Journal of Multivariate Analysis 144 (2016) 25–37

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Testing covariates in high dimension linear regression with latent factors

Wei Lan [a], Yue Ding [b], Zheng Fang [c], Kuangnan Fang [d,*]

[a] School of Statistics and center of Statistical Research, Southwestern University of Finance and Economics, Chengdu, China

[b] School of Economics and Management, Southwest Jiaotong University, Chengdu, China

[c] Business School, Sichuan University, Chengdu, China

[d] School of Economics and MOE Key Laboratory of Econometrics, Xiamen University, Xiamen, China

ARTICLE INFO

Article history:
Received 17 November 2014
Available online 3 November 2015

AMS subject classifications:
62F03

Keywords:
Approximate factor model
Global significance testing
High dimension regression
Individual effect testing

ABSTRACT

We propose here both F-test and z-test (or t-test) for testing global significance and individual effect of each single predictor respectively in high dimension regression model when the explanatory variables follow a latent factor structure (Wang, 2012). Under the null hypothesis, together with fairly mild conditions on the explanatory variables and latent factors, we show that the proposed F-test and t-test are asymptotically distributed as weighted chi-square and standard normal distribution respectively. That leads to quite different test statistics and inference procedures, as compared with that of Zhong and Chen (2011) when the explanatory variables are weakly dependent. Moreover, based on the p-value of each predictor, the method of Storey et al. (2004) can be used to implement the multiple testing procedure, and we can achieve consistent model selection as long as we can select the threshold value appropriately. All the results are further supported by extensive Monte Carlo simulation studies. The practical utility of the two proposed tests are illustrated via a real data example for index funds tracking in China stock market.

1. Introduction

Traditional F-test and z-test (or t-test) are commonly used to detect the relationship between a response variable $Y_i \in \mathbb{R}^1$ and a set of explanatory variables $X_i \in \mathbb{R}^p$ in a linear regression model when the number of explanatory variables p is fixed. By contrast, when p is diverging and much larger than the sample size n, classical statistical inferences (F test and z-test) were not applicable since the resulting ordinary least square (OLS) estimator is no longer computable. To fix the issue, there is a large stream of papers intending to extend the traditional F-test and z-test (or t-test) to accommodate high dimensional settings; see, for example, [22,9,21,12].

The aforementioned testing procedures are quite useful for high dimensional data analyses. However, their applicability is heavily relying on one critical assumption, i.e., the explanatory variables are weakly dependent such that $tr(\Sigma^4) = o\{tr^2(\Sigma^2)\}$, where $\Sigma = \text{cov}(X_i) \in \mathbb{R}^{p\times p}$. For more detailed illustrations for such assumption, we refer to [22,23]. It is remarkable that such assumption is violated if the explanatory variables X_i admit a latent factor structure, which is usually encountered in real practice [6,19]. Specifically, we consider the following data generation process $X_i = \gamma Z_i + \widetilde{X}_i$, where each element of the common factors $Z_i \in \mathbb{R}^d$ and random errors $\widetilde{X}_i$ are all independently generated from a standard

* Corresponding author.
E-mail addresses: facelw@gmail.com (W. Lan), dingyue-300@163.com (Y. Ding), 149281891@qq.com (Z. Fang), xmufkn@xmu.edu.cn (K. Fang).

http://dx.doi.org/10.1016/j.jmva.2015.10.013
0047-259X/

normal distribution, with $d > 0$ is the finite number of common factors. Moreover, the factor loadings $\gamma \in \mathbb{R}^{p\times d}$ satisfy $p^{-1}\gamma^\top\gamma \to I_d$, where I_d represents the identity matrix of dimension d. In this setting, one can verify that $tr(\Sigma^4) = tr(\gamma\gamma^\top)^4\{1+o(1)\} = tr(\gamma^\top\gamma)^4\{1+o(1)\} = p^4 tr(I_d)\{1+o(1)\} = dp^4\{1+o(1)\}$, and $tr(\Sigma^2) = dp^2\{1+o(1)\}$. As a result, we can have $tr(\Sigma^4)/tr^2(\Sigma^2) \to 1/d \neq 0$, which violates condition (2.8) of [22], and condition (C1) of [12]. Consequently, how to construct testing procedures for this special types of explanatory variables is a problem of theoretical demand.

It is also noteworthy that the above testing problems are also empirically motivated. For example, consider the problem of index fund tracking of reproducing the performance of a stock market index. In this particular application, the response of interest is the return on some specific market index, say Shanghai composite index in China stock market, while the explanatory variables can be the return of all the stocks in China stock market. Therefore, the number of explanatory variables may be very large compared with the number of observations; see Section 3.2 of real data analysis for details. For these types of explanatory variables, we cannot expect that the returns across different stocks are weakly dependent. In fact, it has long been recognized empirically and theoretically that there should exist some latent common factors that influence all stock returns [16,4,7,5]. To this end, it is quite natural and reasonable to assume that the explanatory variables X_i follow a latent factor structure so that the condition $tr(\Sigma^4) = o\{tr^2(\Sigma^2)\}$ is violated.

Motivated by the theoretical and practical demand, we intend to construct some testing procedures for the regression coefficients when the explanatory variables admit a latent factor structure [19]. We develop both F-test and z-test (or t-test) for testing global significance and effect of each single predictor respectively in high dimension regression model. Specifically, we revisit the test statistic of [12] used for testing global significance of regression coefficients for weakly dependent explanatory variables, and show that the resulting test statistic is asymptotic weighted chi-square when the explanatory variables follow an approximate factor model under some mild conditions. That leads to quite different test statistics and inference procedures, as compared with that of [22,12], when the explanatory variables are weakly dependent. In addition, after controlling for the latent common effect of the explanatory variables, the remaining factor profiled predictors are weakly dependent [19]. As a consequence, the univariate regression [7] can be used to assess the significance of each variable. Based on the p-value of each predictor, we can then apply the method of [17] to control the false discovery rate (FDR), and the method can achieve consistent model selection as long as we can set the nominal level appropriately. Extensive simulation results and an empirical example on index fund tracking in China stock market confirmed the usefulness of the proposed method.

The remainder of the paper is organized as follows. Section 2 introduces global significance testing, and individual effect testing with FDR control together with their theoretical properties. Numerical studies, including simulation and a real data analysis, are reported in Section 3. Section 4 concludes the article with a short discussion and all the technical details are provided in the Appendix.

2. The methodology

2.1. Model and notations

Let (Y_i, X_i) be the observation collected at ith unit for $1 \le i \le n$, where $Y_i \in \mathbb{R}^1$ is the response value, $X_i = (X_{i1}, \ldots, X_{ip})^\top \in \mathbb{R}^p$ be the p-dimensional explanatory variables with mean $\mathbf{0}$ and covariance matrix $\Sigma = (\sigma_{j_1j_2}) \in \mathbb{R}^{p\times p}$. Unless explicitly stated otherwise, we hereafter assume that $p \gg n$ and n tends to infinity for asymptotic behavior. In addition, we assume that all the explanatory variables have been appropriately standardized such that $E(X_{ij}) = 0$, and $\sigma_{jj} = 1$ for every $1 \le j \le p$. To establish the relationship between Y_i and X_i, we consider the following linear regression model,

$$Y_i = X_i^\top\beta + \varepsilon_i, \tag{2.1}$$

where $\beta = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$ is an unknown vector of regression coefficients, ε_i is the random noise that is independent of X_i, distributed with mean 0 and finite variance $\sigma^2 < \infty$. For notation convenience, define $\mathbb{Y} = (Y_1, \ldots, Y_n)^\top \in \mathbb{R}^n$ be a vector of response variable, $\mathbb{X} = (X_1, \ldots, X_n)^\top \in \mathbb{R}^{n\times p}$ be the design matrix with the jth column $\mathbb{X}_j = (X_{1j}, \ldots, X_{nj})^\top \in \mathbb{R}^n$, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top \in \mathbb{R}^n$.

Since the traditional F-test and z-test (or t-test) are no longer applicable when p is diverging and much larger than the sample size n, there is a large stream of papers intending to extend the traditional F-test and z-test (or t-test) to accommodate high dimensional settings; see, for example, [22,9,21,12]. For the statistical validity of the aforementioned tests, appropriate technical conditions have to be assumed. Among all the conditions, Zhang and Zhang [21] and Lan et al. [12] require that

$$\lambda_{\max}(\Sigma) < \infty, \tag{2.2}$$

where $\lambda_{\max}(A)$ represents for the largest eigenvalues of any arbitrary matrix A. In contrast, Zhong and Chen [22] replaced condition (2.2) by

$$tr(\Sigma^4) = o\big\{tr^2(\Sigma^2)\big\}. \tag{2.3}$$

We find that both (2.2) and (2.3) are sensible if Σ is not highly singular, this should happen if the predictors are weakly correlated. Unfortunately, conditions (2.2) and (2.3) are violated if the explanatory variables X_i are highly correlated that admit a latent factor structure. For illustration, we consider the following data generation process $X_i = \gamma Z_i + \widetilde{X}_i$, where each element of the common factors $Z_i \in \mathbb{R}^d$ and random errors $\widetilde{X}_i$ are all independently generated from a standard normal distribution, with $d > 0$ is the finite number of common factors. Moreover, the factor loadings $\gamma \in \mathbb{R}^{p\times d}$ satisfy $p^{-1}\gamma^\top\gamma \to I_d$, where I_d represents the identity matrix of dimension d. In this setting, we have $tr(\Sigma^4) = tr(\gamma\gamma^\top)^4\{1 + o(1)\} = tr(\gamma^\top\gamma)^4\{1+o(1)\} = p^4 tr(I_d)\{1+o(1)\} = dp^4\{1+o(1)\}$, and $tr(\Sigma^2) = dp^2\{1+o(1)\}$. As a result, we can have $tr(\Sigma^4)/tr^2(\Sigma^2) \to 1/d \neq 0$, which violates conditions (2.2) and (2.3). Consequently, how to construct testing procedure of β for highly correlated predictors is a problem of demanding and interest.

