

The relative effects of dimensionality and multiplicity of hypotheses on the F -test in linear regression

Lukas Steinberger

*Department of Statistics and Operations Research
University of Vienna
Oskar-Morgenstern-Platz 1
1090 Vienna, Austria
e-mail: lukas.steinberger@univie.ac.at*

Abstract: Recently, several authors have re-examined the power of the classical F -test in a non-Gaussian linear regression under a “large- p , large- n ” framework [e.g. 29, 27]. They highlight the loss of power as the number of regressors p increases relative to sample size n . These papers essentially focus only on the overall test of the null hypothesis that all p slope coefficients are equal to zero. Here, we consider the general case of testing q linear hypotheses on the $p + 1$ -dimensional regression parameter vector that includes p slope coefficients and an intercept parameter. In the case of Gaussian design, we describe the dependence of the local asymptotic power function on both the relative number of parameters p/n and the relative number of hypotheses q/n being tested, showing that the negative effect of dimensionality is less severe if the number of hypotheses is small. Using the recent work of Srivastava and Vershynin [23] on high-dimensional sample covariance matrices we are also able to substantially generalize previous results for non-Gaussian regressors.

MSC 2010 subject classifications: Primary 62F03, 62F05; secondary 62J05, 60F05, 60B20.

Keywords and phrases: High-dimensional linear regression, F -Test, multiple hypothesis testing, large- p asymptotics.

Received October 2015.

Contents

1	Introduction	2585
2	Model formulation and main results	2587
2.1	Technical conditions	2592
3	Testing only a small number of hypotheses	2593
4	Discussion of the technical assumptions	2595
5	Numerical results	2597
6	Proofs of main results	2601
6.1	Outline	2601
6.2	Asymptotic normality of the noise term	2604
6.3	Asymptotic χ^2 -distribution of the noise term	2608
6.4	Asymptotic behavior of the non-centrality term	2610

6.4.1	The case of Theorem 2.1(i)	2611
6.4.2	The case of Theorem 2.1(ii)	2612
6.4.3	The cases of Theorem 2.1(iii) and Theorem 3.1	2615
A	Auxiliary results of Section 4	2617
B	Proofs of auxiliary results of Section 6.2	2621
C	Other technical results	2627
	References	2639

1. Introduction

Following a suggestion of Bai and Saranadasa [1] to investigate classical statistical procedures in high-dimensional settings, Wang and Cui [27] re-examine the usual F -test in the linear regression model under a “large p , large n ” asymptotic framework. They derive the asymptotic power in a fairly general, non-Gaussian setting, highlighting the dependence of the local power function on the dimensionality of the problem, i.e., on the limit $\rho = \lim p/n \in (0, 1)$, where n is sample size, and p is the number of regressors in the model. In particular, they find that the rejection probability of the F -test for $H_0 : R\beta = r_0$, where $R = [0, I_q]$ and $p/n \rightarrow \rho$, $q/n \rightarrow \rho$, satisfies

$$\mathbb{P}(F_n > f_{q, n-p-1}^{(1-\nu)}) - \Phi \left(-\zeta_{1-\nu} + \sqrt{n} \Delta_\beta \sqrt{\frac{1-\rho}{2\rho}} \right) \rightarrow 0. \quad (1.1)$$

Here F_n is the usual F -statistic, $f_{q, n-p-1}^{(1-\nu)}$ is the appropriate F -quantile, Φ is the cdf of the standard normal distribution, $\zeta_{1-\nu} = \Phi^{-1}(1-\nu)$ and $\Delta_\beta = (R\beta - r_0)'(R\Sigma^{-1}R')^{-1}(R\beta - r_0)/\sigma^2$ is the scaled distance from the null hypothesis. From this approximation we see that the local asymptotic power of the F -test depends monotonically on the value of ρ and inflates to the nominal significance level ν as ρ increases to one. The result of Wang and Cui [27] is consistent with the derivations of the local asymptotic power in the case of Gaussian errors, as obtained by Zhong and Chen [29]. Both of these studies consider only the overall F -test for the null hypothesis that all, or almost all (cf. Condition (C3) in Wang and Cui [27]) of the p slope coefficients are equal to zero. Also, they do not consider hypotheses involving the intercept parameter. Here, we extend this analysis and study the problem of testing q general linear hypotheses (including also hypotheses on the intercept term), without the restriction that $(p-q)/n \rightarrow 0$. In this sense, we examine the effect of the dimension of the null hypothesis (i.e., the number of linear restrictions being tested) on the local asymptotic rejection probability of the F -test. We find that when testing the null hypothesis $H_0 : R_0\gamma = r_0$, for some $q \times (p+1)$ matrix R_0 of rank $q \leq p+1$, such that $p/n \rightarrow \rho_1$ and $q/n \rightarrow \rho_2 \leq \rho_1$, the rejection probability of the F -test satisfies

$$\mathbb{P}(F_n > f_{q, n-p-1}^{(1-\nu)}) - \Phi \left(-\zeta_{1-\nu} + \sqrt{n} \Delta_\gamma \sqrt{\frac{(1-\rho_1)(1-\rho_1+\rho_2)}{2\rho_2}} \right) \rightarrow 0. \quad (1.2)$$

Now the asymptotic rejection probability depends also on the mean $\mu \in \mathbb{R}^p$ of the random design through $\Delta_\gamma = (R_0\gamma - r_0)'(R_0S^{-1}R_0')^{-1}(R_0\gamma - r_0)/\sigma^2$,

where $\gamma = (\alpha, \beta)'$ is the vector of regression coefficients including an intercept parameter $\alpha \in \mathbb{R}$ and

$$S = \begin{bmatrix} 1 & \mu' \\ \mu & \Sigma + \mu\mu' \end{bmatrix}.$$

This limiting expression coincides with that in (1.1) if $\rho_1 = \rho_2$ and $R_0 = [0, R]$. But (1.2) refines the statement in (1.1) and shows the impact of both the relative number of regressors ρ_1 and the relative number of hypotheses ρ_2 . These quantities affect the asymptotic rejection probability monotonically, which is consistent with small sample analyses in the Gaussian error case [cf. 12]. However, in contrast to the complicated nature of the cdf of the non-central F -distribution as a function in p , q and the non-centrality parameter, our asymptotic approximation to the rejection probability depends on the quantities ρ_1 , ρ_2 and Δ_γ only through elementary operations and the Gaussian cdf, and it is valid for a large class of error distributions. In particular, we see that even if ρ_1 is close to 1, the F -test still has power if ρ_2 is sufficiently small. In a second step, and under slightly more restrictive assumptions on the data generating process, we also investigate the case where only a very small relative number of hypotheses q/n is tested, i.e., q is bounded as $n \rightarrow \infty$ and $\rho_2 = 0$, and the result in (1.2) no longer holds.

Our work heavily builds on the ideas of Wang and Cui [27] (hereafter abbreviated as WC). The first part of the present work is concerned with reproducing their results under substantially more general assumptions. First of all, here we do not require independence between the random design and the error terms, but we assume only the usual first and second order specification of conditional moments of the errors given the design. This extension requires a slight modification of the result of Bhansali, Giraitis and Kokoszka [5] on the asymptotic normality of certain quadratic forms as applied by WC (cf. Lemma 6.1). Furthermore, we do not assume that the $n \times p$ design matrix X , after standardization, consists of i.i.d. components, as is needed for the application of the famous Bai-Yin Theorem [2] used by WC in order to control extreme eigenvalues of large sample covariance matrices. Instead, we apply a recent result of Srivastava and Vershynin [23] which essentially requires only certain moment restrictions on the i.i.d. rows of X . For our extensions, we also develop a novel result on the diagonal entries of a fairly general random projection matrix that might be of interest on its own (see Lemma 6.3). It has the statistical interpretation that in a moderately high-dimensional regression the leverage values h_i , i.e., the diagonal entries of the projection matrix $U(U'U)^{-1}U'$, where $U = [\iota, X]$ is the design matrix including an intercept column, typically behave like p/n . Finally, we point out that since we also consider tests on the intercept parameter, the distribution of the F -statistic, in general, also depends on the mean μ of the random design vectors x_1, \dots, x_n . This causes certain technical complications due to non-centrality issues which are often avoided in the literature on random design regression by restricting to the case $\mu = 0$. Here, we present a detailed treatment of the general case.

The paper is organized as follows. Section 2 introduces the setup and notation and presents our main results in Theorem 2.1 and Corollary 2.2, which provide a precise formulation of the statement in (1.2). In Section 3, we specifically consider the situation where q is fixed and also provide a unifying result that does not distinguish between large or small q . Next, in Section 4 we provide a detailed discussion of our technical assumptions and explain the main differences to those imposed by WC. The results of an extensive numerical study are reported in Section 5. Finally, Section 6 provides the basic steps in the proof of our main results. Some of the more technical arguments are deferred to the appendices.

2. Model formulation and main results

We consider a random array $\{(y_{i,n}, x_{i,n}) : 1 \leq i \leq n, n \geq 1\}$ where, for each $n \in \mathbb{N}$, the pairs $(y_{i,n}, x_{i,n})_{i=1}^n$ are i.i.d. observations of a real valued response variable $y_{1,n}$ and p_n -dimensional random regressors $x_{1,n}$ with $p_n < n - 1$, satisfying $\mathbb{E}[y_{1,n}|x_{1,n}] = \alpha_n + \beta_n' x_{1,n}$ and $\text{Var}[y_{1,n}|x_{1,n}] = \sigma_n^2 \in (0, \infty)$. Equivalently, writing $\varepsilon_{i,n} = y_{i,n} - \mathbb{E}[y_{i,n}|x_{i,n}]$, the observations can be represented as

$$y_{i,n} = \alpha_n + \beta_n' x_{i,n} + \varepsilon_{i,n}, \quad i = 1, \dots, n, \quad (2.1)$$

where the $(\varepsilon_{i,n})_{i=1}^n$ are i.i.d., satisfying $\mathbb{E}[\varepsilon_{i,n}|x_{i,n}] = 0$ and $\text{Var}[\varepsilon_{i,n}|x_{i,n}] = \sigma_n^2$. Note that $\varepsilon_{1,n}$ does not need to be independent of $x_{1,n}$. For identifiability, we also assume that $\Sigma_n := \text{Var}[x_{1,n}]$ is positive definite and we define $\mu_n := \mathbb{E}[x_{1,n}]$. Furthermore, we adopt the matrix notation $Y_n = (y_{1,n}, \dots, y_{n,n})'$, $X_n = [x_{1,n}, \dots, x_{n,n}]'$, $\varepsilon_n = (\varepsilon_{1,n}, \dots, \varepsilon_{n,n})'$, $\gamma_n = (\alpha_n, \beta_n)'$ and $U_n = [\iota_n, X_n]$, where $\iota_n = (1, 1, \dots, 1)' \in \mathbb{R}^n$. For notational convenience we will drop the subscript n whenever there is no risk of confusion, i.e., we write $Y = Y_n$, $X = X_n$, $\alpha = \alpha_n$, $\beta = \beta_n$, etc., keeping in mind that, unless noted otherwise, all quantities to follow depend on sample size n . With this, the model equations in (2.1) become

$$Y = U\gamma + \varepsilon. \quad (2.2)$$

We want to test a general linear hypothesis on the coefficients γ , i.e.,

$$H_0 : R_0\gamma = r_0 \quad \text{vs.} \quad H_1 : R_0\gamma \neq r_0, \quad (2.3)$$

where R_0 is a $q \times (p+1)$ matrix with $\text{rank } R_0 = q \leq p+1$ and $r_0 \in \mathbb{R}^q$. Without restriction we may assume that R_0 has orthonormal rows (premultiply (2.3) by $(R_0 R_0')^{-1/2}$). We test H_0 by use of the F -statistic F_n defined as

$$F_n = \frac{(R_0 \hat{\gamma}_n - r_0)' (R_0 (U'U)^{-1} R_0')^{-1} (R_0 \hat{\gamma}_n - r_0) / q}{(Y - U \hat{\gamma}_n)' (Y - U \hat{\gamma}_n) / (n - p - 1)}, \quad (2.4)$$

provided that all the appearing quantities are well defined, and $F_n = 0$, otherwise. The F -statistic is then compared to the $1 - \nu$ quantile of an F -distribution with q and $n - p - 1$ degrees of freedom, which we denote by $f_{q, n-p-1}^{(1-\nu)}$. Here, $\hat{\gamma}_n = (\hat{\alpha}_n, \hat{\beta}_n)'$ is the OLS estimate in the unrestricted model. We also define

the usual estimator of the error variance $\hat{\sigma}_n^2 = \|Y - U\hat{\gamma}_n\|^2/(n - p - 1)$, that appears in the denominator of the F -statistic.

In Section 6 we prove the following results, involving the scaled distance from the null hypothesis

$$\Delta_\gamma := (R_0\gamma - r_0)'(R_0S^{-1}R_0')^{-1}(R_0\gamma - r_0)/\sigma^2,$$

where

$$S = \begin{bmatrix} 1 & \mu' \\ \mu & \Sigma + \mu\mu' \end{bmatrix} = \mathbb{E} \left[\begin{pmatrix} 1 \\ x_1 \end{pmatrix} (1 \ x_1') \right] = \mathbb{E} [U'U/n].$$

A list of further technical conditions is given below in Section 2.1.

Theorem 2.1. *Consider the linear, homoskedastic model (2.1) and set $s_n = 2 \left(\frac{1}{q_n} + \frac{1}{n-p_n-1} \right)$ and $b_n = \sqrt{\frac{(1-(p_n+1)/n)(1-(p_n+1)/n+q_n/n)}{2q_n/n}}$. Moreover, suppose that $\limsup_n p_n/n < 1$, $q_n \rightarrow \infty$ and $\Delta_\gamma = o(q_n/n)$ as $n \rightarrow \infty$. If either one of the following three cases applies then the F -statistic satisfies*

$$s_n^{-1/2}(F_n - 1) - \sqrt{n}\Delta_\gamma b_n \xrightarrow[n \rightarrow \infty]{w} \mathcal{N}(0, 1). \tag{2.5}$$

- (i) *The assumptions (A1).(a,b,c,d) and (A2) on the random design and on the error distribution, are satisfied, and either $R_0 = I_{p_n+1}$ for all $n \in \mathbb{N}$, or $R_0 = [0, I_{p_n}]$ for all $n \in \mathbb{N}$. (In this case, either $q_n = p_n$ or $q_n = p_n + 1$, and thus $(p_n - q_n)/n \rightarrow 0$ holds.)*
- (ii) *The assumptions (A1).(a,b,c,d,e) and (A2) on the random design and on the error distribution, are satisfied, $(p_n - q_n)/n \rightarrow 0$, and $(R_0\gamma - r_0)'R_0SR_0'(R_0\gamma - r_0)/\sigma^2 = O(\sqrt{q_n}/n)$ holds.¹*
- (iii) *Assumption (A2) on the error distribution is satisfied and the design vectors $x_{1,n}, \dots, x_{n,n}$ are i.i.d. Gaussian with mean $\mu_n \in \mathbb{R}^{p_n}$ and positive definite covariance matrix Σ_n .²*

By a simple argument involving Polya’s theorem this translates into the following corollary on the rejection probability of the F -test.

Corollary 2.2. *If $\limsup_n p_n/n < 1$ and $q_n \rightarrow \infty$, as $n \rightarrow \infty$, and the conclusion of Theorem 2.1 holds, then the rejection probability of the F -test satisfies*

$$\mathbb{P}(F_n > f_{q_n, n-p_n-1}^{(1-\nu)}) - \Phi \left(-\zeta_{1-\nu} + \sqrt{n}\Delta_\gamma \sqrt{\frac{(1 - \frac{p_n+1}{n})(1 - \frac{p_n+1}{n} + \frac{q_n}{n})}{2q_n/n}} \right) \rightarrow 0.$$

Here, $\zeta_{1-\nu} = \Phi^{-1}(1-\nu)$ is the $1-\nu$ quantile of the standard normal distribution and $\nu \in (0, 1)$ does not depend on n .

¹Notice that this additional requirement implies and strengthens the assumption that $\Delta_\gamma = o(q_n/n)$. Simply observe that, by block matrix inversion, $R_0S^{-1}R_0' = (R_0SR_0' - R_0SR_1'(R_1SR_1')^{-1}R_1SR_0')^{-1}$, where R_1 is a $(p_n + 1 - q_n) \times (p_n + 1)$ matrix with orthonormal rows which are also orthogonal to the rows of R_0 . Therefore, $\Delta_\gamma \leq (R_0\gamma - r_0)'R_0SR_0'(R_0\gamma - r_0)/\sigma^2 = O(\sqrt{q_n}/n)$.

²It is easy to see that the normality assumption implies Assumption (A1).

Proof. It is easy to see, using Polya's theorem and Lemma C.8, that $\tilde{f}_n := s_n^{-1/2}(f_{q_n, n-p_n-1}^{(1-\nu)} - 1)$ satisfies $\tilde{f}_n \rightarrow \zeta_{1-\nu}$. Now use the conclusion of Theorem 2.1, Polya's theorem and the Lipschitz continuity of Φ to obtain

$$\begin{aligned} & |\mathbb{P}(F_n > f_{q_n, n-p_n-1}^{(1-\nu)}) - \Phi(-\zeta_{1-\nu} + \eta_n^2)| \\ &= |\mathbb{P}(s_n^{-1/2}(F_n - 1) > \tilde{f}_n) - \Phi(-\zeta_{1-\nu} + \eta_n^2)| \\ &= |\mathbb{P}(s_n^{-1/2}(F_n - 1) - \eta_n^2 \leq \tilde{f}_n - \eta_n^2) - \Phi(\zeta_{1-\nu} - \eta_n^2)| \\ &\leq |\mathbb{P}(s_n^{-1/2}(F_n - 1) - \eta_n^2 \leq \tilde{f}_n - \eta_n^2) - \Phi(\tilde{f}_n - \eta_n^2)| \\ &\quad + |\Phi(\tilde{f}_n - \eta_n^2) - \Phi(\zeta_{1-\nu} - \eta_n^2)| \\ &\leq \sup_{t \in \mathbb{R}} |\mathbb{P}(s_n^{-1/2}(F_n - 1) - \eta_n^2 \leq t) - \Phi(t)| + o(1) \xrightarrow{n \rightarrow \infty} 0, \end{aligned}$$

where $\eta_n^2 := \sqrt{n} \Delta_\gamma b_n$. \square

If $(p_n - q_n)/n \rightarrow 0$ and $0 < \liminf_n q_n/n \leq \limsup_n p_n/n < 1$, then Corollary 2.2 recovers the result of WC under weaker assumptions on the joint distribution of the design and the errors, and for a null hypothesis that possibly restricts also the intercept parameter α (cf. the assumptions of Theorem 2.1(i) and (ii)). In this case the factor b_n above asymptotically reduces to

$$b_n = \sqrt{\frac{(1 - \frac{p_n+1}{n})(1 - \frac{p_n+1}{n} + \frac{q_n}{n})}{2q_n/n}} \sim \sqrt{\frac{1 - \frac{p_n+1}{n}}{2\frac{p_n+1}{n}}},$$

as in (1.1). It highlights the dependence of the power function on the relative number of regressors p_n/n . However, since $(p_n - q_n)/n \rightarrow 0$, the individual roles of p_n and q_n can not be discerned. This shortcoming is removed here, but it comes at the price of a stronger design condition (cf. Theorem 2.1(iii)). It is tempting, however, to conjecture that Assumptions (A1) and (A2) are actually sufficient also for the general case. Corollary 2.2 nicely shows the effect of both the dimension of the parameter space as well as the dimension of the null hypothesis, on the asymptotic power function. In particular, we see that even in a case where the relative number of regressors p_n/n is large, the classical F -test still has power, as long as we are interested in testing only a relatively small number of hypotheses. However, we should make a cautionary remark at this point. In Theorem 2.1 and Corollary 2.2 we have assumed that $q_n \rightarrow \infty$. If the number of hypotheses q being tested is too small, then the asymptotic approximation presented above will not be very accurate, in the same way the χ_q^2 distribution is not very accurately approximated by the normal if q is small. Therefore, in the next section we specifically study the case where q_n is bounded.

