theta_j\| = O(1)$ for $j \in G_1^{op}$ and $\|\hat{\Sigma}_{1j_{\min}}^{-1/2}\theta_{j_{\min}}^{(1)}\| = o_p(1)$ for $j \in G_0^{op}$. Under some conditions, we can select the index $j_{\min} \in G^{op}$ with probability of one such that $\|\hat{\Sigma}_{1j_{\min}}^{-1/2}\theta_{j_{\min}}^{(1)}\| = o_p(1)$. Thus, the FWER can be asymptotically controlled at the prespecified level. In the case that the alternative hypothesis is true, we prove that the ToMin test is asymptotically unbiased and the power converges to one under some conditions.

By applying the ToMax and ToMin tests, we can simultaneously perform hypothesis testing for problems (1) and (2), termed as two directional Simultaneous Inference (TOSI).

2.2. TOSI Inference with $L > 1$

When $L = 1$, we randomly split the data $\{z_i\}_{i=1}^{n}$ into two parts for once, which may cause a loss of efficiency of inference. We consider an improvement that uses a multi-split method, that is, $L > 1$, to make full use of data. Since the multi-split testing method for $H_{0, G^0}$ is the same as that of testing $H_{0, G^{op}}$, we only introduce the method for $H_{0, G^0}$. Specifically, we repeat the data splitting $L$ times and obtain p-values $\{p_{\max}, l = 1, \ldots, L\}$ based on $\{\hat{T}_{l\max}, l = 1, \ldots, L\}$ for problem (1). Through Theorem 1, it asymptotically holds

$$P(\hat{T}_{l\max} \leq u) = u \text{ under } H_{0, G^0}. \quad (3)$$

Illustrated by the idea in Romano and DiCiccio (2019), we propose a testing procedure by defining a rule that we reject $H_{0, G^0}$ if at least $k_{\max}$ out of the $L$ p-values are less than or equal to $\gamma$, where $0 < k_{\max} \leq L$ and $\gamma \in (0, 1)$. According to Markov's inequality, we have

$$P(\text{reject } H_{0, G^0}) = \left(\sum_{l=1}^{L} 1_{\hat{T}_{l\max} \leq \gamma}\right) \geq k_{\max} \leq \frac{P(\hat{T}_{l\max} \leq \gamma)}{k_{\max}}. \quad (4)$$

where $1_{\hat{T}_{l\max} \leq \gamma}$ is the indicator function. Then, combining (3) and (4), it asymptotically holds that $P(\text{reject } H_{0, G^0}) \leq \frac{\gamma}{\alpha}$ under $H_{0, G^0}$, where $r = \frac{k_{\max}}{L}$. Therefore, if we choose $\gamma$ and $r$ such that $\frac{\gamma}{\alpha} = \alpha$, then the Type 1 error is asymptotically controlled below level $\alpha$. Interestingly, we can regard the L times test to be a multiple test on the same null hypothesis $H_{0, G^0}$. When $H_{0, G^0}$ is true, the family wise error rate (FWER) is equal to $P(k_{\max} \geq 1)$ which is the Type 1 error when $r = 1/L$. This is similar with the
Bonferroni correction, which leads us to consider the more powerful method called ToMax(L) by using Bonferroni-Holm (BH) procedure:

Let \( k_{\text{max}} \) be the number of BH-adjusted p-values that are less than \( \alpha \). We reject \( H_{0,G^p} \) if the number of rejections \( k_{\text{max}} \geq 1 \).

The validity of ToMax(L) is ensured by Theorem 3. Clearly, ToMax is a special case of ToMax(L) with \( L = 1 \).

### 3. Theoretical Properties

We now investigate the statistical properties of the TOSI test. Recall \( G_t^\ast = \{ j \in G^p : \theta_j \neq 0 \}, \ G_0^\ast = \{ j \in G^p : \theta_j = 0 \} \), \( G_t^{\text{no}} = \{ j \in G^{\text{no}} : \theta_j \neq 0 \}, \ G_0^{\text{no}} = \{ j \in G^{\text{no}} : \theta_j = 0 \} \) and denote \( G_1 = G_t^\ast \cup G_0^{\text{no}}, \) the nonzero index set, and \( s = |G_1| \). \( a_n \gg b_n \) implies that \( a_n \) dominates \( b_n \) in order. We use \( c \) to represent general positive constant which may be different in different places.

#### 3.1. Conditions and Explanation

We require some conditions for the theoretical properties displayed in Theorems 1–3.

(A1) For each \( j, \sqrt{n}(\hat{\theta}_j - \theta_j) \overset{d}{\rightarrow} N(0, \Sigma_j) \), where \( \Sigma_j \) can be estimated consistently by \( \hat{\Sigma}_j \).

(A2) \( \max_{j \in G_t^\ast} \| \hat{\theta}_j \| \gg \max_{j \in G_0^\ast} \| \theta_j \| \) if \( G_t^\ast \neq \emptyset \).

(A3) \( \min_{j \in G_0^{\text{no}}} \| \hat{\theta}_j \| \gg \min_{j \in G_0^\ast} \| \theta_j \| \) if \( G_0^{\text{no}} \neq \emptyset \).

(A4) \( \lim_{n \to \infty} \inf_{\theta \in \Theta} \| \sqrt{n} \Sigma_j^{-1/2} \theta_j \| > c > 0 \).

Conditions (A1)–(A4) are weak and easily satisfied. Condition (A1) ensures each population parameter \( \theta_j \) has asymptotically normal estimator \( \hat{\theta}_j \). In fact, this condition can be relaxed to any known asymptotic distribution. The relationship between \( p \) and \( n \) is implicitly contained in Conditions (A2)–(A3). For example, in the case with the cardinalities of \( G^p \) and \( G^{\text{no}} \) being the same order of \( p \), that is, \( |G^p| = O(p) \) and \( |G^{\text{no}}| = O(p) \), it can be shown that \( \max_{j \in G_t^\ast} \| \hat{\theta}_j \| = O_p(\sqrt{\frac{\ln(p)}{n}}) \) and \( \max_{j \in G_0^\ast} \| \theta_j - \hat{\theta}_j \| = O_p(\sqrt{\frac{\ln(p)}{n}}) \) under some tail probability restrictions.

Coupling with \( \max_{j \in G_t^\ast} \| \theta_j \| \gg \sqrt{\frac{\ln(p)}{n}} \) due to \( \max_{j \in G_t^\ast} \| \hat{\theta}_j \| \gg \max_{j \in G_t^\ast} \| \theta_j \| \) and \( \min_{j \in G_0^{\text{no}}} \| \theta_j \| \gg \min_{j \in G_0^\ast} \| \hat{\theta}_j \| \) (Condition (A2) holds if \( n \gg \ln(p) \)). Similarly, Condition (A3) holds if \( n \gg \ln(s) \) and \( \min_{j \in G_0^{\text{no}}} \| \theta_j \| \gg \sqrt{\frac{\ln(s)}{n}} \). Condition (A4) is a requirement for the lower bound of signals that is used to prove the unbiasedness of TOSI test. We give two examples to explain Conditions (A1)–(A4), especially for (A2) and (A3).