2.2. A factor model

To model the dependence structure of X_i, we assume that X_i admits the following latent factor structure [5,19],

$$X_i = \gamma Z_i + \widetilde{X}_i, \tag{2.4}$$

where $Z_i = (Z_{i1}, \ldots, Z_{id})^\top$ is the d-dimensional latent factors, $\gamma = (\gamma_{jk}) \in \mathbb{R}^{p\times d}$ is the associated factor loadings. $\widetilde{X}_i = (\widetilde{X}_{i1}, \ldots, \widetilde{X}_{ip})^\top \in \mathbb{R}^p$ is the factor profiled predictor that is independent of Z_i, it represents for the information that contained in X_i but cannot be fully explained by the low dimensional ($d < \infty$) latent factors Z_i. For identifiability purpose, we assume that $\mathrm{cov}(Z_i) = I_d$ throughout the entire article. Moreover, we assume that $\widetilde{\Sigma} = (\widetilde{\sigma}_{j_1 j_2}) = \mathrm{cov}(\widetilde{X}_i)$ is a diagonal matrix, that is, $\mathrm{cov}(\widetilde{X}_{ij_1}, \widetilde{X}_{ij_2}) = 0$ for any $j_1 \neq j_2$. Moreover, we further assume that the diagonal elements of $\widetilde{\Sigma}$ are bounded from zero to infinity such that

$$\sigma_{\min} < \min_j \widetilde{\sigma}_{jj} \le \max_j \widetilde{\sigma}_{jj} < \sigma_{\max} \tag{2.5}$$

for some finite positive constants $0 < \sigma_{\min} \le \sigma_{\max} < \infty$. Further define notation $\widetilde{\mathbb{X}} = (\widetilde{X}_1, \ldots, \widetilde{X}_n)^\top \in \mathbb{R}^{n\times p}$, $\widetilde{\mathbb{X}}_j = (\widetilde{X}_{1j}, \ldots, \widetilde{X}_{nj})^\top \in \mathbb{R}^n$, and $Z = (Z_1, \ldots, Z_n)^\top \in \mathbb{R}^{n\times d}$. Consequently, under the factor model setting (2.4), model (2.1) is reduced to the following matrix form $\mathbb{Y} = Z\gamma^\top\beta + \widetilde{\mathbb{X}}\beta + \varepsilon$. To extract the common effects Z, we multiply $\mathcal{Q}(Z) = I_n - Z(Z^\top Z)^{-1}Z^\top$ by each part of the equation, which leads to $\mathcal{Q}(Z)\mathbb{Y} = \mathcal{Q}(Z)\mathbb{X}\beta + \mathcal{Q}(Z)\varepsilon$. Denote $\bar{\mathbb{Y}} = \mathcal{Q}(Z)\mathbb{Y}$, $\bar{\mathbb{X}} = \mathcal{Q}(Z)\mathbb{X}$, and $\bar{\varepsilon} = \mathcal{Q}(Z)\varepsilon$, we then obtain the following factor profiled regression model [19],

$$\bar{\mathbb{Y}} = \bar{\mathbb{X}}\beta + \bar{\varepsilon}, \tag{2.6}$$

where $\bar{\mathbb{X}} = (\bar{X}_1, \ldots, \bar{X}_n)^\top \in \mathbb{R}^{n\times p}$. The main focus of current article is intending to construct some testing procedures for the regression coefficients β. We consider the following two aspects. *First*, we test the statistical significance of β globally, i.e., the so-called F-test, similar testing procedures were investigated for general weakly correlated predictors; see, for example, [22,12]. *Second*, we consider testing the statistical significance of each single predictor separately together with a FDR controlling procedure, this procedure can help us to identify the relevant predictors if we reject the null hypothesis of global significance in the first step.

2.3. Technical conditions

Before presenting the detailed testing procedures, we need to investigate a number of technical conditions. These conditions are assumed to simplify the theoretical proofs, they are all quite mild and sensible in real practice.

- (C1) Assume the profiled predictors $\widetilde{X}_i$ and latent factors Z_i are all independent and normally distributed.
- (C2) Assume the common factor number d is fixed, while the sample size n goes to infinity. Moreover, there exist some finite positive constants $c_{\min}$, $c_{\max}$ and $\xi > 0$ such that $n^{-\xi}\log p < c_{\max}$ and $p/n \ge c_{\min}$.
- (C3) There exists some positive definite matrix $\Sigma_\gamma \in \mathbb{R}^{d\times d}$ such that $p^{-1}\gamma^\top\gamma \to \Sigma_\gamma$. The eigenvalues of Σ_γ are all bounded from zero to infinity.

Condition (C1) is popularly used in high dimensional regression setting to simplify the theoretical proofs; see, for example, [7,18,19]. Condition (C2) indicates that as the sample size n is diverging, the predictor dimension p can grow at an exponential order of n. As a result, p may be much larger than n. Condition (C3) is also reasonable and commonly assumed in the literature; see, for example, [1,19], it can be satisfied if γ_{jk}s are independently generated from some non-degenerate distribution with finite fourth moment.

2.4. Global significance testing

We firstly consider the problem of testing statistical significance of β globally in this subsection. Accordingly, we consider the following statistical hypotheses,

$$H_0 : \beta = 0, \quad \text{vs.} \quad H_1 : \beta \neq 0. \tag{2.7}$$

Under the null hypothesis, model (2.1) is reduced to $Y_i = \varepsilon_i$, then we can have $E(\mathbb{Y}^\top \mathbb{X}_j) = E(\varepsilon^\top \mathbb{X}_j) = 0$ for every $1 \le j \le p$. As a result, we can expect that its sample counterpart $n^{-1}\mathbb{Y}^\top \mathbb{X}_j$ should be close to 0 as well, such that $n^{-1}\mathbb{Y}^\top \mathbb{X}_j \approx 0$ for every j. Combining the information from every j leads us to consider the following test statistic

$$T_{\text{initial}} = n^{-1}p^{-1}\sum_j \|\mathbb{Y}^\top \mathbb{X}_j\|^2/\hat{\sigma}^2 = n^{-1}p^{-1}\mathbb{Y}^\top \mathbb{X}\mathbb{X}^\top \mathbb{Y}/\hat{\sigma}^2, \tag{2.8}$$

where $\hat{\sigma}^2 = \mathbb{Y}^\top \mathbb{Y}/n$ is the normalizing constant. The asymptotic distribution of T_{initial} is given below, whose proof is relegated to Appendix B.

Theorem 1. *Under the null hypothesis of (2.7), further assume that conditions* (C1)–(C3) *are satisfied and the number of latent factors d is known, then as $\min\{n, p\} \to \infty$, $T_{initial} - \bar{\sigma}^2$ follows a weighted chi-square distribution of $\sum_{i=1}^d \lambda_i \chi_1^2$, where $\bar{\sigma}^2 = p^{-1}tr(\widetilde{\Sigma})$, and λ_i is the ith largest eigenvalues of Σ_γ.*

To make the proposed test practically useful, one needs to estimate unknown parameters λ_i, $\bar{\sigma}^2$ and d. The true number of factors d is unknown albeit fixed. We start with an arbitrary number k for $k < \min(n, p)$. For each k, we estimate Z and γ using the method of principal component [19]. Specifically, we define $\hat{\zeta}_j$ be the jth largest eigenvalues of $(np)^{-1}\mathbb{X}\mathbb{X}^\top$, while $\hat{\varrho}_j$ be the corresponding eigenvector. We next set $\hat{Z}_k = n^{1/2}(\hat{\varrho}_1, \ldots, \hat{\varrho}_k)$ and then $\gamma_k^\top$ can be estimated by $\hat{\gamma}_k^\top = (\hat{Z}_k^\top \hat{Z}_k)^{-1}\hat{Z}_k^\top \mathbb{X}$. Subsequently, the number of common factors d can be selected by minimizing the following objective function

$$PC(k) = n^{-1}p^{-1}tr\left\{(\mathbb{X} - Z_k\gamma_k^\top)^\top(\mathbb{X} - Z_k\gamma_k^\top)\right\} + k\sigma_{\text{pc}}^2\left(\frac{n+p}{np}\right)\log\left(\frac{np}{n+p}\right)$$

with respect to k as suggested by Bai and Ng [2], which immediately leads to $\hat{d} = \text{argmin}_{0\le k\le k_{\max}} PC(k)$ with $\sigma_{\text{pc}}^2 = n^{-1}p^{-1}tr(\mathbb{X}^\top \mathbb{X})$ and $k_{\max}$ be a bounded integer such that $d < k_{\max}$. According to the theoretical results of [2], $\hat{d}$ equals d with probability approaching to 1 under the conditions (C1)–(C3) and assumption (2.5). As a result, in the subsequent article, we assume that the true latent factor number d is known to simplify the theoretical proofs. Finally, the estimated common factor can be then defined as $\hat{Z} = n^{1/2}(\hat{\varrho}_1, \ldots, \hat{\varrho}_{\hat{d}})$ and $\gamma^\top$ can be estimated by $\hat{\gamma}^\top = (\hat{Z}^\top \hat{Z})^{-1}\hat{Z}^\top \mathbb{X}$, $\widetilde{\Sigma}$ can be estimated by $n^{-1}\mathbb{X}^\top \mathcal{Q}(\hat{Z})\mathbb{X}$, and Σ_γ can be estimated by $\hat{\Sigma}_\gamma = p^{-1}\hat{\gamma}^\top\hat{\gamma}$. We next demonstrate the following results.

Proposition 1. *Under the conditions* (C1)–(C3), *we can have* (1.) $n^{-1}p^{-1}tr\left\{\mathbb{X}^\top \mathcal{Q}(\hat{Z})\mathbb{X}\right\} - p^{-1}tr(\widetilde{\Sigma}) \to_p 0$ *and* (2.) $p^{-1}\hat{\gamma}^\top\hat{\gamma} - \Sigma_\gamma \to_p 0$.

Accordingly, we propose the following test statistic

$$T_{\text{final}} = n^{-1}p^{-1}\mathbb{Y}^\top \mathbb{X}\mathbb{X}^\top \mathbb{Y}/\hat{\sigma}^2 - n^{-1}p^{-1}tr\left\{\mathbb{X}^\top \mathcal{Q}(\hat{Z})\mathbb{X}\right\}.$$

T_{final} should follow a weighted chi-square distribution as $\sum_{i=1}^{\hat{d}} \hat{\lambda}_i\chi_1^2$, where $\hat{\lambda}_i$ is the ith largest eigenvalues of $\hat{\Sigma}_\gamma$. Let $\hat{q}_\alpha$ stand for the α-th quantile of the weighted chi-square distribution. We then reject the null hypothesis of (2.7) if $T_{\text{final}} > \hat{q}_\alpha$. Based on the above results, one can calibrate the sizes of the proposed test.

