The Annals of Statistics
2024, Vol. 52, No. 5, 2034–2058
https://doi.org/10.1214/24-AOS2420

TESTING HIGH-DIMENSIONAL REGRESSION COEFFICIENTS IN LINEAR MODELS

BY ALEX ZHAO[1,a], CHANGCHENG LI[2,d], RUNZE LI[1,b] AND ZHE ZHANG[1,c]

[1]*Department of Statistics, Pennsylvania State University at University Park*, [a]*alexzhao@alumni.psu.edu*, [b]*rzli@psu.edu*, [c]*zxz288@psu.edu*

[2]*School of Mathematical Sciences, Dalian University of Technology*, [d]*lichangcheng@dlut.edu.cn*

This paper is concerned with statistical inference for regression coefficients in high-dimensional linear regression models. We propose a new method for testing the coefficient vector of the high-dimensional linear models, and establish the asymptotic normality of our proposed test statistic with the aid of the martingale central limit theorem. We derive the asymptotical relative efficiency (ARE) of the proposed test with respect to the test proposed in Zhong and Chen (*J. Amer. Statist. Assoc.* **106** (2011) 260–274), and show that the ARE is always greater or equal to one under the local alternative studied in this paper. Our numerical studies imply that the proposed test with critical values derived from its asymptotical normal distribution may retain Type I error rate very well. Our numerical comparison demonstrates the proposed test performs better than existing ones in terms of powers. We further illustrate our proposed method with a real data example.

1. Introduction.

Hypothesis testing in high-dimensional data is an important problem, with applications ranging from genomics to finance, as illustrated by Chen et al. (2011) and Chudik, Kapetanios and Pesaran (2018). This paper aims to develop a new approach towards conducting hypothesis testing on the entire coefficient vector in the setting of high-dimensional linear regression models where the dimension size p of the coefficient vector may be larger than the sample size of observations n.

Traditional low-dimensional approaches such as the F-test for regression coefficient testing are unable to be used for inference in a high-dimensional setting due to insufficient degrees of freedom. Hence, new approaches are needed for the problem of high-dimensional hypothesis testing of regression coefficients. To this end, most existing methods treated hypothesis testing on the high-dimensional regression coefficient vector $\boldsymbol{\beta}$ as a large scale multiple hypothesis test, notably Liu and Luo (2014) in the one-sample case and Xia, Cai and Cai (2018) in the two-sample case. There are only a few existing proposals for tests for linear regression coefficients in the high-dimensional setting. Goeman, van de Geer and van Houwelingen (2006) introduced a global test for multiple coefficients constructed as a score test on the hyperparameter of an empirical Bayesian model, specifically on the variance of the distribution where the coefficients are assumed to be from. Zhong and Chen (2011) proposed a U-statistic for tests on high-dimensional regression coefficients that also works with factorial designs. Lan, Wang and Tsai (2014) proposed a new testing procedure for a set of regression coefficients by using the partial covariances between the response variable and the tested covariates. Cui, Guo and Zhong (2018) developed a testing procedure for high-dimensional regression coefficients by constructing a U-statistic of order two and proposed to use refitted cross-validation (RCV) for estimating the error variance σ^2.

In this paper, we propose to formulate the high-dimensional regression coefficient testing problem as an equivalent high-dimensional one-sample mean testing problem using the score

Received July 2022; revised November 2023.
MSC2020 subject classifications. Primary 62J05, 62F03; secondary 62J20.
Key words and phrases. Hotelling T^2-test, power enhancement, test of high-dimensional means.

function. With the aid of this formulation, one can directly construct a high-dimensional regression coefficient testing procedure by utilizing existing high-dimensional mean tests such as Bai and Saranadasa (1996) and Chen and Qin (2010). To improve on the power of existing testing procedures, we propose a new test statistic based on the test of Chen and Qin (2010) for the score function mean with an additional power-enhancement term, which utilizes the correlations among the predictors to improve power. We establish the asymptotic normality of the proposed test statistic with the aid of the martingale central limit theorem. We further show the superior asymptotic relative efficiency of our proposed test with respect to some of the existing ones by incorporating the power-enhancement term. We conduct Monte Carlo simulations to examine Type I error rates. Our numerical results demonstrate that the asymptotic normal distribution approximates the distribution of the proposed test reasonably well, and its percentiles can serve as the critical values of the proposed test. We also conduct Monte Carlo simulations to compare the performance of the newly proposed test with existing ones. Our numerical comparison shows that the newly proposed test outperforms the existing ones in terms of power. We further illustrate the proposed test via an empirical analysis of a breast cancer data set.

The remainder of this paper is organized as follows. In Section 2, we propose a new test statistic for the linear regression coefficient testing problem based on the equivalent mean vector testing problem, and establish the asymptotic normality of our test statistic under both the null and the local alternative hypotheses. Section 3 provides numerical results from Monte Carlo simulations and also a real data analysis. Some concluding remarks are given in Section 4, and additional numerical results are given in the Supplementary Material (Zhao et al. (2024)) of this paper.

2. Testing regression coefficients.

Consider the following linear regression model:

$$y = b_0 + \boldsymbol{x}^\top \boldsymbol{\beta} + \varepsilon, \tag{1}$$

where y is the response, $\boldsymbol{x}$ is a p-dimensional random vector with mean $\boldsymbol{\mu}_X$ and covariance matrix Σ_X, $\boldsymbol{\beta}$ is the regression coefficient vector, b_0 is the intercept, and ε is the random error independent from $\boldsymbol{x}$ with $\mathrm{E}(\varepsilon|\boldsymbol{x}) = 0$ and $\mathrm{Var}(\varepsilon|\boldsymbol{x}) = \sigma^2$.

2.1. *A test with power enhancement.*

Suppose that $\{\boldsymbol{x}_i, y_i\}$, $i = 1, \dots, n$, are independent and identically distributed (i.i.d.) random samples from model (1). Of interest is to test the hypothesis

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \quad \text{versus} \quad H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0, \tag{2}$$

where $\boldsymbol{\beta}_0$ is known. To eliminate the intercept, we may center both $\boldsymbol{x}$ and y, and consider

$$y - \mu_Y = (\boldsymbol{x} - \boldsymbol{\mu}_X)^\top \boldsymbol{\beta} + \varepsilon,$$

where $\boldsymbol{\mu}_X = \mathrm{E}(\boldsymbol{x})$ and $\mu_Y = \mathrm{E}(y)$. Thus, the score equation for estimating $\boldsymbol{\beta}$ is

$$\sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu}_X)\big(y_i - \mu_Y - (\boldsymbol{x}_i - \boldsymbol{\mu}_X)^\top \boldsymbol{\beta}\big) = \mathbf{0}.$$

This motivates us to define $\boldsymbol{z}_i = (\boldsymbol{x}_i - \boldsymbol{\mu}_X)(y_i - \mu_Y - (\boldsymbol{x}_i - \boldsymbol{\mu}_X)^\top \boldsymbol{\beta}_0)$ for the purpose of testing (2). Denote $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{z}_i)$, which gives us

$$\boldsymbol{\mu} = \mathrm{E}\big\{(\boldsymbol{x}_i - \boldsymbol{\mu}_X)\big((\boldsymbol{x}_i - \boldsymbol{\mu}_X)^\top (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \varepsilon_i\big)\big\} = \Sigma_X (\boldsymbol{\beta} - \boldsymbol{\beta}_0).$$

Hence, for nonsingular Σ_X, testing the hypothesis in (2) is equivalent to a one-sample mean testing problem on the mean of the score function

$$H_0 : \boldsymbol{\mu} = \mathbf{0} \quad \text{versus} \quad H_1 : \boldsymbol{\mu} \neq \mathbf{0}. \tag{3}$$

Thus, we may apply one-sample mean tests for z_i, $i = 1, \dots, n$, for testing on regression coefficient vectors.

In practice, $\boldsymbol{\mu}_X$ and μ_Y are unknown. It is natural to use the sample means $\bar{\boldsymbol{x}}$ and $\bar{y}$ to estimate $\boldsymbol{\mu}_X$ and μ_Y, respectively. Thus, with a slight abuse of notation, we define

$$z_i = (\boldsymbol{x}_i - \bar{\boldsymbol{x}})\big(y_i - \bar{y} - (\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\top \boldsymbol{\beta}_0\big)$$

for the purpose of hypothesis testing (3). As shown in Section 2.3.1, the estimate errors due to $\bar{\boldsymbol{x}}$ and $\bar{y}$ are negligible.

While literature on testing high-dimensional regression coefficients is limited, there is a much wider scope of work dedicated to mean vector testing in high-dimensional settings in the last two decades. For an overview, see Huang et al. (2022), a recent review article on this topic, and references therein. Furthermore, existing high-dimensional regression coefficient tests such as Zhong and Chen (2011) and Cui, Guo and Zhong (2018) can be seen as extensions of sum-of-squares high-dimensional mean testing approaches (e.g., Bai and Saranadasa (1996) and Chen and Qin (2010)) to the regression coefficient testing setting.

For the one-sample mean testing problem (3), we define a statistic that we will use to construct a test statistic T_n in (5) below

$$W_n = \frac{2}{n(n-1)} \sum_{i<j} \big[z_i^\top z_j + k_n \boldsymbol{\alpha}^\top z_i z_j^\top \boldsymbol{\alpha}\big], \tag{4}$$

where $\boldsymbol{\alpha}$ is a p-dimensional vector with $\|\boldsymbol{\alpha}\| = 1$, and k_n is a positive number. Conditions on k_n and $\boldsymbol{\alpha}$ can be found in Theorem 1. We conduct some numerical comparison on the different choices of k_n and $\boldsymbol{\alpha}$ in Section 3.2. The first term in W_n is a U-statistic, which is similar to the test proposed for the two-sample high-dimensional mean problem in Chen and Qin (2010) and the test for regression coefficients in high-dimensional linear models in Cui, Guo and Zhong (2018). The second term in W_n is designed to enhance the power of the test. The construction of the second term for power enhancement is partly motivated by Cui et al. (2020), where power-enhancing artificial data points are used in the empirical likelihood test for the one-sample large-dimensional mean problem in which it is assumed that $p/n \to c \in [1, \infty)$. It is noteworthy that the expectation of W_n is zero under the null hypothesis when z is constructed using the true population means and the expectation is asymptotically zero under the null hypothesis when z is constructed using the sample means.

To construct the test statistic based on W_n, we further need to derive its asymptotic variance. In Section 2.3.1, we show that under the null hypothesis and the conditions that $\boldsymbol{x}$ and ε have finite fourth order moments,

$$\operatorname{Var}(W_n) = \sigma_n^2,$$

where

$$\sigma_n^2 = \frac{2}{n^2}\big[\operatorname{tr}\big(\operatorname{Cov}(z)^2\big) + 2k_n \boldsymbol{\alpha}^\top \operatorname{Cov}(z)^2 \boldsymbol{\alpha} + k_n^2 \big(\boldsymbol{\alpha}^\top \operatorname{Cov}(z)\boldsymbol{\alpha}\big)^2\big].$$

Note that if we define

$$\Omega = \Sigma_Z + k_n \Sigma_Z^{1/2} \boldsymbol{\alpha}\boldsymbol{\alpha}^\top \Sigma_Z^{1/2},$$

then $\sigma_n^2 = \frac{2}{n^2} \operatorname{tr}(\Omega^2)$. We further propose our test statistic T_n based on W_n as

$$T_n = \frac{n W_n}{\sqrt{2 \widehat{\operatorname{tr}(\Omega^2)}}}, \tag{5}$$

where $\widehat{\mathrm{tr}(\Omega^2)}$ is defined as

$$\widehat{\mathrm{tr}(\Omega^2)} = \mathrm{tr}(\widehat{\Sigma_Z^2}) + 2k_n\boldsymbol{\alpha}^\top\widehat{\Sigma_Z^2}\boldsymbol{\alpha} + k_n^2(\boldsymbol{\alpha}^\top\hat{\Sigma}_Z\boldsymbol{\alpha})^2,$$

$\hat{\Sigma}_Z$ is the sample covariance of z, and

$$\begin{aligned}\widehat{\Sigma_Z^2} &= \frac{1}{n(n-1)}\sum_{i\neq j}\Big(z_i z_j^\top - \frac{1}{n(n-1)}\sum_{i\neq j} z_i z_j^\top\Big)^2 \\ &= \frac{1}{n(n-1)}\sum_{i\neq j}(z_i z_j^\top)^2 - \Big(\frac{1}{n(n-1)}\sum_{i\neq j} z_i z_j^\top\Big)^2.\end{aligned} \tag{6}$$

The rationale in the derivation of these estimators is similar to that of the derivation of the test statistic W_n. A similar estimation method has also been used by the asymptotic variance estimators in Chen and Qin (2010). Additionally, Section 2.3.1 provides a more detailed derivation, and shows that $\frac{2}{n^2}\widehat{\mathrm{tr}(\Omega^2)}$ is a ratio consistent estimator for σ_n^2 under the local alternative hypothesis (7).

2.2. *Limiting distributions and asymptotic relative efficiency.* We next establish the asymptotic normality of the test statistic T_n under the null and local alternative hypotheses. We impose the following two assumptions to facilitate technical proofs.

A1. $\boldsymbol{x}_i = \boldsymbol{\mu}_X + \Gamma\boldsymbol{x}_{0i}$, where $\boldsymbol{x}_{0i}$, $i = 1, \ldots, n$, are i.i.d. random vectors of length p with elements having mean zero, variance one, and bounded fourth order moments, $\Gamma = \Sigma_X^{1/2}$, $\boldsymbol{x}_{0i} = (x_{0i1}, \ldots, x_{0ip})^\top$ and $\mathrm{E}(x_{0i\alpha_1}^{l_{\alpha_1}} \cdots x_{0i\alpha_s}^{l_{\alpha_s}}) = \mathrm{E}(x_{0i\alpha_1}^{l_{\alpha_1}}) \cdots \mathrm{E}(x_{0i\alpha_s}^{l_{\alpha_s}})$ for $1 \le \alpha_1 < \cdots < \alpha_s \le p$ and $\sum_{i=1}^s l_{\alpha_i} \le 4$.

A2. ε_i, $i = 1, \ldots, n$, are i.i.d. random variables with mean zero and variance σ^2, and bounded fourth order moments.

Assumptions A1 and A2 are quite mild. Assumption A1 is similar to the independent component assumption widely used in high-dimensional literature (Bai and Saranadasa (1996)). Assumption A2 is a mild assumption on the distribution of the random error.

We further assume that the local alternative has the following form:

$$H_a : \boldsymbol{\beta} = \boldsymbol{\beta}_0 + n^{-1/2}\delta\boldsymbol{u} \quad \text{with } |\delta| \le C \text{ and } \|\boldsymbol{u}\| = 1, \tag{7}$$

where C is a nonnegative constant independent of p, and $\|\cdot\|$ is the Euclidean L_2 norm.

THEOREM 1. *Suppose that Assumptions A1 and A2 hold and that* $\lambda_{\max}(\Sigma_X) = o(\sqrt{\min(n, p)})$, *where* $\lambda_{\max}(\Sigma_X)$ *is the largest eigenvalue of* Σ_X. *Further assume that* $\boldsymbol{\alpha}$ *is a* p*-dimensional vector with* $\|\boldsymbol{\alpha}\| = 1$, *that* $k_n\boldsymbol{\alpha}^\top\Sigma_X\boldsymbol{\alpha} = o(\sqrt{\mathrm{tr}(\Sigma_X^2)})$ *and that* $k_n = o(\sqrt{p})$ *as* $n \to \infty$. *Then*:

(i) *under the null hypothesis* H_0, $T_n \to N(0, 1)$ *in distribution as* $n \to \infty$;

(ii) *under the local alternative hypothesis* H_a *in* (7),

$$T_n - \frac{n\|\boldsymbol{\delta}_z\|^2 + nk_n|\boldsymbol{\alpha}^T\boldsymbol{\delta}_z|^2}{\{2\,\mathrm{tr}(\Omega^2)\}^{1/2}} \xrightarrow{d} N(0, 1) \quad \text{as } n \to \infty, \tag{8}$$

where $\boldsymbol{\delta}_z = \Sigma_X(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$, *and* $\|\cdot\|$ *is the Euclidean* L_2 *norm.*

Theorem 1 establishes the asymptotic distribution of T_n under the null hypothesis and the local alternative (7). The proof of Theorem 1 is given in Section 2.3.