In Theorem 2.1, we treat the special cases of $R_0 = I_{p+1}$ and $R_0 = [0, I_p]$ separately, because here it is considerably much easier to deal with the non-centrality term in the decomposition of the F -statistic (see Section 6.4.1). In particular, in this case we do not need to impose further restrictions on the distance from the null Δ_γ other than that it is of order $o(q_n/n)$ as $n \rightarrow \infty$ and we can also work with weaker design conditions (cf. Theorem 2.1(i)).

Remark 2.3 (On the detection boundary of the F -test). In the classical setting, where q and p are fixed, while n goes to infinity, it is well known that the detection boundary of the F -test is $n^{-1/2}$. This means that a violation of the null hypothesis $H_0 : R_0\gamma = r_0$ that is of the order $\|R_0\gamma - r_0\|_2 \asymp n^{-1/2}$ leads to non-trivial asymptotic power, while a slower order yields asymptotic power equal to the size of the test and a larger order yields asymptotic power equal to one. However, in general, when $q = q_n$ and $p = p_n$ are allowed to grow with sample size n , the detection boundary of the F -test is no longer $n^{-1/2}$ but rather $q_n^{1/4}/n^{1/2}$. To see this, first we ignore the influence of the nuisance parameters and set $\mu = 0$, $\Sigma = I_p$ and $\sigma^2 = 1$, so that $S = I_{p+1}$ and $\Delta_\gamma = \|R_0\gamma - r_0\|_2^2$. From Corollary 2.2, we see that in order to obtain non-trivial asymptotic power, i.e., asymptotic power in the open interval $(\nu, 1)$, the non-centrality term

$$\sqrt{n}\Delta_\gamma b_n = \left(\frac{n^{1/2}}{q_n^{1/4}} \|R_0\gamma - r_0\|_2 \right)^2 \sqrt{\frac{1}{2} \left(1 - \frac{p_n + 1}{n} \right) \left(1 - \frac{p_n + 1}{n} + \frac{q_n}{n} \right)},$$

has to stay away from 0 and ∞ . If we exclude the pathological case $\limsup_n p_n/n = 1$, then this requirement is met if the violation of the null hypothesis is of the order $\|R_0\gamma - r_0\|_2 \asymp q_n^{1/4}/n^{1/2}$.

Remark 2.4 (Gaussian errors and fixed design). In the classical case where the error ε follows a spherical normal distribution which is independent of the design, the F -statistic (2.4) follows a non-central F -distribution with q_n and $n - p_n - 1$ degrees of freedom and non-centrality parameter $\zeta_n^2 = n\nabla_n$, conditional on X , where $\nabla_n = (R_0\gamma - r_0)'(R_0(U'U/n)^{-1}R_0')^{-1}(R_0\gamma - r_0)/\sigma^2$ [cf. 22, page 41]. Nevertheless, even in this traditional case, only basic monotonicity results are available for the power function $\mathbb{P}(F_n > f_{q_n, n-p_n-1}^{(1-\nu)} | X)$ as a function of q_n , $n - p_n - 1$ and ζ_n^2 [e.g., 12, 26]. In Section 6.4 we investigate ∇_n as $n \rightarrow \infty$, such that $\limsup_n p_n/n < 1$. Our results provide approximations for the average (or unconditional) rejection probability, i.e., for $\mathbb{E}[\mathbb{P}(F_n > f_{q_n, n-p_n-1}^{(1-\nu)} | X)]$, which are given by the Gaussian cdf applied to an elementary function in p_n/n , q_n/n and $\Delta_\gamma = (R_0\gamma - r_0)'(R_0S^{-1}R_0')^{-1}(R_0\gamma - r_0)/\sigma^2$ and which are therefore easy to interpret (cf. Corollary 2.2).

Remark 2.5 (On omitted variable misspecification). One major motivation for us to extend the result of WC to scenarios where there is some dependence between the design and the errors, and also among the components of the standardized design vectors themselves (see Section 4 for details), was to treat simple sub-models of high-dimensional linear models. These sub-models typically exhibit misspecification due to omitted regressor variables. Consider an i.i.d. sample $(y_i, z_i)_{i=1}^n$ from the high-dimensional linear model

$$y_i = \theta' z_i + u_i,$$

where the z_i are random d -vectors with $d \gg n$ that are independent of the u_i , which satisfy $\mathbb{E}[u_i] = 0$ and $\mathbb{E}[u_i^2] \in (0, \infty)$. Moreover, assume that the marginal distribution of the regressors z_i can be represented as $z_i = \Omega^{1/2}\tilde{z}_i$, where $\Omega^{1/2}$ is

symmetric and positive definite and \tilde{z}_i has a Lebesgue density $f_{\tilde{z}}$ which is such that the components of \tilde{z}_i are independent with zero means, unit variances and bounded 8-th moments, so that in particular $\text{Cov}[z_i] = \Omega$.

Now suppose we want to use only a small number p of the d available regressors, with $p < n$, so that classical regression methods are feasible within this subset regression problem. These working regressors can be described as

$$x_i = M'z_i,$$

where M is a $d \times p$ matrix of full rank $p < d$. For instance, M could be a selection matrix so that x_i consists of a certain choice of p components of z_i . In such a situation, the sample $(y_i, x_i)_{i=1}^n$ need not follow a linear homoskedastic model as in (2.1), because, in general, the conditional expectation $\mathbb{E}[y_i|x_i]$ is not linear in x_i and the conditional variance $\text{Var}[y_i|x_i]$ is not constant if the pair (y_i, x_i) is not jointly Gaussian. However, one can always write

$$y_i = \tilde{\beta}'x_i + \xi_i,$$

where $\tilde{\beta} = \text{argmin}_b \mathbb{E}[(y_i - b'x_i)^2]$ and $\xi_i = y_i - \tilde{\beta}'x_i$. Here, the parameter $\tilde{\beta}$ corresponds to the best linear predictor of y_i given x_i , and it may be of interest to test whether $H_0 : \tilde{\beta} = 0$, i.e., whether the selected regressors $x_i = M'z_i$ have any value for linearly predicting the response variables y_i .

Clearly, in the present setting the errors ξ_i are not independent of the design vectors x_i and, in general, $\mathbb{E}[\xi_i|x_i] \neq 0$ and $\text{Var}[\xi_i|x_i] \neq \text{Var}[\xi_i]$. Now consider the unobservable corrected sample $(y_i^*, x_i)_{i=1}^n$, where

$$y_i^* = \tilde{\beta}'x_i + \sqrt{\frac{\text{Var}[\xi_i]}{\text{Var}[\xi_i|x_i]}}(\xi_i - \mathbb{E}[\xi_i|x_i]).$$

The corrected sample clearly follows a linear homoskedastic model as in (2.1), because $\mathbb{E}[y_i^*|x_i] = \tilde{\beta}'x_i$ and $\text{Var}[y_i^*|x_i] = \text{Var}[\xi_i]$. Moreover, the corrected sample satisfies Assumption (A1) in view of Lemma A.2 applied with $\Gamma = M'\Omega^{1/2}$, $\mu = 0$ and $m_n = d$, and because the design matrix $X = [x_1, \dots, x_n]'$ has a Lebesgue density on $\mathbb{R}^{n \times p}$. Of course, in general, the actual sample $(y_i, x_i)_{i=1}^n$ may be very different from the corrected sample $(y_i^*, x_i)_{i=1}^n$, but the results of Steinberger and Leeb [24] suggest that if $d \gg p$, then $\mathbb{E}[\xi_i|x_i] \approx 0$ and $\text{Var}[\xi_i|x_i] \approx \text{Var}[\xi_i]$, at least for a large collection of selection matrices M . Therefore, the observed sample and the corrected sample should be very similar if d is large relative to p , and one might expect that also for a sufficiently regular statistic T , we have $T((y_i, x_i)_{i=1}^n) \approx T((y_i^*, x_i)_{i=1}^n)$. We suspect that the results of Steinberger and Leeb [24] can also be used to establish the validity of Assumption (A2) for the corrected sample, with $e_{i,n} := \sqrt{\text{Var}[\xi_i]/\text{Var}[\xi_i|x_i]}$ and $\tilde{\varepsilon}_{i,n} := \xi_i - \mathbb{E}[\xi_i|x_i]$. Thus, we expect the F -test for $H_0 : \tilde{\beta} = 0$ based on the sample $(y_i, x_i)_{i=1}^n$ to be asymptotically valid and even possess similar power as the F -test in a correctly specified model, for most choices of M , provided that $d = d_n$ and $p = p_n$ tend to infinity along with n , such that $d_n \gg p_n$. The details of this line of reasoning will be further developed elsewhere.

2.1. Technical conditions

Throughout this paper, the reader will encounter several different norms. For vectors $v \in \mathbb{R}^k$ we write $\|v\| = (\sum_{i=1}^k v_i^2)^{1/2}$ for the usual Euclidean norm, whereas for matrices $M \in \mathbb{R}^{k \times \ell}$ we distinguish between the spectral norm $\|M\|_S = (\lambda_{\max}(M'M))^{1/2}$ and the Frobenius norm $\|M\|_F = (\text{trace}(M'M))^{1/2}$. We write P_M for the matrix of orthogonal projection onto the column span of M . If M satisfies $\text{rank } M = \ell \leq k$, then $P_M = M(M'M)^{-1}M'$. We also make use of the stochastic Landau notation. For a sequence of real random variables z_n , we say that $z_n = O_{\mathbb{P}}(1)$ if the sequence is bounded in probability, i.e., if $\sup_{n \in \mathbb{N}} \mathbb{P}(|z_n| > \delta) \rightarrow 0$ as $\delta \rightarrow \infty$, and we say that $z_n = o_{\mathbb{P}}(1)$ if $z_n \rightarrow 0$ in probability. For a non-stochastic real sequence $a_n \neq 0$ we write $z_n = O_{\mathbb{P}}(a_n)$ if $z_n/a_n = O_{\mathbb{P}}(1)$ and $z_n = o_{\mathbb{P}}(a_n)$ if $z_n/a_n = o_{\mathbb{P}}(1)$.

The following is a list of technical assumptions needed in the proof of Theorem 2.1.

(A1) (a) The design vectors $x_{i,n}$ are linearly generated as follows:

$$x_{i,n} = \mu_n + \Gamma_n z_{i,n},$$

where Γ_n is a $p_n \times m_n$ matrix with $m_n \geq p_n$, such that $\Gamma_n \Gamma_n' = \Sigma_n$. The random m_n -vectors $z_{1,n}, \dots, z_{n,n}$ are i.i.d. and satisfy $\mathbb{E}[z_{1,n}] = 0$, $\mathbb{E}[z_{1,n} z_{1,n}'] = I_{m_n}$.

- (b) The $(n-1) \times (p_n+1)$ matrix $U_{n,-1} = [0, I_{n-1}]U_n$ has rank p_n+1 with probability one. (In particular, $\mathbb{P}(\det(U_n' U_n) = 0) = 0$.)
- (c) For every $n \in \mathbb{N}$, the random m_n -vector $z_{1,n}$ from Assumption (A1).(a) also has the following property. There exist universal positive constants c and C , not depending on n , such that for every orthogonal projection P in \mathbb{R}^{m_n} and for every $t > C \text{rank } P$, we have $\mathbb{P}(\|P z_{1,n}\|^2 > t) \leq C t^{-1-c}$.
- (d) Let $z_{1,n}$ be the random m_n -vector from Assumption (A1).(a). For $\mathcal{L}_{k,n} := \sup_{\|v\|=1} (\mathbb{E}|v' z_{1,n}|^k)^{1/k}$ we have $\mathcal{L}_{4,n} = O(1)$ as $n \rightarrow \infty$. For every symmetric matrix $M_n \in \mathbb{R}^{m_n \times m_n}$ we have $\text{Var}[z_{1,n}' M_n z_{1,n}] = O(\text{trace}(M_n^2)) + (\text{trace}(M_n))^2 o(1)$, as $n \rightarrow \infty$.
- (e) In addition to (A1).(d), we also have $\mathcal{L}_{8,n} = O(1)$ as $n \rightarrow \infty$ and for any projection matrix P_n in \mathbb{R}^{m_n} , $(\mathbb{E}[(z_{1,n}' P_n z_{1,n})^4])^{1/4} = O(\|P_n\|_F^2)$ as $n \rightarrow \infty$.

(A2) The error terms $\varepsilon_{i,n}$ can be written as $\varepsilon_{i,n} = e_{i,n} \tilde{\varepsilon}_{i,n}$, where $e_{i,n}$ is $x_{i,n}$ -measurable and such that $\max_{i=1, \dots, n} e_{i,n} = O_{\mathbb{P}}(1)$, and the $\tilde{\varepsilon}_{i,n}$ have the following properties. There exists a universal constant $\kappa > 0$, not depending on n , such that $\mathbb{E}[(\mathbb{E}[|\tilde{\varepsilon}_{1,n}/\sigma_n|^4 | x_{1,n}])^{1+\kappa}] = O(1)$ as $n \rightarrow \infty$, and $\max_i \mathbb{E}[(\tilde{\varepsilon}_{i,n}/\sigma_n)^4 | x_{i,n}] = o_{\mathbb{P}}(\sqrt{q_n})$.

This set of assumptions is weaker than the analogous conditions imposed by WC to treat the case $(p_n - q_n)/n \rightarrow 0$. A detailed discussion and comparison of the differences between our treatment and the one of WC is deferred to Section 4.

3. Testing only a small number of hypotheses

In Theorem 2.1 above, we needed the assumption that $q_n \rightarrow \infty$ as $n \rightarrow \infty$ in order to achieve asymptotic normality of the F -statistic. Since the asymptotic distribution of the F -statistic is well understood in the Gaussian error and fixed design case (cf. Remark 2.4 and Lemma C.8), we expect the asymptotic distribution of the F -statistic to be χ_q^2 rather than Gaussian if $q_n = q$ is fixed. The following result establishes this asymptotic χ^2 distribution for the F -statistic when the error distribution of the model 2.1 is arbitrary up to bounded fourth moments. For this result we need somewhat stronger assumptions than those of Theorem 2.1. The proof is deferred to Section 6.

Theorem 3.1. *Consider the linear, homoskedastic model (2.1) and assume that the design vectors $x_{1,n}, \dots, x_{n,n}$ are i.i.d. Gaussian with mean $\mu_n \in \mathbb{R}^{p_n}$ and positive definite covariance matrix Σ_n , and that the design $X_n = [x_1, \dots, x_n]'$ is independent of the errors $\varepsilon_n = (\varepsilon_{1,n}, \dots, \varepsilon_{n,n})'$ which satisfy $\mathbb{E}[(\varepsilon_{1,n}/\sigma_n)^4] = O(1)$. Moreover, suppose that $\limsup_n p_n/n < 1$, $\Delta_\gamma = o(q_n/n)$ and that the null hypothesis does not involve a restriction on the intercept parameter (i.e., the first column of R_0 is equal to zero). If $q_n = q \in \mathbb{N}$ does not depend on n , then*

$$s_n^{-1/2}(F_n - 1) - \sqrt{n}\Delta_\gamma b_n \xrightarrow[n \rightarrow \infty]{w} (2q)^{-1/2}\chi_q^2 - \sqrt{q/2}, \quad (3.1)$$

where s_n and b_n are as in Theorem 2.1.

Together with the asymptotic normality of the F -statistic in the case $q_n \rightarrow \infty$ (cf. Theorem 2.1), we can establish a non-central F approximation for the F -statistic for any number of tested hypotheses q_n .

Corollary 3.2. *Consider the linear, homoskedastic model (2.1) and assume that the design vectors $x_{1,n}, \dots, x_{n,n}$ are i.i.d. Gaussian with mean $\mu_n \in \mathbb{R}^{p_n}$ and positive definite covariance matrix Σ_n , and that the design $X_n = [x_1, \dots, x_n]'$ is independent of the errors $\varepsilon_n = (\varepsilon_{1,n}, \dots, \varepsilon_{n,n})'$ which satisfy $\mathbb{E}[(\varepsilon_{1,n}/\sigma_n)^4] = O(1)$. Moreover, suppose that $\limsup_n p_n/n < 1$, $\Delta_\gamma = o(q_n/n)$ and that the null hypothesis does not involve a restriction on the intercept parameter (i.e., the first column of R_0 is equal to zero). Then,*

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(F_n \leq t) - \mathbb{P}(F_{q_n, n-p_n-1}(\lambda_n) \leq t)| \xrightarrow[n \rightarrow \infty]{} 0,$$

where $F_{q_n, n-p_n-1}(\lambda_n)$ denotes a random variable following the non-central F distribution with q_n and $n - p_n - 1$ degrees of freedom and non-centrality parameter $\lambda_n = \Delta_\gamma(n - p_n - 1 + q_n)$.