Remark 1. It is worth mentioning that the above testing procedures can be extended to high dimensional partial F-test [12]. Specifically, if our interest is to test the global effect of X_{ib} after controlling for the effect of X_{ia}, with $X_i = (X_{ia}^\top, X_{ib}^\top)^\top$. Here, the dimension of X_{ib} is ultra high, while the dimension of X_{ia} is fixed or much smaller than n. Then the testing procedure can still be applicable by replacing Y_i with the residual after regressing Y_i on X_{ia} in (2.8). One can verify that the asymptotic distribution cannot be changed, we thus omit it to save space.

Remark 2. The proposed initial test statistic T_{initial} and its asymptotic distribution given in Theorem 1 are identical to the test statistic proposed by Goeman et al. [9]. The difference is that the asymptotic distribution of the test statistic proposed by Goeman et al. [9] rely on the normal error assumption and weakly correlated predictors. Therefore, the asymptotic results obtained in this article primarily extend the result of [9] to non-normal errors while allow the predictors to be highly correlated that admit a latent factor structure.

2.5. Individual effect testing with FDR control

If we reject the null hypothesis of global significance, there should exist some predictors that have nontrivial effect on the response. To this end, it is necessary to assess the effect of each single variable individually, that is, the so-called z-test (or t-test). Specifically, our interest is to test the null hypothesis $H_{0j} : \beta_j = 0$ for some $1 \le j \le p$. Without loss of generality, we only consider testing the significance of the first predictor, that is,

$$H_{01} : \beta_1 = 0 \quad \text{vs.} \quad H_{11} : \beta_1 \neq 0. \tag{2.9}$$

To test $\beta_1 = 0$, the univariate regression of [7] is no longer applicable, since the predictors are highly correlated according to Eq. (2.4). Fortunately, the dependence between predictors is mainly driven by the common factors Z. Therefore, the dependence between predictors can be removed by projecting off the effects of common factors by applying the operator $\mathcal{Q}(Z)$. By doing so, the projected regressors become just the idiosyncratic components in the factor model (2.4), which are uncorrelated by model assumption. This finding motivates us to apply univariate regression based on projected response and regressors to assess the significance effect of each single variable in (2.1). Since the latent factor Z is usually unknown in practice, we use the estimator $\hat{Z}$ proposed in Section 2.4 for instead. To this end, we define $\hat{\mathbb{Y}} = \mathcal{Q}(\hat{Z})\mathbb{Y}$, $\hat{\mathbb{X}} = \mathcal{Q}(\hat{Z})\mathbb{X}$ and $\hat{\varepsilon} = \mathcal{Q}(\hat{Z})\varepsilon$. We next conduct univariate regression for the first predictor by regressing $\hat{\mathbb{Y}}$ on $\hat{\mathbb{X}}_1$, the regression coefficient estimate is given by $\hat{\beta}_1 = (\hat{\mathbb{X}}_1^\top \hat{\mathbb{X}}_1)^{-1}\hat{\mathbb{X}}_1^\top \hat{\mathbb{Y}}$, whose asymptotic distribution is given below.

Theorem 2. *Under the conditions* (C1)–(C3) *and the bounded variance condition* (2.5)*, further assume that* $\sum_j |\beta_j| < C_{\max}$ *for some finite positive constant* $C_{\max} > 0$*. Then, we can have* $n^{1/2}(\hat{\beta}_1 - \beta_1) \to_d N(0, \sigma_{\beta_1}^2)$*, where* $\sigma_{\beta_1}^2 = \tau_{\beta_1}^2/\widetilde{\sigma}_{11}$ *with* $\tau_{\beta_1}^2 = \big\{\sigma^2 + \beta_{1*}^\top\big(\widetilde{\Sigma}_{1*} - \widetilde{\Sigma}_{1*1}\widetilde{\Sigma}_{11}^{-1}\widetilde{\Sigma}_{11*}\big)\beta_{1*}\big\}$ *and* $\beta_{1*} = (\beta_j : j \neq 1)^\top \in \mathbb{R}^{p-1}$, $\widetilde{\Sigma}_{1*} = (\widetilde{\sigma}_{j_1 j_2} : j_1 \neq 1, j_2 \neq 1) \in \mathbb{R}^{(p-1)\times(p-1)}$.

The condition $\sum_j |\beta_j| < C_{\max}$ is sensible in practice and can be satisfied if the number of nonzero coefficients is finite. According to Theorem 2, we can construct the following test statistic

$$\hat{Z}_1 = n^{1/2}\hat{\beta}_1/\hat{\sigma}_{\beta_1}, \tag{2.10}$$

where $\hat{\sigma}_{\beta_1}^2 = \hat{\tau}_{\beta_1}^2\big(n^{-1}\hat{\mathbb{X}}_1^\top\hat{\mathbb{X}}_1\big)^{-1}$, and $\hat{\tau}_{\beta_1}^2 = n^{-1}\hat{\varepsilon}^{*\top}\hat{\varepsilon}^*$ with $\hat{\varepsilon}^*$ is the residual obtained by regressing $\hat{\mathbb{Y}}$ on $\hat{\mathbb{X}}_1$. One can verify that $\hat{Z}_1$ is asymptotic standard normal by employing the Slutsky's theorem. Accordingly, one reject the null hypothesis of (2.9) if $|\hat{Z}_1| > z_{1-\alpha/2}$, where z_α stands for the α-th quantile of a standard normal distribution.

When p is ultra-high, the predictors that need to be tested should be large as well. It is well known that by conducting a large number of testing problems simultaneously, the type I error can get inevitable inflated and thus lead to nontrivial multiple testing effect [3]. To guard against false discoveries, we employ the method of [17] for controlling the false discovery rate. We define $\mathcal{N}_0 = \{j : \beta_j = 0\}$, $\mathcal{N}_1 = \{j : \beta_j \neq 0\}$, and N_0 and N_1 are the cardinality of sets $\mathcal{N}_0$ and $\mathcal{N}_1$, respectively. Denote the p-value obtained by testing each individual null hypothesis, H_{0j}, as $p_j = 2\big\{1 - \Phi(|\hat{Z}_j|)\big\}$, where $\hat{Z}_j$ is the test statistic and can be constructed similarly to that in Eq. (2.10). Moreover, let $V(t) = \#\{j \in \mathcal{N}_0 : p_j \leq t\}$ be the number of falsely rejected hypotheses and $R(t) = \#\{j : p_j \leq t\}$ be the number of totally rejected hypotheses. As a result, for any threshold value $t \in [0, 1]$, the false discovery proportion is defined as $\mathrm{FDP}(t) = V(t)/[R(t) \vee 1]$ and $\mathrm{FDR}(t) = E\big\{\mathrm{FDP}(t)\big\}$ with $R(t) \vee 1 = \max\{V(t), 1\}$. For any pre-chosen level q and a tuning parameter $\lambda \in (0, 1]$, a data-driven threshold for the p-values is determined by

$$t_q\big(\widehat{\mathrm{FDR}}_\lambda\big) = \sup\big\{0 \leq t \leq 1 : \widehat{\mathrm{FDR}}_\lambda(t) \leq q\big\}, \tag{2.11}$$

where $\widehat{\mathrm{FDR}}_\lambda(t)$ is a point estimate of $\mathrm{FDR}(t)$, which is given by

$$\widehat{\mathrm{FDR}}_\lambda(t) = \frac{p\hat{\pi}_0(\lambda)t}{R(t) \vee 1} = \frac{\hat{\pi}_0(\lambda)t}{\{R(t) \vee 1\}/p} \tag{2.12}$$

with $\hat{\pi}_0(\lambda) = \big\{(1-\lambda)p\big\}^{-1}\big\{p - R(\lambda)\big\}$ is an estimate of π_0 for any given λ. We then reject the null hypothesis of $\beta_j = 0$ if its associated p-value p_j is less than or equal to $t_q(\widehat{\mathrm{FDR}}_\lambda)$. Next theorem shows that the FDR can be controlled at the nominal level asymptotically for this special type of threshold values.

Proposition 2. *Assume* $N_1/p \to 0$ *and* $\lim_{n\to\infty} T_{1,n}(t) = T_1(t)$ *for some continuous function* $T_1(t)$*, where* $T_{1,n}(t) = p^{-1}\sum_{j=1}^p P(p_j \leq t)$*, then under the same conditions as that assumed in Theorem 1, we have that* $\limsup_{n\to\infty} FDR\big\{t_q(\widehat{\mathrm{FDR}}_\lambda)\big\} \leq q$.

Remark 3. It is worthy mentioning that the problem of testing the significance of a single regression coefficient in high dimensional linear regression model has been tentatively studied by Lan et al. [13]. Even though both papers try to first remove the dependence among the regressors, the method of [13] is quite different with ours. To access the significance of a single regression coefficient, Lan et al. [13] proposed to first remove the effect of the predictors that are highly correlated with the target predictor. As noted by Lan et al. [13], their method is only applicable when the predictors are weakly correlated; see condition (C2) therein. By contrast, our method is applicable for highly correlated predictors by assuming a latent factor structure. Since the factors are unknown and need to be inferred using principal components, the resulting procedure poses more challenges when investigating the effect of estimation errors.

2.6. Model selection consistency

According to the proceeding proposition, for any pre-specified significance level $q > 0$, FDR can be controlled asymptotically at the nominal level. It further motivates us to investigate the theoretical properties for $q \to 0$, i.e., model selection consistency. In order to get the property of consistent model selection, we need to analyze the power of the resulting multiple testing procedure. To this end, we need to assume the following minimum signal and exponential tail assumptions.

(C4) Assume there exist two positive constants κ and C_β such that $\min_{j\in\mathcal{N}_1} |\beta_j| > C_\beta n^{-\kappa}$ for $\kappa + \xi < 1/2$, where ξ was defined in condition (C2).

(C5) There exists some positive constant C_e such that for any $\ell > 0$ and $1 \le j \le p$, $P\left(n^{-1}|\mathbb{X}_j^\top \varepsilon| > \ell\right) \le \exp(-C_e n \ell^2)$.

Condition (C4) is sensible in practice, the condition has been popularly used for variable selection literature [7,18]. Condition (C5) is also reasonable. Specifically, if the random noise ε is normally distributed, then using the fact that $n^{-1}\|\mathbb{X}_j\|^2 \to 1$, $n^{-1/2}\mathbb{X}_j^\top\varepsilon$ follows a normal distribution with finite variance for $j = 1, \ldots, p$. As a result, the condition (C5) can be satisfied. Under the above conditions, we can demonstrate the following result.