From Theorem 1, the test with the rejection region $\{T_n > z_a\}$ has asymptotic size a, where z_a is the right a quantile of the standard normal distribution. Furthermore, we can obtain that the asymptotic power of the proposed test of size a under the local assumption (7) is

$$\beta(\boldsymbol{\delta}_z) = \Phi\left(-z_a + \frac{n\|\boldsymbol{\delta}_z\|^2 + nk_n|\boldsymbol{\alpha}^T\boldsymbol{\delta}_z|^2}{\{2\operatorname{tr}(\Omega^2)\}^{1/2}}\right), \tag{9}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Comparing the asymptotic power expression (9) with $\beta_{\mathrm{ZC}}(\boldsymbol{\delta}_z)$ in Zhong and Chen (2011),

$$\beta_{\mathrm{ZC}}(\boldsymbol{\delta}_z) = \Phi\left(-z_a + \frac{n\|\boldsymbol{\delta}_z\|^2}{\{2\operatorname{tr}(\Sigma_X^2\sigma^4)\}^{1/2}}\right),$$

we get the asymptotic relative efficiency (ARE) of the proposed test with respect to Zhong and Chen (2011)

$$\begin{aligned}\mathrm{ARE} &= \frac{\|\boldsymbol{\delta}_z\|^2 + k_n|\boldsymbol{\alpha}^\top\boldsymbol{\delta}_z|^2}{\|\boldsymbol{\delta}_z\|^2} \times \frac{\operatorname{tr}(\Sigma_X^2\sigma^4)^{1/2}}{\operatorname{tr}(\Omega^2)^{1/2}} \\ &= \frac{\operatorname{tr}(\Sigma_X^2\sigma^4)^{1/2}}{\operatorname{tr}(\Omega^2)^{1/2}}\left\{1 + \frac{k_n|\boldsymbol{\alpha}^\top\boldsymbol{\delta}_z|^2}{\|\boldsymbol{\delta}_z\|^2}\right\} \\ &\to 1 + \frac{k_n|\boldsymbol{\alpha}^\top\boldsymbol{\delta}_z|^2}{\|\boldsymbol{\delta}_z\|^2}.\end{aligned}$$

The last limit is because $\operatorname{tr}(\Omega^2) = \operatorname{tr}(\Sigma_X^2\sigma^4)$ under the null hypothesis, and that $\operatorname{tr}(\Omega^2) = \operatorname{tr}(\Sigma_X^2\sigma^4)(1 + o(1))$ under the local alternative in (7). Hence, the ARE for the proposed test with respect to the ZC test is always greater than or equal to one under the local alternative (7). Moreover, if k_n and $\boldsymbol{\alpha}$ are chosen such that $\frac{k_n|\boldsymbol{\alpha}^\top\boldsymbol{\delta}_z|^2}{\|\boldsymbol{\delta}_z\|^2}$ is bounded away from zero, then the ARE of the proposed test with respect to the test in Zhong and Chen (2011) is always greater than one. Note that for $\frac{k_n|\boldsymbol{\alpha}^\top\boldsymbol{\delta}_z|^2}{\|\boldsymbol{\delta}_z\|^2}$ to be large for a given k_n, we want the direction of $\boldsymbol{\alpha}$ to be similar to that of $\boldsymbol{\delta}$. In practice, we choose $\boldsymbol{\alpha} = p^{-1/2}(1, 1, \dots, 1)^\top$, and obtain good numerical results. We also experiment with other choices of $\boldsymbol{\alpha}$. Details and discussion of the results are in the numerical studies section.

REMARK 1. One natural question for the proposed method is how to choose tuning parameters such as $\boldsymbol{\alpha}$ and k_n. Note that because Ω depends on $\boldsymbol{\alpha}$ and k_n, it seems that deriving an exact optimal choice of $\boldsymbol{\alpha}$ in terms of reaching the best power based on (9) would be very difficult. It is expected that the exact optimal value of $\boldsymbol{\alpha}$ depends on both k_n and additional unknown parameters and is of a complex expression. Nonetheless, from the asymptotic expression for ARE, we may obtain that under the conditions in Theorem 1, the asymptotically optimal direction for $\boldsymbol{\alpha}$ is $\boldsymbol{\delta}_z/\|\boldsymbol{\delta}_z\|$, where $\boldsymbol{\delta}_z = \Sigma_X(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$, which depends on unknown parameters Σ_X and $\boldsymbol{\beta}$. Thus, if we use the same data to do the calculation of $\boldsymbol{\alpha}$ as well as the hypothesis test, it would make the limiting null distribution of the test statistic quite complicated. Alternatively, one could use a data-splitting procedure, where part of the data is used to estimate $\boldsymbol{\alpha}$ and a different part is used for hypothesis testing. This may result in a loss of power due to the reduction of sample size for testing. Further study on this is beyond the scope of this paper, but it would be an interesting topic of future research.

REMARK 2. From the asymptotic expression for ARE, it seems that larger k_n will lead to larger ARE. However, note that the asymptotic expression for ARE is derived under the conditions of Theorem 1, especially the conditions that $k_n\boldsymbol{\alpha}^\top\Sigma_X\boldsymbol{\alpha} = o(\sqrt{\operatorname{tr}(\Sigma_X^2)})$ and that $k_n = o(\sqrt{p})$ as $n \to \infty$. Also note that when the largest eigenvalue of Σ_X is bounded, the conditions on k_n can be simplified into $k_n = o(\sqrt{p})$.

To study the optimality of the proposed test, we follow the random effects setting in Arias-Castro, Candès and Plan (2011a), and consider a random effects model (REM) with the following prior π on the regression coefficient vector in the alternatives: the regression vector $\boldsymbol{\beta}$ has $S = p^{1-s}$ nonzero coefficients following i.i.d. sub-Gaussian distributions with mean zero and variance γ^2, and the support of $\boldsymbol{\beta}$ is uniformly randomly generated among the size-S subsets of $\{1, 2, \dots, p\}$. We have the following optimality results.

THEOREM 2. *Suppose that Assumptions A1 and A2 hold, and that* $\lambda_{\max}(\Sigma_X) = o(\sqrt{\min(n, p)})$. *Let*

$$\gamma_0^2 = \frac{p}{nS\sqrt{\operatorname{tr}(\Sigma_X^2)}},$$

then under the alternative hypothesis as specified in the REM model with $s \in [0, 1/2)$:

(i) *further assuming that the random error* ϵ_i, $i = 1, \dots, n$, *are i.i.d. Gaussian random variables with mean zero and variance* σ^2, *then all sequences of tests are asymptotically powerless if* $\gamma^2/\gamma_0^2 \to 0$ *as* $n \to \infty$;

(ii) *the proposed test is asymptotically powerful when* $\boldsymbol{\alpha}$ *is a* p-*dimensional vector with* $\|\boldsymbol{\alpha}\| = 1$ *and* $k_n \boldsymbol{\alpha}^\top \Sigma_X \boldsymbol{\alpha} = o(\sqrt{\operatorname{tr}(\Sigma_X^2)})$ *if* $\gamma^2/\gamma_0^2 \to \infty$ *as* $n \to \infty$.

In Theorem 2, the REM is assumed for the derivation of the threshold γ_0^2. Similar models are also used in other studies on optimality, such as Arias-Castro, Candès and Plan (2011a). In the REM model, s measures the sparsity level of the regression coefficient, and γ^2 measures the signal strength. Under the REM with moderate sparsity $s \in [0, \frac{1}{2})$, Theorem 2 establishes a signal strength threshold γ_0^2, and that if the ratio between the signal strength γ^2 and the threshold γ_0^2 tends to zero, then every test is asymptotically powerless; while the ratio γ^2/γ_0^2 tends to infinity, the proposed test is asymptotically powerful. Such results illustrate the optimality of the proposed test. The proof of Theorem 2 is given in Section 2.3.

2.3. *Additional technical details and proof of Theorems 1 and 2.* We now provide additional technical details for theoretical derivations. In Section 2.3.1, we derive the explicit forms of the mean and variance of W_n. The proofs of Theorems 1 and 2 are given in Sections 2.3.2 and 2.3.3, respectively.

2.3.1. *Derivation of mean and variance of* W_n. Let us start with the scenario in which μ_X and μ_Y are known. That is, $z_i = (\boldsymbol{x}_i - \boldsymbol{\mu}_X)(y_i - \mu_Y - (\boldsymbol{x}_i - \boldsymbol{\mu}_X)^\top \boldsymbol{\beta}_0)$. Thus,

$$W_n = \frac{2}{n(n-1)} \sum_{i<j} [z_i^\top z_j + k_n \boldsymbol{\alpha}^\top z_i z_j^\top \boldsymbol{\alpha}].$$

For $\mathrm{E}(W_n)$, we have

$$\mathrm{E}(W_n) = \mathrm{E}(z_i^\top z_j) + k_n \mathrm{E}(\boldsymbol{\alpha}^\top z_i z_j^\top \boldsymbol{\alpha}) = \|\mathrm{E}(z)\|^2 + k_n |\boldsymbol{\alpha}^\top \mathrm{E}(z)|^2.$$

Under the null hypothesis, $\mathrm{E}(z) = \mathbf{0}$. Therefore, it follows that

$$\mathrm{E}(W_n) = 0.$$

Furthermore, we have

$$\begin{aligned}
\operatorname{Var}(W_n) &= \frac{4}{n^2(n-1)^2} \sum_{i<j} \operatorname{Var}(W_{nij}) + \frac{1}{n^2(n-1)^2} \sum_{i \neq j \neq k} \operatorname{Cov}(W_{nij}, W_{nik}) \\
&= \frac{2}{n^2} \operatorname{Var}(W_{n12})(1 + o(1)) + O\left(\frac{1}{n}\right) \operatorname{Cov}(W_{n12}, W_{n13}),
\end{aligned}$$

where $W_{nij} = z_i^\top z_j + k_n \boldsymbol{\alpha}^\top z_i z_j^\top \boldsymbol{\alpha}$. Thus, it follows that

$$\operatorname{Var}(W_{n12}) = \operatorname{Var}(z_1^\top z_2) + 2k_n \operatorname{Cov}(z_1^\top z_2, \boldsymbol{\alpha}^\top z_1 z_2^\top \boldsymbol{\alpha}) + k_n^2 \operatorname{Var}(\boldsymbol{\alpha}^\top z_1 z_2^\top \boldsymbol{\alpha}).$$

From the law of total variance, we further have

$$\operatorname{Var}(z_1^\top z_2) = \operatorname{tr}(\operatorname{Cov}(z)^2) + 2\mathrm{E}(z)^\top \operatorname{Cov}(z)\mathrm{E}(z).$$

Similarly, from the law of total variance, for $\operatorname{Var}(\boldsymbol{\alpha}^\top z_1 z_2^\top \boldsymbol{\alpha})$, we have

$$\operatorname{Var}(\boldsymbol{\alpha}^\top z_1 z_2^\top \boldsymbol{\alpha}) = (\boldsymbol{\alpha}^\top \operatorname{Cov}(z)\boldsymbol{\alpha})^2 + 2\boldsymbol{\alpha}^\top \operatorname{Cov}(z)\boldsymbol{\alpha}(\boldsymbol{\alpha}^\top \mathrm{E}(z))^2.$$

Additionally, for $\operatorname{Cov}(z_1^\top z_2, \boldsymbol{\alpha}^\top z_1 z_2^\top \boldsymbol{\alpha})$, we have

$$\operatorname{Cov}(z_1^\top z_2, \boldsymbol{\alpha}^\top z_1 z_2^\top \boldsymbol{\alpha}) = \boldsymbol{\alpha}^\top \operatorname{Cov}(z)^2 \boldsymbol{\alpha} + 2\mathrm{E}(z)^\top \operatorname{Cov}(z)\boldsymbol{\alpha}\boldsymbol{\alpha}^\top \mathrm{E}(z).$$

Furthermore, under the null hypothesis, it is easy to have that $\operatorname{Cov}(W_{n12}, W_{n13}) = 0$. In sum, under the null hypothesis, we have

$$\operatorname{Var}(W_n) = \sigma_n^2,$$

where

$$\sigma_n^2 = \frac{2}{n^2}\big[\operatorname{tr}(\operatorname{Cov}(z)^2) + 2k_n \alpha^\top \operatorname{Cov}(z)^2 \alpha + k_n^2 (\boldsymbol{\alpha}^\top \operatorname{Cov}(z)\boldsymbol{\alpha})^2\big].$$

Consider the following estimator for σ_n^2:

$$\hat{\sigma}_n^2 = \frac{2}{n^2}\big[\operatorname{tr}(\widehat{\Sigma_Z^2}) + 2k_n \alpha^\top \widehat{\Sigma_Z^2} \alpha + k_n^2 (\boldsymbol{\alpha}^\top \hat{\Sigma}_Z \boldsymbol{\alpha})^2\big],$$

where $\hat{\Sigma}_Z$ is the sample covariance of z, and

$$\begin{aligned}\widehat{\Sigma_Z^2} &= \frac{1}{n(n-1)} \sum_{i \neq j} \Big(z_i z_j^\top - \frac{1}{n(n-1)} \sum_{i \neq j} z_i z_j^\top \Big)^2 \\ &= \frac{1}{n(n-1)} \sum_{i \neq j} (z_i z_j^\top)^2 - \Big(\frac{1}{n(n-1)} \sum_{i \neq j} z_i z_j^\top \Big)^2.\end{aligned}$$

Note that a similar estimation method has also been used by the asymptotic variance estimator derivation in Chen and Qin (2010). In fact, $\operatorname{tr}(\widehat{\Sigma_Z^2})$ and $\boldsymbol{\alpha}^\top \widehat{\Sigma_Z^2} \boldsymbol{\alpha}$ are ratio consistent estimators for $\operatorname{tr}(\operatorname{Cov}(z)^2)$ and $\boldsymbol{\alpha}^\top \operatorname{Cov}(z)^2 \boldsymbol{\alpha}$ under the local alternative hypothesis, respectively. Take $\boldsymbol{\alpha}^\top \widehat{\Sigma_Z^2} \boldsymbol{\alpha}$, for example. Under the local alternative (7), we have

$$\boldsymbol{\alpha}^\top \operatorname{Cov}(z)^2 \boldsymbol{\alpha} = \boldsymbol{\alpha}^\top \big(\mathrm{E}(zz^\top) - \mathrm{E}(z)\mathrm{E}(z)^\top\big)^2 \boldsymbol{\alpha} = \boldsymbol{\alpha}^\top \big(\mathrm{E}(zz^\top)\big)^2 \boldsymbol{\alpha}\big(1 + o(1)\big),$$

where the last equality is from $\lambda_{\min}(\mathrm{E}(zz^\top)) = O(1)$ and that $\|\mathrm{E}(z)\| = o(1)$, which follows from

$$\begin{aligned}\|\mathrm{E}(z)\| &= \big\|\mathrm{E}\big[(\boldsymbol{x} - \boldsymbol{\mu}_X)\big(y - \mu_Y - (\boldsymbol{x} - \boldsymbol{\mu}_X)^\top \boldsymbol{\beta}_0\big)\big]\big\| \\ &= \big\|\mathrm{E}\big[(\boldsymbol{x} - \boldsymbol{\mu}_X)(\boldsymbol{x} - \boldsymbol{\mu}_X)^\top (\boldsymbol{\beta} - \boldsymbol{\beta}_0)\big]\big\| \\ &= O\big(n^{-1/2} \lambda_{\max}(\Sigma_X)\big) = o(1)\end{aligned}$$

under the local alternative (7). Similarly, under the local alternative (7), we have

$$\boldsymbol{\alpha}^\top \widehat{\Sigma_Z^2} \boldsymbol{\alpha} = \frac{1}{n(n-1)} \Big[\sum_{i \neq j} \boldsymbol{\alpha}^\top (z_i z_j^\top)^2 \boldsymbol{\alpha} \Big] \big(1 + o(1)\big).$$

Furthermore, we have $\mathrm{E}(\boldsymbol{\alpha}^\top(z_i z_j^\top)^2\boldsymbol{\alpha}) = \boldsymbol{\alpha}^\top(\mathrm{E}(zz^\top))^2\boldsymbol{\alpha}$. Hence, from the theory of the U-statistic, $\frac{1}{n(n-1)}[\sum_{i\neq j}\boldsymbol{\alpha}^\top(z_i z_j^\top)^2\boldsymbol{\alpha}]$ is a consistent U-statistic estimator for $\boldsymbol{\alpha}^\top(\mathrm{E}(zz^\top))^2\boldsymbol{\alpha}$. Additionally, we can further show that $\hat{\sigma}_n^2$ is a ratio consistent estimator for σ_n^2 under local alternatives.