#### Example 1 (High-dimensional mean models)

In high-dimensional mean models, \( \theta_j = E(z_{ij}) \) with \( q = 1 \), where \( z_{ij} \) is the \( j \)th component of \( z_i \) and \( \inf_j \var{ \hat{\theta}_j } = \sigma > 0 \). We can choose \( \hat{\theta}_j \overset{\text{B.1}}{\rightarrow} \hat{\theta}_j \). Then Conditions (A1)–(A4) are satisfied if: (B1) there exist \( r_1 > 0 \) and \( r_2 > 0 \) such that \( P(|z_{ij}| > t) \leq \exp(-t/r_2)^{r_1} \) for any \( t > 0 \) and \( j \), and (B2): \( \inf_{j \in G_1} | \theta_j | \gg \sqrt{\frac{\ln(p)}{n}} = o(1) \).

Specifically, Condition (B1) ensures the existence of a moment at any order, which leads to Condition (A1) by the central limit theorem. Furthermore, since \( \max_{j \in G_1} | \theta_j | \gg \sqrt{\frac{\ln(p)}{n}} \) and \( \max_{j \in G_0^\ast} | \theta_j | = O_p(\sqrt{\frac{\ln(p)}{n}}) \), Condition (A2) holds. Then, note that \( \min_{j \in G_0^\ast} | \theta_j | \geq \min_{j \in G_1} | \theta_j | - \max_{j \in G_0^\ast} | \theta_j - \theta_j | \). by applying Conditions (B1) and (B2), we have \( \min_{j \in G_0^\ast} | \theta_j | \gg \sqrt{\frac{\ln(p)}{n}} \). Thus, Condition (A3) holds because \( \min_{j \in G_0^\ast} | \theta_j | \) cannot exceed the order \( \sqrt{\frac{\ln(p)}{n}} \). Finally, \( \inf_j \var{ \hat{\theta}_j } > c > 0 \) implies (A4).

The minimum signal assumption (B2) seems stringent. Actually, existing sample-splitting testing methods, by primarily focusing on testing \( H_{0,G^p} \), all required a similar assumption to ensure that the error rate in the variable selection step is ignorable. Otherwise, the Type I error in the inference step cannot be well controlled. For example, Wasserman and Roeder (2009) first proposed a single-split method to obtain the \( p \)-values of regression coefficients in high-dimensional sparse linear regression models, assuming a similar condition. Meinshausen, Meier, and Bühlmann (2009) improved the single-split method by proposing a multi-split method, and assumed a sure screening property: \( \lim_{n \to \infty} P(S \supseteq S) \to 1 \), which is more stringent than the minimum signal assumption (Fan and Lv 2008). Zhang and Cheng (2017) proposed a single-split method, and also required the sure screening property for valid inference.

#### Example 2 (High-dimensional sparse linear regression models)

There exist many classical inference methods for the sparse linear model, including the sample-splitting-based testing methods introduced in Section 1 and non-sample-splitting-based testing methods, where the most popular method is the bias-correction-based method, that is, LASSO-type correction (Zhang and Zhang 2014; De Geer et al. 2014) and ridge-type correction (Bühlmann 2013), that obtains the \( p \)-value for each regression coefficient and may lack power due to the strictness of FWER methods. In addition, Meinshausen (2015) tested a specified group of regression coefficients based on \( \ell_1 \)-norm but required a constraint Gaussian error assumption. In contrast to TOSI, these methods are limited to handle the problem (1). As another research line, Lee et al. (2016) proposed post-selection inference approaches focusing on the confidence interval of each coefficient in the best linear approximation to \( E(y|x) \) given a subset of selected covariates and served a different purpose from TOSI and other aforementioned methods.

In the high-dimensional sparse linear regression model, \( Y = X\theta + \varepsilon \), where \( Y = (y_1, \ldots, y_n)^T \) is a response vector, \( X = (x_1, \ldots, x_n)^T \) is an \( (n \times p) \)-dimensional covariate matrix, error \( \varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T \) with \( E(\varepsilon_i) = 0 \) and \( \var{\varepsilon_i} = \sigma^2 \) is independent of \( X \), and an unknown regression vector \( \theta = (\theta_1, \ldots, \theta_p)^T \).

Suppose \( X \) has iid rows with mean zero and covariance matrix \( \Omega = (\omega_{ij}) \). In this case, we have data \( z_i = (y_i, x_i^T)^T, i = 1, \ldots, n \) with \( q = 1 \). We can choose \( \hat{\theta} \) to be the de-biased LASSO estimator in Zhang and Zhang (2014) and De Geer et al. (2014).

Denote \( \Omega^{-1/2} \hat{\Omega} \hat{\theta} = (\hat{\theta}_j)_{j=1}^p. \) Recall \( z = \{ j : \theta_j \neq 0 \} \) and let \( s_j = |\{ 1 \leq k \leq p : \theta_{jk} \neq 0, k \neq j \}| \). For Example 2, Conditions (A1)–(A4) hold if: (C1) \( x_i \) is a sub-Gaussian random vector; (C2)
the minimum eigenvalue $\lambda_{\min}$ of $\Omega$ satisfies that $\lambda_{\min} > c$ and \[ \max_j \omega_{jj} < C; \] (C3) \[ \frac{\ln(p)}{\sqrt{n}} = o(1) \] and \[ \max_j \frac{\ln(p)}{\sqrt{n}} = o(1); \] (C4) $\epsilon_i$ is a sub-Gaussian random variable; and (C5) min$_j \in G_1 \| \theta_j \| > \sqrt{\frac{\ln(p)}{n}}$.

Condition (C1) is similar to Assumption 2.1 in Zhang and Cheng (2017) and (B1) in De Geer et al. (2014) to control the tail behavior of covariates. Condition (C2) is the same as Assumption 2.2 in Zhang and Cheng (2017), which is used to upper bound the spectral norm of the precision matrix $\Theta$.

Condition (C3) is a standard sparsity assumption for regression coefficients $\Theta$ and the precision matrix $\Theta$, which is also assumed in Theorem 2.4 in De Geer et al. (2014). Condition (C4) constrains the tail behavior of error term, which is also used in the Assumption 2.3(i) in Zhang and Cheng (2017). Condition (C5) is a minimum signal strength assumption to ensure the error rate in the variable selection step is ignorable and similar condition can be found in Assumptions (A2) and (A3) in Wasserstein and Roeder (2009) and assumption (A1) in Meinshausen, Meier, and Bühlmann (2009) since they also adopted the sample-splitting strategy for inference.

### 3.2. Theoretical Results

Denote $\theta_{G^0} = \{ \theta_j : j \in G^0 \}$ and $\theta_{G^{no}} = \{ \theta_j : j \in G^{no} \}$. Let $\beta_{l_{\max}}(\theta_{G^0}) = P(\hat{T}_{\max} > \chi^2_{1-\alpha}(q))$ and $\beta_{R_{\min}}(\theta_{G^{no}}) = P(\hat{R}_{\min} > \chi^2_{1-\alpha}(q))$ be the power functions of ToMax and ToMin tests, respectively, then we present two theorems that ensure the validity of ToMax and ToMin tests.

**Theorem 1.** Suppose that Conditions (A1), (A2), and (A4) are satisfied, we have,

(i) Under $H_{0,G^0} : \theta_j = 0, \forall j \in G^0$, $\hat{T}_{\max}$ is asymptotically distributed as $\chi^2(q)$.