Proof. Suppose the claim does not hold. Then there exists a subsequence n' , such that the supremum converges to a positive number along n' . Then, by compactness of the closed unit interval and the extended real line, we can extract a further subsequence n'' , such that $p_{n''}/n'' \rightarrow \rho_1 \in [0, 1)$, $q_{n''}/n'' \rightarrow \rho_2 \in [0, \rho_1]$,

and $q_{n''} \rightarrow q \in [1, \infty]$, as $n'' \rightarrow \infty$. If $q = \infty$, then we are in the setting of Theorem 2.1(iii) and we obtain asymptotic normality of $s_n^{-1/2}(F_n - 1) - \sqrt{n}\Delta_\gamma b_n$. Since the limiting distribution function is continuous, we get uniform convergence of the corresponding distribution functions, in view of Polya's theorem. Since by Lemma C.8, $s_n^{-1/2}(F_{q_n, n-p_n-1}(\lambda_n) - 1) - \sqrt{n}\Delta_\gamma b_n$ is also asymptotically standard normally distributed, we get a contradiction in that case. If $q < \infty$, then $q_{n''} = q$ for all large n'' , because $q_{n''}$ is integer valued. Thus Theorem 3.1 applies and shows that $s_n^{-1/2}(F_n - 1) - \sqrt{n}\Delta_\gamma b_n$ converges weakly to $(2q)^{-1/2}\chi_q^2 - \sqrt{q/2}$. Since the limiting distribution function is again continuous, and since Lemma C.8 shows that $s_n^{-1/2}(F_{q_n, n-p_n-1}(\lambda_n) - 1) - \sqrt{n}\Delta_\gamma b_n$ has the same asymptotic distribution, we also get a contradiction in this case, upon using the same argument as before, involving Polya's theorem. \square

Corollary 3.2 provides a unified treatment of the asymptotic behavior of the F -statistic without distinguishing between the cases $q_n = q$ and $q_n \rightarrow \infty$. However, this does not immediately lead to a neat formula for the local asymptotic power function of the F -test because of the complicated nature of the cdf of the non-central F -distribution.

Remark 3.3 (Fixed vs. random design). It is instructive to compare the non-central F approximation of Corollary 3.2 to the distribution of the F -statistic in the Gaussian error and fixed design case (cf. Remark 2.4). In the former case the non-centrality parameter is given by $n\Delta_\gamma(n - p_n - 1 + q_n)/n$, while in the latter case it is $n\nabla_n$. Now if the design X is random with i.i.d. rows and $S = \mathbb{E}[U'U/n]$, $U = [\iota, X]$, it turns out that $\Delta_\gamma := (R_0\gamma - r_0)'(R_0S^{-1}R_0')^{-1}(R_0\gamma - r_0)/\sigma^2$ is not a good approximation for $\nabla_n := (R_0\gamma - r_0)'(R_0(U'U/n)^{-1}R_0')^{-1}(R_0\gamma - r_0)/\sigma^2$ if p_n/n is not close to q_n/n , even if n is very large. In fact, we need a correction factor of $\mathbb{E}[\chi_{n-p_n-1+q_n}^2/n] = (n - p_n - 1 + q_n)/n$ in order to account for the additional randomness coming from the design X (cf. Lemma C.5). This correction factor, however, is close to one if $q_n/n \approx p_n/n$. If the number q_n of hypotheses to be tested is much smaller than the number of parameters p_n , this correction is quite significant. Thus, the limiting distribution of the F -statistic under random design with $\mathbb{E}[U'U/n] = S$ is, in general, not equal to the distribution of the F -statistic under Gaussian errors and fixed design X satisfying $U'U/n = S$. This distinction only occurs for $q_n \ll p_n$. In particular, this issue disappears completely if $p_n/n \approx 0$.

Remark 3.4 (On confidence sets for $R_0\gamma$). Corollary 3.2 can immediately be used to construct asymptotically valid confidence sets for $R_0\gamma$. Simply define

$$CI_\nu = \left\{ r \in \mathbb{R}^q : (R_0\hat{\gamma}_n - r)'(R_0(U'U)^{-1}R_0')^{-1}(R_0\hat{\gamma}_n - r) \leq f_{q, n-p-1}^{(1-\nu)} \hat{\sigma}_n^2 \right\},$$

and note that

$$\mathbb{P}(R_0\gamma \notin CI_\nu) = \mathbb{P}(F_n > f_{q, n-p-1}^{(1-\nu)}) \rightarrow \mathbb{P}(F_{q, n-p-1}(0) > f_{q, n-p-1}^{(1-\nu)}) = \nu,$$

where F_n is the F -statistic under the null hypothesis $r_0 = R_0\gamma$.

4. Discussion of the technical assumptions

We pause for a moment to discuss the meaning of our Assumptions (A1) and (A2), as well as the other conditions used in Theorem 2.1, and we comment on the main differences to the conditions imposed in WC.

First of all, Assumption (A1).(a) of linear generation of the design from possibly much higher dimensional random vectors also appears in WC who take it as a modification from Bai and Saranadasa [1]. We point out that this is a straight forward generalization of the case $m = p$, where moment restrictions have to be imposed directly on the design vectors x_i (note that the components of x_1 may not be independent, even after standardization, whereas x_1 can still be linearly generated from a vector z_1 whose components are independent). Moreover, this assumption also allows for the interpretation that there is actually a much higher dimensional set of explanatory variables z_i available whose dimensionality m (possibly $m \gg n$) has already been reduced to $p < n$. See also Remark 2.5.

Together with (A1).(a), our Conditions (A1).(d,e) replace and considerably relax Assumption (C1) in WC, which reads as follows.

- (C1) x_i is linearly generated by a m -variate random vector $z_i = (z_{i1}, \dots, z_{im})'$ so that $x_i = \Gamma z_i + \mu$, where Γ is a $p \times m$ matrix for some $m \geq p$ such that $\Gamma \Gamma' = \Sigma$, each z_{il} has finite 8-th moment, $\mathbb{E}[z_i] = 0$, $\text{Var}[z_i] = I_m$, $E[z_{ik}^4] = 3 + \Delta$ and for any $\sum_{j=1}^d \ell_j \leq 8$, $\mathbb{E}[z_{1i_1}^{\ell_1} z_{1i_2}^{\ell_2} \cdots z_{1i_d}^{\ell_d}] = \mathbb{E}[z_{1i_1}^{\ell_1}] \mathbb{E}[z_{1i_2}^{\ell_2}] \cdots \mathbb{E}[z_{1i_d}^{\ell_d}]$, where Δ is some finite constant.³

In fact, in addition to (C1), WC also need the 8-th moments of z_{ik} to be uniformly bounded so that none of them goes off to infinity as n (and $m = m_n$) increases. The factorization requirement of the 8-th mixed moments in (C1) is a straight forward relaxation of an independence assumption. However, just like the independence assumption, it rules out many spherical distributions (cf. Lemma A.1(i)). Therefore, moment conditions like (A1).(d,e) are much more natural to accommodate both product and spherical distributions. In fact, Condition (C1), together with uniform boundedness of $\mathbb{E}[z_{ik}^8]$, is strictly stronger than our Assumptions (A1).(a,d,e) (cf. Lemma A.1(iii) and Lemma A.2(ii)).

Our Assumption (A1).(b) is important to guarantee that the F -statistic is equal to the expression on the right-hand-side of (2.4), at least with asymptotic probability one, which is used in WC implicitly. The reason that we not only require almost sure invertibility of $U'U$ but also of the design matrix based on $n - 1$ observations is only of a technical nature and plays an important role in the proof of Lemma 6.3 (cf. the end of Subsection 6.2), which is based on leave-one-out ideas. This lemma replaces the strong assumption of WC that there exists a global constant $c_1 > 0$, such that the smallest eigenvalue of the

³Note that this formulation, as it stands, is self-contradictory. Clearly, one has to assume that the indices i_1, \dots, i_d are distinct, or otherwise (C1) implies $1 = \mathbb{E}[z_{11}^2] = \mathbb{E}[z_{11}^1 z_{11}^1] = \mathbb{E}[z_{11}] \mathbb{E}[z_{11}] = 0$.

sample covariance matrix satisfies $\lambda_{\min}(\tilde{X}'\tilde{X}/n) \geq c_1$, almost surely, where $\tilde{X} = (X - \iota\mu')\Sigma^{-1/2}$ is the design matrix based on the standardized regressors (cf. page 147 in WC).⁴

Finally, Assumption (A1).(c) is taken directly from Srivastava and Vershynin [23] to control the extreme eigenvalues of large sample covariance matrices, and different sets of sufficient conditions for (A1).(c) can be found in that reference. In WC, control of extreme eigenvalues is accomplished by use of the celebrated Bai-Yin Theorem of Bai and Yin [2] (cf. Lemma 2 in WC). However, this comes at the price of the implicit assumption that the standardized design vector $\Sigma^{-1/2}(x_1 - \mu)$ has independent components.⁵

Altogether, our design condition (A1) includes linear functions of both, product distributions with uniformly bounded 8-th marginal moments and a large class of spherically symmetric distributions (cf. Lemma A.1 and Lemma A.2 in Appendix A).

The Assumption (A2) on the error distribution extends the fourth moment condition (C2) in WC, which simply states that $\mathbb{E}[(\varepsilon_{1,n}/\sigma_n)^4] = O(1)$, as $n \rightarrow \infty$.⁶ If the errors are independent of the design and $q_n \rightarrow \infty$ (as is the case in WC), then (C2) and (A2) are equivalent. However, Condition (A2) is suited to also allow for some amount of dependence between the errors and the design. This dependence is ruled out in WC, because they use results of asymptotic normality of quadratic forms from Bhansali, Giraitis and Kokoszka [5] that apply only in the independence case (see Lemma 6.1 and the discussion at the beginning of Subsection 6.2). We note that an $(8 + \kappa)$ -th moment condition like, e.g., $\sup_n \mathbb{E}[(\tilde{\varepsilon}_1/\sigma_n)^{8+\kappa}] \leq K$, together with $\max_i e_i = O_{\mathbb{P}}(1)$, is sufficient for (A2), provided that $\liminf_n q_n/n > 0$.

The additional requirement of Theorem 2.1, that $\limsup_n p_n/n < 1$ and $q_n \rightarrow \infty$, simply describes the regime of the relative number of parameters and hypotheses we are interested in. The corresponding assumption (C3) in WC and also parts (i) and (ii) of Theorem 2.1 additionally require that $(p_n - q_n)/n \rightarrow 0$. This is a more serious restriction which is convenient in the present strategy of proof to show that the non-centrality term in the F -statistic under the local alternative degenerates to the correct value (cf. Section 6.4.2). This, however, means that asymptotically we are only dealing with hypotheses where almost all of the p parameters are restricted, since $q_n/p_n \rightarrow 1$ in this regime. It is therefore important to extend the analysis of the rejection probability of the F -test also to the regime where $q_n \ll p_n$ in order to assess the different contributions of the overall dimensionality and the multiplicity of hypothesis testing to the asymptotic rejection probability. This is what we do in Theorem 2.1(iii) and in Section 3 in the Gaussian design setting. The requirement that $q_n \rightarrow \infty$ as

⁴Notice that this assumption rules out, for example, Gaussian design [see, e.g., 11, Theorem 2.1].

⁵This is particularly inconvenient if one is interested in the case where the random vectors z_1, \dots, z_n in (C1) (and (A1).(a)) have independent components. If both z_1 and $\Sigma^{-1/2}(x_1 - \mu) = (\Gamma')^{-1/2}\Gamma z_1$ have independent components and at least two rows of $(\Gamma')^{-1/2}\Gamma$ have only non-zero entries (this can be relaxed even further), then, by the Darmois-Skitovich Theorem [cf. 7, Theorem 5.3.1.], z_1 must already be Gaussian.

⁶In WC it is implicitly assumed that $\liminf_n \sigma_n^2 > 0$.

$n \rightarrow \infty$ is essential in Theorem 2.1 in order to obtain a Gaussian limit. This assumption is dropped in Theorem 3.1 and Corollary 3.2.

Finally, the assumption in Theorem 2.1 that $\Delta_\gamma = o(q_n/n)$, is rather natural in the case where $\liminf_n q_n/n > 0$, where it simply reduces to $\Delta_\gamma = o(1)$. In this case, it says that we study the asymptotic rejection probability only in a shrinking neighborhood of the null hypothesis. If $\liminf_n q_n/n > 0$, we also do not need to specify a rate at which Δ_γ approaches zero. Note, however, that part (ii) of Theorem 2.1 actually does require a specific rate of contraction which is, again, only needed for technical reasons in establishing the asymptotic behavior of the non-centrality term (cf. Section 6.4). The corresponding Assumption (C4) in WC is rather dubious and seems to arise from a miscalculation when dealing with said non-centrality term. In fact, they also need the $O(n^{-1/2})$ rate of $(R_0\gamma - r_0)'R_0SR_0'(R_0\gamma - r_0)/\sigma_n^2$ imposed by our Theorem 2.1(ii) and nothing more.⁷

In the case where $\liminf_n q_n/n = 0$, we need the rate $\Delta_\gamma = o(q_n/n)$ in order to ensure that the mixed term in the expansion of the F -statistic is asymptotically negligible (cf. the discussion following display (6.7)). Note that in the extreme case where $q_n = q$ is fixed, the assumption $\Delta_\gamma = o(q_n/n)$ appears to be somewhat restrictive because it rules out $n\Delta_\gamma \asymp 1$, as required for non-trivial local asymptotic power (cf. Remark 2.3). In the classical case where $q_n = q$ (and $p_n = p$) is fixed, the classical approach based on the asymptotic normality of the OLS estimator $\hat{\gamma}_n$ allows for a much wider range of local alternatives than our present strategy. Of course, the asymptotic normality of the whole vector $\hat{\gamma}_n$ breaks down if the dimensions p_n and q_n are too large relative to sample size n (see, e.g., Portnoy [20, 21]), which is why we here use a different strategy involving the assumption $\Delta_\gamma = o(q_n/n)$. Judging from the simulations of Section 5 below, it seems as if some bound on Δ_γ that is proportional to q_n is essential for the validity of the normal approximation to the power function, because this approximation turns out to be accurate for a larger range of Δ_γ if q gets larger.

5. Numerical results

In order to better understand the quality of the theoretical approximations to the power function of the F -test derived above, we conducted an extensive simulation study. We roughly follow the simulation setup of WC but our focus is more on the role of the number of tested hypotheses q . The linear model we considered was

$$y_i = \beta'x_i + \varepsilon_i,$$

where the ε_i were i.i.d. with mean zero and variance one, independent of the design. We tried different error distributions but found little effect on the power function so we report only the results for $\mathcal{N}(0,1)$ errors and for $t(5)/\sqrt{5/3}$

⁷See the first display on page 146 in WC, where also the matrix $X_2X_2^T$ needs to be standardized. Also, there is a scaling factor of \sqrt{n} missing in that argument, which is necessary to bring the non-centrality term to the same scale as the noise term.

errors.⁸ The p -dimensional design vectors x_i were generated as i.i.d. realizations of a moving average process

$$x_{ij} = \sum_{t=1}^T \alpha_t z_{i(j+t-1)},$$

where $z_i = (z_{i1}, \dots, z_{i(p+T-1)})'$ was either a $(p+T-1)$ -dimensional standard normal vector or generated with i.i.d. $\Gamma(1,1) - 1$ entries⁹ and $T = 10$. The coefficients α_t were generated as i.i.d. uniform from $(0,1)$ only once and then held fixed across all the simulations to follow.

We tested only null hypotheses of the form

$$H_0 : \beta_1 = \dots = \beta_q = 0,$$

at level $\nu = 0.05$, so that the distribution of the F -statistic does neither depend on the mean of the design vectors nor on the intercept parameter, and hence it is no loss of generality that we have omitted both. The signal β was generated such that half of the tested coefficients were equal to one and all the other coefficients were equal to zero. The signal was then scaled appropriately to produce a range of alternatives at which the power function was evaluated numerically.

Since WC have already extensively studied the effect of the dimensionality p on the power of the F -test, we here focus more on the effect of the number of tested restrictions q . For our first set of Montecarlo experiments we fixed $n = 100$ and $p = 60$ and looked at the cases $q = 4$ and $q = 50$. Figure 1 shows the simulated power of the F -test (solid lines) as a function of the scaled distance from the null hypothesis Δ_γ , where for each value of Δ_γ , 10.000 Montecarlo samples were generated. The left column shows the results for $t(5)/\sqrt{5/3}$ distributed errors and a design that was generated from a moving average process with $\Gamma(1,1) - 1$ innovations. The right column was generated with i.i.d. standard Gaussian errors and a Gaussian moving average design. As a first observation we note that the influence of non-Gaussianity on the simulated power (solid lines) is almost negligible at the present sample size of $n = 100$. Now Figure 1 should be inspected from top to bottom. In the first row we clearly see the gain of power as the number of hypotheses q decreases from $q = 50$ to $q = 4$. We also compare the simulated power to the Gaussian approximation from our asymptotic result of Theorem 2.1 (dotted lines). The picture is qualitatively the same as in WC, who considered $q = p - 2$ and who concluded “that there is a good conformity between the empirical power and the theoretical power of the [...] F -test [...]” [27, p. 142]. It seems hard to evaluate the quality of approximation directly from this picture in absolute terms. Moreover, the Gaussian approximation does not seem to become much more accurate when q increases, contrary

⁸Of course, one can always completely change the picture by, e.g., taking an error distribution that does not have a finite variance. In that case none of our theoretical approximations is of much use.

⁹The $\Gamma(1,1)$ -distribution is just the standard exponential distribution, but we keep the Γ notation in the plots to facilitate comparability with WC.