We next turn to the case where μ_X and μ_Y are unknown. We calculate z using the sample means as $z_i = (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(y_i - \bar{y} - (\boldsymbol{x}_i - \bar{\boldsymbol{x}})^\top\boldsymbol{\beta}_0)$. For clarity, here we use z and W_n to refer to z and W_n calculated using the sample means and use z^0 and W_n^0 to refer to z and W_n calculated using the population means. We have

$$\begin{aligned}
W_n - W_n^0 &= (\bar{z} - \bar{z}^0)^\top(\bar{z} + \bar{z}^0) + k_n\boldsymbol{\alpha}^\top(\bar{z} - \bar{z}^0)(\bar{z} + \bar{z}^0)^\top\boldsymbol{\alpha} \\
&\quad + O_p(n^{-1})(\lambda_{\max}(\Sigma_X) + k_n\boldsymbol{\alpha}^\top\Sigma_X\boldsymbol{\alpha}) \\
&= (\bar{z} - \bar{z}^0)^\top(\bar{z} + \bar{z}^0) + k_n\boldsymbol{\alpha}^\top(\bar{z} - \bar{z}^0)(\bar{z} + \bar{z}^0)^\top\boldsymbol{\alpha} + o_p(n^{-1}\sqrt{\mathrm{tr}(\Sigma_X^2)}).
\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
\mathrm{Var}(W_n - W_n^0) &= O(\mathrm{tr}(V_n)) + k_n^2 O(\boldsymbol{\alpha}^\top V_n\boldsymbol{\alpha}) + o(\mathrm{Var}(W_n)) \\
&= O(k_n^2(\boldsymbol{\alpha}^\top\mathrm{E}[\bar{z}^0\bar{z}^{0\top}]\boldsymbol{\alpha})^2) + o(\mathrm{Var}(W_n)),
\end{aligned}$$

where $V_n = \mathrm{E}[(\bar{z} - \bar{z}^0)(\bar{z} - \bar{z}^0)^\top]\mathrm{E}[\bar{z}^0\bar{z}^{0\top}]$. And under the local alternative (7), we have

$$\begin{aligned}
&\boldsymbol{\alpha}^\top\mathrm{E}[\bar{z}^0\bar{z}^{0T}]\boldsymbol{\alpha} \\
&\quad\le \mathrm{E}((\boldsymbol{\alpha}^\top(\boldsymbol{x} - \boldsymbol{\mu}_X)(\boldsymbol{x} - \boldsymbol{\mu}_X)^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0))^2) + n^{-1}\boldsymbol{\alpha}^\top\mathrm{E}((\boldsymbol{x} - \boldsymbol{\mu}_X)(\boldsymbol{x} - \boldsymbol{\mu}_X)^\top)\boldsymbol{\alpha} \\
&\quad = n^{-1}(\boldsymbol{\alpha}^\top\Sigma_X\boldsymbol{\alpha})^2 + n^{-1}\boldsymbol{\alpha}^\top\Sigma_X\boldsymbol{\alpha}.
\end{aligned}$$

Hence, we have $\mathrm{Var}(W_n) = \mathrm{Var}(W_n^0)(1 + o(1))$, and $\hat{\sigma}_n^2$ is still a ratio consistent estimator for $\mathrm{Var}(W_n)$ for W_n constructed with sample means.

2.3.2. *Two lemmas and Proof of Theorem* 1. Let us introduce some notation. Define

$$\begin{aligned}
z_{\boldsymbol{\eta},i} &= (\boldsymbol{x}_i - \boldsymbol{\mu}_X)(y_i - \mu_Y - (\boldsymbol{x}_i - \boldsymbol{\mu}_X)^\top\boldsymbol{\eta}) \\
&= (\boldsymbol{x}_i - \boldsymbol{\mu}_X)((\boldsymbol{x}_i - \boldsymbol{\mu}_X)^\top(\boldsymbol{\beta} - \boldsymbol{\eta}) + \epsilon_i),
\end{aligned}$$

and

$$S_{Z_\eta} = \frac{1}{n}\sum_{i=1}^n z_{\boldsymbol{\eta},i} z_{\boldsymbol{\eta},i}^\top = \frac{1}{n}\sum_{i=1}^n((\boldsymbol{x}_i - \boldsymbol{\mu}_X)^\top(\boldsymbol{\beta} - \boldsymbol{\eta}) + \epsilon_i)^2(\boldsymbol{x}_i - \boldsymbol{\mu}_X)(\boldsymbol{x}_i - \boldsymbol{\mu}_X)^\top.$$

Then we have $z_{\boldsymbol{\beta},i} = (\boldsymbol{x}_i - \boldsymbol{\mu}_X)\epsilon_i$, and $S_{Z_\beta} = \frac{1}{n}\sum_{i=1}^n z_{\boldsymbol{\beta},i} z_{\boldsymbol{\beta},i}^\top = \frac{1}{n}\sum_{i=1}^n \epsilon_i^2(\boldsymbol{x}_i - \boldsymbol{\mu}_X)(\boldsymbol{x}_i - \boldsymbol{\mu}_X)^\top$. We also let $\Sigma_{Z_\beta} = \mathrm{E}S_{Z_\beta} = \sigma^2\Sigma_X$. Further define $\Omega_1 = S_{Z_\beta}^{1/2}(I_p + k_n\boldsymbol{\alpha}\boldsymbol{\alpha}^\top)S_{Z_\beta}^{1/2}$, and $\Omega_2 = (\Sigma_{Z_\beta}^{1/2}(I_p + k_n\boldsymbol{\alpha}\boldsymbol{\alpha}^\top)\Sigma_{Z_\beta}^{1/2})^2$. To prove Theorem 1, we need the following two lemmas, whose proofs are given in the Appendix.

LEMMA 2.1. *Under Assumptions A1 and A2*,

$$\frac{n\|\bar{z}_\beta\|^2 - \mathrm{tr}(S_{Z_\beta})}{\sqrt{2\sigma^4\,\mathrm{tr}(\Sigma_X^2)}} \xrightarrow{d} N(0, 1). \tag{10}$$

LEMMA 2.2. *Under the conditions for Theorem* 1 *and the local alternative condition* (7),

$$\frac{nW_n - n(\|\boldsymbol{\delta}_z\|^2 + k_n|\boldsymbol{\alpha}^T\boldsymbol{\delta}_z|^2)}{\{2\,\mathrm{tr}(\Omega_2)\}^{1/2}} \xrightarrow{d} N(0, 1).$$

PROOF OF THEOREM 1. The null hypothesis H_0 and the first part of Theorem 1 can be seen as a special case of the local alternative hypothesis (7) and the second part of Theorem 1 with $\delta = 0$ and $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Thus, it suffices to show the second part of Theorem 1 for local alternatives.

It is easy to verify that under the null hypothesis or local alternative hypothesis (7), $\mathrm{tr}(\Omega_1) = \mathrm{tr}(\Omega)(1 + o_p(1))$, and $\mathrm{tr}(\Omega_2) = \mathrm{tr}(\Omega^2)(1 + o_p(1))$. In fact, under the null or local alternative (7), we have

$$\begin{aligned}\mathrm{tr}(\Omega_1) - \mathrm{tr}(\Omega) &= \mathrm{tr}\big((I_p + k_n \boldsymbol{\alpha}\boldsymbol{\alpha}^\top) S_{Z_\beta}\big) - \mathrm{tr}\big((I_p + k_n \boldsymbol{\alpha}\boldsymbol{\alpha}^\top) S_{Z_{\beta_0}}\big) + o_p\big(\mathrm{tr}(\Omega)\big)\\ &= \mathrm{tr}\big((I_p + k_n \boldsymbol{\alpha}\boldsymbol{\alpha}^\top)(S_{Z_\beta} - S_{Z_{\beta_0}})\big) + o_p\big(\mathrm{tr}(\Omega)\big)\\ &= \mathrm{tr}(S_{Z_\beta} - S_{Z_{\beta_0}}) + k_n \boldsymbol{\alpha}^\top (S_{Z_\beta} - S_{Z_{\beta_0}})\boldsymbol{\alpha} + o_p\big(\mathrm{tr}(\Omega)\big).\end{aligned}$$

Since

$$\begin{aligned}S_{Z_\beta} - S_{Z_{\beta_0}} &= -\frac{1}{n}\sum_{i=1}^n \big(\big((\boldsymbol{x}_i - \boldsymbol{\mu}_X)^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + \epsilon_i\big)^2 - \epsilon_i^2\big)(\boldsymbol{x}_i - \boldsymbol{\mu}_X)(\boldsymbol{x}_i - \boldsymbol{\mu}_X)^\top\\ &= -\frac{1}{n}\sum_{i=1}^n \big(\big((\boldsymbol{x}_i - \boldsymbol{\mu}_X)^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\big)^2 + 2(\boldsymbol{x}_i - \boldsymbol{\mu}_X)^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\epsilon_i\big)\\ &\quad \times (\boldsymbol{x}_i - \boldsymbol{\mu}_X)(\boldsymbol{x}_i - \boldsymbol{\mu}_X)^\top,\end{aligned}$$

we have $\mathrm{tr}(\Omega_1) = \mathrm{tr}(\Omega)(1 + o_p(1))$ under the local alternative (7). Similarly, we have

$$\begin{aligned}\mathrm{tr}(\Omega_2) - \mathrm{tr}(\Omega^2) &= \mathrm{tr}\big[\big(\Omega_2^{1/2} - \Omega\big)\big(\Omega_2^{1/2} + \Omega\big)\big]\\ &= (1 + k_n) O_p\big(\mathrm{tr}\big(\Omega_2^{1/2} - \Omega\big)\big)\\ &= (1 + k_n) O_p\big(\mathrm{E}\big[\mathrm{tr}\big((I_p + k_n \boldsymbol{\alpha}\boldsymbol{\alpha}^\top)(S_{Z_\beta} - S_{Z_{\beta_0}})\big)\big]\big)\\ &= o_p\big(\mathrm{tr}(\Omega^2)\big).\end{aligned}$$

For $\boldsymbol{\beta} = \boldsymbol{\beta}_0 + n^{-1/2}\delta\boldsymbol{u}$ with $|\delta| \le C$ and $\|\boldsymbol{u}\| = 1$, we have

$$\begin{aligned}&T_n - \frac{n\|\boldsymbol{\delta}_z\|^2 + nk_n|\boldsymbol{\alpha}^T\boldsymbol{\delta}_z|^2}{\{2\,\mathrm{tr}(\Omega_2)\}^{1/2}}\\ &\quad = \frac{\{2\,\mathrm{tr}(\Omega_2)\}^{1/2}}{\{2\widehat{\mathrm{tr}(\Omega^2)}\}^{1/2}} T_n^a + \frac{\mathrm{tr}(\Omega_1) - \widehat{\mathrm{tr}(\Omega)}}{\{2\widehat{\mathrm{tr}(\Omega^2)}\}^{1/2}}\\ &\qquad + \left[\frac{\{2\,\mathrm{tr}(\Omega_2)\}^{1/2}}{\{2\widehat{\mathrm{tr}(\Omega^2)}\}^{1/2}} - 1\right]\frac{n\|\boldsymbol{\delta}_z\|^2 + nk_n|\boldsymbol{\alpha}^T\boldsymbol{\delta}_z|^2}{\{2\,\mathrm{tr}(\Omega_2)\}^{1/2}},\end{aligned}$$

where

$$T_n^a = \frac{nW_n - n(\|\boldsymbol{\delta}_z\|^2 + k_n|\boldsymbol{\alpha}^T\boldsymbol{\delta}_z|^2)}{\{2\,\mathrm{tr}(\Omega_2)\}^{1/2}}.$$

The second and the third terms are of $o_p(1)$ as n and p go to infinity. In fact, it is easy to have $\mathrm{tr}(\Omega_1) - \mathrm{tr}(\Omega) = o_p(\{2\,\mathrm{tr}(\Omega_2)\}^{1/2})$. Furthermore, since $\mathrm{tr}(\Omega) - \widehat{\mathrm{tr}(\Omega)} = o_p(\{2\,\mathrm{tr}(\Omega_2)\}^{1/2})$ (equation (6.10) Chen and Qin (2010)), the second term is of $o_p(1)$. Since $\widehat{\mathrm{tr}(\Omega^2)}$ is a ratio consistent estimator of $\mathrm{tr}(\Omega^2)$, which is itself ratio consistent with $\mathrm{tr}(\Omega_2)$, we have $\{2\,\mathrm{tr}(\Omega_2)\}^{1/2}/\{2\widehat{\mathrm{tr}(\Omega^2)}\}^{1/2} - 1 = o_p(1)$. Hence, the third term is also of $o_p(1)$. Furthermore, since $\{2\,\mathrm{tr}(\Omega_2)\}^{1/2}/\{2\widehat{\mathrm{tr}(\Omega^2)}\}^{1/2}$ converges to one in probability, and Lemma 2.2 establishes that $T_n^a \xrightarrow{d} N(0, 1)$, the proof of Theorem 1 is complete. □

2.3.3. *Proof of Theorem 2*. Without loss of generality, we assume that $\boldsymbol{\beta}_0 = \mathbf{0}$ in the null hypothesis, and we also assume that $\mathrm{Var}(y_i|\boldsymbol{x}_i) = \sigma^2 = 1$. Then the average likelihood ratio corresponding to the prior π on the alternatives is given by

$$W(\boldsymbol{X}, \boldsymbol{y}) = \mathrm{E}_\pi\big[\exp(\boldsymbol{y}^\top \boldsymbol{X}\boldsymbol{\beta} - \|\boldsymbol{X}\boldsymbol{\beta}\|^2/2)|\boldsymbol{X}, \boldsymbol{y}\big], \tag{11}$$

where $\boldsymbol{X}$ is the $n \times p$ design matrix, $\boldsymbol{y}$ is the response vector, and $\mathrm{E}_\pi[\cdot|\boldsymbol{X}, \boldsymbol{y}]$ is the conditional expectation regarding the prior π on the alternatives given $\boldsymbol{X}$ and $\boldsymbol{y}$. By the Neyman-Pearson lemma, the likelihood ratio test $T = \{W(\boldsymbol{X}, \boldsymbol{y}) > 1\}$ minimizes the average risk $\mathrm{Risk}_\pi(T) = P_0(T = 1) + \mathrm{E}_\pi[I(T = 0)]$, where $P_0(T = 1)$ is the type I error and $\mathrm{E}_\pi[I(T = 0)]$ is the type II error under the prior π on the alternatives. To show all sequences of tests are powerless, we only need to show that $\limsup \mathrm{E}(W(\boldsymbol{X}, \boldsymbol{y})^2) = 1$ (Section A.1, Arias-Castro, Candès and Plan (2011b)). Note that $\mathrm{E}(W(\boldsymbol{X}, \boldsymbol{y})) \geq 1$ from the Jensen inequality and the convexity of the exp function. As a result, we only need to show that $\mathrm{E}(W(\boldsymbol{X}, \boldsymbol{y})^2) \leq 1 + o(1)$ to prove the first part of Theorem 2.