(ii) Under $H_{1,G^0} : \theta_j = 0, \forall j \in G^0$, for a prefixed significance level $\alpha$, $\beta_{l_{\max}}(\theta_{G^0}) \geq \alpha$, when $n \to \infty$. In particular, if inf$_j \in G_1 \| \sqrt{n} \Sigma_j^{-1/2} \theta_j \| \to \infty$, then $\beta_{l_{\max}}(\theta_{G^0}) \to 1$.

**Theorem 2.** Suppose that Conditions (A1), (A3), and (A4) hold, we have

(i) Under $\hat{H}_{0,G^{no}} : \theta_j = 0, \forall j \in G^{no}$, $\hat{R}_{\min}$ is also asymptotically distributed as $\chi^2(q)$.

(ii) Under $\hat{H}_{1,G^{no}} : \theta_j = 0, \forall j \in G^{no}$, then for a prefixed significance level $\alpha$, $\beta_{R_{\min}}(\theta_{G^{no}}) \geq \alpha$, when $n \to \infty$. In particular, if inf$_j \in G_1 \| \sqrt{n} \Sigma_j^{-1/2} \theta_j \| \to \infty$, then $\beta_{R_{\min}}(\theta_{G^{no}}) \to 1$.

We also perform simulation studies to verify the asymptotically distributions of $\hat{T}_{\max}$ and $\hat{R}_{\min}$ in Theorems 1 and 2. Figures 1(a) and (b) show the QQ plots of the empirical distribution under high-dimensional sparse linear regression models, see Experiment 1 in Section 6.1, which confirms the conclusion in Theorems 1 and 2.

We formally present the validity of ToMax($L$) through the following theorem.

**Theorem 3.** Under Conditions (A1), (A2), and (A4), for testing procedure ToMax($L$), it asymptotically holds that $\Pr(\text{reject } H_{0,G^0} | H_{0,G^0} \text{ is true}) \leq \alpha$.

**Remark 4.** It is worth to be noted that Theorem 3 automatically produces a combined $p$-value except of giving a decision of rejection or acceptance. In particular, we denote the BH-correction $p$-values to $\hat{p}_{l_{\max}}^{\text{adj}}, i = 1, \ldots, L$, then the final combined $p$-value is $\hat{p}_{\max}^{\text{adj}} = \min_i \hat{p}_{l_{\max}}^{\text{adj}}$. If $\hat{p}_{\max}^{\text{adj}} < \alpha$, then we reject the null hypothesis.

Similarly, we can obtain multi-split version and related theoretical properties for testing $\hat{H}_{0,G^{no}}$, named by ToMin($L$). ToMax($L$)/ToMin($L$) with $L > 1$ is a conservative method in the sense that it controls the Type I error to not exceed the nominal level $\alpha$ rather than equal to $\alpha$. However, with multiple splits, each individual can be used for both Stages I and II if $L$ is sufficiently large, thus, the data are used more efficiently and the power of ToMax($L$)/ToMin($L$) increases with increasing $L$, which is confirmed by our extensive simulation studies for high-dimensional sparse regression models, sparse latent factor models, and high-dimensional mean models in Section 6 and supplementary materials. However, when $L$ is sufficiently large, sufficient information has been used such that continually increasing $L$ cannot improve the power. Hence, in practice, we
can choose a larger $L$ so that a stable conclusion can be obtained based on the resulting $p$-value.

4. Example 3. Sparse Latent Factor Models

In this section, we introduce another example to illustrate the application of TOSI method to latent factor models. The simultaneous inference in latent factor model can be used to select the important variables contributing to latent factors, such as cell-type-relevant genes in the area of genomics. Taking the single cell RNA sequencing (scRNA-seq) data as an example, the scRNA-seq data are measured on tens/hundreds of thousands of cells and tens of thousands of genes. The normalized data can be modeled by a linear factor model since cells often occupy a limited number of cell types (Hou et al. 2020), where $\mathbf{h}_i$ is interpreted as the cell-type-related latent factors and $\mathbf{b}_j = 0$ means gene $j$ has no expression in all considered cell types. By selecting the genes with $\mathbf{b}_j \neq 0$, we achieved the variable (genes) selection for the downstream analyses.

Suppose that the observations $\mathbf{x}_i = (x_{i1}, \ldots, x_{iq})^T$ are correlated because they share a latent factor $\mathbf{h}_i = (h_{i1}, \ldots, h_{iq})^T$ with $q \ll p$. We consider the model,

\[
x_{ij} = \begin{cases} 
\mathbf{b}_j^T \mathbf{h}_i + u_{ij}, & j \in J, \\
u_{ij}, & j \notin J,
\end{cases} 
\]

(5)

where $J = \{ j : \mathbf{b}_j \neq 0, j = 1, \ldots, p \}$, $J^c$ is the complement of $J$, $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_p)^T$ is a $p \times q$ deterministic matrix ($\mathbf{b}_j = 0$ if $j \notin J^c$), $\mathbf{u}_i = (u_{i1}, \ldots, u_{ip})^T$ is an error term independent of $\mathbf{h}_i$, $E(\mathbf{u}_i) = 0$ and $\text{var}(\mathbf{u}_i) = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2)$. By identifying the set $J$ in (5), we can investigate which features contribute to the latent factors $\mathbf{h}_i$. Denote $\mathbf{b}_j = (b_{j1}, \ldots, b_{jq})^T$; we further allow some of $b_{jk}, k = 1, \ldots, q$ for $j \in J$ to be zero. By identifying $b_{jk}$ to be either zero or nonzero, we can investigate whether variable $j$ is associated with the $k$th component of the latent factor. Thus, we can explicitly explore the path between the high-dimensional observed variables and the latent factor to achieve the interpretability of the latent factor. The high-dimensional sparse latent factor model (5) is substantially different from the manifest models in Examples 1–2 because of the unobserved latent factors $\mathbf{h}_i$.

Denote $\mathbf{H} = (\mathbf{h}_1, \ldots, \mathbf{h}_n)^T$. Since $\mathbf{b}_j^T \mathbf{h}_i = (\mathbf{M}^T \mathbf{b}_j)^T(\mathbf{M}^{-1} \mathbf{h}_i)$ for any invertible matrix $\mathbf{M} \in \mathbb{R}^{q \times q}$, model (5) is not identifiable. To make it identifiable, similar to Bai and Ng (2013) and Jiang, Ma, and Wei (2019), we assume (E1) $n^{-1} \mathbf{H}^T \mathbf{H} = \mathbf{I}_q$; (E2) $\mathbf{B}^T \mathbf{B}$ is diagonal with decreasing diagonal elements and the first nonzero element in each column of $\mathbf{B}$ is positive. Furthermore, for simplicity we assume that the means of $x_{ij}$s and $\mathbf{h}_i$s have already been removed, namely, $E(x_{ij}) = 0$ and $E(\mathbf{h}_i) = 0$.

We propose a new Non-Iterative Two-Step estimation (NITS) in Appendix C.1 of supplementary materials for estimating $\mathbf{H}$ and $\mathbf{B}$, which obtains a sparse solution of $\mathbf{B}$. The resulting estimators $\hat{\mathbf{H}}$ and $\hat{\mathbf{B}}$ have closed forms; hence, the computation and implementation are simple. The large sample properties, including the identifiability of models and the oracle property of the NITS estimators, and their proofs are deferred to the supplementary materials.