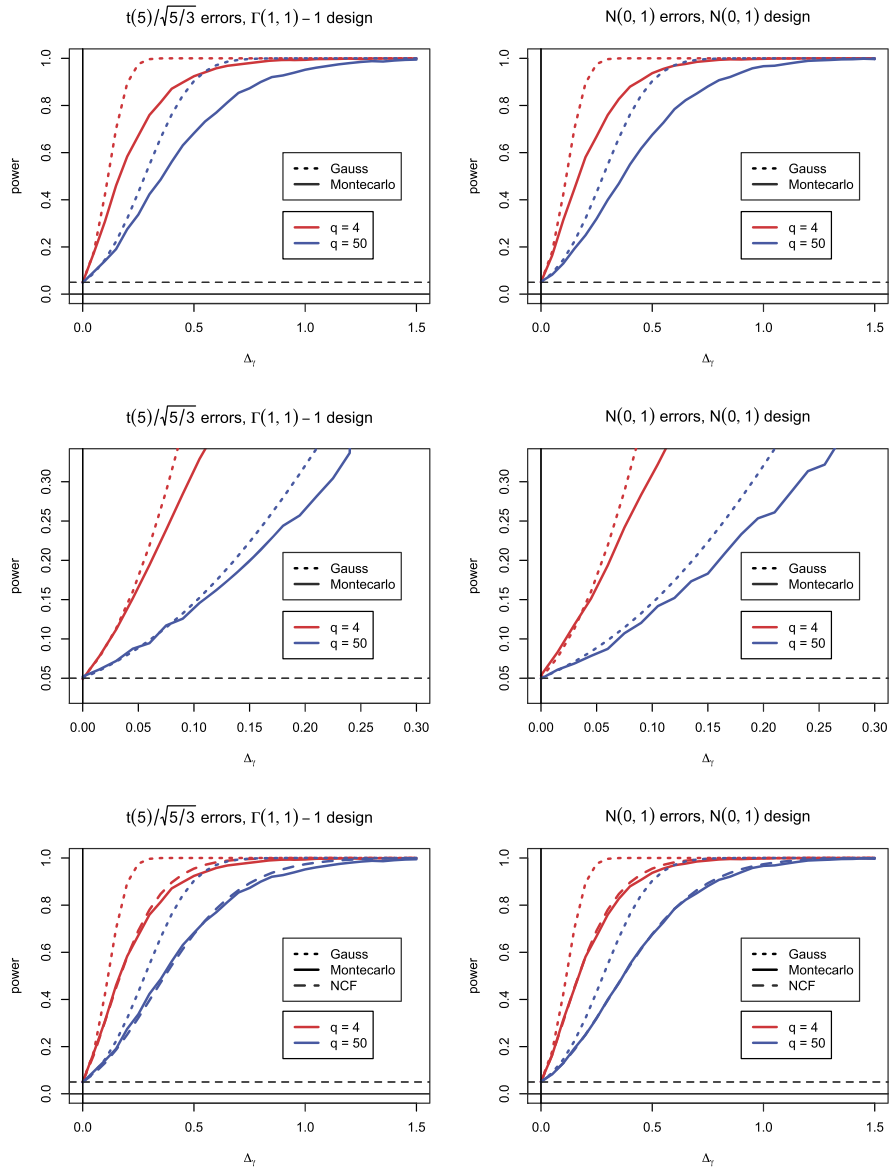


FIG 1. Simulated power functions of the F-test (Montecarlo) at level $\nu = 0.05$, compared to the Gaussian approximation of Corollary 2.2 (Gauss) and the non-central F approximation of Corollary 3.2 (NCF).

to what was suggested by Theorem 2.1. In fact, however, Theorem 2.1 says that we should look at the Gaussian approximation only locally, for values of Δ_γ that are of a smaller order than q/n . Indeed, if we look at the power function in a smaller neighborhood around the null (cf. the second row of Figure 1) we see

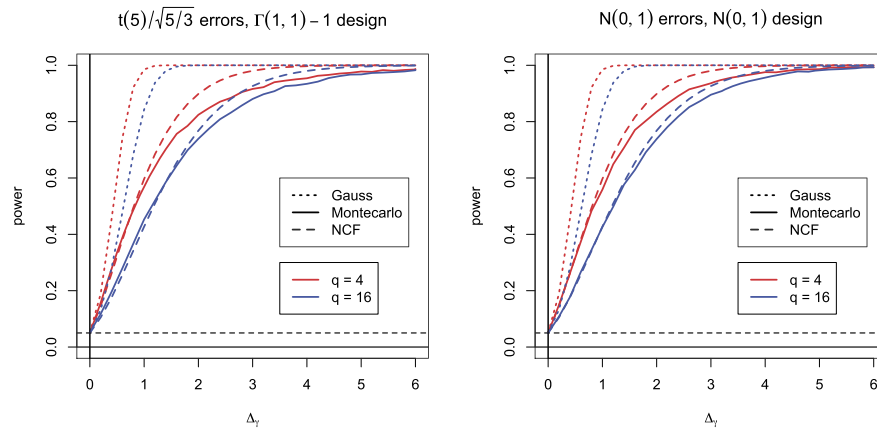


FIG 2. Simulated power functions of the F -test (Montecarlo) at level $\nu = 0.05$, Gaussian approximation (Gauss) and the non-central F approximation (NCF) for sample size $n = 30$ and $p = 20$ regressor variables.

a much better agreement between the simulated true power and the Gaussian approximation. We also see that when q/n is larger, the approximation is accurate on a larger interval around the null, as predicted by the theory. However, a global Gaussian approximation to the power of the F -test seems to be too much to ask for. Finally, the last row of Figure 1 is identical to the first row, except that we have added the theoretical approximation based on the non-central F -distribution as in Corollary 3.2 (NCF). Compared to the Gaussian approximation, the non-central F approximation is remarkably accurate over the entire range of alternatives considered, not just locally, and for both large and small values of q .

To investigate also the quality of our approximations in small samples, we have repeated the simulations above with $n = 30$ and $p = 20$. We present only one instance of this second round of simulations in Figure 2 to discuss the main differences to the case where $n = 100$. We still find that the non-central F approximation is much better than the Gaussian approximation, but clearly also the quality of the former deteriorates considerably compared to the case $n = 100$. However, the local behavior of the power function is still picked up quite well even in the small sample scenario. Moreover, it is remarkable how similar the picture with $t(5)$ -errors is to the picture with normal errors already for $n = 30$. This suggests that the dominant reason for the imperfect approximation by the non-central F -distribution is the randomness of the design rather than the non-Gaussianity of the errors (cf. Remark 3.3).

Finally, as in WC, we also investigate the distribution of the F -statistic under the null hypothesis $H_0 : \beta_1 = \dots = \beta_q = 0$. For different choices of n , p and q , we have generated 10.000 Montecarlo samples of $s_n^{-1/2}(F_n - 1)$ as before, but with vanishing true signal $\beta = 0$, and for $t(5)/\sqrt{5/3}$ -errors and regressors generated from $\Gamma(1, 1) - 1$. To investigate also the impact of a non-symmetric error distribu-

tion we have repeated all the simulations also with the distribution of the errors and the design interchanged.¹⁰ The plots in Figure 3 were generated by applying the R-function ‘density’ [25] to the samples of standardized F -statistics with default settings. As before, the dashed lines correspond to the (appropriately scaled and centered) asymptotic non-central F -approximation of Corollary 3.2 and the black dotted line is the standard normal density. The overall picture is the same as for the power function. The non-central F approximation is remarkably accurate even for moderate sample sizes like $n = 50$. As predicted by the theory, for large n and large q the null distribution is well approximated by the normal, whereas for small q the normal approximation fails. We also note that, again, the approximation accuracy appears to be rather insensitive to changes of the error and design distributions.

6. Proofs of main results

In this section we give a high-level description of the proofs of both Theorem 2.1 and Theorem 3.1. The more technical parts of the argument are collected in the appendices. The following outline section pertains to both proofs and uses only assumptions that are invoked by both theorems. Note that because of compactness and, in particular, $\limsup_n p_n/n < 1$, it is no restriction to assume that p_n/n and q_n/n are convergent sequences with limits $\rho_1 \in [0, 1)$ and $\rho_2 \in [0, \rho_1]$, respectively.

6.1. Outline

In addition to the general model assumptions of Section 2, the present outline section 6.1 only uses the conditions $\Delta_\gamma = o(q_n/n)$, $\mathbb{P}(\det(U'U) = 0) = 0$ and $\sqrt{n}(\hat{\sigma}_n^2/\sigma_n^2 - 1) = O_{\mathbb{P}}(1)$. All of these are satisfied under the assumptions of any part of Theorem 2.1 as well as under the assumptions of Theorem 3.1, in view of Lemma C.1.

The first part of Section 6.1 closely follows the classical approach for the decomposition of the F -statistic as described, e.g., in Rao and Toutenburg [22, Chapter 3.7]. These arguments are kept to a minimum but are included nonetheless to make the notation more intelligible.

Recall the F -statistic F_n as defined in (2.4). For the following preliminary consideration, we work only on the event $C_n = \{\omega : \hat{\sigma}_n^2(\omega) > 0, \det(U'U(\omega)) \neq 0\}$, where $\hat{\sigma}_n^2 = \|Y - U\hat{\gamma}_n\|^2/(n - p - 1)$. On this event, F_n is given by

$$F_n = \frac{(R_0\hat{\gamma}_n - r_0)'(R_0(U'U)^{-1}R_0')^{-1}(R_0\hat{\gamma}_n - r_0)/q \frac{\sigma_n^2}{\hat{\sigma}_n^2}}{\sigma_n^2}. \quad (6.1)$$

Setting $\delta_\gamma = (R_0\gamma - r_0)/\sigma_n$, we have

$$(R_0\hat{\gamma}_n - r_0)/\sigma_n = (R_0(U'U)^{-1}U'Y - r_0)/\sigma_n = R_0(U'U)^{-1}U'(\varepsilon/\sigma_n) + \delta_\gamma,$$

¹⁰We do not report any results for Gaussian errors here because in that case the F -statistic follows exactly a central $F_{q,n-p-1}$ -distribution under the null hypothesis, irrespective of the design (cf. Remark 2.4).

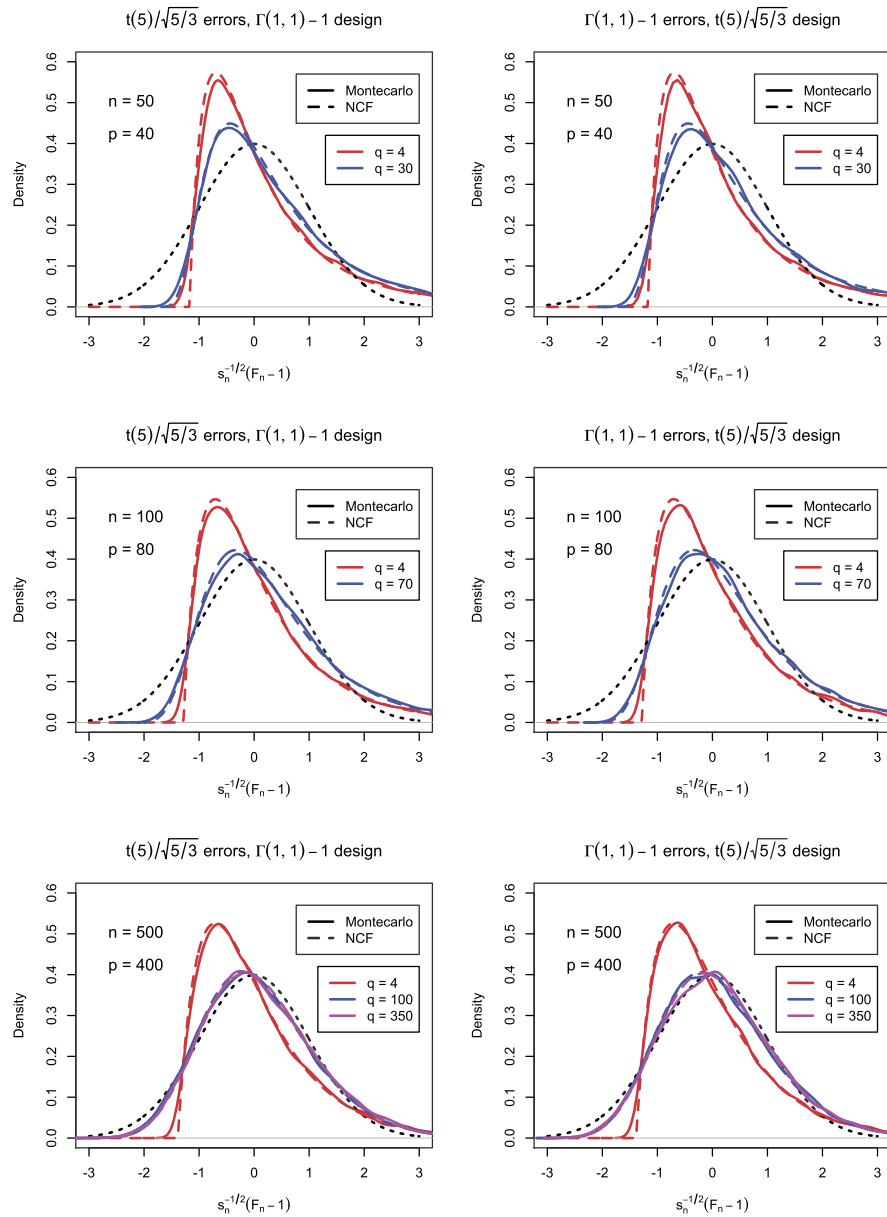


FIG 3. Kernel density estimates of simulated standardized F -statistics under the null hypothesis (solid lines), appropriately standardized non-central F -approximation of Corollary 3.2 (dashed lines) and the standard normal density (black dotted lines) for different choices of n , p and q , and different design and error distributions.

and thus, the first fraction in (6.1) reads

$$\begin{aligned} & (\varepsilon/\sigma_n)'U(U'U)^{-1}R'_0(R_0(U'U)^{-1}R'_0)^{-1}R_0(U'U)^{-1}U'(\varepsilon/\sigma_n)/q \\ & + 2(\varepsilon/\sigma_n)'U(U'U)^{-1}R'_0(R_0(U'U)^{-1}R'_0)^{-1}\delta_\gamma/q \\ & + \delta'_\gamma(R_0(U'U)^{-1}R'_0)^{-1}\delta_\gamma/q. \end{aligned}$$

Next, if $q < p + 1$, choose a $(p + 1 - q) \times p$ matrix R_1 , whose rows form an orthonormal basis for the orthogonal complement of the rows of R_0 . Recall that R_0 was chosen such that $R_0R'_0 = I_q$. Hence, $T := [R'_0, R'_1]'$ is a $(p + 1) \times (p + 1)$ orthogonal matrix. Partitioning $UT' = [U_0, U_1]$, where $U_0 = UR'_0$ and $U_1 = UR'_1$, and using block matrix inversion, we see that

$$\begin{aligned} R_0(U'U)^{-1}R'_0 &= [I_q, 0]T(U'U)^{-1}T'[I_q, 0]' \\ &= [I_q, 0](TU'UT')^{-1}[I_q, 0]' \\ &= (U'_0(I_n - P_{U_1})U_0)^{-1}. \end{aligned}$$

Similarly, we get

$$\begin{aligned} R_0(U'U)^{-1}U' &= [I_q, 0]T(U'U)^{-1}T'TU' \\ &= (U'_0(I_n - P_{U_1})U_0)^{-1}U'_0(I_n - P_{U_1}). \end{aligned}$$

Now, by writing $W = (I_n - P_{U_1})U_0$, on C_n , we can simplify the F -statistic to read

$$(\hat{\sigma}_n^2/\sigma_n^2)F_n = (\varepsilon/\sigma_n)'P_W(\varepsilon/\sigma_n)/q + 2(\varepsilon/\sigma_n)'W\delta_\gamma/q + \delta'_\gamma W'W\delta_\gamma/q.$$

The above representation remains correct also in the case where $q = p + 1$ provided that the matrix U_1 is removed wherever it appears, i.e., $W = U_0$ in this case. The correct centering and scaling of F_n is $s_n^{-1/2}(F_n - 1)$, for $s_n = 2(1/q + 1/(n - p - 1))$ (cf. Lemma C.8). After noting that $P_W = P_U - P_{U_1}$ and abbreviating $M_n = (P_U - P_{U_1})/q - (I_n - P_U)/(n - p - 1)$, we obtain

$$s_n^{-1/2}(F_n - 1) = s_n^{-1/2}(\varepsilon/\sigma_n)'M_n(\varepsilon/\sigma_n)\frac{\sigma_n^2}{\hat{\sigma}_n^2} \quad (6.2)$$

$$+ s_n^{-1/2}(2(\varepsilon/\sigma_n)'W\delta_\gamma/q + \delta'_\gamma W'W\delta_\gamma/q)\frac{\sigma_n^2}{\hat{\sigma}_n^2}, \quad (6.3)$$

on the event C_n . Now, to get rid of the restriction to C_n , define G_n by

$$G_n = s_n^{-1/2}(\varepsilon/\sigma_n)'M_n(\varepsilon/\sigma_n) + 2s_n^{-1/2}(\varepsilon/\sigma_n)'W\delta_\gamma/q + s_n^{-1/2}\delta'_\gamma W'W\delta_\gamma/q, \quad (6.4)$$

and note that this is well defined everywhere. It is now elementary to verify that we can study the asymptotic behavior of G_n instead of $s_n^{-1/2}(F_n - 1)$. Simply note that if $G_n - \eta_n^2$ converges weakly to some limiting distribution \mathcal{L} , for an appropriate centering sequence η_n^2 with $\eta_n^2 = o(\sqrt{n})$, then, on C_n ,

$$s_n^{-1/2}(F_n - 1) - \eta_n^2 = (G_n - \eta_n^2)(\sigma_n^2/\hat{\sigma}_n^2) + \eta_n^2(\sigma_n^2/\hat{\sigma}_n^2 - 1) \xrightarrow[n \rightarrow \infty]{w} \mathcal{L}, \quad (6.5)$$

and $\mathbb{P}(C_n) \rightarrow 1$ as $n \rightarrow \infty$, because $\mathbb{P}(\det(U'U) = 0) = 0$, $\sqrt{n}(\hat{\sigma}_n^2/\sigma_n^2 - 1) = O_{\mathbb{P}}(1)$ and $\mathbb{P}(\hat{\sigma}_n^2 = 0) \leq \mathbb{P}(|\hat{\sigma}_n^2/\sigma_n^2 - 1| > 1/2) \rightarrow 0$, as claimed at the beginning of this section.

In what follows, we will establish that the first term on the right of the equal sign in (6.4), which we denote by $Q_n := s_n^{-1/2}\varepsilon'M_n\varepsilon/\sigma_n^2$, satisfies $Q_n \rightarrow \mathcal{L}$, weakly, for an appropriate limit distribution \mathcal{L} . The last summand in (6.4) can be abbreviated to $s_n^{-1/2}n\nabla_n/q$, where $\nabla_n = \delta'_\gamma W'W\delta_\gamma/n = \delta'_\gamma(R_0(U'U/n)^{-1}R'_0)^{-1}\delta_\gamma$ (as in Remark 2.4). It will play the role of a non-centrality term and it will be shown to be asymptotically non-random. Note that if we can also show $s_n^{-1/2}n\nabla_n/q = o_{\mathbb{P}}(\sqrt{q})$, then the mixed term in (6.4) satisfies $s_n^{-1/2}(\varepsilon/\sigma_n)'W\delta_\gamma/q = o_{\mathbb{P}}(1)$. Indeed, the conditional mean of the latter expression given X is equal to zero, and its conditional variance is equal to $s_n^{-1}n\nabla_n/q^2 = (s_n^{-1/2}n\nabla_n/q)/(\sqrt{q}\sqrt{s_nq}) = o_{\mathbb{P}}(1)$, provided that $s_n^{-1/2}n\nabla_n/q = o_{\mathbb{P}}(\sqrt{q})$.