By Fubini's theorem, first integrating with regards to $\boldsymbol{y}$, we have

$$\mathrm{E}\big(W(\boldsymbol{X}, \boldsymbol{y})^2\big) = \mathrm{E}\big[\mathrm{E}_\pi[\exp(\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}')|\boldsymbol{X}]\big],$$

where $\boldsymbol{\beta}'$ is another vector from the prior π, that is, $\boldsymbol{\beta}, \boldsymbol{\beta}' \overset{i.i.d.}{\sim} \pi$, and $\mathrm{E}_\pi[\cdot|\boldsymbol{X}]$ is the conditional expectation regarding the prior π on the alternatives given $\boldsymbol{X}$. Let $R = \sum_{j:\beta_j \neq 0} R_j$, where $R_j = \boldsymbol{\beta}_j \boldsymbol{X}_j^\top \boldsymbol{X}\boldsymbol{\beta}'$, and $\boldsymbol{X}_j$ is the jth column of the design matrix $\boldsymbol{X}$. Then we have

$$\mathrm{E}\big(W(\boldsymbol{X}, \boldsymbol{y})^2\big) = \mathrm{E}\big[\mathrm{E}_\pi[\exp(R)|\boldsymbol{X}]\big] = \mathrm{E}\Big[\mathrm{E}_\pi\Big[\exp\Big(\sum_{j:\beta_j \neq 0} R_j\Big)\Big|\boldsymbol{X}\Big]\Big].$$

Denote $\boldsymbol{\beta}$'s and $\boldsymbol{\beta}'$'s supports as $\mathcal{J}$ and $\mathcal{J}'$, respectively. Without loss of generality, we assume that $\boldsymbol{\mu}_X = \mathbf{0}$ in Assumption A1. We also assume that the diagonal elements of Σ_X are bounded from above. Note that $R_j = a_j \beta_j$, $j \in \mathcal{J}$, $a_j = \boldsymbol{X}_j^\top \boldsymbol{X}\boldsymbol{\beta}'$, are mean-zero sub-exponential random variables, and they are jointly independent given $\boldsymbol{X}$ and $\boldsymbol{\beta}'$. Then with Bernstein's inequality (Theorem 2.8.2, Vershynin (2018)), we have

$$\begin{aligned} P\big(|R| \geq t|\boldsymbol{X}, \boldsymbol{\beta}'\big) &= P\Big(\Big|\sum_{j \in \mathcal{J}} a_j \beta_j\Big| \geq t\Big|\boldsymbol{X}, \boldsymbol{\beta}'\Big) \\ &\leq 2\exp\Big(-c \min\Big(\frac{t^2}{\gamma^2 \sum_{j \in \mathcal{J}} a_j^2}, \frac{t}{\gamma \max_{j \in \mathcal{J}} |a_j|}\Big)\Big) \\ &\leq 2\exp\Big(-\frac{ct^2}{\gamma^2 \sum_{j \in \mathcal{J}} a_j^2}\Big) + 2\exp\Big(-\frac{ct}{\gamma \max_{j \in \mathcal{J}} |a_j|}\Big), \end{aligned} \tag{12}$$

with some absolute constant c. Note that because the function $f(x) = \exp(-1/x)$ is concave in $x > 0$, then from Jensen's inequality we have

$$\begin{aligned} P\big(|R| \geq t|\boldsymbol{X}\big) &= \mathrm{E}_\pi\big[P\big(|R| \geq t|\boldsymbol{X}, \boldsymbol{\beta}'\big)|\boldsymbol{X}\big] \\ &\leq 2\exp\Big(-\frac{ct^2}{\gamma^2 \mathrm{E}_\pi[\sum_{j \in \mathcal{J}} a_j^2|\boldsymbol{X}]}\Big) + 2\exp\Big(-\frac{ct}{\gamma \mathrm{E}_\pi[\max_{j \in \mathcal{J}} |a_j||\boldsymbol{X}]}\Big). \end{aligned} \tag{13}$$

For $\gamma^2 \mathrm{E}_\pi[\sum_{j \in \mathcal{J}} a_j^2|\boldsymbol{X}]$ in (13), we have $\mathrm{E}_\pi[R^2|\boldsymbol{X}] = \gamma^2 \mathrm{E}_\pi[\sum_{j \in \mathcal{J}} a_j^2|\boldsymbol{X}]$. And we have

$$\begin{aligned} \mathrm{E}_\pi[R^2|\boldsymbol{X}] &= \mathrm{E}_\pi[\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}'\boldsymbol{\beta}'^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}|\boldsymbol{X}] = \mathrm{E}_\pi[\mathrm{tr}(\boldsymbol{\beta}'\boldsymbol{\beta}'^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X})|\boldsymbol{X}] \\ &= \mathrm{tr}\big[\mathrm{E}_\pi(\boldsymbol{\beta}'\boldsymbol{\beta}'^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X}|\boldsymbol{X})\big] = \mathrm{tr}\big[\mathrm{E}_\pi(\boldsymbol{\beta}\boldsymbol{\beta}^\top \boldsymbol{X}^\top \boldsymbol{X}|\boldsymbol{X})^2\big]. \end{aligned}$$

From the prior π of $\boldsymbol{\beta}$, we have

$$\begin{aligned}\mathrm{E}_{\pi}[R^2|\boldsymbol{X}] = \mathrm{tr}[\mathrm{E}_{\pi}(\boldsymbol{\beta}\boldsymbol{\beta}^{\top}\boldsymbol{X}^{\top}\boldsymbol{X}|\boldsymbol{X})^2] &= \mathrm{tr}\Big(\frac{S\gamma^2}{p}D_X + \frac{S^2\gamma^2}{p^2}(\boldsymbol{X}^{\top}\boldsymbol{X} - D_X)\Big)^2 \\ &\le \frac{n^2S^2\gamma^4}{p^2}\mathrm{tr}\big((\boldsymbol{X}^{\top}\boldsymbol{X}/n)^2\big) = o_p(1),\end{aligned}$$

where D_X is the diagonal matrix of $\boldsymbol{X}^{\top}\boldsymbol{X}$, and the last equality is from $\mathrm{tr}((\boldsymbol{X}^{\top}\boldsymbol{X}/n)^2) = O_p(\mathrm{tr}(\Sigma_X^2))$ as n and p go to infinity, and that $\gamma^2/\gamma_0^2 = o(1)$ in the first part of Theorem 2. Then for the second term in (13), we have

$$\exp\Big(-\frac{ct^2}{\gamma^2\mathrm{E}_{\pi}[\sum_{j\in\mathcal{J}}a_j^2|\boldsymbol{X}]}\Big) \le \exp(-t^2/K_1),$$

with $K_1 = o_p(1)$. Now we consider how to control the second term in (13). We have $a_j = \sum_{j'\in\mathcal{J}'}\boldsymbol{X}_j^{\top}\boldsymbol{X}_{j'}\boldsymbol{\beta}'_{j'}$. Then we use Bernstein's inequality (Theorem 2.8.2, Vershynin (2018)) again for a_j, and get

$$\begin{aligned}P\big(|a_j| \ge t|\boldsymbol{X}, \mathcal{J}, \mathcal{J}'\big) &\le 2\exp\Big(-c\min\Big(\frac{t^2}{\gamma^2\sum_{j'\in\mathcal{J}'}(\boldsymbol{X}_j^{\top}\boldsymbol{X}_{j'})^2}, \frac{t}{\gamma\max_{j'\in\mathcal{J}'}|\boldsymbol{X}_j^{\top}\boldsymbol{X}_{j'}|}\Big)\Big) \\ &\le 2\exp\Big(-c_1\min\Big(\frac{t^2}{\gamma^2\sum_{j'\in\mathcal{J}'}(\boldsymbol{X}_j^{\top}\boldsymbol{X}_{j'})^2}, \frac{t}{n\gamma}\Big)\Big) \\ &\le 2\exp\Big(-\frac{c_1t^2}{\gamma^2\sum_{j'\in\mathcal{J}'}(\boldsymbol{X}_j^{\top}\boldsymbol{X}_{j'})^2}\Big) + 2\exp\Big(-\frac{c_1t}{n\gamma}\Big),\end{aligned}$$

for some bounded positive constant c_1. Further, from Jensen's inequality and the concavity of $f(x) = \exp(-1/x)$, we have

$$\begin{aligned}P\big(|a_j| \ge t|\boldsymbol{X}\big) &\le 2\exp\Big(-\frac{c_1t^2}{\gamma^2\mathrm{E}_{\pi}[\sum_{j'\in\mathcal{J}'}(\boldsymbol{X}_j^{\top}\boldsymbol{X}_{j'})^2|\boldsymbol{X}]}\Big) + 2\exp\Big(-\frac{c_1t}{n\gamma}\Big) \\ &\le 2\exp\Big(-\frac{c_1t^2p^2}{S\gamma^2\mathrm{tr}((\boldsymbol{X}^{\top}\boldsymbol{X})^2)}\Big) + 2\exp\Big(-\frac{c_1t}{n\gamma}\Big).\end{aligned}$$

Note that this holds for any j, and we get

$$P\Big(\max_{j\in\mathcal{J}}|a_j| \ge t\big|\boldsymbol{X}\Big) \le 2\exp\Big(-\frac{c_1t^2p^2}{S\gamma^2\mathrm{tr}((\boldsymbol{X}^{\top}\boldsymbol{X})^2)\log(S)}\Big) + 2\exp\Big(-\frac{c_1t}{n\gamma\log(S)}\Big).$$

Hence, we have

$$\begin{aligned}\gamma\mathrm{E}_{\pi}\Big[\max_{j\in\mathcal{J}}|a_j||\boldsymbol{X}\Big] &= O\Big(\sqrt{\frac{n^2S\gamma^4\mathrm{tr}((\boldsymbol{X}^{\top}\boldsymbol{X}/n)^2)\log(S)}{p^2}}\Big) + O\big(n\gamma^2\log(S)\big) \\ &= \sqrt{G}O\big(\sqrt{\log(S)/S}\big) + \sqrt{G}O\big(\sqrt{p}\log(S)/S\big) \\ &= o_p(1),\end{aligned}$$

where

$$G = \frac{n^2S^2\gamma^4\mathrm{tr}(\Sigma_X^2)}{p^2} \times \frac{p}{\mathrm{tr}(\boldsymbol{X}^{\top}\boldsymbol{X}/n)} = o_p(1).$$

Then for the second term in (13), we have

$$\exp\left(-\frac{ct}{\gamma \mathrm{E}_\pi[\max_{j\in\mathcal{J}} |a_j||\boldsymbol{X}]}\right) \le \exp(-t^2/K_2) + \exp(-t/K_2),$$

with $K_2 = o_p(1)$. In sum, from (13), we get

$$\begin{aligned} P\big(|R| \ge t|\boldsymbol{X}\big) &\le 2\exp\left(-\frac{ct^2}{\gamma^2 \mathrm{E}_\pi[\sum_{j\in\mathcal{J}} a_j^2|\boldsymbol{X}]}\right) + 2\exp\left(-\frac{ct}{\gamma \mathrm{E}_\pi[\max_{j\in\mathcal{J}} |a_j||\boldsymbol{X}]}\right) \\ &\le 4\exp\big(-t^2/K_3\big) + 4\exp(-t/K_3), \end{aligned}$$

where $K_3 = \max(K_1, K_2) = o_p(1)$. Then we get

$$\begin{aligned} \mathrm{E}_\pi\big[\exp(R)|\boldsymbol{X}\big] &\le \int_0^{1+\varepsilon} P\big(\exp(|R|) \ge t|\boldsymbol{X}\big)\,dt + \int_{1+\varepsilon}^{\infty} P\big(|R| \ge \log(t)|\boldsymbol{X}\big)\,dt \\ &\le 1 + \varepsilon + 4\int_{1+\varepsilon}^{\infty} \big[\exp\big(-\log(t)^2/K_3\big) + \exp\big(-\log(t)/K_3\big)\big]\,dt \\ &= 1 + \varepsilon + o_p(1), \end{aligned}$$

for any $\varepsilon > 0$. Since the inequality holds for any $\varepsilon > 0$, we have

$$\mathrm{E}_\pi\big[\exp(R)|\boldsymbol{X}\big] \le 1 + o_p(1).$$

In sum, we get that

$$\mathrm{E}\big(W(\boldsymbol{X}, \boldsymbol{y})^2\big) = \mathrm{E}\big[\mathrm{E}_\pi\big[\exp(R)|\boldsymbol{X}\big]\big] \le 1 + o(1),$$

which completes the proof of the first part of Theorem 2.

For the second part of Theorem 2, note that Theorem 1 already shows that the proposed test statistic T_n has a standard normal distribution asymptotically under the null, thus we only need to show that the proposed test statistic goes to infinity in probability under the alternative to establish the asymptotic powerfulness of the proposed test in the second part of the theorem. Recall that our proposed test statistic T_n is defined in (5). We assume that $\mu_X = \mathbf{0}$ without loss of generality. From the derivation of $\mathrm{E}(W_n)$ in Section 2.3.1, we have

$$\begin{aligned} \mathrm{E}_\pi\big[\mathrm{E}(W_n|\boldsymbol{\beta})\big] &= \mathrm{E}_\pi\big[\|\mathrm{E}(z|\boldsymbol{\beta})\|^2\big] + k_n \mathrm{E}_\pi\big[|\boldsymbol{\alpha}^\top \mathrm{E}(z|\boldsymbol{\beta})|^2\big] \\ &= \mathrm{E}_\pi\big[\|\Sigma_X\boldsymbol{\beta}\|^2\big] + \mathrm{E}_\pi\big[|\boldsymbol{\alpha}^\top \Sigma_X\boldsymbol{\beta}|^2\big] \\ &= \frac{S\gamma^2}{p}\operatorname{tr}\big(\Sigma_X^2\big) + k_n\frac{S\gamma^2}{p}\|\Sigma_X\boldsymbol{\alpha}\|^2, \end{aligned}$$

where E_π is the expectation regarding the prior π on the alternatives, and $\mathrm{E}(\cdot|\boldsymbol{\beta})$ is the conditional expectation given $\boldsymbol{\beta}$. From the derivation of $\mathrm{Var}(W_n)$ in Section 2.3.1, we have

$$\mathrm{Var}(W_n|\boldsymbol{\beta}) = O\left(\frac{1}{n^2}\right)\mathrm{Var}(W_{n12}|\boldsymbol{\beta}) + O\left(\frac{1}{n}\right)\mathrm{Cov}(W_{n12}, W_{n13}|\boldsymbol{\beta}),$$

where $\mathrm{Var}(\cdot|\boldsymbol{\beta})$ is the conditional variance given $\boldsymbol{\beta}$, that is, $\mathrm{Var}(T|\boldsymbol{\beta}) = \mathrm{E}(T^2|\boldsymbol{\beta}) - \mathrm{E}(T|\boldsymbol{\beta})^2$ for any statistic T, $W_{nij} = z_i^\top z_j + k_n\boldsymbol{\alpha}^\top z_i z_j^\top \boldsymbol{\alpha}$. Note that we use $a = O(b)$ to mean that a/b is bounded as n and p go to infinity, and the bound does not depend on n or p. We have that

$$\begin{aligned} \mathrm{Var}(W_{n12}|\boldsymbol{\beta}) &= O\big(\operatorname{tr}\big(\mathrm{Cov}(z|\boldsymbol{\beta})^2\big)\big) + O\big(\mathrm{E}(z|\boldsymbol{\beta})^\top \mathrm{Cov}(z|\boldsymbol{\beta})\mathrm{E}(z|\boldsymbol{\beta})\big) \\ &\quad + k_n^2 O\big(\big(\boldsymbol{\alpha}^\top \mathrm{Cov}(z|\boldsymbol{\beta})\boldsymbol{\alpha}\big)^2\big) + k_n^2 O\big(\boldsymbol{\alpha}^\top \mathrm{Cov}(z|\boldsymbol{\beta})\boldsymbol{\alpha}\big(\boldsymbol{\alpha}^\top \mathrm{E}(z|\boldsymbol{\beta})\big)^2\big) \\ &= O\big(\operatorname{tr}\big(\mathrm{Cov}(z|\boldsymbol{\beta})^2\big)\big) + k_n^2 O\big(\big(\boldsymbol{\alpha}^\top \mathrm{Cov}(z|\boldsymbol{\beta})\boldsymbol{\alpha}\big)^2\big) + O\big(\|\mathrm{E}(z|\boldsymbol{\beta})\|^4\big). \end{aligned}$$