4.1. Inference on $\mathbf{B}$

We are interested in making inferences on two aspects of two problems: (a) whether $\mathbf{b}_j$ identified as zero is indeed zero; whether $\mathbf{b}_j$ identified as nonzero are significantly different from zero; (b) the same problems for $b_{jk}$. Therefore, we can fully identify which variables are associated with the latent factor $\mathbf{h}_i$ to improve the interpretability. The inference on entries $b_{jk}$ is similar with that for $b_j$ and is omitted here.

Using the estimation and variable selection described in the supplementary materials, we can partition all $p \mathbf{b}_j$s into two groups: $G^0$, the all unimportant index set and $G^{0,0}$, the active index set. We are interested in the following hypotheses,

\[
H_{0, G^0} : \mathbf{b}_j = 0, \forall j \in G^0 \quad \text{versus} \quad H_{1, G^0} : \mathbf{b}_j \neq 0, \exists j \in G^0, \quad (6)
\]

\[
H_{0, G^{0,0}} : \mathbf{b}_j = 0, \exists j \in G^{0,0} \quad \text{versus} \quad H_{1, G^{0,0}} : \mathbf{b}_j \neq 0, \forall j \in G^{0,0}. \quad (7)
\]

To obtain the test statistics for (6) and (7), we consider $\tilde{\mathbf{B}} = (\mathbf{b}_1, \ldots, \mathbf{b}_p)^T$ from

\[
(\tilde{\mathbf{H}}, \tilde{\mathbf{B}}) = \arg\min_{\mathbf{B}, \mathbf{H}, \lambda} \| \mathbf{X} - \mathbf{HB}^T \|_F^2. \quad (8)
\]

Under Conditions (D1)–(D5) in Section 4.2, we can show that $\tilde{\mathbf{b}}_j$ satisfies $\sqrt{n}(\tilde{\mathbf{b}}_j - \mathbf{b}_j) \sim N(0, \sigma_j^2 \mathbf{I}_q)$, where $\sigma_j^2$ is the variance of $u_{ij}$ and can be estimated by $\hat{\sigma}_j^2 = n^{-1} \sum_{i=1}^{n} (x_{ij} - \tilde{h}_j \tilde{\mathbf{b}}_j)^2$. See Lemma 2 in the supplementary materials. The asymptotic normality of $\tilde{\mathbf{b}}_j$ is also given in Bai and Ng (2013) in the context of high-dimensional panel data. With $\hat{\theta}_j$ and $\hat{\beta}_j$ replaced by $\mathbf{b}_j$ and $\tilde{\mathbf{b}}_j$, respectively, the TOSI test described in Section 2.1 can also be obtained for problems (6) and (7).

4.2. Asymptotic Properties

We now establish the validity of the proposed TOSI procedure. To establish the asymptotic properties, we need the regularity Conditions (D1)–(D6) given in Appendix C.1.3 of supplementary materials. Conditions (D1)–(D5) yields the asymptotic normality of $\mathbf{b}_j$, which implies Condition (A1). Conditions (A2) and (A3) can be proved by considering Condition (D6). Finally, Condition (D6.3) implies (A4). Then we give the following theorem whose proofs are deferred to Appendix C.2 in supplementary materials.

Theorem 4. Under Conditions (D1)–(D6), the same conclusions in Theorems 1–3 can be obtained.

5. Application to the Penalty Parameter Selection

As mentioned in Section 1, the TOSI method can select the penalty parameter $\lambda$ so that the resulting sets of zeros ($G^0(\lambda)$) and nonzeros ($G^{0,0}(\lambda)$) are statistically insignificant and significant, respectively. Hence, the $\lambda$ selected using TOSI is meaningful to identify important and unimportant variables. Here, we take the sparse linear regression model as an example to illustrate the selection of penalty parameter based on TOSI. We first randomly split the data $\{\mathbf{x}_i\}_{i=1}^{n}$ into two parts $\mathcal{D}$ and $\mathcal{D}_s$ for inference and variable selection, respectively. We design a bisection method for searching $\lambda$ according to the inference.
results from TOSI. To start the searching, we set an initial searching domain \( H_0 = [\lambda_{\text{lower}}^{(1)}, \lambda_{\text{upper}}^{(1)}], \) and use the K-fold cross-validation (CV) to select an initial penalty parameter \( \lambda^{(1)} \) in \( H_0 \), the estimated sets of zeros \( G^o(\lambda^{(1)}) \) and non-zeros \( G^{no}(\lambda^{(1)}) \) using LASSO based on the sample \( D_s \). In the \( h \)th iteration, we update the inference and estimators as follows.

Step a. Based on sample \( D_2 \), we use TOSI to test \( H_{0,G^o}\) : \( \theta_j = 0, \forall j \in G^o(\lambda^{(1)}) \) and \( \tilde{H}_{0,G^o}\) : \( \theta_j = 0, \exists j \in G^{no}(\lambda^{(1)}) \), resulting in four cases: (a) both are rejected; (b) both are accepted; (c) the former is accepted and the latter is rejected; and (d) the former is rejected and the latter is accepted. Due to the variable selection consistency based on LASSO, (d) is rare and ignored here.

- Case (a) implies there exists nonzero elements in \( G^o(\lambda^{(1)}) \), thus, we move \( \lambda^{(1)} \) to a smaller one \( \lambda^{(1+1)} = (\lambda^{(1)} + \lambda_{\text{lower}}^{(1)})/2 \), and set \( \lambda_{\text{upper}}^{(1)} = \lambda^{(1)} \) and \( \lambda_{\text{lower}}^{(1)} = \lambda^{(1)} \).
- Case (b) indicates there exists zero index in \( G^{no}(\lambda^{(1)}) \), then we move \( \lambda^{(1)} \) to a larger one with \( \lambda^{(1+1)} = (\lambda^{(1)} + \lambda_{\text{upper}}^{(1)})/2 \), and set \( \lambda_{\text{lower}}^{(1+1)} = \lambda^{(1)} \) and \( \lambda_{\text{upper}}^{(1+1)} = \lambda^{(1)} \).
- In the case (c), the searching process is stopped and \( \lambda^{(1)} \) is regarded as optimal.

Step b. Based on sample \( D_s \), we apply LASSO with the penalty parameter \( \lambda^{(1+1)} \) to determine sets \( G^o(\lambda^{(1+1)}) \) and \( G^{no}(\lambda^{(1+1)}) \).

Repeat the iteration until the searching is stopped. The simulation studies in Section 6.1 show that the bisection method based on TOSI performs better than the existing methods for tuning penalty parameter, including K-fold CV, AIC, BIC, and scaled LASSO, as shown in Table 2.

### 6. Numerical Studies

In this section, we conduct simulation studies to assess the finite-sample performance of the proposed TOSI method in comparison with the existing simultaneous inference methods and \( p \)-value-adjusted methods. For testing \( H_{0,G^o} \), if there exists an adjusted \( p \)-value less than \( \alpha \), then \( p \)-value-adjusted methods reject it. We use testing size and power to evaluate the performance of the inference methods. The resulting size and power are obtained based on the empirical average from 500 repeats. We also investigate the performance of the TOSI in guiding the selection of penalty parameters.