Suppose, for now, that we have already established both, the weak convergence

$$Q_n \xrightarrow[n \rightarrow \infty]{w} \mathcal{L}, \quad (6.6)$$

and also the fact that

$$s_n^{-1/2}n\nabla_n/q - \sqrt{n}\Delta_\gamma b_n \xrightarrow[n \rightarrow \infty]{i.p.} 0, \quad (6.7)$$

where b_n is as in Theorem 2.1. Then, because $\Delta_\gamma = o(q/n)$, we have $\eta_n^2 := \sqrt{n}\Delta_\gamma b_n = o(\sqrt{q}) = o(\sqrt{n})$, as required for the argument in (6.5). It also follows that $s_n^{-1/2}n\nabla_n/q = o(\sqrt{q}) + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(\sqrt{q})$, so we have asymptotic negligibility of the mixed term in (6.4) by the argument in the previous paragraph. Altogether, we arrive at

$$G_n - \eta_n^2 = Q_n + o_{\mathbb{P}}(1) \xrightarrow[n \rightarrow \infty]{w} \mathcal{L},$$

which establishes the conclusion of Theorem 2.1 and Theorem 3.1, for an appropriate choice of \mathcal{L} , provided that (6.6) and (6.7) hold.

For Theorem 2.1, we will prove the weak convergence (6.6) with $\mathcal{L} = \mathcal{N}(0, 1)$, under the general Assumptions (A1).(a,b,c,d) and (A2) in Section 6.2, and the convergence in (6.7) under each of the sets of assumptions of Theorem 2.1(i), 2.1(ii) and 2.1(iii), respectively (cf. Section 6.4).

For Theorem 3.1, we will prove the convergence (6.6) with $\mathcal{L} = (2q)^{-1/2}\chi_q^2 - \sqrt{q}/2$ in Section 6.3, and the convergence in (6.7) is established by the same argument as in the case of Theorem 2.1(iii) in Section 6.4, which does not require $q_n \rightarrow \infty$ nor Assumption (A2), so that it goes through also in the setting of Theorem 3.1.

6.2. Asymptotic normality of the noise term

This section establishes the weak convergence in (6.6) with $\mathcal{L} = \mathcal{N}(0, 1)$. For this claim we only use the Assumptions (A1).(a,b,c,d), (A2), as well as $p_n/n \rightarrow \rho_1 \in$

$[0, 1)$, $q_n/n \rightarrow \rho_2 \in [0, \rho_1]$ and $q_n \rightarrow \infty$. The following lemma is a variation of Theorem 2.1 in Bhansali, Giraitis and Kokoszka [5] on the asymptotic normality of quadratic forms for the case where the matrix and the enclosing vectors may exhibit a certain dependence between each other. Its proof is deferred to Appendix B.

Lemma 6.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the common probability space on which all the random quantities below are defined. For every $n \in \mathbb{N}$, let $\mathcal{G}_n \subseteq \mathcal{F}$ be a sub-sigma algebra, let $A_n = (a_{ij,n})_{i,j=1}^n$ be a real random symmetric $n \times n$ matrix that is \mathcal{G}_n measurable and such that $A_n(\omega) \neq 0, \forall \omega \in \Omega$. Let $Z_{1,n}, \dots, Z_{n,n}$ be real random variables that are conditionally independent, given \mathcal{G}_n , and such that for $i \leq n$, almost surely, $\mathbb{E}[Z_{i,n}|\mathcal{G}_n] = 0, \mathbb{E}[Z_{i,n}^2|\mathcal{G}_n] = 1$ and $\mathbb{E}[|Z_{i,n}|^4|\mathcal{G}_n] < \infty$. Moreover, assume that, as $n \rightarrow \infty$,*

$$\begin{aligned} \frac{\|A_n\|_S^2}{\|A_n\|_F^2} \max_{j=1, \dots, n} \mathbb{E}[Z_{j,n}^4|\mathcal{G}_n] &\xrightarrow[n \rightarrow \infty]{i.p.} 0, \\ \frac{\max_j (A_n^2)_{jj}}{\|A_n\|_F^2} \left(\max_{j=1, \dots, n} \mathbb{E}[Z_{j,n}^4|\mathcal{G}_n] \right)^2 &\xrightarrow[n \rightarrow \infty]{i.p.} 0, \\ \text{and} \quad \sum_{j=1}^n \frac{a_{jj,n}^2 \mathbb{E}[Z_{j,n}^4|\mathcal{G}_n]}{\|A_n\|_F^2} &\xrightarrow[n \rightarrow \infty]{i.p.} 0. \end{aligned}$$

Then, for $Z_n = (Z_{1,n}, \dots, Z_{n,n})'$, we have

$$\frac{Z_n' A_n Z_n - \mathbb{E}[Z_n' A_n Z_n|\mathcal{G}_n]}{\sqrt{2}\|A_n\|_F} \xrightarrow[n \rightarrow \infty]{w} \mathcal{N}(0, 1).$$

Remark 6.2. The proof of Lemma 6.1 essentially follows the rationale of Bhansali, Giraitis and Kokoszka [5] with the obvious modification that all the moments of $Z_{i,n}$ have to be replaced by conditional moments. Note that if the $Z_{1,n}, \dots, Z_{n,n}$ are the first n elements of a sequence of i.i.d. random variables and A_n is non-random, as in Bhansali, Giraitis and Kokoszka [5], then the assumptions of Lemma 6.1 reduce to those imposed by Theorem 2.1(iii) in that reference, except for the additional requirement that $\mathbb{E}[Z_1^4] < \infty$, as needed here. By the method of Bhansali, Giraitis and Kokoszka [5] we can not get rid of this additional requirement because their truncation argument does not apply in the case of dependence between A_n and Z_n .

With Lemma 6.1 at hand, we can proof the asymptotic normality of

$$Q_n = s_n^{-1/2}(\varepsilon/\sigma_n)' M_n(\varepsilon/\sigma_n).$$

Under the linear model (2.1), the components $\varepsilon_1, \dots, \varepsilon_n$ of ε are conditionally independent given the design X , with $\mathbb{E}[\varepsilon_i/\sigma_n|X] = 0, \mathbb{E}[(\varepsilon_i/\sigma_n)^2|X] = 1$ and $\mathbb{E}[(\varepsilon_i/\sigma_n)^4|X] < \infty$, almost surely, in view of Assumption (A2). Moreover, the random matrix $M_n = (P_U - P_{U_1})/q - (I_n - P_U)/(n - p - 1)$ is $\sigma(X)$ -measurable and satisfies $\text{trace}(M_n) = 0$ and

$$\|M_n\|_F^2 = \text{trace}(M_n^2) = \text{trace}([(P_U - P_{U_1})/q^2 + (I_n - P_U)/(n - p - 1)^2])$$

$$= 1/q + 1/(n - p - 1),$$

with probability one, in view of (A1).(b). Also, $\|M_n\|_F^2 \geq 1/(n-p-1) \geq 1/n$, everywhere, because $\text{trace}(P_U) \leq p+1$. With this and in view of $\mathbb{E}[\varepsilon' M_n \varepsilon | X] / \sigma_n^2 = \text{trace}(M_n) = 0$, almost surely, we see that

$$Q_n = s_n^{-1/2} \frac{\varepsilon' M_n \varepsilon}{\sigma_n^2} = \frac{(\varepsilon/\sigma_n)' M_n (\varepsilon/\sigma_n) - \mathbb{E}[(\varepsilon/\sigma_n)' M_n (\varepsilon/\sigma_n) | X]}{\sqrt{2} \|M_n\|_F},$$

at least on a set of probability one, and it remains to verify the convergence conditions of Lemma 6.1. For the first one, note that $\|M_n\|_S^2 \leq (1/q + 1/(n - p - 1))^2$ and hence, almost surely,

$$\frac{\|M_n\|_S^2}{\|M_n\|_F^2} \max_{j=1, \dots, n} \mathbb{E}[(\varepsilon_j/\sigma_n)^4 | X] \leq \left(1 + \frac{q}{n - p - 1}\right) \frac{\max_{j=1, \dots, n} \mathbb{E}[(\varepsilon_j/\sigma_n)^4 | X]}{q}.$$

But clearly $\max_j \mathbb{E}[(\varepsilon_j/\sigma_n)^4 | X] / q \leq O_{\mathbb{P}}(1) \max_j \mathbb{E}[(\tilde{\varepsilon}_j/\sigma_n)^4 | x_j] / q = o_{\mathbb{P}}(q^{-1/2})$ under Assumption (A2). Therefore, the upper bound in the previous display converges to zero in probability. For the second condition, since the diagonal entries of a projection matrix are between 0 and 1, we see that $(M_n^2)_{jj} \leq 1/q^2 + 1/(n - p - 1)^2$, and thus

$$\begin{aligned} & \frac{\max_j (M_n^2)_{jj}}{\|M_n\|_F^2} \left(\max_{j=1, \dots, n} \mathbb{E}[(\varepsilon_j/\sigma_n)^4 | X] \right)^2 \\ & \leq \frac{1 + q^2/(n - p - 1)^2}{1 + q/(n - p - 1)} (\max_i e_i)^8 \frac{(\max_j \mathbb{E}[(\tilde{\varepsilon}_j/\sigma_n)^4 | x_j])^2}{q}, \end{aligned}$$

which converges to zero in probability under Assumption (A2) and $q_n \rightarrow \infty$. Establishing the validity of the last condition is slightly more involved. Since $\|M_n\|_F^2$ is of order $1/q$, we have to show that

$$q \sum_{j=1}^n m_{jj}^2 \mathbb{E}[(\varepsilon_j/\sigma_n)^4 | X] \leq (\max_i e_i)^4 q \sum_{j=1}^n m_{jj}^2 \mathbb{E}[(\tilde{\varepsilon}_j/\sigma_n)^4 | X] = o_{\mathbb{P}}(1),$$

where $M_n = (m_{ij})_{i,j=1}^n$. By Assumption (A2), $(\max_i e_i)^4 = O_{\mathbb{P}}(1)$. Now, take expectation and use Hölder's inequality with $a, b > 1$ such that $1/a + 1/b = 1$ to obtain

$$\begin{aligned} & \mathbb{E} \left(q \sum_{j=1}^n m_{jj}^2 \mathbb{E}[(\tilde{\varepsilon}_j/\sigma_n)^4 | X] \right) \\ & \leq (\mathbb{E}[(\mathbb{E}[(\tilde{\varepsilon}_1/\sigma_n)^4 | x_1])^b])^{1/b} q \sum_{j=1}^n (\mathbb{E}[m_{jj}^{2a}])^{1/a}. \end{aligned} \tag{6.8}$$

Now choose $b = 1 + \kappa$ and invoke Assumption (A2) to show that the b -th moment of the conditional expectation in (6.8) is $O(1)$. To establish that

$$q \sum_{j=1}^n (\mathbb{E}[m_{jj}^{2a}])^{1/a} \xrightarrow[n \rightarrow \infty]{i.p.} 0, \tag{6.9}$$

we distinguish between the cases $\rho_2 > 0$ and $\rho_2 = 0$.

If $\rho_2 > 0$, we write the diagonal elements of $M_n = (P_U - P_{U_1})/q - (I_n - P_U)/(n - p - 1)$ as

$$\begin{aligned} m_{jj} &= \frac{(P_U)_{jj} - (P_{U_1})_{jj}}{q} - \frac{1 - (P_U)_{jj}}{n - p - 1} \\ &= \frac{(P_U)_{jj} - \frac{p+1}{n} + \frac{p+1-q}{n} - (P_{U_1})_{jj}}{q} + \frac{(P_U)_{jj} - \frac{p+1}{n}}{n - p - 1}, \end{aligned}$$

and note that $(P_U)_{jj} - (p+1)/n \in [-1, 1]$ and $(P_{U_1})_{jj} - (p+1-q)/n \in [-1, 1]$, in order to get the bound

$$\begin{aligned} & (\mathbb{E}[m_{jj}^{2a}])^{1/a} \\ &= \frac{1}{n^2} \left(\mathbb{E} \left[\left\{ \left((P_U)_{jj} - \frac{p+1}{n} \right) \left(\frac{n}{q} + \frac{n}{n-p-1} \right) + \right. \right. \right. \\ & \quad \left. \left. \left. \left(\frac{p+1-q}{n} - (P_{U_1})_{jj} \right) \frac{n}{q} \right\}^{2a} \right] \right)^{1/a} \\ &\leq \frac{1}{n^2} \left(2^{2a-1} \left\{ \left(\frac{n}{q} + \frac{n}{n-p-1} \right)^{2a} \mathbb{E} \left[\left((P_U)_{jj} - \frac{p+1}{n} \right)^{2a} \right] + \right. \right. \\ & \quad \left. \left. \left(\frac{n}{q} \right)^{2a} \mathbb{E} \left[\left((P_{U_1})_{jj} - \frac{p+1-q}{n} \right)^{2a} \right] \right\} \right)^{1/a}, \end{aligned}$$

where we have used the inequality $(c + d)^{2a} \leq 2^{2a-1}(c^{2a} + d^{2a})$ for $c, d \in \mathbb{R}$. Hence, if we can show that the $(P_U)_{jj}$, for $j = 1, \dots, n$, and also the $(P_{U_1})_{jj}$, for $j = 1, \dots, n$, are identically distributed, then

$$\begin{aligned} n \sum_{j=1}^n (\mathbb{E}[m_{jj}^{2a}])^{1/a} &\leq 2^{(2a-1)/a} O(1) \times \\ &\quad \times (\mathbb{E} [|(P_U)_{11} - (p+1)/n|^{2a}] + \mathbb{E} [|(P_{U_1})_{11} - (p+1-q)/n|^{2a}])^{1/a}, \end{aligned}$$

because $n/q \rightarrow 1/\rho_2$. Since $a = (1 + \kappa)/\kappa$ is fixed, it then remains to show that $|(P_U)_{11} - (p+1)/n| \rightarrow 0$ and $|(P_{U_1})_{11} - (p+1-q)/n| \rightarrow 0$, in probability. The desired properties of the diagonal entries of P_U and P_{U_1} are now established by the following lemma, which applies under the Assumptions (A1).(a,b,d), and whose proof is deferred to Appendix B.

Lemma 6.3. *For every $n \in \mathbb{N}$, let $x_{1,n}, \dots, x_{n,n}$ be i.i.d. random p_n -vectors that satisfy Assumptions (A1).(a,b) with $\mu_n \in \mathbb{R}^{p_n}$ and positive definite covariance matrix Σ_n . Moreover, suppose that the random vector $z_{1,n}$ from Assumption (A1).(a) also satisfies $\text{Var}[z'_{1,n} M_n z_{1,n}] = O(\text{trace}(M_n^2) + (\text{trace}(M_n))^2 o(1))$, as $n \rightarrow \infty$, for every symmetric $m_n \times m_n$ matrix M_n . Let R_n be a non-random $(p_n + 1) \times k_n$ matrix such that $\text{rank } R_n = k_n \leq p_n + 1$ and define $X_n = [x_{1,n}, \dots, x_{n,n}]'$ and $W_n = [\iota, X_n] R_n$, where $\iota = (1, \dots, 1)' \in \mathbb{R}^n$. Furthermore, let $h_{1,n}, \dots, h_{n,n}$ denote the diagonal entries of the projection matrix*

P_{W_n} . Then, the $(h_{j,n})_{j=1}^n$ are exchangeable random variables and

$$|h_{1,n} - k_n/n| \xrightarrow[n \rightarrow \infty]{} 0,$$

in probability.

Remark 6.4. Note that for $R_n = I_{p_n+1}$, the h_1, \dots, h_n in Lemma 6.3 are the leverage values of the regression with design matrix $U = [t, X]$.

Altogether, we see that in the case $\rho_2 > 0$, (6.9) holds and the weak convergence

$$Q_n = s_n^{-1/2} \frac{\varepsilon' M_n \varepsilon}{\sigma_n^2} \xrightarrow[n \rightarrow \infty]{w} \mathcal{N}(0, 1),$$

follows, as required in (6.6).

To treat the case $\rho_2 = 0$, we recall from Section 6.1 that $P_U - P_{U_1} = P_W = U(U'U)^{-1}R'_0(R_0(U'U)^{-1}R'_0)^{-1}R_0(U'U)^{-1}U'$, almost surely, by Assumption (A1).(b), and write the diagonal elements of $M_n = P_W/q - (I_n - P_U)/(n - p - 1)$ as before as

$$m_{jj} = \frac{(P_W)_{jj} - \frac{q}{n}}{q} + \frac{(P_U)_{jj} - \frac{p+1}{n}}{n - p - 1}.$$

Note that the $(P_W)_{jj} = (1, x'_j)(U'U)^{-1}R'_0(R_0(U'U)^{-1}R'_0)^{-1}R_0(U'U)^{-1}(1, x'_j)'$, for $j = 1, \dots, n$, are exchangeable random variables, because x_1, \dots, x_n are i.i.d. and $U'U = \sum_{j=1}^n (1, x'_j)'(1, x'_j)$ is a function in x_1, \dots, x_n that is invariant under permutations of its arguments. Therefore,

$$\begin{aligned} q \sum_{j=1}^n (\mathbb{E}[m_{jj}^{2a}])^{1/a} &\leq q \sum_{j=1}^n \left(\mathbb{E} \left[\left(\frac{|(P_W)_{jj} - \frac{q}{n}|}{q} + \frac{1}{n - p - 1} \right)^{2a} \right] \right)^{1/a} \\ &= \left(\mathbb{E} \left[\left(\sqrt{\frac{n}{q}} \left| (P_W)_{11} - \frac{q}{n} \right| + \frac{\sqrt{nq}}{n - p - 1} \right)^{2a} \right] \right)^{1/a}. \end{aligned}$$

By boundedness of the diagonal entries of a projection matrix, it remains to show that $H_n := \sqrt{n/q}((P_W)_{11} - q/n) \rightarrow 0$, in probability, as $n \rightarrow \infty$. By exchangeability, and Assumption (A1).(b), it follows that $q = \text{trace}(P_W) = \mathbb{E}[\text{trace}(P_W)] = \sum_{j=1}^n \mathbb{E}[(P_W)_{jj}] = n\mathbb{E}[(P_W)_{11}]$, almost surely, and thus $\mathbb{E}[H_n] = 0$. Moreover, $H_n \geq -\sqrt{q/n}$, and for $\varepsilon > 0$, $0 = \mathbb{E}[H_n] \geq \mathbb{E}[H_n \mathbf{1}_{\{H_n > \varepsilon\}}] - \sqrt{q/n}$, which implies that for n large (such that $\sqrt{q/n} < \varepsilon$), $\varepsilon \mathbb{P}(|H_n| > \varepsilon) \leq \varepsilon \mathbb{P}(H_n > \varepsilon) + \varepsilon \mathbb{P}(H_n < -\varepsilon) \leq \mathbb{E}[H_n \mathbf{1}_{\{H_n > \varepsilon\}}] + 0 \leq \sqrt{q/n}$. This establishes the asymptotic normality required in (6.6) also in the case $\rho_2 = 0$.