Furthermore, from Assumptions A1 and A2, we get

$$\begin{aligned}
&\mathrm{Var}(W_{n12}|\boldsymbol{\beta}) \\
&= O\big((\boldsymbol{\beta}^\top \Sigma_X^2 \boldsymbol{\beta})^2 + \mathrm{tr}(\Sigma_X^2)\big) + O\big(k_n^2\big)\big((\boldsymbol{\alpha}^\top \Sigma_X \boldsymbol{\beta})^4 + (\boldsymbol{\alpha}^\top \Sigma_X \boldsymbol{\alpha})^2\big) + O\big(\|\mathrm{E}(z|\boldsymbol{\beta})\|^4\big) \\
&= O\big(\mathrm{E}(W_n|\boldsymbol{\beta})^2 + \mathrm{tr}(\Sigma_X^2)\big).
\end{aligned}$$

For $\mathrm{Cov}(W_{n12}, W_{n13}|\boldsymbol{\beta})$, we have

$$\begin{aligned}
&\mathrm{Cov}(W_{n12}, W_{n13}|\boldsymbol{\beta}) \\
&= \mathrm{Cov}\big(z_1^\top z_2, z_1^\top z_3|\boldsymbol{\beta}\big) + 2k_n \mathrm{Cov}\big(z_1^\top z_2, \boldsymbol{\alpha}^\top z_1 z_3^\top \boldsymbol{\alpha}|\boldsymbol{\beta}\big) \\
&\quad + k_n^2 \mathrm{Cov}\big(\boldsymbol{\alpha}^\top z_1 z_2^\top \boldsymbol{\alpha}, \boldsymbol{\alpha}^\top z_1 z_3^\top \boldsymbol{\alpha}|\boldsymbol{\beta}\big) \\
&= \mathrm{E}(z|\boldsymbol{\beta})^\top \mathrm{Cov}(z|\boldsymbol{\beta})\mathrm{E}(z|\boldsymbol{\beta}) + 2k_n \mathrm{E}(z|\boldsymbol{\beta})^\top \mathrm{Cov}(z|\boldsymbol{\beta})\boldsymbol{\alpha}\boldsymbol{\alpha}^\top \mathrm{E}(z|\boldsymbol{\beta}) \\
&\quad + k_n^2 \boldsymbol{\alpha}^\top \mathrm{Cov}(z|\boldsymbol{\beta})\boldsymbol{\alpha}\big(\boldsymbol{\alpha}^\top \mathrm{E}(z|\boldsymbol{\beta})\big)^2 \\
&= O\big(\mathrm{E}(z|\boldsymbol{\beta})^\top \mathrm{Cov}(z|\boldsymbol{\beta})\mathrm{E}(z|\boldsymbol{\beta})\big) + O\big(k_n^2\big)\boldsymbol{\alpha}^\top \mathrm{Cov}(z|\boldsymbol{\beta})\boldsymbol{\alpha}\big(\boldsymbol{\alpha}^\top \mathrm{E}(z|\boldsymbol{\beta})\big)^2.
\end{aligned}$$

We further have

$$\begin{aligned}
&\mathrm{Cov}(W_{n12}, W_{n13}|\boldsymbol{\beta}) \\
&= O\big(\boldsymbol{\beta}^\top \Sigma_X (\Sigma_X \boldsymbol{\beta}\boldsymbol{\beta}^\top \Sigma_X + \Sigma_X)\Sigma_X \boldsymbol{\beta}\big) + k_n^2 O\big((\boldsymbol{\alpha}^\top \Sigma_X \boldsymbol{\beta})^2\big)\boldsymbol{\alpha}^\top (\Sigma_X \boldsymbol{\beta}\boldsymbol{\beta}^\top \Sigma_X + \Sigma_X)\boldsymbol{\alpha} \\
&= O\big((\boldsymbol{\beta}^\top \Sigma_X^2 \boldsymbol{\beta})^2 + \boldsymbol{\beta}^\top \Sigma_X^3 \boldsymbol{\beta}\big) + O\big(k_n^2\big)(\boldsymbol{\alpha}^\top \Sigma_X \boldsymbol{\beta})^2\big((\boldsymbol{\alpha}^\top \Sigma_X \boldsymbol{\beta})^2 + \boldsymbol{\alpha}^\top \Sigma_X \boldsymbol{\alpha}\big) \\
&= O\big(\mathrm{E}(W_n|\boldsymbol{\beta})^2)\big) + O\big(\boldsymbol{\beta}^\top \Sigma_X^3 \boldsymbol{\beta}\big).
\end{aligned}$$

For $\boldsymbol{\beta}^\top \Sigma_X^3 \boldsymbol{\beta}$, we have that

$$\begin{aligned}
\mathrm{E}_\pi[\boldsymbol{\beta}^\top \Sigma_X^3 \boldsymbol{\beta}] = \frac{S}{p}\gamma^2 \mathrm{tr}(\Sigma_X^3) &= o\left(n\left(\frac{S}{p}\gamma^2 \mathrm{tr}(\Sigma_X^2)\right)^2\right) + O\left(\frac{1}{n}\mathrm{tr}(\Sigma_X^2)\right) \\
&= o\big(n\mathrm{E}_\pi[\mathrm{E}(W_n|\boldsymbol{\beta})^2]\big) + O\left(\frac{1}{n}\mathrm{tr}(\Sigma_X^2)\right).
\end{aligned}$$

Note that we use $a = o(b)$ to mean that a/b goes to zero as n and p go to infinity. In sum,

$$\begin{aligned}
\mathrm{E}_\pi[\mathrm{Var}(W_n|\boldsymbol{\beta})] &= O\left(\frac{1}{n}\right)\big(\mathrm{E}_\pi[\mathrm{E}(W_n|\boldsymbol{\beta})^2] + \mathrm{E}_\pi[\boldsymbol{\beta}^\top \Sigma_X^3 \boldsymbol{\beta}]\big) + O\left(\frac{1}{n^2}\right)\mathrm{tr}(\Sigma_X^2) \\
&= o\big(\mathrm{E}_\pi[\mathrm{E}(W_n|\boldsymbol{\beta})^2]\big) + O\left(\frac{1}{n^2}\mathrm{tr}(\Sigma_X^2)\right).
\end{aligned}$$

Since $\gamma^2/\gamma_0^2 \to \infty$ as n and p go to infinity in the second part of Theorem 2, we know that

$$\frac{\mathrm{E}_\pi[\mathrm{E}(nW_n|\boldsymbol{\beta})]}{\sqrt{\mathrm{E}_\pi[\mathrm{Var}(nW_n|\boldsymbol{\beta})]}} \to \infty,$$

which means that under the alternative prior π, we have $nW_n \sim \mathrm{E}_\pi[\mathrm{E}(nW_n|\boldsymbol{\beta})]$ as n and p go to infinity, where $x \sim y$ means that x and y are of the same order asymptotically. Recall the definition of $\widehat{\mathrm{tr}(\Omega^2)}$:

$$\widehat{\mathrm{tr}(\Omega^2)} = \mathrm{tr}\big(\widehat{\Sigma_Z^2}\big) + 2k_n \boldsymbol{\alpha}^\top \widehat{\Sigma_Z^2}\boldsymbol{\alpha} + k_n^2\big(\boldsymbol{\alpha}^\top \hat{\Sigma}_Z \boldsymbol{\alpha}\big)^2,$$

where $\hat{\Sigma}_Z$ is the sample covariance of z and $\widehat{\Sigma_Z^2}$ is as defined in (6). Then we have that

$$\begin{aligned}\widehat{\operatorname{tr}(\Omega^2)} &= O_p(\operatorname{tr}(\operatorname{Cov}(z|\boldsymbol{\beta})^2)) + O_p(k_n)\boldsymbol{\alpha}^\top \operatorname{Cov}(z|\boldsymbol{\beta})^2\boldsymbol{\alpha} + O_p(k_n^2)(\boldsymbol{\alpha}^\top \operatorname{Cov}(z|\boldsymbol{\beta})\boldsymbol{\alpha})^2 \\ &= O_p((\boldsymbol{\beta}^\top \Sigma_X^2 \boldsymbol{\beta})^2) + O_p(\operatorname{tr}(\Sigma_X^2)) + O_p(k_n^2)(\boldsymbol{\alpha}^\top(\Sigma_X \boldsymbol{\beta}\boldsymbol{\beta}^\top \Sigma_X + \Sigma_X)\boldsymbol{\alpha})^2 \\ &= O_p(\mathrm{E}(W_n|\boldsymbol{\beta})^2) + O_p(\operatorname{tr}(\Sigma_X^2)).\end{aligned}$$

Note that we use $a = O_p(b)$ to mean that a/b is bounded in probability as n and p go to infinity, and the bound does not depend on n or p. Under the prior of the alternative, we have

$$\sqrt{2\widehat{\operatorname{tr}(\Omega^2)}} = O_p(\mathrm{E}_\pi[\mathrm{E}(W_n|\boldsymbol{\beta})]) + O_P(\sqrt{\operatorname{tr}(\Sigma_X^2)}).$$

In sum, we get that $T_n = \frac{nW_n}{\sqrt{2\widehat{\operatorname{tr}(\Omega^2)}}}$ goes to infinity in probability as n and p go to infinity, which completes the proof of Theorem 2.

3. Numerical studies. In this section, we examine the finite sample performance of the proposed test and illustrate the proposed procedure via an empirical analysis of a breast cancer data set provided from Prat et al. (2014). In our numerical study, we compare the proposed test to the ones put forward by Zhong and Chen (2011) and Cui, Guo and Zhong (2018), which are labelled as ZC and CGZ throughout this section, respectively.

3.1. *Practical implementation issue with the estimation of* $\operatorname{tr}(\Omega^2)$. In Section 2, we proposed an estimator of $\operatorname{tr}(\Omega^2)$ as follows:

$$\widehat{\operatorname{tr}(\Omega^2)} = \operatorname{tr}(\widehat{\Sigma_Z^2}) + 2k_n\boldsymbol{\alpha}^\top\widehat{\Sigma_Z^2}\boldsymbol{\alpha} + k_n^2(\boldsymbol{\alpha}^\top\hat{\Sigma}_Z\boldsymbol{\alpha})^2.$$

Direct calculation of $\widehat{\Sigma_Z^2}$ based on (6) may lead to a very expensive computational cost with computational complexity $O(n^2p^4)$. Similarly, the computational complexity of $\hat{\Sigma}_Z$ is $O(np^2)$. Furthermore, both $z_iz_j^\top$ and $\hat{\Sigma}_Z$ are $p \times p$ matrices, and therefore direct calculation based on (6) requires an extremely large amount of memory for large p.

To address the computational issues, it is of interest to derive an effective way to calculate the estimate of $\operatorname{tr}(\Omega^2)$. It is noteworthy that we need to obtain the scalar values of $\operatorname{tr}(\widehat{\Sigma_Z^2})$, $\boldsymbol{\alpha}^\top\widehat{\Sigma_Z^2}\boldsymbol{\alpha}$, and $\boldsymbol{\alpha}^\top\hat{\Sigma}_Z\boldsymbol{\alpha}$ instead of calculating and saving the large dimensional matrices. This observation helps us to derive an equivalent but easy-to-compute formula for $\operatorname{tr}(\widehat{\Sigma_Z^2})$ as follows:

$$\begin{aligned}\operatorname{tr}(\widehat{\Sigma_Z^2}) = {}& \frac{1}{n(n-1)}\left(1 - \frac{1}{n(n-1)}\right)\sum_{i,j}(z_i^\top z_j)^2 - \frac{1}{n(n-1)}\sum_{i=1}^n (z_i^\top z_i)^2 \\ & - \frac{n^2}{(n-1)^2}\|\bar{z}\|_2^4 + \frac{2}{(n-1)^2}\sum_{i=1}^n (z_i^\top \bar{z})^2,\end{aligned} \tag{14}$$

where $\bar{z} = n^{-1}\sum_{i=1}^n z_i$. This significantly reduces the computational cost to the order of $O(n^2p)$, and avoids storing the $p \times p$ matrices. Similarly, define $\boldsymbol{v}_i = \boldsymbol{\alpha}^\top z_i z_i$, and it follows that

$$\boldsymbol{\alpha}^\top\widehat{\Sigma_Z^2}\boldsymbol{\alpha} = \frac{n}{n-1}\bar{\boldsymbol{v}}^\top\bar{\boldsymbol{v}} - \frac{1}{n(n-1)}\sum_{i=1}^n \boldsymbol{v}_i^\top\boldsymbol{v}_i - \frac{1}{(n(n-1))^2}\|n^2\boldsymbol{\alpha}^\top\bar{z}\bar{z}^\top - n\bar{\boldsymbol{v}}\|_2^2,$$

and

$$\boldsymbol{\alpha}^\top\hat{\Sigma}_Z\boldsymbol{\alpha} = \frac{1}{n-1}\sum_{i=1}^n \|\boldsymbol{\alpha}^\top(z_i - \bar{z})\|_2^2.$$

Using these equivalent formulas, we can significantly reduce the computational cost for calculating $\widehat{\mathrm{tr}(\Omega^2)}$.

3.2. *Monte Carlo simulation study.* In our simulation study, we generate data from

$$y = \boldsymbol{x}^\top \boldsymbol{\beta} + \varepsilon,$$

where $\boldsymbol{x}_i$ follows either a multivariate normal $\mathrm{N}(\mathbf{0}, \Sigma)$ or $(\{4/6\}^{1/2})\mathrm{t}_6(\mathbf{0}, \Sigma)$ distribution. Note that $\mathrm{t}_\nu(\boldsymbol{\mu}, \Sigma)$ represents a multivariate t-distribution with ν degrees of freedom, mean $\boldsymbol{\mu}$, and covariance structure $\frac{\nu}{\nu-2}\Sigma$. We consider autoregressive (AR) structures for $\Sigma = (\sigma_{ij})$ with $\sigma_{ij} = \rho^{|i-j|}$. In our simulations, we set $\rho = 0.25, 0.5$ or 0.75. Due to page limits, we have included the simulation results where Σ satisfies the block compound structure (BCS) in Section S.2.

The error terms ε were each drawn from either a standard normal distribution $\mathrm{N}(0, 1)$ or $(4/6)^{1/2}\mathrm{t}(6)$, where $\mathrm{t}(6)$ is the univariate t-distribution with 6 degrees of freedom. For the purposes of our hypothesis testing scenarios, we set $\boldsymbol{\beta}_0 = \mathbf{0}$.

In our simulation, we set $(n, p) = (200, 300), (400, 600)$ and $(800, 1200)$. We set $k_n = (p/\ln(p))^{1/2}$, and $\boldsymbol{\alpha} = (1, \dots, 1)^\top/\sqrt{p}$. We will study the performance of the proposed test with different values of k_n and different choices of $\boldsymbol{\alpha}$ later.

We first examine whether the proposed test can retain an adequate type I error rate and the performance of the proposed test under local alternatives. To make simulation settings challenging to testing procedures included in this comparison, we set $\boldsymbol{\beta}$ as follows: Only 25% of the p elements in $\boldsymbol{\beta}$ are nonzero and the nonzero elements are randomly located within the coefficient vector. We consider two cases for setting values of nonzero elements in $\boldsymbol{\beta}$.

Case 1. *Fixed signals*: all nonzero elements equal $\delta/\sqrt{n}$,
Case 2. *Random signals*: Nonzero elements are set to be $U(0, 1)\delta/\sqrt{n}$.

The cases with $\delta = 0$ are used to examine the type I error rate and the other values are used for local power. We examined type I error rates and local power performance based on 1000 replications for each combination involving the sample size n, dimension p, Σ with different ρ's, errors drawn from both the standard normal and $(4/6)^{1/2}t(6)$ distributions, and fixed and random signal cases. Simulation results are summarized in Tables 1–3 for $\boldsymbol{x} \sim N(\mathbf{0}, \Sigma)$ and Tables S.1–S.3 for $\boldsymbol{x} \sim (4/6)^{1/2}\mathrm{t}_6(\mathbf{0}, \Sigma)$. Throughout this paper, the proposed test is labelled as New when shown against comparison methods.

Across Tables 1–3 and Tables S.1–S.3, it can be seen that all methods control Type I error rate well. Tables 1–3 and Tables S.1–S.3 imply that the newly proposed test has higher power than the ZC and CGZ tests. The ZC and CGZ tests perform similarly.