Theorem 3. *Under conditions* (C1)–(C5) *and bounded variance assumption* (2.5)*, further assume* $\sum_j |\beta_j| < C_{\max}$ *for some finite positive constant* $C_{\max}$*, there should exist a sequence of significance levels* $\alpha_n \to 0$ *such that* $P(\hat{\mathcal{S}}^{\alpha_n} = \mathcal{N}_1) \to 1$*, where* $\hat{\mathcal{S}}^{\alpha_n} = \{1 \le j \le p : p_j \le \alpha_n\}$.

The proof is provided in Appendix F. According to the theorem proof, one can select α_n at the level of $\alpha_n = 2\{1 - \Phi(n^\jmath)\}$ with $\xi < \jmath < 1/2 - \kappa$. For this special sequences of α_n and under the minimum signal assumption (C4), the power of the test can approach to 1 while preserve reasonable type I error and false discoveries. Compared with the variable screening method of [19], the proposed testing procedure is able to control the false discovery rate and the type I error for the sequence of nominal level α_n. This finding is quite important especially in finite samples; see, for example, [20,15] and for detailed discussions.

3. Numerical studies

3.1. Simulation studies

To demonstrate the finite sample performance of the proposed testing procedures, we present here two simulation examples including three different sample sizes ($n = 100, 200, 400$), two different dimensions of explanatory variables ($p = 500, 1,000$), and two different dimensions of common factors ($d = 1, 3$) for the purpose of illustration. For each fixed parameter setting (i.e., n, p and d), all simulation results were conducted by 1,000 realizations, the nominal levels of the F-test, z-test and the FDR level were set to be 5%. To access the finite sample performance of the proposed testing procedures, we evaluate the size of the proposed F-test (FS). Moreover, we measure the performance of the proposed z-test (or t-test) by the average empirical size (AES). Specifically, let p_{rj} be the individual test p-value for testing the significance of the jth explanatory variable in the rth simulation. Hence, AES can be defined as $\text{AES} = |\mathcal{N}_0|^{-1}\sum_{j\in\mathcal{N}_0}\text{ERP}_j$, where $\mathcal{N}_0 = \{1 \le j \le p : \beta_j = 0\}$, and the empirical rejection probability (ERP) for the jth explanatory variable test as $\text{ERP}_j = 1000^{-1}\sum_{r=1}^{1000} I(p_{rj} < \alpha)$. To assess the effect of model selection consistency, we report the average true rate $\text{TR} = |\mathcal{S}^\alpha \cap \mathcal{N}_1|/|\mathcal{N}_1|$ and the average false rate $\text{FR} = |\mathcal{S}^\alpha \cap \mathcal{N}_0|/|\mathcal{N}_0|$. Intuitively, if the testing procedure can identify significant predictors consistently, the FR should approach to 0 while the TR approaching to 1 as the sample size $n \to \infty$. To guard against the false discoveries, the empirical FDR based on the procedure proposed by Storey et al. [17] with $\lambda = 1/2$ is also reported.

Example 3.1. The predictors X_i is simulated according to (2.4), where each element of Z_i, γ and $\widetilde{X}_{ij}$ were independently generated from a standard normal distribution. Moreover, the response Y_i is generated according to (2.1), with ε_i is independently generated from either a t distribution with 3 degrees of freedom $t(3)$ or a mixture distribution $0.1N(0, 3^2) + 0.9N(0, 1)$. All of the regression coefficients were set to be zero so that $|\mathcal{N}_1| = 0$ and $|\mathcal{N}_0| = p$. For the sake of comparison, the method proposed by Zhong and Chen [22] is also included, we name it as ZC-test. The simulation results based on $d = 1$ were summarized in Table 1.

A well behaved test should have an empirical size around 5%. As a result, we can expect that both FS and AES should be close to 5% in this simulation setting. According to the results of Table 1, our proposed F-test totally dominates ZC-test. Both the proposed F-test and z-test (or t-test) can control the type I error well at the nominal level, regardless of the error distributions, which are consistent with the theoretical findings in Theorems 1 and 2. By contrast, the size of the ZC test is alarmingly larger than the nominal significance level. Such a finding is not surprising, since the ZC-test is designed mainly for weakly dependent data, not for the data with a latent factor structure.

Example 3.2. For the purpose of illustration, we only consider $d = 3$ in this example, since the results for other settings are similar. Similar to Example 3.1, each element of Z_i and γ are independently generated from a standard normal distribution. We next generate $\widetilde{X}_i$. We consider the following two scenarios. In scenario 1, each element of $\widetilde{X}_i$ was independently generated

Table 1
Simulation results for Example 1 with $d = 1$.

		t Distribution				Mixture distribution			
		ZC-test		FA-test		ZC-test		FA-test	
n	p	FS	AES	FS	AES	FS	AES	FS	AES
100	500	0.083	–	0.043	0.052	0.088	–	0.058	0.054
	1000	0.087	–	0.054	0.055	0.075	–	0.050	0.054
200	500	0.082	–	0.047	0.053	0.078	–	0.051	0.052
	1000	0.084	–	0.044	0.052	0.074	–	0.040	0.052
400	500	0.077	–	0.045	0.051	0.075	–	0.047	0.051
	1000	0.078	–	0.048	0.051	0.069	–	0.040	0.051

Table 2
Simulation results for Example 2 with $d = 3$.

		Normal distribution					Mixture distribution				
n	p	FS	AES	TR	FR	FDR	FS	AES	TR	FR	FDR
Scenario 1											
100	500	0.992	0.058	0.968	0.029	0.091	0.991	0.058	0.969	0.029	0.107
	1000	0.994	0.058	0.962	0.058	0.106	0.991	0.057	0.970	0.057	0.117
200	500	0.999	0.054	1.000	0.027	0.071	0.999	0.054	0.998	0.027	0.067
	1000	0.997	0.054	0.998	0.054	0.074	1.000	0.054	0.999	0.054	0.074
400	500	1.000	0.052	1.000	0.026	0.061	0.999	0.053	1.000	0.026	0.063
	1000	1.000	0.052	1.000	0.052	0.053	1.000	0.053	1.000	0.053	0.062
Scenario 2											
100	500	0.991	0.058	0.970	0.029	0.105	0.989	0.058	0.971	0.029	0.110
	1000	0.995	0.054	0.999	0.055	0.087	0.993	0.058	0.968	0.058	0.115
200	500	0.999	0.055	1.000	0.027	0.081	0.999	0.055	0.999	0.027	0.075
	1000	0.997	0.055	0.998	0.055	0.079	0.999	0.054	0.998	0.054	0.719
400	500	1.000	0.053	1.000	0.026	0.070	1.000	0.054	1.000	0.027	0.078
	1000	0.998	0.052	1.000	0.052	0.069	0.999	0.053	1.000	0.053	0.068

from a standard normal distribution. In scenario 2, $\widetilde{X}_i$ follows a multivariate normal distribution with mean $\mathbf{0}$ and covariance $\mathrm{cov}(\widetilde{X}_i) = (\sigma_{j_1 j_2}) \in R^{p\times p}$ with $\sigma_{j_1 j_2} = 0.1^{|j_1 - j_2|}$. As a result, $\widetilde{X}_{ij_1}$ and $\widetilde{X}_{ij_2}$ are allowed to be weakly correlated. For the two scenarios, X_i were generated according to (2.4), Y_i was generated by (2.1) with ε_i independently following either a $N(0, 1)$ or a mixture distribution $0.1N(0, 3^2) + 0.9N(0, 1)$. The regression coefficient vector β is $\beta_1 = 5$, $\beta_4 = 3$, $\beta_7 = 2$, and $\beta_j = 0$ for any $j \notin \{1, 4, 7\}$. All the simulation results were summarized in Table 2.

According to Table 2, when $\widetilde{X}_i$s are independently generated according to scenario 1 and ε is normal, FSs are steadily increasing to 1 as the sample size increases. Similar results can be found when the distribution of ε is mixture, which indicate that the proposed global significance test is indeed consistent. Moreover, all AES and FDR values are around 0.05 as the sample size getting large, which suggests that the proposed individual coefficient test performs robustly well. The results are similar when the $\widetilde{X}_i$s are independently generated according to scenario 2. Lastly, for the two scenarios, the TR tends to 1 and FR tends to 0 as the sample size increases while ε follows both normal and mixture distributions, which are all consistent with the theoretical findings of Theorem 3.

3.2. *Real data analysis*

To further illustrate the practical usefulness of the proposed method, we present here a real data analysis. The data contains a total of $n = 409$ observations, where the response of interest (Y_i) is the daily return of Shanghai Stock Exchange Composite index and the explanatory variables $(X_1, \ldots, X_p)$ are $p = 757$ returns of individual stocks listed in Shanghai stock market. The sample period of the study is from 2011/1/3 to 2012/9/28. All data are from the CSMR database, which is one of the most popularly used and authoritative database in China. The main focus of current analysis is intending to identify the stocks which are significantly associated with the index Y_i by assuming a linear relationship exists between Y_i and X_i such that $Y_i = X_i^\top \beta + \varepsilon_i$ for some random error ε_i, and a portfolio will be constructed by selecting a small number of significant stocks to track the performance of Shanghai Stock Exchange Composite index.

We first need to estimate the dimension of the latent factors. The criteria proposed by Bai and Ng [2], namely, $\hat{d} = \mathrm{argmin}_{0\le k\le k_{\max}} PC(k)$ for $k_{\max} = 8$, was used to determine the dimension of latent factors, which suggest the presence of $d = 1$ factor. Consequently, in the rest of this real data analysis, we will treat the number of common factors $d = 1$.

The p-value of the proposed F-test is 0. As a result, we next consider how to identify the relevant explanatory variables that are associated with Y_i using the proposed t-test. For each explanatory variable, we consider the following test for every $j = 1, \ldots, p$,

$$H_{0j} : \beta_j = 0 \quad \text{vs.} \quad H_{1j} : \beta_j \neq 0. \tag{3.1}$$

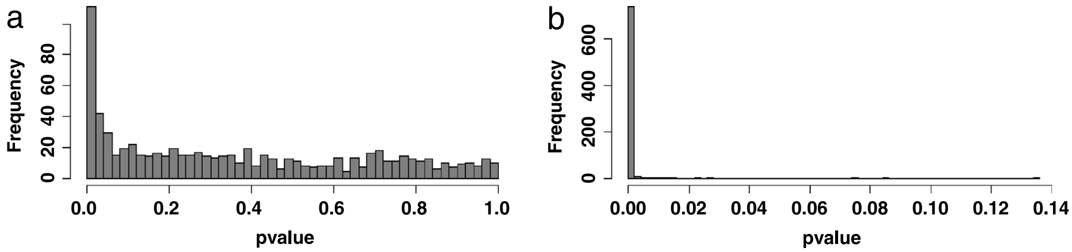

Fig. 1. The histograms of the p-values of the proposed t-test and MUR for all of the stocks.