6.3. Asymptotic χ^2 -distribution of the noise term

In this section we show that under the assumptions of Theorem 3.1, (6.6) holds with $\mathcal{L} = (2q)^{-1/2} \chi_q^2 - \sqrt{q/2}$. First note that

$$Q_n = s_n^{-1/2} \frac{\varepsilon' M_n \varepsilon}{\sigma_n^2} = s_n^{-1/2} \frac{\varepsilon' P_W \varepsilon}{\sigma_n^2 q} - s_n^{-1/2} \frac{\hat{\sigma}_n^2}{\sigma_n^2}.$$

By Lemma C.1, we have $s_n^{-1/2} \hat{\sigma}_n^2 / \sigma_n^2 \rightarrow \sqrt{q/2}$, in probability. Since we do not test the intercept parameter α , we can write $R_0 = [0, T_0]$, for some $q \times p$ matrix T_0 , and

$$R_1 = \begin{pmatrix} 1 & 0 \cdots 0 \\ 0 & \\ \vdots & T_1 \\ 0 & \end{pmatrix}$$

for some $(p - q) \times p$ matrix T_1 , such that $\bar{T} := [T_0', T_1']'$ is $p \times p$ orthogonal. If $q = p$, then $T_0 = I_p$ and $R_1 = (1, 0, \dots, 0)$. With this notation, we get $U = [\iota, X]$, $U_0 = UR_0' = XT_0'$ and $U_1 = UR_1' = [\iota, XT_1']$. From this we see that the distribution of $W = (I_n - P_{U_1})U_0 = (I_n - P_{[\iota, (I_n - P_\iota)XT_1']})XT_0'$ does not depend on μ , and without loss of generality we may assume that $\mu = 0$. Moreover, by standard properties of orthogonal projections,

$$\begin{aligned} P_W &= P_U - P_{U_1} = P_X + P_{(I_n - P_X)\iota} - \left(P_{XT_1'} + P_{(I_n - P_{XT_1'})\iota} \right) \\ &= P_{X\bar{T}'} - P_{XT_1'} + P_{(I_n - P_X)\iota} - P_{(I_n - P_{XT_1'})\iota} \\ &= P_{(I_n - P_{XT_1'})XT_0'} + P_{(I_n - P_X)\iota} - P_{(I_n - P_{XT_1'})\iota}. \end{aligned}$$

Now, abbreviate $A = (I_n - P_{XT_1'})XT_0'$ and $B = P_{(I_n - P_X)\iota} - P_{(I_n - P_{XT_1'})\iota}$, and note that P_A is uniformly distributed on the Grassmann manifold of $n \times n$ projection matrices of rank q , because for an $n \times n$ orthogonal matrix O we have $OX = (OX\Sigma^{-1/2})\Sigma^{1/2} \sim X$, $OP_AO' = OA(A'O'OA)^{-1}A'O'$, and

$$\begin{aligned} OA &= (I_n - OXT_1'(T_1X'O'OX T_1')^{-1}T_1X'O')OXT_0' \\ &\sim (I_n - XT_1'(T_1X'XT_1')^{-1}T_1X')XT_0' = A. \end{aligned}$$

Moreover, $\text{trace}(B) = 0$, almost surely, and thus $\mathbb{E}[\varepsilon' B \varepsilon / \sigma_n^2 | X] = 0$ and by standard calculations using independence (cf. the proof of Lemma C.1) and the fact that $P_X = P_{X\bar{T}'} = P_{[XT_0', XT_1']}$,

$$\begin{aligned} \text{Var}[\varepsilon' B \varepsilon / \sigma_n^2 | X] &= 2\text{trace}(B^2) + \sum_{i=1}^n (\mathbb{E}[(\varepsilon_i / \sigma_n)^4] - 3) B_{ii}^2 \\ &\leq (2 + \mathbb{E}[(\varepsilon_1 / \sigma_n)^4]) \text{trace}(B^2) \\ &= O(1) \left(2 - 2 \frac{\iota'(I_n - P_X)(I_n - P_{XT_1'})\iota \iota'(I_n - P_{XT_1'})(I_n - P_X)\iota}{\iota'(I_n - P_X)\iota \iota'(I_n - P_{XT_1'})\iota} \right) \\ &= O(1) \left(1 - \frac{\iota'(I_n - P_{X\bar{T}'})\iota}{\iota'(I_n - P_{XT_1'})\iota} \right). \end{aligned}$$

In view of Lemma C.7 with $\mu = 0$ and using the first two moments of the χ^2 distribution, we see that the expressions $\iota'(I_n - P_{X\bar{T}'})\iota/n$ and $\iota'(I_n - P_{XT_1'})\iota/n$ both converge to $1 - \rho_1$, in probability, and thus the whole expression on the

last line of the previous display converges to zero in probability. Altogether, we see that

$$Q_n = (s_n q^2)^{-1/2} \frac{\varepsilon' P_A \varepsilon}{\sigma_n^2} - \sqrt{q/2} + o_{\mathbb{P}}(1).$$

Because P_A is uniformly distributed on the Grassmann manifold, it can be stochastically represented as $P_A \sim CC'$, where C is a random $n \times q$ matrix that is uniformly distributed on the Stiefel manifold of order $n \times q$, i.e., V has orthonormal columns and its distribution is both left and right invariant under the action of the appropriate orthogonal group. Since C and ε are independent, the so called Diaconis-Freedman effect as described in Dümbgen and Conte-Zerial [9, Theorem 2.1] entails that the conditional distribution of $C'\varepsilon/\sigma_n$ given C , converges weakly in probability to a q -dimensional standard normal distribution because $\|\varepsilon/\sigma_n\|^2/n \rightarrow 1$ in probability and $\varepsilon'\bar{\varepsilon}/(n\sigma_n^2) \rightarrow 0$ in probability, where $\bar{\varepsilon}$ is an independent copy of ε . Weak convergence in probability implies convergence of the conditional characteristic functions of $C'\varepsilon/\sigma_n$ given C , in probability, which, by boundedness, implies convergence of the unconditional characteristic functions. Consequently, we obtain the weak convergence

$$Q_n \sim (s_n q^2)^{-1/2} \|C'\varepsilon/\sigma_n\|^2 - \sqrt{q/2} + o_{\mathbb{P}}(1) \xrightarrow[n \rightarrow \infty]{w} (2q)^{-1/2} \chi_q^2 - \sqrt{q/2}.$$

This establishes (6.6) with \mathcal{L} as claimed.

6.4. Asymptotic behavior of the non-centrality term

Finally, we have to establish the convergence in (6.7) in the three cases of Theorem 2.1(i), 2.1(ii) and 2.1(iii) as well as under the assumptions of Theorem 3.1. We begin by a representation of $s_n^{-1/2} n \nabla_n / q$ that pertains to all of these cases. In this preliminary consideration we only use Assumptions (A1).(a) and (b) which are assumed to hold in each of the cases under investigation. Recall the conventions and definitions of Section 6.1, in particular, $U = [\iota, X]$, $T = [R'_0, R'_1]'$, $U_0 = UR'_0$, $U_1 = UR'_1$, $W = (I_n - P_{U_1})U_0$, $\delta_\gamma = (R_0\gamma - r_0)/\sigma_n$, $\Delta_\gamma = \delta'_\gamma (R_0 S^{-1} R'_0)^{-1} \delta_\gamma$, and $\nabla_n = \delta'_\gamma W' W \delta_\gamma / n = \delta'_\gamma (R_0 (U'U/n)^{-1} R'_0)^{-1} \delta_\gamma$. Write $\hat{\mu}_n = \sum_{i=1}^n x_i / n = X'\iota/n$ and $\hat{\Sigma}_n = X'X/n - \hat{\mu}_n \hat{\mu}'_n = X'(I_n - P_\iota)X/n$ and partition the $(p+1) \times (p+1)$ orthogonal matrix T as

$$T = \begin{pmatrix} R_0 \\ R_1 \end{pmatrix} = \begin{pmatrix} t_0 & T_0 \\ t_1 & T_1 \end{pmatrix},$$

where $t_0 \in \mathbb{R}^q$ and $T_0 \in \mathbb{R}^{q \times p}$. Since

$$(U'U/n)^{-1} = \begin{pmatrix} 1 & \hat{\mu}'_n \\ \hat{\mu}_n & \hat{\Sigma}_n + \hat{\mu}_n \hat{\mu}'_n \end{pmatrix}^{-1} = \begin{pmatrix} 1 + \hat{\mu}'_n \hat{\Sigma}_n^{-1} \hat{\mu}_n & -\hat{\mu}'_n \hat{\Sigma}_n^{-1} \\ -\hat{\Sigma}_n^{-1} \hat{\mu}_n & \hat{\Sigma}_n^{-1} \end{pmatrix},$$

almost surely, by (A1).(b), we have

$$s_n^{-1/2}n\nabla_n/q = \begin{cases} \sqrt{\frac{n}{q}}(s_nq)^{-1/2}\sqrt{n}\delta'_\gamma(U'_0U_0/n)\delta_\gamma, & \text{if } q = p + 1, \\ \sqrt{\frac{n}{q}}(s_nq)^{-1/2}\sqrt{n}\delta'_\gamma U'_0(I_n - P_{U_1})U_0\delta_\gamma/n, & \text{if } q \leq p, \\ \sqrt{\frac{n}{q}}(s_nq)^{-1/2}\sqrt{n}\delta'_\gamma(T_0\hat{\Sigma}_n^{-1}T'_0)^{-1}\delta_\gamma, & \text{if } t_0 = 0. \end{cases} \quad (6.10)$$

Notice that the last two cases are not mutually exclusive, but the case $t_0 = 0$ is a sub-case of the case $q \leq p$. The representation of ∇_n in the case $t_0 = 0$ will come in handy. With the notation

$$\Omega = \begin{bmatrix} \Omega_{00} & \Omega_{01} \\ \Omega_{10} & \Omega_{11} \end{bmatrix} = \begin{bmatrix} R_0SR'_0 & R_0SR'_1 \\ R_1SR'_0 & R_1SR'_1 \end{bmatrix} = TST', \quad \text{where} \\ S = \begin{pmatrix} 1 & \mu' \\ \mu & \Sigma + \mu\mu' \end{pmatrix} = \mathbb{E} \left[\begin{pmatrix} 1 \\ x_1 \end{pmatrix} (1 \quad x'_1) \right] = \mathbb{E}[U'U/n],$$

under (A1).(a), and by the simple block matrix inversion argument $R_0S^{-1}R'_0 = [I_q, 0](TST')^{-1}[I_q, 0]' = (\Omega_{00} - \Omega_{01}\Omega_{11}^{-1}\Omega_{10})^{-1}$, involving the orthogonality of T , we analogously get

$$\Delta_\gamma = \begin{cases} \delta'_\gamma\Omega_{00}\delta_\gamma, & \text{if } q = p + 1, \\ \delta'_\gamma(\Omega_{00} - \Omega_{01}\Omega_{11}^{-1}\Omega_{10})\delta_\gamma, & \text{if } q \leq p, \\ \delta'_\gamma(T_0\Sigma^{-1}T'_0)^{-1}\delta_\gamma, & \text{if } t_0 = 0. \end{cases} \quad (6.11)$$

6.4.1. The case of Theorem 2.1(i)

Under the assumptions of Theorem 2.1(i), choose b_n as in the theorem, which can be written as $b_n = \sqrt{(1 - (p + 1)/n)(1 - (p + 1)/n + q/n)/(2q/n)} = \sqrt{n/q}(s_nq)^{-1/2}(n - p - 1 + q)/n$. To establish the convergence in (6.7) in the case of $q = p + 1$, first note that now $b_n = \sqrt{n/q}(s_nq)^{-1/2}$, and consider

$$s_n^{-1/2}n\nabla_n/q - \sqrt{n}\Delta_\gamma b_n = b_n\sqrt{n}\delta'_\gamma(U'_0U_0/n - \Omega_{00})\delta_\gamma, \quad (6.12)$$

which has mean zero. For the variance, we observe that $\text{Var}[\sqrt{n}\delta'_\gamma(U'_0U_0/n)\delta_\gamma] = n^{-1}\sum_{i=1}^n \text{Var}[\delta'_\gamma R_0(1, x'_i)'(1, x'_i)R'_0\delta_\gamma] \leq \mathbb{E}[|\delta'_\gamma R_0(1, x'_1)'|^4] = O(|\delta'_\gamma\Omega_{00}\delta_\gamma|^2) = O(\Delta_\gamma^2)$, in view of Lemma C.3(i) and Assumption (A1).(d), and $b_n\Delta_\gamma = O(1)\sqrt{n/q}\Delta_\gamma = o(1)$, by assumption. This clearly covers also the case $R_0 = I_{p+1}$. If $R_0 = [0, I_p]$, then $t_0 = 0$, $T_0 = I_p$, $q = p$ and the difference in (6.7) reads

$$s_n^{-1/2}n\nabla_n/q - \sqrt{n}\Delta_\gamma b_n = \sqrt{\frac{n}{q}}(s_nq)^{-1/2}\sqrt{n}\delta'_\gamma \left(\hat{\Sigma}_n - \Sigma \frac{n-1}{n} \right) \delta_\gamma.$$

The mean of this expression is, again, equal to zero. For its variance we find that $\text{Var}[\sqrt{n}\delta'_\gamma\hat{\Sigma}_n\delta_\gamma] = O(|\delta'_\gamma\Sigma\delta_\gamma|^2)$, in view of Lemma C.3(ii) together with Assumption (A1).(d), and $\sqrt{n/q}(s_nq)^{-1/2}\delta'_\gamma\Sigma\delta_\gamma = O(1)\sqrt{n/q}\Delta_\gamma \rightarrow 0$, by assumption. This finishes the proof of Theorem 2.1(i).

6.4.2. The case of Theorem 2.1(ii)

For part (ii) we only need to consider the case where $q \leq p$, as the case $q = p + 1$ has already been treated above (simply restrict to the subsequence n' such that $q_{n'} \leq p_{n'}$). We establish the convergence in (6.7) for $b_n = \sqrt{n/q}(s_n q)^{-1/2}$ rather than b_n as in the Theorem. It should be clear, however, that this is no restriction. [Indeed, if (6.7) holds with $b_n = \sqrt{n/q}(s_n q)^{-1/2}$ and $\tilde{b}_n = b_n(n - p - 1 + q)/n$ is as in the theorem, then $\sqrt{n}\Delta_\gamma b_n - \sqrt{n}\Delta_\gamma \tilde{b}_n = (b_n - \tilde{b}_n)\sqrt{n}\Delta_\gamma = ((p + 1)/n - q/n)b_n\sqrt{n}O(\sqrt{q}/n)$, by the additional assumption of Theorem 2.1(ii). Since in the present case $(p - q)/n \rightarrow 0$, the previous expression converges to zero as $n \rightarrow \infty$.] Now, since $q \leq p$, the quantity of interest reads

$$s_n^{-1/2}n\nabla_n/q - \sqrt{n}\Delta_\gamma b_n = b_n\sqrt{n}\delta'_\gamma(U'_0U_0/n - \Omega_{00})\delta_\gamma + b_n\sqrt{n}\delta'_\gamma\left(\Omega_{01}\Omega_{11}^{-1}\Omega_{10} - \frac{U'_0U_1}{n}\left(\frac{U'_1U_1}{n}\right)^{-1}\frac{U'_1U_0}{n}\right)\delta_\gamma. \tag{6.13}$$

The first term on the right-hand-side has already been studied in (6.12), and the same argument applies, except that now $\delta'_\gamma\Omega_{00}\delta_\gamma \neq \Delta_\gamma$, in general. But $b_n\delta'_\gamma\Omega_{00}\delta_\gamma = O(1)\sqrt{n/q}\delta'_\gamma R_0 S R'_0 \delta_\gamma = O(n^{-1/2}) \rightarrow 0$ by the additional assumption of Theorem 2.1(ii). For the remaining term in (6.13), as in WC, we begin by approximating U'_1U_1/n by Ω_{11} . This can only be successful because here we are dealing with a sample covariance matrix of dimension $p + 1 - q$, based on n independent observations and we assume that $(p + 1 - q)/n \rightarrow 0$. We abbreviate $\tilde{U}_0 = U_0\Omega_{00}^{-1/2}$ and $\tilde{U}_1 = U_1\Omega_{11}^{-1/2}$ and consider the absolute difference

$$\begin{aligned} & \left| \sqrt{n}\delta'_\gamma \left[\frac{U'_0U_1}{n} \left(\frac{U'_1U_1}{n} \right)^{-1} \frac{U'_1U_0}{n} - \frac{U'_0U_1}{n} \Omega_{11}^{-1} \frac{U'_1U_0}{n} \right] \delta_\gamma \right| \\ &= \left| \sqrt{n}\delta'_\gamma \Omega_{00}^{1/2} \frac{\tilde{U}'_0\tilde{U}_1}{n} \left[\left(\frac{\tilde{U}'_1\tilde{U}_1}{n} \right)^{-1} - I_{p+1-q} \right] \frac{\tilde{U}'_1\tilde{U}_0}{n} \Omega_{00}^{1/2} \delta_\gamma \right| \\ &\leq \left\| \left(\frac{\tilde{U}'_1\tilde{U}_1}{n} \right)^{-1} - I_{p+1-q} \right\|_S \left\| \frac{\tilde{U}'_1\tilde{U}_1}{n} \right\|_S \left\| \frac{\tilde{U}'_0\tilde{U}_0}{n} \right\|_S \sqrt{n}\delta'_\gamma\Omega_{00}\delta_\gamma. \end{aligned} \tag{6.14}$$

Now, Lemma C.3(iii) with $k_n = p + 1 - q$ shows that $\|\tilde{U}'_1\tilde{U}_1/n - I_{p+1-q}\|_S \rightarrow 0$ in probability, since $k_n/n \rightarrow 0$, and it also establishes the boundedness in probability of $\|\tilde{U}'_1\tilde{U}_1/n\|_S\|\tilde{U}'_0\tilde{U}_0/n\|_S$. The assumptions of this lemma are clearly satisfied under (A1).(a,c,d). Now the convergence in spectral norm implies the convergence of the extreme eigenvalues of $\tilde{U}'_1\tilde{U}_1/n$ to 1, and thus, also the extreme eigenvalues of the inverse converge to 1, which means that $\|(\tilde{U}'_1\tilde{U}_1/n)^{-1} - I_{p+1-q}\|_S \rightarrow 0$, in probability. Since $b_n\sqrt{n}\delta'_\gamma\Omega_{00}\delta_\gamma = O(1)$, by the additional assumption of Theorem 2.1(ii), the upper bound in (6.14) converges to zero in probability. Thus, we have shown that we can replace U'_1U_1/n in (6.13) by Ω_{11} , without changing the limit.