From these tables, we can see that the distributions from which $\boldsymbol{x}$ and ε are generated seem to have little impact on the empirical power of the proposed test and ZC and CGZ tests. Across the different simulation scenarios presented in these tables, the empirical power increases as n and p increase proportionally while other conditions (signal strength, ρ, ε_i and $\boldsymbol{x}_i$ generation) are held constant. Indeed, for all such other conditions being held constant, the relative δ value necessary to achieve empirical power values close to 1 decreases as n increases from 200 to 800. We can also see an increase in the power as ρ increases from 0.25 to 0.75. Since under the local alternative the asymptotic power of our proposed test relies on $\Sigma_X(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ and its Euclidean norm, such an increase is to be expected. Indeed, we anticipate the new method being more powerful for data generated from a stronger covariance structure. Similar performance can be seen from both Zhong and Chen (2011) and Cui, Guo and Zhong (2018), consistent with the asymptotic power of those tests. As seen in the simulation results from the tables, the proposed test would perform better than ZC and CGZ tests

TABLE 1
Type I error rate and power with $\boldsymbol{x}_i \sim \mathrm{N}(\mathbf{0}, \Sigma)$ and $\rho = 0.25$

(n, p)	δ	$\varepsilon_i \sim N(0,1)$			$\varepsilon_i \sim (4/6)^{1/2}\mathrm{t}(6)$		
		New	ZC	CGZ	New	ZC	CGZ
				Case 1: Fixed Signals			
(200, 300)	0	0.065	0.053	0.057	0.048	0.039	0.041
	0.27	0.225	0.061	0.068	0.239	0.072	0.083
	0.53	0.735	0.2	0.206	0.761	0.207	0.212
	0.8	0.973	0.554	0.564	0.971	0.561	0.557
	1.06	0.998	0.865	0.87	1	0.863	0.864
(400, 600)	0	0.044	0.057	0.061	0.047	0.057	0.058
	0.25	0.356	0.093	0.096	0.382	0.067	0.067
	0.5	0.935	0.294	0.299	0.949	0.305	0.312
	0.75	0.998	0.75	0.753	1	0.751	0.755
	1	1	0.979	0.979	1	0.972	0.975
(800, 1200)	0	0.051	0.05	0.05	0.045	0.045	0.044
	0.25	0.65	0.086	0.088	0.641	0.092	0.091
	0.49	0.997	0.504	0.508	1	0.504	0.51
	0.74	1	0.955	0.955	1	0.955	0.953
	0.99	1	1	1	1	1	1
				Case 2: Random Signals			
(200, 300)	0	0.059	0.051	0.053	0.046	0.046	0.046
	0.71	0.426	0.125	0.125	0.417	0.109	0.114
	1.41	0.952	0.54	0.55	0.928	0.534	0.534
	2.12	0.999	0.91	0.916	0.997	0.911	0.912
	2.83	1	0.993	0.993	1	0.991	0.994
(400, 600)	0	0.046	0.054	0.057	0.047	0.049	0.048
	0.5	0.371	0.089	0.09	0.424	0.086	0.086
	1	0.957	0.411	0.414	0.946	0.399	0.408
	1.5	1	0.863	0.866	1	0.873	0.875
	2	1	0.992	0.992	1	0.991	0.992
(800, 1200)	0	0.045	0.054	0.054	0.053	0.053	0.052
	0.35	0.375	0.063	0.068	0.348	0.064	0.065
	0.71	0.952	0.274	0.276	0.941	0.279	0.277
	1.06	1	0.741	0.746	1	0.731	0.729
	1.41	1	0.977	0.978	1	0.976	0.977

across different methods of generating $\boldsymbol{x}_i$ (both from a multivariate normal and a multivariate t), different assumed distributions for the ε terms (normal or t-distribution), the covariance matrix structures with $\rho^{|i-j|}$ for ρ at different sizes from 0.25 to 0.75, as well as for signal strengths for $\boldsymbol{\beta}$.

We next examine the performance of the proposed tests with different choices of $\boldsymbol{\alpha}$, different values of k_n and different sparsity under local alternatives. To this end, we let $n = 400$, $\boldsymbol{x}_i \sim N(\mathbf{0}, \Sigma)$ with (i, j)-element of Σ being $(0.50^{|i-j|})$, $\varepsilon \sim N(0, 1)$. We consider the following cases.

Case C1: The default parameters used in Tables 1–3 and Tables S.1–S.3. That is, $\boldsymbol{\alpha} = (1, \dots, 1)^\top/\sqrt{p}$, $k_n = (p/\ln(p))^{1/2}$ and 25% of elements of the $\boldsymbol{\beta}$ are nonzero and equal $\delta/\sqrt{n}$ under the local alternative.

Case C2: We set $\boldsymbol{\alpha}$ to be a uniformly distributed random direction with $||\boldsymbol{\alpha}|| = 1$, set $k_n = (p/\ln(p))^{1/2}$ and 25% of elements of the $\boldsymbol{\beta}$ are nonzero and equal $\delta/\sqrt{n}$ under the local alternative.

TABLE 2
Type I error rate and power with $\boldsymbol{x}_i \sim \mathrm{N}(\mathbf{0}, \Sigma)$ and $\rho = 0.5$

(n, p)	δ	$\varepsilon_i \sim N(0, 1)$			$\varepsilon_i \sim (4/6)^{1/2}\mathrm{t}(6)$		
		New	ZC	CGZ	New	ZC	CGZ
				Case 1: Fixed Signals			
(200, 300)	0	0.06	0.057	0.062	0.046	0.043	0.042
	0.13	0.137	0.06	0.065	0.151	0.046	0.05
	0.27	0.44	0.106	0.114	0.486	0.098	0.102
	0.4	0.773	0.233	0.239	0.813	0.241	0.244
	0.53	0.951	0.496	0.503	0.952	0.493	0.502
(400, 600)	0	0.047	0.045	0.049	0.046	0.054	0.052
	0.12	0.195	0.056	0.065	0.221	0.063	0.064
	0.25	0.673	0.132	0.136	0.68	0.121	0.12
	0.38	0.952	0.323	0.324	0.967	0.351	0.357
	0.5	0.997	0.653	0.656	0.998	0.662	0.665
(800, 1200)	0	0.044	0.049	0.047	0.054	0.055	0.059
	0.12	0.388	0.069	0.067	0.386	0.064	0.066
	0.25	0.938	0.207	0.209	0.923	0.185	0.186
	0.37	0.997	0.551	0.555	1	0.551	0.552
	0.49	1	0.898	0.9	1	0.892	0.897
				Case 2: Random Signals			
(200, 300)	0	0.064	0.049	0.054	0.046	0.043	0.043
	0.35	0.24	0.076	0.081	0.23	0.063	0.068
	0.71	0.701	0.232	0.241	0.717	0.225	0.231
	1.06	0.961	0.555	0.56	0.943	0.562	0.565
	1.41	0.999	0.859	0.864	0.994	0.848	0.848
(400, 600)	0	0.054	0.057	0.058	0.045	0.049	0.049
	0.25	0.202	0.061	0.066	0.234	0.054	0.057
	0.5	0.663	0.168	0.167	0.688	0.142	0.143
	0.75	0.961	0.416	0.417	0.957	0.418	0.421
	1	0.999	0.741	0.749	0.999	0.747	0.747
(800, 1200)	0	0.04	0.05	0.05	0.051	0.035	0.038
	0.12	0.105	0.05	0.052	0.124	0.04	0.042
	0.25	0.379	0.054	0.061	0.362	0.069	0.066
	0.37	0.747	0.117	0.117	0.708	0.111	0.111
	0.49	0.947	0.222	0.222	0.936	0.227	0.234

Case C3: We set $k_n = (1/2)(p\ln(p))^{1/2}$, $\boldsymbol{\alpha} = (1, \dots, 1)^\top/\sqrt{p}$ and 25% of elements of the $\boldsymbol{\beta}$ are nonzero and equal $\delta/\sqrt{n}$ under the local alternative.

Case C4: We set $k_n = 2(p\ln(p))^{1/2}$, $\boldsymbol{\alpha} = (1, \dots, 1)^\top/\sqrt{p}$ and 25% of elements of the $\boldsymbol{\beta}$ are nonzero and equal $\delta/\sqrt{n}$ under the local alternative.

Case C5: We set one-tenth of elements in $\boldsymbol{\beta}$ to be nonzero and equal $\delta/\sqrt{n}$, and set $\boldsymbol{\alpha} = (1, \dots, 1)^\top/\sqrt{p}$ and $k_n = (p/\ln(p))^{1/2}$.

Case C6: We set one-third of elements in $\boldsymbol{\beta}$ to be nonzero and equal $\delta/\sqrt{n}$, and set $\boldsymbol{\alpha} = (1, \dots, 1)^\top/\sqrt{p}$ and $k_n = (p/\ln(p))^{1/2}$.

Case C1 can be viewed as the reference case for the rest of this section. Case C2 is designed to examine whether the proposed test is sensitive to the choice of $\boldsymbol{\alpha}$. Cases C3 and C4 are designed to investigate the robustness of the proposed test to k_n. Cases C5 and C6 are used to examine the performance of the proposed test under different sparsity levels in $\boldsymbol{\beta}$ under the local alternative. Simulation results are summarized in Table 4, from which we can see that the proposed test controls the Type I error rate well for all cases. Comparing Cases C1

TABLE 3
Type I error rate and power with $\boldsymbol{x}_i \sim \mathrm{N}(\mathbf{0}, \Sigma)$ and $\rho = 0.75$

		$\varepsilon_i \sim N(0, 1)$			$\varepsilon_i \sim (4/6)^{1/2}\mathrm{t}(6)$		
(n, p)	δ	New	ZC	CGZ	New	ZC	CGZ
				Case 1: Fixed Signals			
(200, 300)	0	0.057	0.05	0.052	0.052	0.039	0.044
	0.13	0.289	0.092	0.094	0.318	0.09	0.087
	0.27	0.811	0.35	0.357	0.839	0.37	0.369
	0.4	0.985	0.771	0.776	0.981	0.796	0.797
	0.53	0.999	0.961	0.963	1	0.966	0.962
(400, 600)	0	0.049	0.048	0.049	0.035	0.048	0.054
	0.12	0.465	0.099	0.103	0.484	0.108	0.111
	0.25	0.968	0.467	0.474	0.979	0.485	0.481
	0.38	1	0.911	0.915	1	0.921	0.923
	0.5	1	0.998	0.998	1	0.996	0.996
(800, 1200)	0	0.049	0.047	0.05	0.05	0.056	0.059
	0.12	0.776	0.161	0.161	0.783	0.133	0.132
	0.25	0.999	0.727	0.727	1	0.716	0.72
	0.37	1	0.993	0.994	1	0.992	0.992
	0.49	1	1	1	1	1	1
				Case 2: Random Signals			
(200, 300)	0	0.054	0.051	0.051	0.043	0.033	0.038
	0.18	0.166	0.066	0.068	0.159	0.049	0.047
	0.35	0.488	0.168	0.17	0.506	0.148	0.153
	0.53	0.831	0.389	0.395	0.823	0.383	0.389
	0.71	0.973	0.684	0.683	0.958	0.673	0.681
(400, 600)	0	0.049	0.05	0.056	0.053	0.046	0.048
	0.12	0.15	0.059	0.061	0.166	0.059	0.062
	0.25	0.449	0.117	0.118	0.506	0.118	0.113
	0.38	0.807	0.247	0.256	0.825	0.277	0.277
	0.5	0.972	0.521	0.517	0.968	0.53	0.537
(800, 1200)	0	0.04	0.056	0.057	0.048	0.043	0.043
	0.12	0.251	0.06	0.063	0.249	0.068	0.068
	0.25	0.793	0.155	0.153	0.762	0.161	0.164
	0.37	0.984	0.413	0.414	0.98	0.399	0.401
	0.49	1	0.782	0.784	1	0.752	0.756

and C2, Table 4 implies that the uniform random direction is less powerful than the default one. Comparing Cases C1, C3, and C4, we can see from Table 4 that the differences in power for the three different k_n methods are moderate. This implies that the performance of the

TABLE 4
The empirical power of T_n with different $\boldsymbol{\alpha}$, k_n and sparsity of signals

δ	C1	C2	C3	C4	C5	C6
0	0.047	0.038	0.055	0.055	0.047	0.047
0.125	0.195	0.055	0.233	0.246	0.067	0.328
0.250	0.673	0.140	0.734	0.737	0.128	0.888
0.375	0.952	0.337	0.970	0.967	0.299	0.997
0.500	0.997	0.653	0.999	0.998	0.491	1.000

TABLE 5

Summary of Significant(*S*) *and nonsignificant*(*NS*) *results for GO groupings regressed against* 204531_*s_at* *expression at significance level* 0.05

Bonferroni correction	No		Yes	
	New (S)	New (NS)	New (S)	New (NS)
ZC (S)	5	9	0	0
ZC (NS)	5	118	2	135

proposed test is robust to different values of k_n. Comparing Cases C1, C5, and C6, we see that the proposed test has the highest power in Case C6, and has the lowest power in Case C5. This is expected since the signals in Case C6 are the strongest, while the signals in Case C5 are the weakest.

3.3. *Real data analysis.* In this section, we demonstrate the proposed methodology via an application to the transNOAH breast cancer trial (GEO series GSE50948), which can be found at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE50948 (Prat et al. (2014)). The data consists of genome-wide gene expression profiling from biopsies from 114 pre-treated patients with HER2 positive tumors, along with 42 patients with HER2-negative tumors as a control cohort. From this data, we used the two Affymetrix Probe Set IDs associated with BRCA1, 204531_s_at and 211851_x_at, as our two response variables, while grouping the remaining Probe Set IDs by Gene Ontology (GO) Molecular Function IDs to provide a robust 137 sets of predictors for testing. Since this paper focuses on high-dimensional regression problems, only GO IDs with more than 158 Probe Set IDs were used in this empirical analysis. Our goal here is to determine if, for any of these covariates, the coefficients in the vector $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_p\}$ are significantly different than zero, and thus intend to test

$$H_0 : \boldsymbol{\beta} = \mathbf{0} \quad \text{versus} \quad H_1 : \boldsymbol{\beta} \neq \mathbf{0}. \tag{15}$$

To construct the test statistic T_n, we first create the necessary score equation vectors $\boldsymbol{z} = (\boldsymbol{x} - \bar{\boldsymbol{x}})\{(y - \bar{y} - (\boldsymbol{x} - \bar{\boldsymbol{x}})^\top \boldsymbol{\beta}_0\} = \boldsymbol{x} y$ under the null hypothesis. As in our simulation study in the previous section, we set $\boldsymbol{\alpha} = 1_p/\sqrt{p}$ and $k_n = (p/\ln p)^{1/2}$. For 211851_x_at, both our method and Zhong and Chen found statistical significance for all sets of predictors, and thus the results are excluded here. For 204531_s_at, Table 5 provides the number of predictor sets on which the new method and ZC test agreed and disagreed. Specific test statistics results for each GO set can be found in Table S.8 in the Supplementary Material (Zhao et al. (2024)). The new method identifies as significant two GO groups with Bonferroni correction, while the ZC method fails to detect these two GO groups.

4. Concluding remarks. Testing on high-dimensional regression coefficients in linear models plays an important role in the analysis of high-dimensional data. In this paper, we develop a new procedure for testing high-dimensional regression coefficients. We further establish the asymptotic normality of the proposed test statistic. Our simulation study implies that the limiting null distribution can be directly used for calculating the critical value of the proposed test. Our numerical comparison demonstrates that the proposed test outperforms the existing ones in terms of power.

APPENDIX: PROOFS OF LEMMAS 2.1 AND 2.2

PROOF OF LEMMA 2.1. Note that the left-hand side in (10) is invariant under an orthogonal transformation on the regression coefficient vector and $\boldsymbol{x}$ in the sense that $\boldsymbol{\beta}_* = Q\boldsymbol{\beta}$,

$\boldsymbol{\beta}_{0*} = Q\boldsymbol{\beta}_0$ and $\boldsymbol{x}_{i*} = Q\boldsymbol{x}_i$, which implies that $\operatorname{cov}(\boldsymbol{x}_{i*}) = Q\Sigma_X Q^\top$. Thus, without loss of generality, assume that Σ_X is a diagonal matrix, and denote $\Sigma_X = \operatorname{diag}(\sigma_1^2, \dots, \sigma_p^2)$.