For comparison purpose, we also report the results without adjusting for any latent factors, i.e., using marginal univariate regression (MUR) directly. Fig. 1 depicts the histogram of the p-values of the proposed t-test and MUR for all of the stocks. As we can see from the histograms, the proposed t-test yields a very flat histogram except for a bump around the point 0. Among all the 757 stocks, there are 167 stocks with p-values less than the significance level at 5%. Moreover, after controlling the false discoveries rate for the p-values from the proposed test at 5%, 68 stocks were declared statistically significant. This finding is quite reasonable. In fact, if the underlying coefficient structure is highly sparse, most of the p-values are computed for the case that the null hypothesis is true, and the asymptotic distribution of these p-values should be uniform in $[0, 1]$. In contrast, the histogram of the MUR is extremely skew, most of the p-values are very small, even after controlling the false discovery rate, there were a total of 754 stocks that are significant, which indicates that the resulting p-values may not be reliable.

Based on the selected stocks of the two methods, we constructed two portfolios (CF-portfolio and NoCF-portfolio) to mimic the Shanghai composite index by linear regression. CF-portfolio consists of the selected stocks by the proposed t-test, and NoCF-portfolio consists of the selected stocks by the marginal univariate regression without considering any latent factors. To evaluate the out-of-sample performance of the two portfolios, monthly rolling forecasting process is used with a constant rolling window of 12 months. For each month, we use the information of its previous 12 months to select stocks and estimate the resulting regression coefficients, then we form a portfolio consisting of the selected stocks with the portfolio weights determined by the estimate regression coefficients, and evaluate the return of this portfolio in that month. Such process which continues until 2012/9 is predicted. As a result, the rolling window length is fixed and we update the selected stocks and regression coefficients once a month. To measure how closely the portfolio follows the benchmark index, we calculate the tracking error $TE_t = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}(TD_t - \overline{TD})^2}$, where $TD_t = R_{pt} - R_{mt}$ is the tracking difference, R_{pt} and R_{mt} are the returns of portfolio and Shanghai Composite index respectively at time t and T is the total number of replications. The tracking error of the CF-portfolio and NoCF-portfolio are 0.73% and 0.91% respectively, which indicates that the proposed method in this article is more efficient than the method without adjusting for any factors. In addition, we also tried the scenario of $d = 3$, the results are robust and the corresponding tracking error of CF-portfolio is 0.77%. In real practice, the decision makers could take into account more information and transaction cost to further minimize the tracking errors.

4. Conclusion

In linear regression model, we develop both F-test and z-test (or t test) for testing global significance and individual effect of each single variable respectively in ultra high dimension regression when the explanatory variables follow a latent factor model. Under the null hypothesis, together with fairly mild conditions on the explanatory variables and latent factors, we show that the proposed F-test and t-test are asymptotically distributed as weighted chi-square and standard normal distribution respectively, which are quite different with the results derived by Zhong and Chen [22] when the explanatory variables are weakly dependent. Moreover, based on the p-value of each predictor, the method of [17] can be used to implement the multiple testing procedure, and we can achieve consistent model selection as long as we can select the threshold value appropriately.

To conclude the article, we point out a few possible research avenues in this discussion. Since we only established testing procedures for linear regression model, it would be interesting to extent the entire testing procedure to other regression models, which includes generalized linear model [8] and various semiparametric models [10]. Moreover, it would also be quite practical demanding to allow the explanatory variables to be discrete or ordinal. Lastly, it is also quite interesting to allow the profiled predictors $\widetilde{X}_i$ to be weakly correlated that admit an approximate factor model [5].

Acknowledgments

We thank the editor and reviewers for careful review. This study was supported by the National Natural Science Foundation of China (71471152, 71201139, 11401482, 11471264, 71303200, 71490722, 71572154) and National Social Science Foundation of China (13&ZD148, 13CTJ001), and Fundamental Research Funds for the Central Universities (268SWJTU15WBR01).

Appendix A. Some useful lemmas

Before proving the theoretical results, we present the following two useful lemmas. Lemma 1 can be proved in a similar manner by Theorem 2 of [19] under the current conditions (C1)–(C3) together with bounded variance assumption (2.5), while Lemmas 2 and 3 can be directly found in [12]. As a result, their proofs are omitted to save space.

Lemma 1. *Under conditions* (C1)–(C3) *and the latent factor number d is known, together with the bounded variance condition* (2.5), *we can have that $n^{-1}tr\{Z^\top \mathcal{Q}(\hat{Z})Z\} = O_p(n^{-1})$ and $tr\{\mathcal{Q}(\hat{Z}) - \mathcal{Q}(Z)\}^2 = O_p(n^{-1})$, where $\hat{Z}$ is defined in Section* 2.3.

Lemma 2. *Let $V = (V_1, \ldots, V_m)^\top \in \mathbb{R}^m$ be a multivariate normal vector with $E(V) = 0$ and $cov(V) = I_m$. Then, for any symmetric $m \times m$ matrix A, we have that* (i.) $E(V^\top A_1 V) = tr(A)$; (ii.) $E(V^\top A_1 V)^2 = tr^2(A_1) + 2tr(A_1^2)$.

Lemma 3. *Let $(U_1, U_2, U_3, U_4)^\top \in \mathbb{R}^4$ be a 4-dimensional normal random vector with $E(U_j) = 0$ and $var(U_j) = 1$ for $1 \le j \le 4$. We then have $E(U_1U_2U_3U_4) = \delta_{12}\delta_{34} + \delta_{13}\delta_{24} + \delta_{14}\delta_{23}$, where $\delta_{ij} = E(U_iU_j)$.*

Appendix B. Proof of Theorem 1

Note that $\mathbb{X} = Z\gamma^\top + \widetilde{\mathbb{X}}$. As a result, T_{initial} can be decomposed into the following three parts as $T_{\text{initial}} = \Lambda_1 + 2\Lambda_2 + \Lambda_3$ with $\Lambda_1 = n^{-1}p^{-1}\mathbb{Y}^\top Z\gamma^\top\gamma Z^\top\mathbb{Y}/\hat{\sigma}^2$, $\Lambda_2 = n^{-1}p^{-1}\mathbb{Y}^\top Z\gamma^\top\widetilde{\mathbb{X}}^\top\mathbb{Y}/\hat{\sigma}^2$ and $\Lambda_3 = n^{-1}p^{-1}\mathbb{Y}^\top\widetilde{\mathbb{X}}\widetilde{\mathbb{X}}^\top\mathbb{Y}/\hat{\sigma}^2$. We then consider the three parts separately. We firstly consider Λ_3. By Lemma L3 in [19], we have $p^{-1}\widetilde{\mathbb{X}}\widetilde{\mathbb{X}}^\top = \bar{\sigma}^2 I_T\{1 + o_p(1)\}$ with $\bar{\sigma}^2 = p^{-1}tr(\widetilde{\Sigma})$. Consequently, we can obtain $\Lambda_3 = \bar{\sigma}^2 n^{-1}\mathbb{Y}^\top\mathbb{Y}/\hat{\sigma}^2\{1 + o_p(1)\}$, which immediately leads to $\Lambda_3 = \bar{\sigma}^2\{1 + o_p(1)\}$. We next consider Λ_2. By Lemma L5 in [19], we have $\lambda_{\max}(p^{-1}Z\gamma^\top\widetilde{\mathbb{X}}^\top) = o_p(1)$. As a result,

$$\Lambda_2 = n^{-1}\mathbb{Y}^\top\{p^{-1}Z\gamma^\top\widetilde{\mathbb{X}}^\top\}\mathbb{Y}/\hat{\sigma}^2 \le \lambda_{\max}(p^{-1}Z\gamma^\top\widetilde{\mathbb{X}}^\top)n^{-1}\mathbb{Y}^\top\mathbb{Y}/\hat{\sigma}^2 = o_p(1).$$

We lastly consider Λ_1, we show that $n^{-1}p^{-1}\mathbb{Y}^\top Z\gamma^\top\gamma Z^\top\mathbb{Y}/\hat{\sigma}^2$ is asymptotically weighted chi-square. By condition (C3), $p^{-1}\gamma^\top\gamma \to \Sigma_\gamma$, we have $n^{-1}p^{-1}\mathbb{Y}^\top Z\gamma^\top\gamma Z^\top\mathbb{Y}/\hat{\sigma}^2 = \mathbb{Y}^\top\{n^{-1}Z\Sigma_\gamma Z^\top\}\mathbb{Y}/\hat{\sigma}^2\{1 + o_p(1)\}$. Moreover, note that $n^{-1}Z^\top Z \to I_d$ by model specification assumption. Thus, the eigenvalues of $n^{-1}Z\Sigma_\gamma Z^\top$ should be equal to the eigenvalues of $n^{-1}\Sigma_\gamma Z^\top Z = \Sigma_\gamma\{1 + o_p(1)\}$. Let λ_i be the ith largest eigenvalue of Σ_γ. Consequently, $n^{-1}p^{-1}\mathbb{Y}^\top Z\gamma^\top\gamma Z^\top\mathbb{Y}/\hat{\sigma}^2$ is asymptotically distributed as $\sum_{i=1}^d \lambda_i\chi_1^2$, which completes the entire proof.