We considered three experiments corresponding to three models which are high-dimensional mean models, sparse linear regression models and sparse factor models, with \( G^o \) and \( G^{no} \) set as follows. For evaluating the testing size, we set \( G^o \) to \( G_1 = \{p-1, p\}, G_2 = \{p/2, \ldots, p\} \) and \( G_3 = \{s+1, \ldots, p\} \) and \( G^{no} \) to \( G_1 = \{p-1, p\}, G_2 = \{s+1, \ldots, p\} \) and \( G_3 = \{1, \ldots, p\} \). For evaluating the testing size, we set \( G^o \) to \( G_4 = \{2, s+1\}, G_5 = \{3, s+1, \ldots, p\} \) and \( G_6 = \{3, 4, s+1, \ldots, p\} \) and \( G^{no} \) to \( G_7 = \{1, 2\}, G_8 = \{1, \ldots, 4\} \) and \( G_9 = \{1, \ldots, s\} \). To save space, the results of high-dimensional mean models are deferred to Appendix D in supplementary materials.

#### 6.1. Experiment 1: High-Dimensional Sparse Regression Models

We consider a high-dimensional regression model with the same setting as Zhang and Cheng (2017). In detail, \( y_i = x_i^T \beta + \epsilon_i, \beta = (\beta_1, \ldots, \beta_{50}), i = 1, \ldots, n = 50 \) or 100, where \( x_i \sim N(0, \Sigma^o) \) with \( \Sigma^o = 0.8^{j-1}I, \epsilon_i \sim t(4)/\sqrt{2}, t(4) \) is Student’s \( t \)-distribution with degrees of freedom equal to four. We set \( \beta_j = \rho z \) for \( j \leq s \) and \( \beta_j = 0 \) for \( j > s \) with \( s = 5 \), where \( z \) is a random variable following uniform distribution \( U[0, 2] \). We consider three signal-noise-ratio settings by taking \( \rho = 0.3, 0.5 \) and 0.8, respectively. \( \beta \) is fixed after being generated.

TOSI adopts the debiased estimator (Zhang and Zhang 2014) for high-dimensional sparse regression models in Stages I and II, where 10-fold cross-validation is used to select the tuning parameters in the nodewise LASSO; see Zhang and Zhang (2014) for details.

To benchmark the testing performance of TOSI in testing \( H_{0,G^o} \), we compare it with five methods: (a) one-step procedure using test statistic max \( |\beta_j - \hat{\beta}_j| \) proposed in Zhang and Cheng (2017), denoted as ZC1-17; (b) three-step procedure based single sample splitting in Zhang and Cheng (2017), denoted as ZC3-17; (c) Benjamini-Yekutieli \( p \)-value-adjusted method (Benjamini and Yekutieli 2001) based on the \( p \)-values obtained from Zhang and Zhang (2014), denoted as ZZ-14; (d) \( p \)-values corrected method based on the sample multi-splitting approach in Meinshausen, Meier, and Bühlmann (2009), denoted as MMB-09; (e) Holm \( p \)-value-adjusted method (Holm 1979) based on the \( p \)-values obtained from Bühlmann (2013), denoted as B-13. The results for different settings from 500 replicates are presented in Table 1. We conclude (a) our ToMax/ToMin can asymptotically control the Type I error at the nominal level if sample size is adequate. (b) ToMax(L)/ToMin(L) has a conservative size below the significance level 0.05, which is consistent with Theorem 3, and has higher power than ToMax/ToMin by increasing \( L \). (c) The difference in power between ToMax(L) and ZC1-17/ZC3-17 decreases as \( L \) increases, and the ToMax(8) outperforms ZC1-17 and ZC3-17. (d) ToMax(L) outperforms the three \( p \)-values-adjusted methods (ZZ-14, B-13 and MMB-09) in terms of testing size and power, especially for testing sets with smaller cardinalities. The \( p \)-values-adjusted methods control Type I error too conservatively and hence have lower testing powers. ZC1-17 somewhat fails to control the Type I error \( (0.07 \sim 0.13) \) when sample size is not sufficiently large. (e) Larger \( n \) or signal-noise-ratio improves the size and power due to the stronger signal. The results for different signal-noise-ratio settings are referred to Table S2 in supplementary materials.

In addition to linear regression, we also showcase the application of TOSI on a nonlinear sparse logistic regression model, see Table S3 for results in supplementary materials, which suggests that TOSI outperforms the existing methods and similar conclusions as those for sparse linear regression models can be obtained.

TOSI successfully guided the selection of the penalty parameters. To illustrate this result, we compare TOSI(\( L = 1 \)) with the cross-validation based on LASSO regression (CV LASSO), BIC based on LASSO (BIC), AIC based on LASSO (AIC) and scaled LASSO for Experiment 1 with \( s = 3 \) and \( \rho = 0.3 \). First, we generate an independent sample \( D_s \) with sample size...
variables are included in the selected model (IN), and the percentage of occasions when exactly select important variables (CS) over 500 replications. We observe that both methods can select important variables, however, CV LASSO, AIC, and scaled LASSO usually over-selects the variables and exactly selects important variables (CS) at a frequency of 10%, 36%, and 8% on average, respectively. By contrast, TOSI achieves the highest frequency of 95% on average. Thus, TOSI can more accurately identify the model structure. Finally, by setting three different nominal levels ($\alpha = 0.1, 0.05,$ and 0.01), we verify that TOSI is robust to the prespecified nominal level in identifying the model structure; see Table S4 in supplementary materials.

### Table 1: Results of Experiment 1: high-dimensional sparse linear regression models.

| Method      | $n = 50$ | $n = 100$ |
|-------------|----------|-----------|
|             | $G_{11}$ | $G_{12}$ | $G_{13}$ | $G_{11}$ | $G_{12}$ | $G_{13}$ |
| Size        |          |          |          |          |          |          |
| ToMax(1)    | 0.030    | 0.010    | 0.015    | 0.050    | 0.045    | 0.045    |
| ToMax(2)    | 0.015    | 0.030    | 0.040    | 0.030    | 0.035    | 0.060    |
| ToMax(5)    | 0.015    | 0.020    | 0.045    | 0.030    | 0.045    | 0.045    |
| ToMin(8)    | 0.000    | 0.020    | 0.040    | 0.025    | 0.040    | 0.015    |
| ZC-17       | 0.070    | 0.100    | 0.130    | 0.050    | 0.065    | 0.070    |
| ZC-17(1/5)  | 0.029    | 0.037    | 0.039    | 0.049    | 0.041    | 0.031    |
| ZC-17(1/3)  | 0.031    | 0.031    | 0.034    | 0.052    | 0.026    | 0.030    |
| ZZ-14       | 0.001    | 0.010    | 0.005    | 0.025    | 0.005    | 0.000    |
| B-13        | 0.000    | 0.002    | 0.006    | 0.004    | 0.004    | 0.006    |
| MMB-09      | 0.000    | 0.000    | 0.000    | 0.000    | 0.000    | 0.000    |