Let $S_{n,p} = n(n\|\bar{z}_\beta\|^2 - \operatorname{tr}(S_{Z_\beta}))$, we have

$$S_{n,p} = \sum_{k=1}^{p} \sigma_k^2 \left(\sum_{i=1}^{n} \epsilon_i X_{0ik} \right)^2 - \sum_{k=1}^{p} \sigma_k^2 \left(\sum_{i=1}^{n} \epsilon_i^2 X_{0ik}^2 \right),$$

where p is the number of covariates in X which grows with n.

We want to establish the asymptotic normality of $S_{n,p}$ using the central limit theorem (CLT) for martingales. We divide our proof into three steps. Step 1 shows that $\{S_{n,p}, n \geq 1\}$ is a zero-mean square-integrable martingale with filtration $\mathcal{F}_n$, where $\mathcal{F}_n$ is the σ-field generated by the random variables ϵ_i, $i = 1, 2, \dots, n$, and X_{0ik}, $i = 1, 2, \dots, n$, $k = 1, 2, \dots, p$. Steps 2 and 3 check the conditions of the CLT for $S_{n,p}$. Namely, Step 2 shows the variance of $S_{n,p}$, s_n^2, to be

$$s_n^2 = 2n(n-1)\operatorname{tr}(\Sigma_{Z_\beta}^2) = 2n(n-1)\sigma^4 \sum_{k=1}^{p} \sigma_k^4,$$

and that $\frac{d_n^2}{s_n^2} = 1$, where d_n^2 is the quadratic variation which is defined as $d_n^2 := \sum_{i=1}^{n} \mathrm{E}((S_{i,p} - S_{i-1,p})^2 | \mathcal{F}_{i-1})$. Step 3 checks Lyapunov's condition for $S_{n,p}$

$$\frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} \mathrm{E}\big((S_{i,p} - S_{i-1,p})^{2+\delta}\big) \to 0$$

as $n \to \infty$ for $\delta = 2$.

Step 1: We have $\mathcal{F}_i \subseteq \mathcal{F}_j$ for $i < j$, and

$$\mathrm{E}(S_{n,p} - S_{n-1,p} | \mathcal{F}_{n-1}) = \mathrm{E}(U_n | \mathcal{F}_{n-1}),$$

where

$$U_n = \sum_{k=1}^{p} \sigma_k^2 \left[\left(\sum_{i=1}^{n} \epsilon_i X_{0ik} \right)^2 - \left(\sum_{i=1}^{n-1} \epsilon_i X_{0ik} \right)^2 \right] - \epsilon_n^2 \sum_{k=1}^{p} \sigma_k^2 X_{0nk}^2.$$

Furthermore, we have

$$\begin{aligned}
\mathrm{E}(U_n | \mathcal{F}_{n-1}) &= \mathrm{E}\left(\sum_{k=1}^{p} \sigma_k^2 \left[\left(\sum_{i=1}^{n} \epsilon_i X_{0ik} \right)^2 - \left(\sum_{i=1}^{n-1} \epsilon_i X_{0ik} \right)^2 \right] \middle| \mathcal{F}_{n-1} \right) \\
&\quad - \mathrm{E}\left(\epsilon_n^2 \sum_{k=1}^{p} \sigma_k^2 X_{0nk}^2 \middle| \mathcal{F}_{n-1} \right) \\
&= \mathrm{E}\left(\sum_{k=1}^{p} \sigma_k^2 \left[\epsilon_n^2 X_{0nk}^2 + 2\epsilon_n X_{0nk} \left(\sum_{i=1}^{n-1} \epsilon_i X_{0ik} \right) \right] \middle| \mathcal{F}_{n-1} \right) \\
&\quad - \mathrm{E}\left(\epsilon_n^2 \sum_{k=1}^{p} \sigma_k^2 X_{0nk}^2 \middle| \mathcal{F}_{n-1} \right) \\
&= \mathrm{E}\left(\sum_{k=1}^{p} \sigma_k^2 \epsilon_n^2 X_{0nk}^2 \middle| \mathcal{F}_{n-1} \right) - \mathrm{E}\left(\epsilon_n^2 \sum_{k=1}^{p} \sigma_k^2 X_{0nk}^2 \middle| \mathcal{F}_{n-1} \right) \\
&= 0.
\end{aligned}$$

Moreover, since ϵ_i and X_{0ik} have bounded fourth order moments, $S_{n,p}$ is a square-integrable martingale with respect to the σ-field $\mathcal{F}_n$.

Step 2: We first calculate the variance of $S_{n,p}$. Since $s_n^2 = \mathrm{Var}(S_{n,p}) = \mathrm{E}(S_{n,p}^2)$, we have

$$s_n^2 = \sum_{k=1}^p \sigma_k^4 \mathrm{E}\Biggl[\Biggl(\sum_{i=1}^n \epsilon_i X_{0ik}\Biggr)^2 - \sum_{i=1}^n \epsilon_i^2 X_{0ik}^2\Biggr]^2 + 2 \sum_{1\le k_1<k_2\le p} \sigma_{k_1}^2\sigma_{k_2}^2 \mathrm{E}(t_{nk_1k_2}),$$

where

$$t_{nk_1k_2} = \Biggl[\Biggl(\sum_{i=1}^n \epsilon_i X_{0ik_1}\Biggr)^2 - \sum_{i=1}^n \epsilon_i^2 X_{0ik_1}^2\Biggr]\Biggl[\Biggl(\sum_{i=1}^n \epsilon_i X_{0ik_2}\Biggr)^2 - \sum_{i=1}^n \epsilon_i^2 X_{0ik_2}^2\Biggr],$$

and we have

$$\mathrm{E}(t_{nk_1k_2}) = \mathrm{E}\bigl[\mathrm{E}(t_{nk_1k_2}|\epsilon_i, i = 1, 2, \dots, n)\bigr] = 0.$$

Hence,

$$\begin{aligned} s_n^2 &= \sum_{k=1}^p \sigma_k^4 \mathrm{E}\Biggl[\Biggl(\sum_{i=1}^n \epsilon_i X_{0ik}\Biggr)^2 - \sum_{i=1}^n \epsilon_i^2 X_{0ik}^2\Biggr]^2 = \sum_{k=1}^p \sigma_k^4 \mathrm{E}\Biggl[2 \sum_{1\le i_1<i_2\le n} \epsilon_{i_1}\epsilon_{i_2} X_{0i_1k} X_{0i_2k}\Biggr]^2 \\ &= \sum_{k=1}^p 4\sigma_k^4 \Biggl\{\sum_{1\le i_1<i_2\le n} \mathrm{E}\bigl[\epsilon_{i_1}^2\epsilon_{i_2}^2 X_{0i_1k}^2 X_{0i_2k}^2\bigr]\Biggr\} = \sum_{k=1}^p 2n(n-1)\sigma_k^4\sigma^4. \end{aligned}$$

For $d_n^2 = \sum_{i=1}^n \mathrm{E}((S_{i,p} - S_{i-1,p})^2|\mathcal{F}_{i-1})$, we have

$$d_n^2 = \sum_{i=1}^n \mathrm{E}\bigl((S_{i,p} - S_{i-1,p})^2|\mathcal{F}_{i-1}\bigr) = \sum_{i=1}^n \mathrm{E}\bigl(U_i^2|\mathcal{F}_{i-1}\bigr).$$

As a result, we have that

$$\begin{aligned} \mathrm{E}\bigl(U_n^2|\mathcal{F}_{n-1}\bigr) &= 4\mathrm{E}\Biggl(\Biggl[\sum_{k=1}^p \sigma_k^2\Biggl(\epsilon_n X_{0nk}\Biggl(\sum_{i=1}^{n-1} \epsilon_i X_{0ik}\Biggr)\Biggr)\Biggr]^2 |\mathcal{F}_{n-1}\Biggr) \\ &= 4\sigma^2 \sum_{k=1}^p \sigma_k^4\Biggl[\Biggl(\sum_{i=1}^{n-1} \epsilon_i X_{0ik}\Biggr)^2\Biggr]. \end{aligned}$$

To establish the condition for the central limit theorem for the martingale $S_{n,p}$, we need to check that $\frac{d_n^2}{\mathrm{E}d_n^2} \xrightarrow{p} 1$. To show $\frac{d_n^2}{\mathrm{E}d_n^2} \xrightarrow{p} 1$, we only need to show $\frac{\mathrm{Var}(d_n^2)}{(\mathrm{E}d_n^2)^2} \to 0$. It is easy to have

$$(\mathrm{E}d_n^2)^2 \sim n^4\Biggl(\sum_{k=1}^p \sigma_k^4\Biggr)^2 \sim n^4 \operatorname{tr}^2(\Sigma_X^2),$$

where $m \sim n$ means that m and n are of the same order asymptotically. For $\mathrm{Var}(d_n^2)$, we have

$$\begin{aligned} \mathrm{Var}\bigl(d_n^2\bigr) &= \mathrm{Var}\Biggl(\sum_{i=1}^n \mathrm{E}\bigl(U_i^2|\mathcal{F}_{i-1}\bigr)\Biggr) = O\bigl(n^2\bigr)\mathrm{Var}\bigl(\mathrm{E}\bigl(U_n^2|\mathcal{F}_{n-1}\bigr)\bigr) \\ &= O\bigl(n^2\bigr)\mathrm{Var}\Biggl(\sigma^2 \sum_{k=1}^p \sigma_k^4\Biggl[\Biggl(\sum_{i=1}^{n-1} \epsilon_i X_{0ik}\Biggr)^2\Biggr]\Biggr) \qquad (16) \end{aligned}$$

$$
\begin{aligned}
&= O(n^2) \sum_{k=1}^{p} \sigma_k^8 \operatorname{Var}\left(\left[\left(\sum_{i=1}^{n-1} \epsilon_i X_{0ik}\right)^2\right]\right) \\
&\quad + O(n^2) \sum_{k<j}^{p} \sigma_k^4 \sigma_j^4 \operatorname{Cov}\left(\left(\sum_{i=1}^{n-1} \epsilon_i X_{0ik}\right)^2, \left(\sum_{i=1}^{n-1} \epsilon_i X_{0ij}\right)^2\right).
\end{aligned}
$$

Furthermore, we have

$$
\operatorname{Var}\left(\left[\left(\sum_{i=1}^{n-1} \epsilon_i X_{0ik}\right)^2\right]\right) \le \mathrm{E}\left[\left(\sum_{i=1}^{n-1} \epsilon_i X_{0ik}\right)^4\right] = O(n^2).
$$

Also, for $1 \le k < j \le p$, we have

$$
\begin{aligned}
&\operatorname{Cov}\left(\left(\sum_{i=1}^{n-1} \epsilon_i X_{0ik}\right)^2, \left(\sum_{i=1}^{n-1} \epsilon_i X_{0ij}\right)^2\right) \\
&\quad = \operatorname{Cov}\left(\sum_{i=1}^{n-1} \epsilon_i^2 X_{0ik}^2 + 2\sum_{i<m}^{n-1} \epsilon_i X_{0ik} \epsilon_m X_{0mk}, \sum_{i=1}^{n-1} \epsilon_i^2 X_{0ij}^2 + 2\sum_{i<m}^{n-1} \epsilon_i X_{0ik} \epsilon_m X_{0mj}\right) \\
&\quad = \sum_{i=1}^{n-1} \operatorname{Cov}(\epsilon_i^2 X_{0ik}^2, \epsilon_i^2 X_{0ij}^2) = O(n).
\end{aligned}
$$

In sum, we have that

$$
\begin{aligned}
\operatorname{Var}(d_n^2) &= O\left(n^4 \sum_{k=1}^{p} \sigma_k^8\right) + O\left(n^3 \left(\sum_{k=1}^{p} \sigma_k^4\right)^2\right) = O(n^4 \operatorname{tr}(\Sigma_X^4)) + O(n^3 \operatorname{tr}^2(\Sigma_X^2)) \\
&= o\big((\mathrm{E} d_n^2)^2\big).
\end{aligned}
$$

Thus, we have shown that $\frac{d_n^2}{\mathrm{E} d_n^2} \xrightarrow{p} 1$, which finishes the proof of Step 2.

Step 3: In this step, we check Lyapunov's condition for $S_{n,p}$. We want to show that $\frac{1}{s_n^{2+\delta}} \sum_{i=1}^{n} \mathrm{E}((S_{i,p} - S_{i-1,p})^{2+\delta}) \to 0$ as $n \to \infty$ for $\delta = 2$. We have

$$
\frac{1}{s_n^4} \sum_{i=1}^{n} \mathrm{E}\big((S_{i,p} - S_{i-1,p})^4\big) = \frac{1}{s_n^4} \sum_{i=1}^{n} S_{i1},
$$

where S_{i1} is denoted as S_{n1} below for ease of notation, and

$$
S_{n1} = 16\mathrm{E}\left(\left[\sum_{k=1}^{p} \sigma_k^2 \varepsilon_n X_{0nk} \left\{\sum_{i=1}^{n-1} \varepsilon_i X_{0ik}\right\}\right]^4\right).
$$

We want to show that $\sum_{i=1}^{n} S_{i1} = o(s_n^4)$. For S_{n1}, we have that

$$
\begin{aligned}
&16\mathrm{E}\left(\left[\sum_{k=1}^{p} \sigma_k^2 \varepsilon_n X_{0nk} \left\{\sum_{i=1}^{n-1} \varepsilon_i X_{0ik}\right\}\right]^4\right) \\
&\quad = 16\mathrm{E}(\varepsilon_n^4)\mathrm{E}\left(\sum_{k=1}^{p} \sigma_k^8 X_{0nk}^4 \left[\sum_{i=1}^{n-1} \varepsilon_i X_{0ik}\right]^4\right) \\
&\qquad + 16\mathrm{E}(\varepsilon_n^4)\mathrm{E}\left(6 \sum_{k \ne j}^{p} \sigma_k^4 X_{0nk}^2 \left[\sum_{i=1}^{n-1} \varepsilon_i X_{0ik}\right]^2 \sigma_j^4 X_{0nj}^2 \left[\sum_{i=1}^{n-1} \varepsilon_i X_{0ij}\right]^2\right)
\end{aligned}
$$

$$= 16\mathrm{E}(\varepsilon_n^4)(\sum_{k=1}^{p}\sigma_k^8\mathrm{E}[X_{0nk}^4]\left(\sum_{i=1}^{n-1}\mathrm{E}[\varepsilon_i^4]\mathrm{E}[X_{0ik}^4] + 6\sum_{i\neq j}^{n-1}\mathrm{E}[\varepsilon_i^2]\mathrm{E}[X_{0ik}^2]\mathrm{E}[\varepsilon_j^2]\mathrm{E}[X_{0jk}^2]\right)$$

$$+ 96\mathrm{E}(\varepsilon_n^4)\left[\sum_{k\neq j}^{p}\sigma_k^4\sigma_j^4\left(\sum_{i=1}^{n-1}\mathrm{E}(\varepsilon_i^4) + (n-1)(n-2)\sigma^4 + 4\sum_{i\neq h}^{n-1}\sigma^4\right)\right].$$

Furthermore, since the fourth-order moments of ϵ_i and X_{0ik} are bounded, we get that

$$S_{n1} = O\left(n^2\left(\sum_{k=1}^{p}\sigma_k^8\right)\right) + O\left(n^2\left(\sum_{k=1}^{p}\sigma_k^4\right)^2\right) = O(n^2\,\mathrm{tr}(\Sigma_X^4)) + O(n^2\,\mathrm{tr}^2(\Sigma_X^2)).$$

Then we have $\sum_{i=1}^{n} S_{i1} = o(s_n^4)$, which completes the proof of Lemma 2.1. □