Appendix C. Proof of Proposition 1

Without loss of generality, we only present the proof of the first part, while the second part can be proved in a similar manner. Note that $\mathbb{X} = Z\gamma^\top + \widetilde{\mathbb{X}}$, then we can have $n^{-1}p^{-1}tr\{\mathbb{X}^\top\mathcal{Q}(\hat{Z})\mathbb{X}\} = n^{-1}p^{-1}\{\gamma Z^\top\mathcal{Q}(\hat{Z})Z\gamma^\top\} + n^{-1}p^{-1}tr\{\widetilde{\mathbb{X}}^\top\mathcal{Q}(\hat{Z})\widetilde{\mathbb{X}}\} + 2n^{-1}p^{-1}tr\{\widetilde{\mathbb{X}}^\top\mathcal{Q}(\hat{Z})Z\gamma^\top\} \doteq \Delta_1 + \Delta_2 + \Delta_3$. We next consider the three parts separately. By Lemma 1 and condition (C3), we can have that $n^{-1}p^{-1}\{\gamma Z^\top\mathcal{Q}(\hat{Z})Z\gamma^\top\} \le n^{-1}tr\{Z^\top\mathcal{Q}(\hat{Z})Z\}tr(p^{-1}\gamma^\top\gamma) = O_p(n^{-1}) = o_p(1)$. We next consider Δ_2. Note that d is fixed constant, one can verify that $n^{-1}p^{-1}tr\{\widetilde{\mathbb{X}}^\top\mathcal{Q}(Z)\widetilde{\mathbb{X}}\} = p^{-1}tr(\widetilde{\Sigma})\{1 + o_p(1)\}$. In addition, by Lemma 1 again, we have $n^{-1}p^{-1}tr[\widetilde{\mathbb{X}}^\top\{\mathcal{Q}(Z) - \mathcal{Q}(\hat{Z})\}\widetilde{\mathbb{X}}] \le [tr\{\mathcal{Q}(Z) - \mathcal{Q}(\hat{Z})\}^2]^{1/2}n^{-1}p^{-1}tr(\widetilde{\mathbb{X}}^\top\widetilde{\mathbb{X}}) = O_p(n^{-1}) = o_p(1)$. Consequently, we have $\Delta_2 = p^{-1}tr(\widetilde{\Sigma})\{1 + o_p(1)\}$. We lastly consider Δ_3. Employing Lemma L5 in [19], we have $\lambda_{\max}(p^{-1}Z\gamma^\top\widetilde{\mathbb{X}}^\top) = o_p(1)$, which immediately leads to $\Delta_3 = o_p(1)$. Combining the results above, we have $n^{-1}p^{-1}tr\{\mathbb{X}^\top\mathcal{Q}(\hat{Z})\mathbb{X}\} - p^{-1}tr(\widetilde{\Sigma}) \to 0$, which completes the entire proof.

Appendix D. Proof of Theorem 2

Note that $\hat{\mathbb{Y}} = \hat{\mathbb{X}}\beta + \hat{\varepsilon}$. Thus, $\hat{\beta}_1$ can be decomposed into two parts, $\hat{\beta}_1 = T^1 + T^2$ with $T^1 = (\hat{\mathbb{X}}_1^\top\hat{\mathbb{X}}_1)^{-1}\hat{\mathbb{X}}_1^\top\varepsilon$ and $T_2 = (\hat{\mathbb{X}}_1^\top\hat{\mathbb{X}}_1)^{-1}\hat{\mathbb{X}}_1^\top\hat{\mathbb{X}}_{1*}\beta_{1*}$, where $\hat{\mathbb{X}}_{1*} = (\hat{\mathbb{X}}_j : j \ne 1) \in \mathbb{R}^{n\times(p-1)}$. Consequently, to prove the theorem, it suffices to show that (T^1, T^2) is asymptotical bivariate-normal since T^1 and T^2 are un-correlated. To this end, we consider T^1 and T^2 separately. We firstly consider T^1. Note that $\hat{\mathbb{X}}_1^\top\hat{\mathbb{X}}_1 = \mathbb{X}_1^\top\mathcal{Q}(\hat{Z})\mathbb{X}_1 = \gamma_1^\top Z^\top\mathcal{Q}(\hat{Z})Z\gamma_1 + 2\gamma_1^\top Z^\top\mathcal{Q}(\hat{Z})\widetilde{\mathbb{X}}_1 + \widetilde{\mathbb{X}}_1^\top\mathcal{Q}(\hat{Z})\widetilde{\mathbb{X}}_1$. We further consider the three parts of $\hat{\mathbb{X}}_1^\top\hat{\mathbb{X}}_1$ separately. Firstly by Lemma 1 and condition (C3), we have $\gamma_1^\top Z^\top\mathcal{Q}(\hat{Z})Z\gamma_1 \le \|\gamma_1\|^2 tr\{Z^\top\mathcal{Q}(\hat{Z})Z\} = O_p(1)$. Moreover, employing Lemma L5 in [19], we can obtain that $\gamma_1^\top Z^\top\mathcal{Q}(\hat{Z})\widetilde{\mathbb{X}}_1 = O_p(1)$. Lastly, one can verify that $\widetilde{\mathbb{X}}_1^\top\mathcal{Q}(\hat{Z})\widetilde{\mathbb{X}}_1 = \widetilde{\mathbb{X}}_1^\top\mathcal{Q}(Z)\widetilde{\mathbb{X}}_1\{1 + o_p(1)\}$ by Lemma 1, further noting that $\widetilde{\mathbb{X}}_1^\top\mathcal{Q}(Z)\widetilde{\mathbb{X}}_1/\widetilde{\sigma}_{11}$ follows a chi-square distribution of degree $n - d$. Combining these results above, we can have $n^{-1}\hat{\mathbb{X}}_1^\top\hat{\mathbb{X}}_1 = \widetilde{\sigma}_{11} + o_p(1)$. Moreover,

$\hat{\mathbb{X}}_1^\top \varepsilon = \mathbb{X}_1^\top \mathcal{Q}(\hat{Z})\varepsilon = \gamma_1^\top Z^\top \mathcal{Q}(\hat{Z})\varepsilon + \widetilde{\mathbb{X}}_1^\top \varepsilon$. Since $\text{var}\{\gamma_1^\top Z^\top \mathcal{Q}(\hat{Z})\varepsilon\} = O(1)$ and $\text{var}\{\widetilde{\mathbb{X}}_1^\top \varepsilon\} = O(n)$. As a result, $\widetilde{\mathbb{X}}_1^\top \varepsilon$ should dominate $\gamma_1^\top Z^\top \mathcal{Q}(\hat{Z})\varepsilon$, which leads to $\hat{\mathbb{X}}_1^\top \varepsilon = \widetilde{\mathbb{X}}_1^\top \varepsilon\{1 + o_p(1)\}$. Consequently, we have

$$n^{1/2}T^1 = n^{-1/2}\widetilde{\mathbb{X}}_1^\top \varepsilon/\widetilde{\sigma}_{11}\{1 + o_p(1)\} = \{1 + o_p(1)\}n^{-1/2}\sum_{i=1}^n \widetilde{X}_{i1}\varepsilon_1/\widetilde{\sigma}_{11}.$$

We next consider T^2. Applying the same arguments as those given above, we can obtain that

$$n^{1/2}T^2 = \{1 + o_p(1)\}n^{-1/2}\sum_{i=1}^n \widetilde{X}_{i1}\widetilde{X}_{i1*}\beta_{1*}/\widetilde{\sigma}_{11}.$$

Let $\delta_i = \widetilde{X}_{i1}\varepsilon_1$ and $\eta_i = \widetilde{X}_{i1}\widetilde{X}_{i1*}\beta_{1*}$. Then it can be shown that $E(\delta_i) = 0$ and $\text{var}(\delta_i) = \sigma^2\widetilde{\sigma}_{11}$. Moreover, we have $n^{1/2}E(\eta_i) = n^{1/2}\widetilde{\sigma}_{11*}\beta_{1*} = 0$. Subsequently, $\text{var}(\eta_i) \to E(\eta_i^2) = \widetilde{\sigma}_{11}\beta_{1*}^\top E\{\widetilde{X}_{i1*}\widetilde{X}_{i1*}^\top|\widetilde{X}_{i1}\}\beta_{1*} \to \widetilde{\sigma}_{11}\beta_{1*}^\top\{\widetilde{\Sigma}_{1*} - \widetilde{\Sigma}_{1*1}\widetilde{\Sigma}_{11*}/\widetilde{\sigma}_{11}\}\beta_{1*}$. Consequently, by the bivariate Central Limit Theorem, we can thus obtain

$$\left(n^{1/2}T^1, n^{1/2}T^2\right)^\top = \{1 + o_p(1)\}\left\{n^{-1/2}\sum_{i=1}^n(\delta_i, \eta_i)^\top/\widetilde{\sigma}_{11}\right\}$$

is asymptotically normal with mean 0 and covariance $V = \text{diag}(V_{ii})$. In addition, $V_{11} = \sigma^2/\widetilde{\sigma}_{11}$ and $V_{22} = \beta_{1*}^\top\{\widetilde{\Sigma}_{1*} - \widetilde{\Sigma}_{1*1}\widetilde{\Sigma}_{11*}/\widetilde{\sigma}_{11}\}\beta_{1*}/\widetilde{\sigma}_{11}$. Consequently, $n^{1/2}(T^1 + T^2)$ is asymptotical normal with mean $\mathbf{0}$ and covariance $V_{11} + V_{22}$, which completes the entire proof.

Appendix E. Proof of Proposition 2

According to the proof in [17], to prove the proposition, it suffices to demonstrate the following two results,

$$\frac{1}{p}\sum_{j=1}^p I(p_j \le t) - T_{1,n}(t) \to 0 \text{ a.s.} \quad \text{and} \tag{E.1}$$

$$\frac{1}{N_0}\sum_{j\in\mathcal{N}_0} I(p_j \le t) - G_{0,n}(t) \to 0 \text{ a.s.}, \tag{E.2}$$

as $p \to \infty$, where $G_{0,n}(t) = N_0^{-1}\sum_{j\in\mathcal{N}_0} P(p_j \le t)$. Since the proofs for (E.1) and (E.2) are quite similar, we only verify (E.1). By the law of large numbers [14], it is enough to show that

$$\text{var}\left\{\frac{1}{p}\sum_{j=1}^p I(p_j \le t)\right\} = O(p^{-\delta}) \quad \text{for any } \delta > 0. \tag{E.3}$$

Note that the left part of (E.3) is equivalent to

$$\text{var}\left\{\frac{1}{p}\sum_{j=1}^p I(|Z_j| \ge z_{1-t/2})\right\} = \frac{1}{p^2}\sum_{j=1}^p \text{var}\left\{I(|Z_j| \ge z_{1-t/2})\right\} + \frac{2}{p^2}\sum_{j_1\ne j_2}\text{cov}\left\{I(|Z_{j_1}| \ge z_{1-t/2}), I(|Z_{j_2}| \ge z_{1-t/2})\right\}. \tag{E.4}$$

The first part of (E.4) is $O(p^{-1})$ by the fact that $\text{var}\{I(|Z_j| \ge z_{1-t/2})\} \le 1$. We next consider the second part, the covariance is given by

$$P\left(|Z_{j_1}| \ge z_{1-t/2}, |Z_{j_2}| \ge z_{1-t/2}\right) - P\left(|Z_{j_1}| \ge z_{1-t/2}\right)P\left(|Z_{j_1}| \ge z_{1-t/2}\right).$$