#### 6.2. Experiment 2: High-Dimensional Latent Factor Models

Let $b_k$ be the $k$th column of $B = (b_1, \ldots, b_p)^\top$, $p = 150$. To construct the sparse matrix $B$, for $k < q$, we set the $k$th component of $b_k$ to be nonzero for $j \in A_k = \{(k-1)s_0 + 1, \ldots, ks_0\}$, and the components of $b_q$ in locations $A_k = \{(q-1)s_0 + 1, \ldots, s\}$ to be nonzeros, where $s_0 = \lfloor s/q \rfloor$ and $s$ are the number of $b$s such that $b_k \neq 0$, $|x|$ is the largest integer less than $x$. Hence, $f = \bigcup_{k=1}^q A_k = [s]$ and the $k$th factor is contributed by the variables $x_j$s with $j \in A_k$. We set $s = \lceil 3p/4 \rceil$ and randomly generate the nonzeros of $b_k$ from $\rho (1.5 - 0.24(k-1) + z)$, where $z \sim U[0,1]$ and $\rho = 0.3$ controls the strength of the signal. Clearly, $B$ satisfies the identifiability condition (E2). We independently generate $h_i$, $i = 1, \ldots, n$ from a multivariate normal distribution with mean zero and covariance matrix $\sigma_{ij} = 0.5^{i-j}$. Then, we center and normalize $H = (h_1, \ldots, h_n)^\top$ so that $H$ satisfies the identifiability condition (E1). We consider $q = 1$. The results for the NITs estimator are deferred to Appendix C.3 of supplementary materials. Table 3 shows testing size and power of TOSI under various settings, the similar conclusion with those for Table 1 can be drawn. In Table 2, we also compare ToMax(L) with a newly developed $p$-values-adjusted method (UY-21) by Uematsu and Yamagata (2021). From Table 3, we observe UY-21, with FDR 0.05, controls Type I error too conservative and has a lower power than ToMax(L).

### Table 2: Results of Experiment 1: high-dimensional sparse linear regression models.

| Method      | $n = 50$ | $n = 100$ |
|-------------|----------|-----------|
|             | $G_{11}$ | $G_{12}$ | $G_{13}$ | $G_{11}$ | $G_{12}$ | $G_{13}$ |
| Power       |          |          |          |          |          |          |
| ToMax(1)    | 0.115    | 0.115    | 0.210    | 0.215    | 0.250    | 0.400    |
| ToMax(2)    | 0.135    | 0.175    | 0.275    | 0.195    | 0.350    | 0.355    |
| ToMax(5)    | 0.180    | 0.200    | 0.355    | 0.250    | 0.405    | 0.620    |
| ToMax(8)    | 0.205    | 0.205    | 0.400    | 0.260    | 0.440    | 0.655    |
| ZC-17       | 0.170    | 0.200    | 0.350    | 0.200    | 0.230    | 0.585    |
| ZC-17(1/5)  | 0.206    | 0.146    | 0.340    | 0.264    | 0.140    | 0.500    |
| ZC-17(1/3)  | 0.208    | 0.170    | 0.336    | 0.248    | 0.168    | 0.464    |
| ZZ-14       | 0.150    | 0.100    | 0.255    | 0.275    | 0.160    | 0.500    |
| B-13        | 0.000    | 0.010    | 0.036    | 0.010    | 0.022    | 0.102    |
| MMB-09      | 0.005    | 0.120    | 0.295    | 0.035    | 0.190    | 0.466    |

#### 7. Real Data Analysis

In this section, we apply TOSI to a liquor sales dataset and a criminal dataset by using high-dimensional linear regression models, where the debiased estimator in Zhang and Zhang (2014) is used to construct the testing procedure.

### 7.1. Liquor Sales Data

TOSI is now applied to analyze a liquor sales dataset of Jiangsu province from one of China’s largest liquor companies. The purpose of the analysis is investigating factors that are associated with the monthly sales of liquor in Jiangsu province. The dataset includes $monthly sales y_t$ and covariates information for $n = 280$ observations in Jiangsu province from 2011 to 2018. After data preprocessing, there are 249 covariates which include four parts: (a) the company’s product information such as brand promotion and advertising investment, (b) brewing industry information such as monthly liquor yields and monthly beer...
yields, (c) economic information of related cities and towns such as per capita GDP, per capita disposable income and consumer price index, and (d) geographic information such as monthly average temperature and monthly average relative humidity. Log transformation is taken to response variable and all of covariates are standardized. A histogram of monthly sales is shown in Figure 2(a) and the final response variable is took logarithm of monthly sales (see Figure 2(b)). Then, we applied the TOSI method based on high-dimensional regression models to explore the influencing factors for sales of liquor.

First, we used lasso regression, implemented via the glmnet package in R, to roughly separate the important variables and unimportant variables based on the first 100 samples, where the penalty parameter \( \lambda_{opt} = 0.1866 \) is selected by 10-fold cross-validation; see Figure 2(c). Then, we obtained nine important variables and 240 unimportant variables.

With some abuse of notation, we denote the index set with unimportant variables as \( G^n \) and the index set with important variables as \( G^o \). Given \( G^n \), we used the rest samples to test whether there was nonzeros in \( G^o \) using ToMax, ZC1-17, \( p \)-values correction method based on ridge regression (Bühlmann 2013) (B-13), and \( p \)-values correction method based on Lasso regression (Zhang and Zhang 2014; De Geer et al. 2014) (ZZ-14) and all methods failed to reject the null hypothesis. The TOSI method could further test whether there were zeros in \( G^{no} \), but the other methods could not. The \( p \)-value of ToMin is 0.3018 and the adjusted \( p \)-values of ToMin(5) is 0.7546, indicating that the null hypothesis could not be rejected; that is, there might be zeros in \( G^{no} \). This is consistent with the simulation results that CV LASSO tended to over-select features. Therefore, we increased the value of penalty parameter by finely tuning within \{0.2239, 0.3172, 0.3545\} until the null hypothesis \( H_{0,G^{no}} \) was rejected and the null hypothesis \( H_{0,G^n} \) was not rejected. Finally, we identified six important variables, as presented in Table 4(a). Furthermore, we conducted another check whether the selected 243 variables were truly unimportant via ZC1-17, B-13, and ZZ-14, and none of these methods could reject the null hypothesis. Thus, the analysis results obtained via TOSI were more interpretable than those from CV LASSO.

Table 3. Results of Experiment 2: high-dimensional sparse factor models.

| Method | \( G_{11} \) | \( G_{12} \) | \( G_{13} \) | \( G_{14} \) | \( G_{15} \) | \( G_{16} \) |
|--------|--------------|--------------|--------------|--------------|--------------|--------------|
| ToMin(1) | 0.060 | 0.062 | 0.072 | 0.052 | 0.034 | 0.046 |
| ToMin(2) | 0.072 | 0.060 | 0.066 | 0.056 | 0.052 | 0.066 |
| ToMin(8) | 0.052 | 0.068 | 0.064 | 0.050 | 0.048 | 0.058 |
| ToMin(15) | 0.060 | 0.066 | 0.070 | 0.044 | 0.044 | 0.054 |
| ToMin(20) | 0.054 | 0.064 | 0.062 | 0.050 | 0.040 | 0.054 |
| UY-21 | 0.005 | 0.020 | 0.045 | 0.005 | 0.035 | 0.065 |
| ToMin(1) | 0.062 | 0.046 | 0.046 | 0.052 | 0.038 | 0.038 |
| ToMin(5) | 0.060 | 0.062 | 0.062 | 0.046 | 0.040 | 0.040 |
| ToMin(8) | 0.062 | 0.054 | 0.054 | 0.038 | 0.048 | 0.048 |
| ToMin(15) | 0.046 | 0.074 | 0.074 | 0.034 | 0.052 | 0.052 |
| ToMin(20) | 0.046 | 0.078 | 0.078 | 0.034 | 0.046 | 0.046 |
| ToMin(1) | 0.050 | 0.054 | 0.054 | 0.072 | 0.060 | 0.066 |
| ToMin(5) | 0.060 | 0.070 | 0.070 | 0.028 | 0.090 | 0.090 |
| ToMin(8) | 0.070 | 0.081 | 0.093 | 0.028 | 0.090 | 0.090 |
| ToMin(15) | 0.070 | 0.085 | 0.098 | 0.054 | 0.078 | 1.000 |
| ToMin(20) | 0.070 | 0.089 | 1.000 | 0.064 | 0.086 | 1.000 |