PROOF OF LEMMA 2.2. For the first term nW_n in the numerator, we have

$$nW_n = n\tilde{b}_n^2 - n(\tilde{b}_n^2 - b_n^2) - \mathrm{tr}(\Omega_1),$$

where $b_n = \{\|\bar{z}\|^2 + k_n|\boldsymbol{\alpha}^T\bar{z}|^2\}^{1/2}$, $\tilde{b}_n = \{\|\bar{z}_{\beta_0}\|^2 + k_n|\boldsymbol{\alpha}^T\bar{z}_{\beta_0}|^2\}^{1/2}$, and $\|\cdot\|$ is the Euclidean L_2 norm. Furthermore, under the local alternative hypothesis (7), we have that

$$\begin{aligned}
|b_n^2 - \tilde{b}_n^2| &\le \big|\|\bar{z}\|^2 - \|\bar{z}_{\beta_0}\|^2\big| + k_n\big(|\boldsymbol{\alpha}^T\bar{z}_{\beta_0}|^2 + |\boldsymbol{\alpha}^T\bar{z}|^2\big)\\
&= \big|(\bar{z} - \bar{z}_{\beta_0})^\top(\bar{z} + \bar{z}_{\beta_0})\big| + o_p\big(\sqrt{\mathrm{tr}(\Sigma_X^2)}/n\big)\\
&\le \big|(\bar{z} - \bar{z}_{\beta_0})^\top(\mathrm{E}(z) + \mathrm{E}(z_{\beta_0}))\big| + o_p(\|\bar{z} - \bar{z}_{\beta_0}\|) + o_p\big(\sqrt{\mathrm{tr}(\Sigma_X^2)}/n\big)\\
&= o_p(\|\bar{z} - \bar{z}_{\beta_0}\|) + o_p\big(\sqrt{\mathrm{tr}(\Sigma_X^2)}/n\big),
\end{aligned}$$

where the last equality is from $\|\mathrm{E}(z) + \mathrm{E}(z_{\beta_0})\| = O_p(n^{-1/2})$. From the definition of z and z_{β_0}, we get

$$\bar{z} - \bar{z}_{\beta_0} = (\bar{\boldsymbol{x}} - \boldsymbol{\mu}_X)(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_X)^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + (\boldsymbol{\mu}_X - \bar{\boldsymbol{x}})\bar{\varepsilon},$$

and

$$\|\bar{z} - \bar{z}_{\beta_0}\| \le \|(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_X)(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_X)^\top(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\| + \|(\boldsymbol{\mu}_X - \bar{\boldsymbol{x}})\bar{\varepsilon}\| = O_p\big(\sqrt{\mathrm{tr}(\Sigma_X^2)}/n\big).$$

Hence, $n(\tilde{b}_n^2 - b_n^2) = o_p\big(\sqrt{\mathrm{tr}(\Sigma_X^2)}\big)$. Additionally, we have

$$nW_n = n\tilde{b}_n^2 - \mathrm{tr}(\Omega_1) + o_p\big(\sqrt{\mathrm{tr}(\Sigma_X^2)}\big).$$

We also have

$$nW_n - n(\|\boldsymbol{\delta}_z\|^2 + k_n|\boldsymbol{\alpha}^T\boldsymbol{\delta}_z|^2) = n\tilde{b}_n^2 - n(\|\boldsymbol{\delta}_z\|^2 + k_n|\boldsymbol{\alpha}^T\boldsymbol{\delta}_z|^2) - \mathrm{tr}(\Omega_1) + o_p\big(\sqrt{\mathrm{tr}(\Sigma_X^2)}\big).$$

For $n\tilde{b}_n^2 - n(\|\boldsymbol{\delta}_z\|^2 + k_n|\boldsymbol{\alpha}^T\boldsymbol{\delta}_z|^2)$, we have

$$\begin{aligned}
&n\tilde{b}_n^2 - n(\|\boldsymbol{\delta}_z\|^2 + k_n|\boldsymbol{\alpha}^T\boldsymbol{\delta}_z|^2)\\
&\quad= n(\|\bar{z}_{\beta_0}\|^2 + k_n|\boldsymbol{\alpha}^T\bar{z}_{\beta_0}|^2) - n(\|\boldsymbol{\delta}_z\|^2 + k_n|\boldsymbol{\alpha}^T\boldsymbol{\delta}_z|^2)\\
&\quad= n(\|\bar{z}_{\beta_0} - \boldsymbol{\delta}_z\|^2 + k_n|\boldsymbol{\alpha}^T(\bar{z}_{\beta_0} - \boldsymbol{\delta}_z)|^2) + 2n\{(\bar{z}_{\beta_0} - \boldsymbol{\delta}_z)^\top\boldsymbol{\delta}_z + k_n\boldsymbol{\alpha}^\top(\bar{z}_{\beta_0} - \boldsymbol{\delta}_z)\boldsymbol{\delta}_z^\top\boldsymbol{\alpha}\}.
\end{aligned}$$

The second term $2n\{(\bar{z}_{\beta_0} - \boldsymbol{\delta}_z)^\top\boldsymbol{\delta}_z + k_n\boldsymbol{\alpha}^\top(\bar{z}_{\beta_0} - \boldsymbol{\delta}_z)\boldsymbol{\delta}_z^\top\boldsymbol{\alpha}\}$ is of order $o_p\big(\sqrt{\mathrm{tr}(\Sigma_X^2)}\big)$. In fact, we have

$$\mathrm{E}\big((2n(\bar{z}_{\beta_0} - \boldsymbol{\delta}_z)^\top\boldsymbol{\delta}_z)^2\big) = O_p(n^2)\mathrm{E}[(\bar{z}_\beta^\top\boldsymbol{\delta}_z)^2] = O_p(n\boldsymbol{\delta}_z^\top\Sigma_X\boldsymbol{\delta}_z) = o_p(\mathrm{tr}(\Sigma_X^2)),$$

and

$$\mathrm{E}\big((2nk_n\boldsymbol{\alpha}^\top(\bar{\mathbf{z}}_{\beta_0}-\boldsymbol{\delta}_z)\boldsymbol{\delta}_z^\top\boldsymbol{\alpha})^2\big)=O_p\big(nk_n^2(\boldsymbol{\alpha}^\top\boldsymbol{\delta}_z)^2\boldsymbol{\alpha}^\top\Sigma_X\boldsymbol{\alpha}\big)=o_p\big(\mathrm{tr}(\Sigma_X^2)\big).$$

Hence, the second term is of order $o_p(\sqrt{\mathrm{tr}(\Sigma_X^2)})$. Furthermore, for the first term we have

$$\begin{aligned}
&n\big(\|\bar{\mathbf{z}}_{\beta_0}-\boldsymbol{\delta}_z\|^2+k_n|\boldsymbol{\alpha}^T(\bar{\mathbf{z}}_{\beta_0}-\boldsymbol{\delta}_z)|^2\big)\\
&\quad=n\big(\|\bar{\mathbf{z}}_\beta\|^2+k_n|\boldsymbol{\alpha}^T\bar{\mathbf{z}}_\beta|^2\big)+n(\bar{\mathbf{z}}_{\beta_0}-\bar{\mathbf{z}}_\beta-\boldsymbol{\delta}_z)^\top(\bar{\mathbf{z}}_{\beta_0}+\bar{\mathbf{z}}_\beta-\boldsymbol{\delta}_z)\\
&\qquad+k_nn\boldsymbol{\alpha}^\top(\bar{\mathbf{z}}_{\beta_0}-\bar{\mathbf{z}}_\beta-\boldsymbol{\delta}_z)(\bar{\mathbf{z}}_{\beta_0}+\bar{\mathbf{z}}_\beta-\boldsymbol{\delta}_z)^\top\boldsymbol{\alpha}\\
&\quad=n\big(\|\bar{\mathbf{z}}_\beta\|^2+k_n|\boldsymbol{\alpha}^T\bar{\mathbf{z}}_\beta|^2\big)+o_p\big(\sqrt{\mathrm{tr}(\Sigma_X^2)}\big).
\end{aligned}$$

The last equality holds because $\|\bar{\mathbf{z}}_{\beta_0}-\bar{\mathbf{z}}_\beta-\boldsymbol{\delta}_z\|=O_p(n^{-1/2})$ and $\|\bar{\mathbf{z}}_{\beta_0}+\bar{\mathbf{z}}_\beta-\boldsymbol{\delta}_z\|=O_p(1)$. In sum, we have that

$$\begin{aligned}
&\frac{nW_n-n(\|\boldsymbol{\delta}_z\|^2+k_n|\boldsymbol{\alpha}^T\boldsymbol{\delta}_z|^2)}{\{2\,\mathrm{tr}(\Omega_2)\}^{1/2}}\\
&\quad=\frac{n(\|\bar{\mathbf{z}}_\beta\|^2+k_n|\boldsymbol{\alpha}^T\bar{\mathbf{z}}_\beta|^2)-\mathrm{tr}(\Omega_1)}{\{2\,\mathrm{tr}(\Omega_2)\}^{1/2}}+\frac{o_p(\sqrt{\mathrm{tr}(\Sigma_X^2)})}{\{2\,\mathrm{tr}(\Omega_2)\}^{1/2}}\\
&\quad=\frac{n(\|\bar{\mathbf{z}}_\beta\|^2+k_n|\boldsymbol{\alpha}^T\bar{\mathbf{z}}_\beta|^2)-\mathrm{tr}(\Omega_1)}{\{2\,\mathrm{tr}(\Omega_2)\}^{1/2}}+o_p(1).
\end{aligned}$$

Let $T_{n1}=n(\|\bar{\mathbf{z}}_\beta\|^2+k_n|\boldsymbol{\alpha}^T\bar{\mathbf{z}}_\beta|^2)-\mathrm{tr}(\Omega_1)$. Similar to Step 2 in the proof of Lemma 2.1, we can show that $\mathrm{E}(T_{n1})=0$ and $\mathrm{Var}(T_{n1})=2\frac{n-1}{n}\,\mathrm{tr}(\Omega_2)$. T_{n1} can be further decomposed as $T_{n1}=T_{n2}+T_{n3}-T_{n4}$, where $T_{n2}=n\|\bar{\mathbf{z}}_\beta\|^2-\mathrm{tr}(S_{Z_\beta})$, $T_{n3}=nk_n|\boldsymbol{\alpha}^T\bar{\mathbf{z}}_\beta|^2$, and $T_{n4}=k_n\boldsymbol{\alpha}^\top S_{Z_\beta}\boldsymbol{\alpha}$. From Lemma 2.1, we know that $\frac{T_{n2}}{\sqrt{2\sigma^4\,\mathrm{tr}(\Sigma_X^2)}}\xrightarrow{d}N(0,1)$. Since T_{n3} and T_{n4} are both nonnegative, and it is easy to verify that $\mathrm{E}(T_{n3})=\mathrm{E}(T_{n4})=o_p(\sqrt{\mathrm{tr}(\Sigma_X^2)})$, both T_{n3} and T_{n4} are of order $o_p(\sqrt{\mathrm{tr}(\Sigma_X^2)})$. Hence, we have

$$\frac{T_{n1}}{\sqrt{2\sigma^4\,\mathrm{tr}(\Sigma_X^2)}}=\frac{T_{n2}}{\sqrt{2\sigma^4\,\mathrm{tr}(\Sigma_X^2)}}+\frac{T_{n3}-T_{n4}}{\sqrt{2\sigma^4\,\mathrm{tr}(\Sigma_X^2)}}\xrightarrow{d}N(0,1).$$

Furthermore, we have

$$\mathrm{tr}(\Omega_2)=\sigma^4\,\mathrm{tr}(\Sigma_X^2)+2k_n\sigma^4\boldsymbol{\alpha}^\top\Sigma_X^2\boldsymbol{\alpha}+k_n^2\sigma^4(\boldsymbol{\alpha}^\top\Sigma_X\boldsymbol{\alpha})^2=\sigma^4\,\mathrm{tr}(\Sigma_X^2)+o\big(\mathrm{tr}(\Sigma_X^2)\big).$$

Hence, $\frac{\sqrt{2\sigma^4\,\mathrm{tr}(\Sigma_X^2)}}{\sqrt{2\,\mathrm{tr}(\Omega_2)}}\to 1$ in probability. In sum, we have that

$$\frac{T_{n1}}{\sqrt{2\,\mathrm{tr}(\Omega_2)}}=\frac{\sqrt{2\sigma^4\,\mathrm{tr}(\Sigma_X^2)}}{\sqrt{2\,\mathrm{tr}(\Omega_2)}}\frac{T_{n1}}{\sqrt{2\sigma^4\,\mathrm{tr}(\Sigma_X^2)}}\xrightarrow{d}N(0,1),$$

which completes the proof of Lemma 2.2. □

Acknowledgments. Alex Zhao and Changcheng Li are the first authors of this paper. The authors would like to thank the Associate Editor and reviewers for their constructive comments, which lead to significant improvement of this work.

Funding. The research of R. Li and was supported by NIH Grants R01AI170249 and R01AI136664. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

SUPPLEMENTARY MATERIAL

Supplement to "Testing high-dimensional regression coefficients in linear models" (DOI: 10.1214/24-AOS2420SUPP; .pdf). The supplement contains additional numerical results.

REFERENCES

ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011a). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.* **39** 2533–2556. MR2906877 https://doi.org/10.1214/11-AOS910

ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011b). Supplement to "Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism.". https://doi.org/10.1214/11-AOS910SUPP

BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305

CHEN, L. S., PAUL, D., PRENTICE, R. L. and WANG, P. (2011). A regularized Hotelling's T^2 test for pathway analysis in proteomic studies. *J. Amer. Statist. Assoc.* **106** 1345–1360. MR2896840 https://doi.org/10.1198/jasa.2011.ap10599

CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697 https://doi.org/10.1214/09-AOS716

CHUDIK, A., KAPETANIOS, G. and PESARAN, M. H. (2018). A one covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models. *Econometrica* **86** 1479–1512. MR3843496 https://doi.org/10.3982/ECTA14176

CUI, H., GUO, W. and ZHONG, W. (2018). Test for high-dimensional regression coefficients using refitted cross-validation variance estimation. *Ann. Statist.* **46** 958–988. MR3797993 https://doi.org/10.1214/17-AOS1573

CUI, X., LI, R., YANG, G. and ZHOU, W. (2020). Empirical likelihood test for a large-dimensional mean vector. *Biometrika* **107** 591–607. MR4138978 https://doi.org/10.1093/biomet/asaa005

GOEMAN, J. J., VAN DE GEER, S. A. and VAN HOUWELINGEN, H. C. (2006). Testing against a high dimensional alternative. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 477–493. MR2278336 https://doi.org/10.1111/j.1467-9868.2006.00551.x

HUANG, Y., LI, C., LI, R. and YANG, S. (2022). An overview of tests on high-dimensional means. *J. Multivariate Anal.* **188** Paper No. 104813, 14 pp. MR4353841 https://doi.org/10.1016/j.jmva.2021.104813

LAN, W., WANG, H. and TSAI, C.-L. (2014). Testing covariates in high-dimensional regression. *Ann. Inst. Statist. Math.* **66** 279–301. MR3171406 https://doi.org/10.1007/s10463-013-0414-0

LIU, W. and LUO, S. (2014). Hypothesis testing for high-dimensional regression models. Technical report. Available at https://math.sjtu.edu.cn/faculty/weidongl/Publication/regression.pdf.

PRAT, A., BIANCHINI, G., THOMAS, M., BELOUSOV, A., CHEANG, M. C., KOEHLER, A., GÓMEZ, P., SEMIGLAZOV, V., EIERMANN, W. et al. (2014). Based PAM50 subtype predictor identifies higher responses and improved survival outcomes in HER2-positive breast cancer in the NOAH study. *Clin. Cancer Res.* **20** 511–521.

VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science. Cambridge Series in Statistical and Probabilistic Mathematics* **47**. Cambridge Univ. Press, Cambridge. MR3837109 https://doi.org/10.1017/9781108231596

XIA, Y., CAI, T. and CAI, T. T. (2018). Two-sample tests for high-dimensional linear regression with an application to detecting interactions. *Statist. Sinica* **28** 63–92. MR3752252

ZHAO, A., LI, C., LI, R. and ZHANG, Z. (2024). Supplement to "Testing high-dimensional regression coefficients in linear models." https://doi.org/10.1214/24-AOS2420SUPP

ZHONG, P.-S. and CHEN, S. X. (2011). Tests for high-dimensional regression coefficients with factorial designs. *J. Amer. Statist. Assoc.* **106** 260–274. MR2816719 https://doi.org/10.1198/jasa.2011.tm10284