By the proof of Theorem 2, uniformly for any $j = 1, \ldots, p$, we have $P(Z_j > t) = \Phi(-t)\{1 + o_p(1)\}$, where $\Phi(\cdot)$ present the cumulative distribution function of a standard normal distribution. Thus, by the bivariate large deviation result [23] and the similar argument of [13], to prove the theorem, it suffices to show that

$$\sum_{j_2=1}^p \rho_{j_1j_2} = o(p),$$

for any $j_1 = 1, \ldots, p$, where $\rho_{j_1j_2} = \text{cov}(Z_{j_1}, Z_{j_2})$. By the proof of Theorem 2, Z_{j_1} can be written as

$$Z_{j_1} = \{1 + o_p(1)\}n^{-1/2}\sum_{i=1}^n\left(\widetilde{X}_{ij_1}\varepsilon_i + \widetilde{X}_{ij_1}\widetilde{X}_{ij_1^*}\beta_{j_1^*}\right).$$

Consequently, for any $j_1 \neq j_2$, we can have

$$\rho_{j_1j_2} = \{1+o_p(1)\}n^{-1}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\Big\{\text{cov}(\widetilde{X}_{i_1j_1}\varepsilon_{i_1}, \widetilde{X}_{i_2j_2}\varepsilon_{i_2}) + \text{cov}(\widetilde{X}_{i_1j_1}\widetilde{X}_{i_1j_1^*}\beta_{j_1^*}, \widetilde{X}_{i_2j_2}\varepsilon_{i_2})$$
$$+\,\text{cov}(\widetilde{X}_{i_2j_2}\widetilde{X}_{i_2j_2^*}\beta_{j_2^*}, \widetilde{X}_{i_1j_1}\varepsilon_{i_1}) + \text{cov}(\widetilde{X}_{i_1j_1}\widetilde{X}_{i_1j_1^*}\beta_{j_1^*}, \widetilde{X}_{i_2j_2}\widetilde{X}_{i_2j_2^*}\beta_{j_2^*})\Big\}$$
$$= \{1+o_p(1)\}n^{-1}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\Big\{\text{cov}(\widetilde{X}_{i_1j_1}\varepsilon_{i_1}, \widetilde{X}_{i_2j_2}\varepsilon_{i_2}) + \text{cov}(\widetilde{X}_{i_1j_1}\widetilde{X}_{i_1j_1^*}\beta_{j_1^*}, \widetilde{X}_{i_2j_2}\widetilde{X}_{i_2j_2^*}\beta_{j_2^*})\Big\}.$$

We next consider the above two terms separately. One can have

$$n^{-1}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\text{cov}(\widetilde{X}_{i_1j_1}\varepsilon_{i_1}, \widetilde{X}_{i_2j_2}\varepsilon_{i_2}) = n^{-1}\sigma^2\sum_{i=1}^{n}\text{cov}(\widetilde{X}_{ij_1}, \widetilde{X}_{ij_2}) = \widetilde{\sigma}_{j_1j_2} = 0.$$

We next consider the term $n^{-1}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\text{cov}(\widetilde{X}_{i_1j_1}\widetilde{X}_{i_1j_1^*}\beta_{j_1^*}, \widetilde{X}_{i_2j_2}\widetilde{X}_{i_2j_2^*}\beta_{j_2^*})$. Note that

$$n^{-1}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\text{cov}(\widetilde{X}_{i_1j_1}\widetilde{X}_{i_1j_1^*}\beta_{j_1^*}, \widetilde{X}_{i_2j_2}\widetilde{X}_{i_2j_2^*}\beta_{j_2^*}) = n^{-1}\sum_{i=1}^{n}\text{cov}(\widetilde{X}_{ij_1}\widetilde{X}_{ij_1^*}\beta_{j_1^*}, \widetilde{X}_{ij_2}\widetilde{X}_{ij_2^*}\beta_{j_2^*}).$$

Moreover, we can also have $\text{cov}(\widetilde{X}_{ij_1}\widetilde{X}_{ij_1^*}\beta_{j_1^*}, \widetilde{X}_{ij_2}\widetilde{X}_{ij_2^*}\beta_{j_2^*}) = E(\widetilde{X}_{ij_1}\widetilde{X}_{ij_1^*}\beta_{j_1^*}\widetilde{X}_{ij_2}\widetilde{X}_{ij_2^*}\beta_{j_2^*}) - E(\widetilde{X}_{ij_1}\widetilde{X}_{ij_1^*}\beta_{j_1^*})E(\widetilde{X}_{ij_2}\widetilde{X}_{ij_2^*}\beta_{j_2^*})$. By condition (C1) and Lemma 3, together with some algebraic simplifications, we can obtain that

$$\text{cov}(\widetilde{X}_{ij_1}\widetilde{X}_{ij_1^*}\beta_{j_1^*}, \widetilde{X}_{ij_2}\widetilde{X}_{ij_2^*}\beta_{j_2^*}) = \sum_{k_1\in\mathcal{N}_1\neq j_1}\sum_{k_2\in\mathcal{N}_1\neq j_2}(\widetilde{\sigma}_{k_1k_2}\widetilde{\sigma}_{j_1j_2} + \widetilde{\sigma}_{k_1j_2}\widetilde{\sigma}_{k_2j_1})\beta_{k_1}\beta_{k_2}.$$

As a result, we can have $\text{cov}(\widetilde{X}_{ij_1}\widetilde{X}_{ij_1^*}\beta_{j_1^*}, \widetilde{X}_{ij_2}\widetilde{X}_{ij_2^*}\beta_{j_2^*}) = \sum_{j_1\neq j_2}\widetilde{\sigma}_{j_1j_1}\widetilde{\sigma}_{j_2j_2}\beta_{j_1}\beta_{j_2}$ if $(j_1 \neq j_2) \in \mathcal{N}_1$, and $\text{cov}(\widetilde{X}_{ij_1}\widetilde{X}_{ij_1^*}\beta_{j_1^*}, \widetilde{X}_{ij_2}\widetilde{X}_{ij_2^*}\beta_{j_2^*}) = 0$ otherwise. Consequently, we can have that for any $j_1 = 1, \ldots, p$,

$$\sum_{j_2=1}^{p}\text{cov}(\widetilde{X}_{ij_1}\widetilde{X}_{ij_1^*}\beta_{j_1^*}, \widetilde{X}_{ij_2}\widetilde{X}_{ij_2^*}\beta_{j_2^*}) = \sum_{j_1\in\mathcal{N}_1}\sum_{j_2\neq j_1}\widetilde{\sigma}_{j_1j_1}\widetilde{\sigma}_{j_2j_2}\beta_{j_1}\beta_{j_2} = O(N_1) = o(p).$$

Combining these results above, we can thus obtain

$$\sum_{j_2=1}^{p}\rho_{j_1j_2} = o(p),$$

which completes the entire proof.

Appendix F. Proof of Theorem 3

Let p_j and Z_j be the corresponding p-values and test statistic for testing $\beta_j = 0$. Then, to prove the theorem, it suffices to show that the theorem holds for a special sequence of $\alpha_n \to 0$ such that $\alpha_n = 2\{1-\Phi(n^{\jmath})\}$ for some $\xi < \jmath < 1/2-\kappa$. We only need to verify that

$$\lim_{n,p\to\infty}P\Big\{V(\alpha_n) > 0\Big\} \to 0, \quad \text{and} \quad \lim_{n,p\to\infty}P\Big\{S(\alpha_n)/N_1 = 1\Big\} \to 1,$$

where $S(\alpha_n) = \sum_{j\in\mathcal{N}_1}I\{p_j \le \alpha_n\}$. We prove the above parts in the following two steps accordingly.

Step I. We firstly show that $P\{V(\alpha_n) > 0\} \to 0$. Note that $\tau^2_{\beta_j} \ge \sigma^2/\widetilde{\sigma}_{jj}$. Further assume $\widetilde{\sigma}_{jj} = 1$ for notation convenience. Then, we can obtain

$$|Z_j| = \Big\{\Big|(\hat{\mathbb{X}}_j^\top\hat{\mathbb{X}}_j)^{-1}\hat{\mathbb{X}}_j^\top\varepsilon + (\hat{\mathbb{X}}_j^\top\hat{\mathbb{X}}_j)^{-1}\hat{\mathbb{X}}_j^\top\mathbb{X}\beta\Big|\Big\}\Big/(n^{1/2}\hat{\sigma}_{\beta_j})$$
$$\le \sigma^{-1}\Big|(\mathbb{X}_j^\top\mathcal{Q}(\hat{Z})\mathbb{X}_j)^{-1/2}\mathbb{X}_j^\top\mathcal{Q}(\hat{Z})\varepsilon\Big| + \sigma^{-1}\Big|(\mathbb{X}_j^\top\mathcal{Q}(\hat{Z})\mathbb{X}_j)^{-1}\mathbb{X}_j^\top\mathcal{Q}(\hat{Z})\mathbb{X}\beta\Big|.$$

Consequently, by Bonferroni inequality, we obtain

$$P\Big\{V(\alpha_n) > 0\Big\} = P\Big(\max_{j\in\mathcal{N}_0}|Z_j| > z_{1-\alpha_n/2}\Big)$$
$$\le P\Big(\max_{j\in\mathcal{N}_0}\big|(\mathbb{X}_j^\top\mathcal{Q}(\hat{Z})\mathbb{X}_j)^{-1/2}\mathbb{X}_j^\top\mathcal{Q}(\hat{Z})\varepsilon/\sigma\big| > n^{\jmath}/2\Big)$$
$$+\,P\Big(\max_{j\in\mathcal{N}_0}\big|(\mathbb{X}_j^\top\mathcal{Q}(\hat{Z})\mathbb{X}_j)^{-1}\mathbb{X}_j^\top\mathcal{Q}(\hat{Z})\mathbb{X}\beta\big| > \sigma n^{\jmath}/2\Big). \tag{F.5}$$