Table 4(b) presents the estimated coefficients for the six important covariates: the sales in the past months (SL_lag1 and SL_lag12), GDP from primary industry, mainly agriculture, in the last year (gdp1_lastyear), brand promotion expenses for the past six months (pptg_six_m), the expense in giving products as gifts to customers in the past 12 months (kq_twelve_m), and the number of transactions of the stock 600779 in the past 10 months (Stkcd600779_lag10), where 600779 is the stock code of another liquor company. From Table 4(b), we could draw following conclusions. First, the positive coefficients of the sales in the past months (SL_lag1 and SL_lag12) indicates that the larger the sales in the past, the larger the sales in the current month. This is consistent with intuition since the larger sales in the past could make customers trust this product more. Moreover, the coefficient of SL_lag1 is much greater than that of SL_lag12, which means SL_lag1 has greater influence on monthly sales. Second, GDP from primary industry, mainly agriculture, in last year (gdp1_lastyear) had negative effect on sales, which

![Figure 2](image URL)  
Figure 2. (a) and (b) Histograms of the target variable monthly sales and its log-transformation in liquor sales data; (c) 10-fold cross-validation error versus lambda in liquor sales data.
Table 4. Results for liquor sales data: (a) The results of the TOSI method under different penalty parameters with $L = 5$; (b) Estimates of coefficients for the six important variables, where $SL_{lag1}$ and $SL_{lag12}$ are the sales in the past one month and 12 months, respectively, $gdp1_{lastyear}$ is the GDP from primary industry, mainly agriculture, in the last year, $pptg\_six\_m$ is the brand promotion expense for the past six months, and $kq\_twelve\_m$ is the expense in giving products as gifts to customers in the past 12 months, and $Stkcd600779\_lag10$ is the number of transactions of the stock 600779 in the past 10 months, where 600779 is the stock code of another liquor company.

(a) The testing results

| $\lambda$   | $|G^0|$ | $|G^{0\circ}|$ | p-value ($H_0|G^0)$ | p-value ($\tilde{H}_0|G^{0\circ}$) |
|-------------|--------|---------------|----------------------|----------------------|
| 0.1866      | 240    | 9             | ToMax 0.5832          | ToMin 0.3018          |
| 0.2239      | 241    | 8             | ToMax(L) 0.5895       | ToMin(L) 0.7545       |
| 0.3172      | 243    | 6             | ToMax 0.2682          | ToMin 0.3699          |
| 0.3545      | 244    | 5             | ToMax(L) 0.2009       | ToMin(L) 0.3953       |
|             |        |               | ToMax(L) 0.0684       | ToMin(L) 0.0431       |

(b) The estimated coefficients of significant variables

| var. name | $SL_{lag1}$ | $SL_{lag12}$ | $gdp1_{lastyear}$ |
|-----------|------------|-------------|-------------------|
| coef. est.| 0.1615     | 0.0785      | $-0.0574$         |
| var. name | $pptg\_six\_m$ | $kq\_twelve\_m$ | $Stkcd600779\_lag10$ |
| coef. est.| 0.0331     | 0.1204      | $-0.0236$         |

Figure 3. (a) Histogram of the response variable monthly sales in criminal data; (b) 10-fold cross-validation error versus lambda in criminal data.

may be caused by the fact that areas with high agricultural output usually have low commercial operation ability. Third, the positive coefficient of brand promotion expenses for the past six months ($pptg\_six\_m$) shows that the higher the brand promotion expenses, the larger the sales. Porto et al. (2017) reported that promotional materials could generate a positive effect on product sales. Fourth, the positive coefficient of the expense in giving products as gifts to customers in the past 12 months ($kq\_twelve\_m$) suggests the expense in gift giving could improve the monthly sales since this gift-giving behavior can attract consumers’ interest of the product. Lastly, the number of transactions of the stock 600779 in the past 10 months ($Stkcd600779\_lag10$) has a negative coefficient that indicates the transactions of stocks of the competitor could impair the monthly sales.

7.2. Criminal Data

In this section, we analyze a criminal dataset (Redmond and Baveja 2002) to demonstrate the usefulness of TOSI. This dataset is collected from 200 communities within the United States and combines socio-economic data from the 1990 U.S. Census, law enforcement data from the 1990 U.S. LEMAS survey, and crime data from the 1995 FBI UCR. After data preprocessing, we removed variables with seriously missing values and zero variance and obtained 101 variables for each community. The response variable of interest is the total number of violent crimes per 100K population ($ViolentCrimesPerPop$), which describes the severity of crime in the community. A histogram of $ViolentCrimesPerPop$ is shown in Figure 3(a) and the final response variable is took logarithm of $ViolentCrimesPerPop$.

The remaining 100 covariates, including population of the community, mean number of people per household, percentage of males who are divorced, and percentage of kids aged 12–17 years in two-parent households, were used as predictors. We were interested in which variables had an impact on $ViolentCrimesPerPop$. To illustrate the application in high-dimensional regression models, we applied our TOSI method to solve this problem.

Similarly, we used lasso regression to roughly determine the important variables and unimportant variables based on the first 100 samples, where the penalty parameter $\lambda_{opt} = 0.1778$ is selected; see Figure 3(b). Then, we obtained nine important variables and 91 unimportant variables.

We denote the index set with unimportant variables as $G^0$ and the index set with important variables as $G^{0\circ}$. Given $G^0$, we
used the last 100 samples to test whether there were nonzeros in $G^0$ using ToMax, ZC1-17, B-13, and ZZ-14. All methods failed to reject the null hypothesis, so $G^0$ may contained all zeros. In contrast to other methods, TOSI method could further test whether there were zeros in $G^0$. The $p$-value of ToMin is 0.1432 and the adjusted $p$-values of ToMin(5) is 0.2864, indicating that the null hypothesis could not be rejected; that is, there might be zeros in $G^0$. This is consistent with the simulation results that CV LASSO tended to over-select features. Therefore, we increased the value of penalty parameter by finely tuning within [0.1828, 0.1878] until the null hypothesis $H_{0,G^0}$ was rejected and the null hypothesis $H_{0,G^0}$ was not rejected at significance level 0.05. Finally, we identified six important variables, as presented in Table 5(a). Furthermore, we conducted another check whether the selected 94 variables were truly unimportant via ZC1-17, B-13, and ZZ-14, and none of these methods could reject the null hypothesis. Therefore, the analysis results obtained via TOSI were more interpretable than those from CV LASSO.