To finish part (ii) it remains to show that $b_n B_n$ converges to zero in probability, where

$$B_n := \sqrt{n} \delta'_\gamma \left[\frac{U'_0 U_1}{n} \Omega_{11}^{-1} \frac{U'_1 U_0}{n} - \Omega_{01} \Omega_{11}^{-1} \Omega_{10} \right] \delta_\gamma.$$

To evaluate its expectation, write

$$B_n = \frac{1}{n^2} \sum_{i,j=1}^n \sqrt{n} \delta'_\gamma [R_0(1, x'_i)'(1, x'_i)R'_1 \Omega_{11}^{-1} R_1(1, x'_j)'(1, x'_j)R'_0 - \Omega_{01} \Omega_{11}^{-1} \Omega_{10}] \delta_\gamma.$$

Since $\mathbb{E}[R_0(1, x'_i)'(1, x'_i)R'_1] = R_0 S R'_1 = \Omega_{01}$, all summands in $\mathbb{E}[B_n]$ with distinct indices $i \neq j$ disappear. Using parts (i) and (iv) of Lemma C.3, which apply in view of Assumptions (A1).(a,d), we arrive at

$$\begin{aligned} |\mathbb{E}[B_n]| &= \left| \frac{1}{n} \sqrt{n} \delta'_\gamma \mathbb{E}[R_0(1, x'_1)'(1, x'_1)R'_1 \Omega_{11}^{-1} R_1(1, x'_1)'(1, x'_1)R'_0] \delta_\gamma \right. \\ &\quad \left. - \frac{1}{n} \sqrt{n} \delta'_\gamma \Omega_{01} \Omega_{11}^{-1} \Omega_{10} \delta_\gamma \right| \\ &\leq n^{-1/2} \sqrt{\mathbb{E}[|\delta'_\gamma R_0(1, x'_1)'|^4] \mathbb{E}[|(1, x'_1)R'_1 \Omega_{11}^{-1} R_1(1, x'_1)'|^2]} \\ &\quad + n^{-1/2} \delta'_\gamma \Omega_{01} \Omega_{11}^{-1} \Omega_{10} \delta_\gamma \\ &= n^{-1/2} \sqrt{O(|\delta'_\gamma R_0 S R'_0 \delta_\gamma|^2) O((p+1-q)^2)} + n^{-1/2} \delta'_\gamma \Omega_{01} \Omega_{11}^{-1} \Omega_{10} \delta_\gamma \\ &= O(\sqrt{n} \delta'_\gamma \Omega_{00} \delta_\gamma (p+1-q)/n) + n^{-1/2} \delta'_\gamma \Omega_{01} \Omega_{11}^{-1} \Omega_{10} \delta_\gamma. \end{aligned}$$

After multiplying by b_n , the upper bound still converges to zero, because, first, $b_n \sqrt{n} \delta'_\gamma \Omega_{00} \delta_\gamma = O(1)$ and $(p+1-q)/n \rightarrow 0$, by the additional assumptions of Theorem 2.1(ii), and because $b_n n^{-1/2} \delta'_\gamma \Omega_{01} \Omega_{11}^{-1} \Omega_{10} \delta_\gamma \leq b_n n^{-1/2} \delta'_\gamma \Omega_{00} \delta_\gamma = O(n^{-1})$, where the inequality holds in view of the second case in (6.11).

In order to show that the distribution of B_n also concentrates around its mean, we make use of the Efron-Stein inequality.¹¹ We use the abbreviations $D = \delta'_\gamma \Omega_{01} \Omega_{11}^{-1} \Omega_{10} \delta_\gamma$, $L_i = \delta'_\gamma R_0(1, x'_i)'$, $Q_{ij} = (1, x'_i)R'_1 \Omega_{11}^{-1} R_1(1, x'_j)'$ and define the functions $g : \mathbb{R}^{p \times n} \rightarrow \mathbb{R}$ and $g_k : \mathbb{R}^{p \times (n-1)} \rightarrow \mathbb{R}$, for $k = 1, \dots, n$, by $g(x_1, \dots, x_n) = n^{-3/2} \sum_{i,j=1}^n L_i Q_{ij} L_j - \sqrt{n} D = B_n$ and

$$g_k(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n) = n^{-3/2} \sum_{\substack{i,j=1 \\ i \neq k, j \neq k}}^n L_i Q_{ij} L_j - \sqrt{n} D.$$

By the Efron-Stein inequality [18, Theorem 9],

$$\text{Var}[B_n] \leq \sum_{k=1}^n \mathbb{E}[(g(x_1, \dots, x_n) - g_k(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n))^2]. \quad (6.15)$$

¹¹An explicit argument for the concentration aspect of B_n is missing in WC.

Now, for $k \in \{1, \dots, n\}$, $g(x_1, \dots, x_n)$ can be expressed as

$$n^{-3/2} \left[\sum_{\substack{i,j=1 \\ i \neq k, j \neq k}}^n L_i Q_{ij} L_j + \sum_{\substack{i=1 \\ i \neq k}}^n L_i Q_{ik} L_k + \sum_{\substack{i=1 \\ i \neq k}}^n L_k Q_{ki} L_i + L_k Q_{kk} L_k \right] - \sqrt{n} D.$$

Using the fact that $Q_{ij} = Q'_{ij} = Q_{ji}$, the differences $g - g_k$ in (6.15), are equal to

$$n^{-3/2} \left[2 \sum_{\substack{i=1 \\ i \neq k}}^n L_i Q_{ik} L_k + L_k Q_{kk} L_k \right]. \tag{6.16}$$

We need to bound the expectation of the squared expression. To this end, we calculate the expectation of $L_i Q_{ik} L_k L_j Q_{jk} L_k$ for arbitrary indices i, j, k , as well as in the special case where $i \neq k, j \neq k$ and $i \neq j$. Observe that Q_{ik} is the inner product of $\Omega_{11}^{-1/2} R_1(1, x'_i)'$ and $\Omega_{11}^{-1/2} R_1(1, x'_k)'$ and therefore, by Cauchy-Schwarz inequalities in both Euclidean and L_p space, satisfies $\mathbb{E}[|Q_{ik}|^4] \leq \mathbb{E}[\|\Omega_{11}^{-1/2} R_1(1, x'_i)'\|^4] \|\Omega_{11}^{-1/2} R_1(1, x'_k)'\|^4] \leq \mathbb{E}[\|\Omega_{11}^{-1/2} R_1(1, x'_i)'\|^8] = \mathbb{E}[Q_{11}^4]$. Moreover, parts (i) and (v) of Lemma C.3, whose assumptions are implied by the conditions (A1).(a,d,e), establish the facts $\mathbb{E}[L_1^4] = O(|\delta'_\gamma \Omega_{00} \delta_\gamma|^2)$, $\mathbb{E}[L_1^8] = O(|\delta'_\gamma \Omega_{00} \delta_\gamma|^4)$ and $\mathbb{E}[Q_{11}^4] = O(|p + 1 - q|^4)$, provided that at least one of the assumptions (a) or (b) of Lemma C.3(v) holds. But this follows from Lemma C.4, because if $t_0 = 0$, then from the representations of ∇_n and Δ_γ in (6.10) and (6.11), we see that the distribution of the quantity of interest does not depend on μ and we may restrict to $\mu = 0$, whereas, if $t_0 \neq 0$, Lemma C.4 shows that T_1 has full rank.¹² With this, in general, we obtain

$$\begin{aligned} |\mathbb{E}[L_i Q_{ik} L_k L_j Q_{jk} L_k]| &= |\mathbb{E}[L_i L_j L_k^2 Q_{ik} Q_{jk}]| \leq \sqrt{\mathbb{E}[L_i^2 L_j^2 L_k^4] \mathbb{E}[Q_{ik}^2 Q_{jk}^2]} \\ &\leq (\mathbb{E}[L_1^8])^{1/2} (\mathbb{E}[Q_{ik}^4])^{1/4} (\mathbb{E}[Q_{jk}^4])^{1/4} \leq (\mathbb{E}[L_1^8])^{1/2} (\mathbb{E}[Q_{11}^4])^{1/2} \\ &= O(|\delta'_\gamma \Omega_{00} \delta_\gamma|^2) O(|p + 1 - q|^2) = O(q|p + 1 - q|^2/n^2). \end{aligned}$$

If $i \neq k, j \neq k$ and $i \neq j$, using the abbreviations $v = R'_0 \delta_\gamma$ and $M = R'_1 \Omega_{11}^{-1} R_1$, we get the smaller bound

$$\begin{aligned} |\mathbb{E}[L_i Q_{ik} L_k L_j Q_{jk} L_k]| &= |\mathbb{E}[\mathbb{E}[L_i Q_{ik} | x_k, x_j] L_j Q_{jk} L_k^2]| \\ &= |\mathbb{E}[v' S M(1, x'_k)' \mathbb{E}[L_j Q_{jk} | x_k] L_k^2]| = |\mathbb{E}[(v' S M(1, x'_k)')^2 L_k^2]| \\ &\leq \sqrt{\mathbb{E}[(v' S M(1, x'_k)')^4] \mathbb{E}[L_k^4]} = \sqrt{O(|v' S M S M S v|^2) O(|\delta'_\gamma \Omega_{00} \delta_\gamma|^2)} \\ &= O(\delta'_\gamma \Omega_{01} \Omega_{11}^{-1} \Omega_{10} \delta_\gamma) O(\delta'_\gamma \Omega_{00} \delta_\gamma) \leq O(|\delta'_\gamma \Omega_{00} \delta_\gamma|^2) = O(q/n^2), \end{aligned}$$

¹²It should be noted that the case $t_0 = 0$ corresponds to a null-hypothesis that does not involve a restriction on the intercept parameter α , i.e., $H_0 : R_0 \gamma = r_0$ can be expressed as $H_0 : T_0 \beta = r_0$, in this case.

where we have used Lemma C.3 and $\delta'_\gamma \Omega_{01} \Omega_{11}^{-1} \Omega_{10} \delta_\gamma \leq \delta'_\gamma \Omega_{00} \delta_\gamma$ again. It is now easy to bound the expectations in (6.15). When squaring the expression in (6.16) we first note the leading factor n^{-3} . Next, we expand the square of the bracket term in (6.16) and take expectation. From the previous considerations we see that those summands in the resulting sum involving $L_k Q_{kk} L_k$ are of order $O(q|p+1-q|^2/n^2)$, and there are $O(n)$ of them. Together with the leading factor n^{-3} and the summation in (6.15) we arrive at a total contribution of $O(q|p+1-q|^2/n^3)$ from all those summands involving $L_k Q_{kk} L_k$. This expression has to be multiplied by $b_n^2 = O(n/q)$ to yield $O(|p+1-q|^2/n^2) = o(1)$. The remaining terms are of the form $|\mathbb{E}[L_i Q_{ik} L_k L_j Q_{jk} L_k]|$ with $i \neq k$ and $j \neq k$. Of those, there are a number of $O(n)$ summands where $i = j$, but they are again of order $O(q|p+1-q|^2/n^2)$ and therefore, as in the case before, their total contribution to (6.15) is asymptotically negligible, even after multiplying by b_n^2 . Finally, there is a number of $O(n^2)$ remaining summands as above, but with $i \neq k, j \neq k$ and $i \neq j$. Therefore, by the refined bound above, they are of order $O(q/n^2)$, so that their total contribution to the variance bound in (6.15) is $O(q/n^2)$. Together with the factor b_n^2 we arrive at an additional term of order $O(1/n) = o(1)$. Hence, we see that the variance of $b_n B_n$ goes to zero as $n \rightarrow \infty$ and the proof of Theorem 2.1(ii) is finished.

6.4.3. The cases of Theorem 2.1(iii) and Theorem 3.1

In this section we establish (6.7) under the assumptions that the design vectors x_1, \dots, x_n are Gaussian, $p_n/n \rightarrow \rho_1 \in [0, 1)$, $q_n/n \rightarrow \rho_2 \in [0, \rho_1]$ (cf. the beginning of Section 6) and $\Delta_\gamma = o(q_n/n)$.

We restrict to $q \leq p$, because this is assumed in Theorem 3.1 and the case $q = p + 1$ of Theorem 2.1(iii) has already been treated above in Section 6.4.1, in much higher generality. We use different arguments for the two cases $t_0 = 0$ and $t_0 \neq 0$. If $t_0 = 0$, then, by (6.10) and (6.11), we see that the quantity of interest is given by

$$s_n^{-1/2} n \nabla_n / q - \sqrt{n} \Delta_\gamma b_n = \sqrt{\frac{n}{q}} (s_n q)^{-1/2} \sqrt{n} \left(\delta'_\gamma (T_0 \hat{\Sigma}_n^{-1} T'_0)^{-1} \delta_\gamma - \delta'_\gamma (T_0 \Sigma^{-1} T'_0)^{-1} \delta_\gamma \frac{n - (p + 1) + q}{n} \right).$$

By Lemma C.5, this expression has mean zero and variance equal to

$$\frac{n}{s_n q^2} (\delta'_\gamma (T_0 \Sigma^{-1} T'_0)^{-1} \delta_\gamma)^2 \frac{2(n - (p + 1) + q)}{n} = \frac{n}{q} \Delta_\gamma^2 O(1) = o(1).$$

For the case $t_0 \neq 0$ we recall $T = (R'_0, R'_1)'$ and introduce the matrix Σ_T by

$$\Sigma_T = \text{Var}[T(1, x'_1)'] = T \begin{pmatrix} 0 & 0 \\ 0 & \Sigma \end{pmatrix} T' = \begin{pmatrix} T_0 \Sigma T'_0 & T_0 \Sigma T'_1 \\ T_1 \Sigma T'_0 & T_1 \Sigma T'_1 \end{pmatrix} = \begin{pmatrix} \Sigma_{00} & \Sigma_{01} \\ \Sigma_{10} & \Sigma_{11} \end{pmatrix},$$

and note that by Lemma C.4 the sub matrix $\Sigma_{11} = \text{Var}[R_1(1, x'_1)']$ of order $(p + 1 - q)$ is regular. In the present case the difference in (6.7) is given by

$$s_n^{-1/2} n \nabla_n / q - \sqrt{n} \Delta_\gamma b_n$$

$$\begin{aligned}
 &= (s_n q)^{-1/2} \sqrt{\frac{n}{q}} \times \\
 &\quad \sqrt{n} \delta'_\gamma \left(\frac{U'_0(I_n - P_{U_1})U_0}{n} - (\Omega_{00} - \Omega_{01}\Omega_{11}^{-1}\Omega_{10}) \frac{n - (p + 1) + q}{n} \right) \delta_\gamma.
 \end{aligned} \tag{6.17}$$

The distribution of this quantity is slightly more complicated than that of the corresponding object in the case $t_0 = 0$, because now we have to deal with non-centrality issues due to $\mu \neq 0$. We take a closer look at the random part. The joint distribution of the first row of U_0 and the first row of U_1 is $T(1, x'_1)' \sim \mathcal{N}(T(1, \mu')', \Sigma_T)$. Therefore, the conditional distribution of $R_0(1, x'_1)'$ given $R_1(1, x'_1)'$ is given by $\mathcal{N}(\tilde{\mu} + \Sigma_{01}\Sigma_{11}^{-1}R_1(1, x'_1)', \Sigma_{00.1})$, where $\tilde{\mu} = R_0(1, \mu')' - \Sigma_{01}\Sigma_{11}^{-1}R_1(1, \mu')'$ and $\Sigma_{00.1} = \Sigma_{00} - \Sigma_{01}\Sigma_{11}^{-1}\Sigma_{10}$. Hence, conditional on U_1 , the rows of $U_{0.1} := U_0 - U_1\Sigma_{11}^{-1}\Sigma_{10}$ are i.i.d. $\mathcal{N}(\tilde{\mu}, \Sigma_{00.1})$ and since this distribution is free of U_1 , this also means that U_1 and $U_{0.1}$ are independent. Also, we clearly have $U'_0(I_n - P_{U_1})U_0 = U'_{0.1}(I_n - P_{U_1})U_{0.1}$. So if V is a random $n \times q$ matrix independent of U_1 , whose rows are i.i.d. $\mathcal{N}(0, \Sigma_{00.1})$, then $U'_0(I_n - P_{U_1})U_0$ has the same distribution as $(V + i\tilde{\mu}')'(I_n - P_{U_1})(V + i\tilde{\mu}') = V'(I_n - P_{U_1})V + \tilde{\mu}'(I_n - P_{U_1})V + V'(I_n - P_{U_1})i\tilde{\mu}' + \tilde{\mu}\tilde{\mu}'i'(I_n - P_{U_1})i$. For the Schur complement of Ω_{11} in Ω we use the representation given by Lemma C.6. Plugging this back into (6.17) and removing the leading $(s_n q)^{-1/2}$ term that converges to a positive constant, it remains to study the limiting behavior of

$$\sqrt{n} \delta'_\gamma \left(\frac{V'(I_n - P_{U_1})V}{n} - \Sigma_{00.1} \frac{n - (p + 1) + q}{n} \right) \delta_\gamma \tag{6.18}$$

$$+ 2\sqrt{n} \delta'_\gamma \tilde{\mu}' \frac{i'(I_n - P_{U_1})V \delta_\gamma}{n} \tag{6.19}$$

$$+ \frac{(\delta'_\gamma \tilde{\mu})^2}{1 + \nu} \sqrt{n} \left(\frac{i'(I_n - P_{U_1})i(1 + \nu)}{n} \frac{n}{n - (p + 1) + q} - 1 \right) \frac{n - (p + 1) + q}{n}, \tag{6.20}$$