We then consider the above two parts separately. We first consider the first part of (F.5). According to the results of the proof of Theorems 1 and 2, one can verify that $\mathbb{X}_j^\top \mathcal{Q}(\hat{Z})\mathbb{X}_j = \mathbb{X}_j^\top \mathcal{Q}(Z)\mathbb{X}_j\{1+o_p(1)\}$ and $\mathbb{X}_j^\top \mathcal{Q}(\hat{Z})\varepsilon = \mathbb{X}_j^\top \mathcal{Q}(Z)\varepsilon\{1+o_p(1)\}$ uniformly for any j. As a result, by conditions (C4)–(C5) and Bonferroni inequality again, we have

$$\begin{aligned} P\Big(\max_{j\in\mathcal{N}_0}\big|(\mathbb{X}_j^\top \mathcal{Q}(\hat{Z})\mathbb{X}_j)^{-1/2}\mathbb{X}_j^\top \mathcal{Q}(\hat{Z})\varepsilon/\sigma\big| > n^\jmath/2\Big) &\le P\Big(\max_{j\in\mathcal{N}_0}\big|(\mathbb{X}_j^\top \mathcal{Q}(Z)\mathbb{X}_j)^{-1/2}\mathbb{X}_j^\top \mathcal{Q}(Z)\varepsilon/\sigma\big| > n^\jmath/4\Big) \\ &\le \sum_{j\in\mathcal{N}_0} P\Big(n^{-1/2}\big|(\mathbb{X}_j^\top \mathcal{Q}(Z)\mathbb{X}_j)^{-1/2}\mathbb{X}_j^\top \mathcal{Q}(Z)\varepsilon/\sigma\big| > n^{\jmath-1/2}/4\Big) \\ &\le p\exp(-C_\varepsilon n^{2\jmath}/16) \le \exp(-C_\varepsilon n^{2\jmath}/16 + \nu n^\xi). \end{aligned}$$

Note that $\xi < 2\jmath$ by definition. Thus, the first term of the above equation $-C_\varepsilon n^{2\jmath}/4$ should dominate the second term νn^ξ, which immediately leads to

$$\lim_{n,p\to\infty} P\Big(\max_{j\in\mathcal{N}_0}\big|(\mathbb{X}_j^\top \mathcal{Q}(\hat{Z})\mathbb{X}_j)^{-1/2}\mathbb{X}_j^\top \mathcal{Q}(\hat{Z})\varepsilon/\sigma\big| > n^\jmath/2\Big) \to 0.$$

Furthermore, similar to the proof of Theorem 2, uniformly for any $j = 1, \ldots, p$, we can have

$$\begin{aligned} (\mathbb{X}_j^\top \mathcal{Q}(\hat{Z})\mathbb{X}_j)^{-1/2}\mathbb{X}_j^\top \mathcal{Q}(\hat{Z})\mathbb{X}\beta &= (\mathbb{X}_j^\top \mathcal{Q}(Z)\mathbb{X}_j/n)^{-1/2} n^{1/2}\sum_{j'\ne j}\hat{\varrho}_{jj'}(Z)\beta_{j'}\{1+o_p(1)\} \\ &\le C_{\max}\Big\{\min_j n^{-1}\mathbb{X}_j^\top \mathcal{Q}(Z)\mathbb{X}_j\Big\}^{-1/2}\max_{j'\ne j}|n^{1/2}\hat{\varrho}_{jj'}(Z)|, \end{aligned}$$

where the last inequality is due to the condition that $\sum_j|\beta_j| < C_{\max}$, and $\hat{\varrho}_{jj'}(Z)$ is the sample partial covariance of X_{ij} and $X_{ij'}$ after controlling for the effect of Z. We next consider the above two parts separately. Firstly, by Bonferroni inequality and condition (C1), we have $\max_j|n^{-1}\mathbb{X}_j^\top \mathcal{Q}(Z)\mathbb{X}_j - \widetilde{\sigma}_{jj}| = o_p(1)$. Consequently, we have $\{\min_j n^{-1}\mathbb{X}_j^\top \mathcal{Q}(Z)\mathbb{X}_j\}^{-1/2} = O_p(1)$. We next consider $\max_{j'\ne j}|n^{1/2}\hat{\varrho}_{jj'}(Z)|$. By model assumption that $\mathrm{cov}(\widetilde{X}_{ij'}, \widetilde{X}_{ij}) = 0$ for any $j' \ne j$, we can verify that $\max_{j'\ne j}|n^{1/2}\varrho_{jj'}(Z)| = 0$, where $\varrho_{jj'}(Z)$ is the partial covariance of X_{ij} and $X_{ij'}$ after controlling for the effect of Z. Consequently, using Corollary 1 of Kalisch and Buhlmann [11] together with Bonferroni inequality, we can immediately obtain

$$\max_{j,j'} P\Big\{\max_{j'\ne j}|n^{1/2}\hat{\varrho}_{jj'}(Z)| > O(n^{b/2})\Big\} \to 0$$

for every $\xi < b < 1$. We set $b = (\xi + \jmath)/2$, then we can have $\max_{j'\ne j}|n^{1/2}\hat{\varrho}_{jj'}(Z)| = o(n^{b/2}) = o(n^\jmath)$. This together with the results that $\{\min_j n^{-1}\mathbb{X}_j^\top \mathcal{Q}(Z)\mathbb{X}_j\}^{-1/2} = O_p(1)$ leads to

$$P\Big(\max_{j\in\mathcal{N}_0}\big|(\mathbb{X}_j^\top \mathcal{Q}(Z)\mathbb{X}_j)^{-1}\mathbb{X}_j^\top \mathcal{Q}(Z)\mathbb{X}\beta\big| > \sigma n^\jmath/2\Big) \to 0.$$

Combining these results above, we have completed the first part of (F.5).

Step II. We next consider the second part of (F.5). By definition, we have

$$N_1^{-1}S(\alpha_n) = N_1^{-1}\sum_{j\in\mathcal{N}_1} I\Big(|n^{1/2}\hat{\beta}_j/\hat{\sigma}_{\beta_j}| > n^\jmath\Big).$$

According to the proof of the first part of (F.5), we have $\max_j|n^{1/2}(\hat{\beta}_j - \beta_j)/\hat{\sigma}_{\beta_j}| = o(n^\jmath)$. Moreover, according to theorem assumption that $\min_{j\in\mathcal{N}_1}|\beta_j| \ge C_\beta n^{-\kappa}$ for some constants $C_\beta > 0$, we can immediately have $\min_{j\in\mathcal{N}_1}|n^{1/2}\beta_j/\hat{\sigma}_{\beta_j}| = O(|n^{1/2-\kappa}|)$. Consequently, by Bonferroni inequality and the fact that $\jmath + \kappa < 1/2$, we can obtain

$$\begin{aligned} P\Big(N_1^{-1}S(\alpha_n) = 1\Big) &= P\Big(\min_{j\in\mathcal{N}_1}|n^{1/2}(\hat{\beta}_j - \beta_j)/\hat{\sigma}_{\beta_j} + n^{1/2}\beta_j/\hat{\sigma}_{\beta_j}| > n^\jmath\Big) \\ &\ge P\Big(\min_{j\in\mathcal{N}_1}|n^{1/2}\beta_j/\hat{\sigma}_{\beta_j}| > n^\jmath\Big) - P\Big(\max_{j\in\mathcal{N}_1}|n^{1/2}(\hat{\beta}_j - \beta_j)/\hat{\sigma}_{\beta_j}| > 2n^\jmath\Big) \to 1, \end{aligned}$$

which completes the second part of (F.5). Combining these results above, we have completed the entire proof of Theorem 3.

References

[1] J. Bai, Inferential theory for factor models of large dimensions, Econometrica 71 (2003) 135–171.
[2] J. Bai, S. Ng, Determining the number of factors in approximate factor models, Econometrica 70 (2002) 191–221.
[3] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B 57 (1995) 289–300.
[4] E.F. Fama, K.R. French, Common risk factors in the return on stocks and bonds, J. Financ. Econ. 33 (1993) 3–56.
[5] J. Fan, X. Han, W. Gu, Estimating false discovery proportion under arbitrary covariance dependence, J. Amer. Statist. Assoc. 107 (2012) 1019–1035.
[6] J. Fan, Y. Liao, M. Mincheva, High dimensional covariance matrix estimation in approximate factor models, Ann. Statist. 39 (2011) 3320–3356.
[7] J. Fan, J. Lv, Sure independence screening for ultra-high dimensional feature space (with discussion), J. R. Stat. Soc. Ser. B 70 (2008) 849–911.

[8] J. Fan, R. Song, Sure independence screening in generalized linear models with NP-dimensionality, Ann. Statist. 38 (2010) 3567–3604.
[9] J. Goeman, V. Houwelingen, L. Finos, Testing against a high dimensional alternative in the generalized linear model: asymptotic type I error control, Biometrika 98 (2011) 381–390.
[10] W. Härdle, H. Liang, J. Gao, Partially Linear Models, Springer, Heidelberg, 2000.
[11] M. Kalisch, P. Buhlmann, Estimating high-dimensional directed acyclic graphs with the PC-algorithm, J. Mach. Learn. Res. 8 (2007) 613–636.
[12] W. Lan, H. Wang, C.L. Tsai, Testing covariates in high dimensional regression, Ann. Inst. Statist. Math. 66 (2014) 279–301.
[13] W. Lan, P. Zhong, R. Li, H. Wang, C.L. Tsai, Testing a single regression coefficient in high dimensional linear models, Working paper, 2013.
[14] R. Lyons, Strong laws of large numbers for weakly correlated random variables, Michigan Math. J. 35 (1988) 353–359.
[15] N. Meinshausen, L. Meier, P. Bühlmann, *P*-values for high-dimensional regression, J. Amer. Statist. Assoc. 104 (2009) 1671–1681.
[16] W.F. Sharpe, Capital asset prices: A theory of market equilibrium under conditions of risk, J. Finance 19 (1964) 425–442.
[17] J.D. Storey, J.E. Taylor, D. Siegmund, Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach, J. R. Stat. Soc. Ser. B 66 (2004) 187–205.
[18] H. Wang, Forward regression for ultra-high dimensional variable screening, J. Amer. Statist. Assoc. 104 (2009) 1512–1524.
[19] H. Wang, Factor profiled independence screening, Biometrika 99 (2012) 15–28.
[20] L. Wasserman, K. Roeder, High dimensional variable selection, Ann. Statist. 37 (2009) 2178–2201.
[21] C.H. Zhang, S.S. Zhang, Confidence intervals for low dimensional parameters in high dimensional linear models, J. R. Stat. Soc. Ser. B 76 (2014) 217–242.
[22] P.S. Zhong, S.X. Chen, Tests for high dimensional regression coefficients with factorial designs, J. Amer. Statist. Assoc. 106 (2011) 260–274.
[23] P.S. Zhong, S.X. Chen, M. Xu, Tests alternative to higher criticism for high dimensional means under sparsity and column-wise dependence, Ann. Statist. 41 (2013) 2820–2851.