Table 5(b) presents the estimated coefficients for the six important covariates: the percentage of the population that is Caucasian ($racePctWhite$), the percentage of population that is of hispanic heritage ($racePctHisp$), the number of people under the poverty level ($NumUnderPov$), the percentage of females who are divorced ($FemalePctDiv$), the percentage of kids in family housing with two parents ($PctKids2Par$) and the percentage of people in owner occupied households ($PctPersOwnOccup$). From Table 5(b), we concluded that $racePctWhite$, $PctKids2Par$, and $PctPersOwnOccup$ had negative effects on $ViolentCrimesPerPop$, while $racePctHisp$, $NumUnderPov$, and $FemalePctDiv$ had positive effects on $ViolentCrimesPerPop$. These results were easy to understand and explain. First, intuitively, the greater the proportion of people below the poverty level is, the more serious the violent crime in the community is. Second, divorce increases the likelihood of violent crime. Moreover, having two parents at home has a substantial inhibitory effect on juvenile violent crime.

8. Discussion

In this article, we introduced a two-directional simultaneous inference framework for high-dimensional manifest and latent models. With TOSI, we can fully identify the zero and nonzero parameters, resulting in more interpretable and simpler models. The simultaneous inference procedure achieves the prespecified significance level asymptotically and has power tending to one. Three typical models are considered as examples to illustrate the application of our TOSI framework, and the corresponding theoretical properties are established. Simulation studies and two real high-dimensional data examples are used to verify the performance and effectiveness of the estimation and inference, and the results are satisfactory.

In this article, we focus on the two testing problems for $G^0$ and $G^{00}$. In fact, TOSI can be applied to any two subsets $G_1$ and $G_2$ of $[p]$ and all the theoretical properties hold. However, TOSI has several potential weaknesses. The single-split version of TOSI can control the Type I error at the prespecified significance level but lacks efficiency. Although the multi-split version of TOSI can mitigate this problem, it controls the Type I error tightly which leads to some loss of power. How to find a statistic in the multi-split method that can control the Type I error exactly at the prespecified significance level, is a potential direction for future research.

Supplementary Materials

The online supplementary materials contain all technical proofs, related materials on sparse latent factor models, and additional simulation results.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

The research were partially supported by National Key R&D Program of China (no. 2022YFA1003702) and National Natural Science Foundation of China (nos. 11931014).

References

Bai, J., and Ng, S. (2013), "Principal Components Estimation and Identification of Static Factors," Journal of Econometrics, 176, 18–29. [303]
Barber, R. F., and Candès, E. J. (2019), "A Knockoff Filter for High-Dimensional Selective Inference," The Annals of Statistics, 47, 2504–2537. [299]
Benjamini, Y., and Yekutieli, D. (2001), "The Control of the False Discovery Rate in Multiple Testing Under Dependency," The Annals of Statistics, 29, 1165–1188. [299,304]
Bühlmann, P. (2013), "Statistical Significance in High-Dimensional Linear Models," Bernoulli, 19, 1212–1242. [301,302,306]
Chernozhukov, V., Chetverikov, D., and Kato, K. (2013), "Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-Dimensional Random Vectors," The Annals of Statistics, 41, 2786–2819. [299]
De Geer, S. V., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014), "On Asymptotically Optimal Confidence Regions and Tests for High-Dimensional Models," The Annals of Statistics, 42, 1166–1202. [298,301,302,306]
Dezeure, R., Bühlmann, P., and Zhang, C.-H. (2017), "High-Dimensional Simultaneous Inference with the Bootstrap," Test, 26, 685–719. [299]
Fan, J., and Li, R. (2001), "Variable Selection via Nonconvex Penalized Likelihood and its Oracle Properties," Journal of the American Statistical Association, 96, 1348–1360. [298]
Fan, J., and Lv, J. (2008), “Sure Independence Screening for Ultrahigh Dimensional Feature Space,” *Journal of the Royal Statistical Society, Series B*, 70, 849–911. [298,301]

Holm, S. (1979), “A Simple Sequentially Rejective Multiple Test Procedure,” *Scandinavian Journal of Statistics*, 6, 65–70. [299,304]

Hou, W., Ji, Z., Ji, H., and Hicks, S. C. (2020), “A Systematic Evaluation of Single-Cell RNA-Seq Imputation Methods,” *Genome Biology*, 21, 1–30. [303]

Jiang, F., Ma, Y., and Wei, Y. (2019), “Sufficient Direction Factor Model and its Application to Gene Expression Quantitative Trait Loci Discovery,” *Biometrika*, 106, 417–432. [303]

Jiang, T. (2004), “The Asymptotic Distributions of the Largest Entries of Sample Correlation Matrices,” *The Annals of Applied Probability*, 14, 865–880. [299]

Lee, J. D., Sun, D. L., Sun, Y., and Taylor, J. E. (2016), “Exact Post-Selection Inference, with Application to the Lasso,” *The Annals of Statistics*, 44, 907–927. [301]

Lou, Z., and Wu, W. B. (2017), “Simultaneous Inference for High Dimensional Mean Vectors,” arXiv preprint arXiv:1704.04806. [299]

Ma, Y., Li, Y., Lin, H., and Li, Y. (2017), “Concordance Measure-based Feature Screening and Variable Selection,” *Statistica Sinica*, 27, 1967–1985. [298]

Mandozzi, J., and Bühlmann, P. (2016), “Hierarchical Testing in the High-Dimensional Setting with Correlated Variables,” *Journal of the American Statistical Association*, 111, 331–343. [299]

Meinshausen, N. (2015), “Group Bound: Confidence Intervals for Groups of Variables in Sparse High Dimensional Regression Without Assumptions on the Design,” *Journal of the Royal Statistical Society, Series B*, 77, 923–945. [301]

Meinshausen, N., Meier, L., and Bühlmann, P. (2009), “P-values for High-Dimensional Regression,” *Journal of the American Statistical Association*, 104, 1671–1681. [299,301,302,304]

Porto, R. B., Costa, R. d. R., and Watanabe, E. A. d. M. (2017), “The Multilevel Effect of Marketing Activities on Sales, Revenue and Profitability in a Micro-Enterprise,” *Revista Brasileira de Gestão de Negócios*, 19, 432–452. [307]

Redmond, M., and Baveja, A. (2002), “A Data-Driven Software Tool for Enabling Cooperative Information Sharing Among Police Departments,” *European Journal of Operational Research*, 141, 660–678. [307]

Rinaldo, A., Wasserman, L., and Gsell, M. (2019), “Bootstrapping and Sample Splitting for High-Dimensional, Assumption-Lean Inference,” *The Annals of Statistics*, 47, 3438–3469. [299]

Romano, J. P., and DiCiccio, C. (2019), *Multiple Data Splitting for Testing*, Department of Statistics, Stanford University. [300]

Tibshirani, R. (1996), “Regression Shrinkage and Selection via the Lasso,” *Journal of the Royal Statistical Society, Series B*, 58, 267–288. [298]

Uematsu, Y., and Yamagata, T. (2021), “Inference in Sparsity-Induced Weak Factor Models,” *Journal of Business & Economic Statistics*, 41, 126–139. [305]

Wasserman, L., and Roeder, K. (2009), “High Dimensional Variable Selection,” *The Annals of Statistics*, 37, 2178–2201. [299,301,302]

Zhang, C., and Zhang, S. (2014), “Confidence Intervals for Low Dimensional Parameters in High Dimensional Linear Models,” *Journal of The Royal Statistical Society, Series B*, 76, 217–242. [298,301,304,305,306]

Zhang, X., and Cheng, G. (2017), “Simultaneous Inference for High-Dimensional Linear Models,” *Journal of the American Statistical Association*, 112, 757–768. [299,301,302,304,305]