multiplied by $\sqrt{n/q}$, where $\nu = (1, \mu')R'_1\Sigma_{11}^{-1}R_1(1, \mu)'$ is defined as in Lemma C.6. From that lemma we also see that $\delta'_\gamma \Sigma_{00.1} \delta_\gamma + (\delta'_\gamma \tilde{\mu})^2 / (1 + \nu) = \delta'_\gamma \Omega_{00.1} \delta_\gamma = \Delta_\gamma = o(q/n)$. Since $\Sigma_{00.1}$ is the Schur complement of the positive definite matrix Σ_{11} within the positive semidefinite matrix Σ_T , it follows that $\Sigma_{00.1}$ is itself positive semidefinite (consider the minimizer of the quadratic form $u \mapsto (v', u')' \Sigma_T (v', u)'$ in those variables $u \in \mathbb{R}^{p+1-q}$ corresponding to the block Σ_{11}). Consequently, $\delta'_\gamma \Sigma_{00.1} \delta_\gamma \geq 0$ and both $\delta'_\gamma \Sigma_{00.1} \delta_\gamma$ and $(\delta'_\gamma \tilde{\mu})^2 / (1 + \nu)$ are bounded by $\delta'_\gamma \Omega_{00.1} \delta_\gamma = \Delta_\gamma = o(q/n)$. Now, we first show that the quantity in (6.20) converges to zero in probability. By Lemma C.7 with $\lambda_n = n\nu$ and $k = p + 1 - q$, we have

$$\begin{aligned}
 &\sqrt{n} \left(\frac{i'(I_n - P_{U_1})i(1 + \nu)}{n} \frac{n}{n - (p + 1) + q} - 1 \right) \\
 &= \sqrt{n} \left(\frac{\xi / (n - (p + 1) + q)}{(\xi + \zeta) / (n(1 + \nu))} - 1 \right)
 \end{aligned}$$

$$= \frac{\sqrt{n} (\xi/(n - (p + 1) + q) - 1) + \sqrt{n} (1 - (\xi + \zeta)/(n(1 + \nu)))}{(\xi + \zeta)/(n(1 + \nu))},$$

where $\xi \sim \chi_{n-(p+1)+q}^2$ independent of $\zeta \sim \chi_{p+1-q}^2(n\nu)$. The term in the denominator has mean $(n + n\nu)/(n(1 + \nu)) = 1$ and variance $2(n + 2n\nu)/(n^2(1 + \nu)^2) = O(1/n)$, by independence, and thus, converges to 1 in probability. From the form of these moments we also conclude that the term $\sqrt{n} (1 - (\xi + \zeta)/(n(1 + \nu)))$ is $O_{\mathbb{P}}(1)$. Moreover, it is easy to see that also $\sqrt{n} (\xi/(n - (p + 1) + q) - 1)$ is $O_{\mathbb{P}}(1)$, which entails that the entire expression in the previous display is of order $O_{\mathbb{P}}(1)$, and hence, the expression in (6.20) converges to zero in probability even after multiplying by $\sqrt{n/q}$. Next, the mixed term in (6.19) is easily seen to have conditional distribution $\mathcal{N}(0, 4\delta'_{\gamma}\Sigma_{00.1}\delta_{\gamma}l'(I_n - P_{U_1})\iota(\delta'_{\gamma}\tilde{\mu})^2/n)$ given U_1 . By the previous considerations, the conditional variance is equal to

$$4\delta'_{\gamma}\Sigma_{00.1}\delta_{\gamma}l'(I_n - P_{U_1})\iota(\delta'_{\gamma}\tilde{\mu})^2/n = O(\Delta_{\gamma}^2)\frac{l'(I_n - P_{U_1})\iota(1 + \nu)}{n} = O_{\mathbb{P}}(\Delta_{\gamma}^2),$$

which converges to zero in probability, even after multiplying by n/q , and this implies convergence to zero of the scaled mixed term itself. Finally, for the expression in (6.18), we note that conditional on U_1 , $V'(I_n - P_{U_1})V$ has a Wishart distribution with scale matrix $\Sigma_{00.1}$ and $n - (p + 1) + q$ degrees of freedom. Therefore, (6.18) has mean zero and $\delta'_{\gamma}V'(I_n - P_{U_1})V\delta_{\gamma} \sim \delta'_{\gamma}\Sigma_{00.1}\delta_{\gamma}\chi_{n-(p+1)+q}^2$ [cf. 19, Theorem 3.4.2], which entails that the variance of (6.18), multiplied by n/q , is $(n/q)(\delta'_{\gamma}\Sigma_{00.1}\delta_{\gamma})^2 2(n - (p + 1) + q)/n = o(q/n)$, since $\delta'_{\gamma}\Sigma_{00.1}\delta_{\gamma} \leq \Delta_{\gamma} = o(q/n)$. This finishes the proofs of Theorem 2.1 and Theorem 3.1. \square

Appendix A: Auxiliary results of Section 4

Lemma A.1. For $m \in \mathbb{N}$, let $Z = (Z_1, \dots, Z_m)'$ be a spherically symmetric random vector in \mathbb{R}^m such that $\mathbb{E}[ZZ'] = I_m$, let $V \sim \mathcal{N}(0, I_m)$ and let z_1, \dots, z_n be i.i.d. copies of Z . For $p \leq m$, let Γ be a $p \times m$ matrix of full rank p , let $\mu \in \mathbb{R}^p$ and define $x_i = \Gamma z_i + \mu$.

- (i) Fix $r \in \mathbb{N}$. If $\mathbb{E}[\|Z\|^{2r}] < \infty$ and for every choice of non-negative integers ℓ_j with $\sum_{j=1}^m \ell_j \leq 2r$, we have $\mathbb{E}[\prod_{j=1}^m Z_j^{\ell_j}] = \prod_{j=1}^m \mathbb{E}[Z_j^{\ell_j}]$, then $\mathbb{E}[Z_1^{2l}] = \mathbb{E}[V_1^{2l}]$ and $\mathbb{E}[\|Z\|^{2l}] = \mathbb{E}[\|V\|^{2l}]$ for every $l = 1, \dots, r$.¹³
- (ii) If $2 \leq p \leq n - 2$ and Z also satisfies $\mathbb{P}(\|Z\| = 0) = 0$, then the random vectors x_1, \dots, x_n satisfy Assumptions (A1).(a,b). Moreover, if also $\mathbb{E}[\|Z\|^8] < \infty$, $\mathbb{E}[\|Z\|^4]/\mathbb{E}[\|V\|^4] \rightarrow 1$ and $\mathbb{E}[\|Z\|^8]/\mathbb{E}[\|V\|^8] = O(1)$, as $m \rightarrow \infty$, then also Assumptions (A1).(d,e) hold.
- (iii) If Z follows the uniform distribution on the ball (of appropriate radius $\sqrt{m + 2}$, to ensure $\mathbb{E}[ZZ'] = I_m$) and $2 \leq p \leq n - 2$, then the x_i , for $i = 1, \dots, n$, satisfy the full Assumption (A1), but not Assumption (C1).

¹³Due to symmetry, we always have $\mathbb{E}[Z_1^l] = 0 = \mathbb{E}[V_1^l]$ if l is odd and the former moment exists.

Proof. We make use of the well known fact that any spherical distribution can be represented as $Z = b\|Z\|$, where b and $\|Z\|$ are independent, and b is uniformly distributed on the unit m -sphere \mathcal{S}^{m-1} [cf. 8].

For part (i), set $\ell = \sum_{j=1}^m \ell_j$ and let $e_i \in \mathbb{R}^m$ denote the i -th element of the standard basis in \mathbb{R}^m and note that

$$\begin{aligned} \mathbb{E} \left[\prod_{j=1}^m Z_j^{\ell_j} \right] &= \mathbb{E} \left[\prod_{j=1}^m (e'_j Z)^{\ell_j} \right] = \mathbb{E} \left[\|Z\|^\ell \prod_{j=1}^m (e'_j b)^{\ell_j} \right] \\ &= \mathbb{E} [\|Z\|^\ell] \mathbb{E} \left[\prod_{j=1}^m (e'_j b)^{\ell_j} \right]. \end{aligned}$$

Of course, the same argument can be carried through for the spherical vector $V \sim \mathcal{N}(0, I_m)$, so that we have

$$\frac{\mathbb{E} \left[\prod_{j=1}^m Z_j^{\ell_j} \right]}{\mathbb{E} \left[\prod_{j=1}^m V_j^{\ell_j} \right]} = \frac{\mathbb{E} [\|Z\|^\ell]}{\mathbb{E} [\|V\|^\ell]}, \tag{A.1}$$

provided that all the ℓ_j are even, so that $\mathbb{E}[\prod_{j=1}^m V_j^{\ell_j}] \neq 0$. Now choose the ℓ_j to be either equal to 2 or 0, such that ℓ is any even number from 2 to $2r$. Therefore, since $\mathbb{E}[Z_1^2] = 1 = \mathbb{E}[V_1^2]$ and by our factorization assumption, the left-hand-side of (A.1) is equal to one, so that we have established the equality of even moments of $\|Z\|$ and $\|V\|$. To see that also the even moments of Z_1 and V_1 coincide, simply choose $\ell_1 = \ell = 2l$, for some $l \in \{1, \dots, k\}$ and $\ell_j = 0$, if $j \neq 1$.

For part (ii), to establish Assumption (A1).(b), first note that the column span of $U_{-1} = [\iota, [z_2, \dots, z_n]' \Gamma' + \iota \mu']$ does not depend on $\mu \in \mathbb{R}^p$, where $\iota = (1, \dots, 1)' \in \mathbb{R}^{n-1}$. So we may assume without restriction that $\mu = 0$. Moreover, $[\iota, [z_2, \dots, z_n]' \Gamma']$ and

$$[\iota, [z_2, \dots, z_n]' \Gamma'] \begin{bmatrix} 1 & 0 \\ 0 & \Sigma^{-1/2} \end{bmatrix}$$

also have the same column span, such that it suffices to determine the rank of $[\iota, [\tilde{x}_2, \dots, \tilde{x}_n]']$, where $\tilde{x}_i = (\Gamma \Gamma')^{-1/2} \Gamma z_i$ is spherically symmetric with $\mathbb{P}(\|\tilde{x}_i\| = 0) = 0$. Next, we claim that the matrix $M_k = [\tilde{x}_2, \dots, \tilde{x}_{k+1}]'$ has full rank p , almost surely, provided that $k \geq p$. To see this, simply write $M_k = D_1 D_2^{-1} \Lambda$, almost surely, where D_1 is $k \times k$ diagonal with entries $\|\tilde{x}_2\|, \dots, \|\tilde{x}_{k+1}\|$, D_2 is $k \times k$ diagonal with i.i.d. χ_p entries and Λ is $k \times p$ and has i.i.d. $\mathcal{N}(0, 1)$ entries. Now it is easy to see that $D_1 D_2^{-1}$ is almost surely of full rank, and thus, $\mathbb{P}(\text{rank}(M_k) = p) = \mathbb{P}(\det(\Lambda' \Lambda) \neq 0) = 1$, where the last equality follows from the well known fact that the zero-set of a non-constant polynomial is a Lebesgue null-set. It remains to show that the event $A = \{\iota \in \text{span}(M_{n-1})\}$ has probability zero. Define $B = \{\text{rank}(M_p) = p\}$, $\iota_k = (1, \dots, 1)' \in \mathbb{R}^k$ and the function $v : \mathbb{R}^{p \times p} \rightarrow \mathbb{R}^p$ by $v(M) = M^{-1} \iota_p$, if $\det(M) \neq 0$, and $v(M) = 0$, else.

Note that $\mathbb{P}(B) = 1$ and v is Borel measurable. Since $p + 2 \leq n$, we see that $A \cap B$ is a subset of the event where both $\iota_p = M_p v(M_p)$ and $1 = \tilde{x}'_{p+2} v(M_p)$, the probability of which is clearly bounded by

$$\begin{aligned} \mathbb{P}(1 = \tilde{x}'_{p+2} v(M_p)) &= \mathbb{P}(\|\tilde{x}_{p+2}\|^{-1} = (\tilde{x}'_{p+2}/\|\tilde{x}_{p+2}\|)v(M_p), \|\tilde{x}_{p+2}\| \neq 0) \\ &= \mathbb{E}[\mathbb{P}(\|\tilde{x}_{p+2}\|^{-1} = (\tilde{x}'_{p+2}/\|\tilde{x}_{p+2}\|)v(M_p), \|\tilde{x}_{p+2}\| \neq 0 \mid M_p, \|\tilde{x}_{p+2}\|)]. \end{aligned}$$

But the conditional probability in the previous display is equal to zero, almost surely, because $v(M_p)$, $\tilde{x}_{p+2}/\|\tilde{x}_{p+2}\|$ and $\|\tilde{x}_{p+2}\|$ are independent and $\tilde{x}_{p+2}/\|\tilde{x}_{p+2}\|$ is uniformly distributed on the unit sphere in \mathbb{R}^p , and therefore its inner product with any fixed vector has a Lebesgue density on \mathbb{R} provided that $p \geq 2$. For Assumptions (A1).(d,e), recall the moments of the χ^2 -distribution with m -degrees of freedom $\mathbb{E}[\|V\|^{2k}] = \prod_{j=0}^{k-1} (m+2j)$ [cf. 15]. The same reasoning as in part (i) yields

$$\mathbb{E}[|v'Z|^8] = \mathbb{E}[|v'b|^8] \mathbb{E}[\|Z\|^8] = \frac{\mathbb{E}[\|Z\|^8]}{\mathbb{E}[\|V\|^8]} \mathbb{E}[|v'V|^8] = O(1),$$

uniformly in $v \in \mathcal{S}^{m-1}$. Similarly, for a symmetric matrix $M \in \mathbb{R}^{m \times m}$,

$$\mathbb{E}[(Z'MZ)^2] = \mathbb{E}[(b'Mb)^2] \mathbb{E}[\|Z\|^4] = \frac{\mathbb{E}[\|Z\|^4]}{\mathbb{E}[\|V\|^4]} \mathbb{E}[(V'MV)^2],$$

and one easily calculates $\mathbb{E}[(V'MV)^2] = (\text{trace}(M))^2 + 2\text{trace}(M^2)$. Therefore,

$$\text{Var}[Z'MZ] = (\text{trace}(M))^2 \left(\frac{\mathbb{E}[\|Z\|^4]}{\mathbb{E}[\|V\|^4]} - 1 \right) + 2 \frac{\mathbb{E}[\|Z\|^4]}{\mathbb{E}[\|V\|^4]} \text{trace}(M^2).$$

Finally, for a projection matrix $P \in \mathbb{R}^{m \times m}$, $V'PV$ follows a χ^2 -distribution with rank $P = \|P\|_F^2$ degrees of freedom, and thus

$$\begin{aligned} (\mathbb{E}[(Z'PZ)^4])^{1/4} &= \left(\frac{\mathbb{E}[\|Z\|^8]}{\mathbb{E}[\|V\|^8]} \right)^{1/4} (\mathbb{E}[(V'PV)^4])^{1/4} \\ &= \left(\frac{\mathbb{E}[\|Z\|^8]}{\mathbb{E}[\|V\|^8]} \right)^{1/4} \left(\prod_{j=0}^3 (\|P\|_F^2 + 2j) \right)^{1/4} = O(\|P\|_F^2). \end{aligned}$$

To establish part (iii), we first verify the conditions of part (ii). The finiteness of the 8-th moment of the radial component and $\mathbb{P}(\|Z\| = 0) = 0$ are immediate. It is also elementary to calculate the higher non-central moments $\mathbb{E}[\|Z\|^{2k}] = (m+2)^k m / (m+2k)$. [Use, for example, the formula for the volume of the m -ball of radius $r > 0$ to obtain $\mathbb{P}(\|Z\| \leq x) = (x/\sqrt{m+2})^m$, for $x \in [0, \sqrt{m+2}]$.] Comparing this to the moments of the χ_m^2 distribution $\mathbb{E}[\|V\|^{2k}] = \prod_{j=0}^{k-1} (m+2j)$ for $k = 2, 4$, we see that for $m \rightarrow \infty$ the moment ratios behave as desired. Therefore, Assumptions (A1).(a,b,d,e) hold in this case. Finally, the validity of Assumption (A1).(c) follows from Srivastava and Vershynin [23, Section 1.4]. But Condition (C1) can not be satisfied in view of part (i) (with $r = 4$) and the fact that $\mathbb{E}[\|Z\|^4] \neq \mathbb{E}[\|V\|^4]$. \square

Lemma A.2. Let $Z = (Z_1, \dots, Z_m)'$ be a random m -vector with $\mathbb{E}[Z] = 0$ and $\mathbb{E}[ZZ'] = I_m$ and let z_1, \dots, z_n be i.i.d. copies of Z . For $p \leq m$, let Γ be a $p \times m$ matrix of full rank p , let $\mu \in \mathbb{R}^p$ and define $x_i = \Gamma z_i + \mu$.

- (i) If Z has independent components, whose 8-th moments are uniformly bounded, then the x_i satisfy Assumptions (A1).(a,c,d,e).
- (ii) If z_1, \dots, z_n are as in Condition (C1) and, in addition, the components of Z have 8-th moments that are uniformly bounded, then the x_i satisfy Assumptions (A1).(a,d,e).

Proof. To establish part (i), we use the results of Whittle [28]. Theorem 2 in that reference shows that for a unit vector $v = (v_1, \dots, v_m)' \in \mathbb{R}^m$,

$$\mathbb{E}[|v'Z|^8] \leq C \left(\sum_j v_j^2 (\mathbb{E}[|Z_j|^8])^{1/4} \right)^4 \leq C \max_j \mathbb{E}[|Z_j|^8],$$

for some numerical constant $C > 0$, and thus, $\mathcal{L}_8 = O(1)$ as $m \rightarrow \infty$, in view of uniform boundedness of $\mathbb{E}[|Z_j|^8]$. Next, for a symmetric matrix $M \in \mathbb{R}^{m \times m}$, the same theorem yields

$$\text{Var}[Z'MZ] \leq C \sum_{j,k} M_{jk}^2 \sqrt{\mathbb{E}[|Z_j|^4] \mathbb{E}[|Z_k|^4]} \leq C \max_j \mathbb{E}[|Z_j|^4] \text{trace}(M^2),$$

and, for a projection matrix $P \in \mathbb{R}^{m \times m}$,

$$\begin{aligned} \left| (\mathbb{E}[(Z'PZ)^4])^{1/4} - \mathbb{E}[Z'PZ] \right| &\leq (\mathbb{E}[(Z'PZ - \mathbb{E}[Z'PZ])^4])^{1/4} \\ &\leq C^{1/4} \left(\sum_{j,k} P_{jk}^2 (\mathbb{E}[|Z_j|^8] \mathbb{E}[|Z_k|^8])^{1/4} \right)^{1/2} \\ &\leq \left(C \max_j \mathbb{E}[|Z_j|^8] \right)^{1/4} \|P\|_F, \end{aligned}$$

where the first inequality is the reverse triangle inequality for the L^4 -norm. Now the previous chain of inequalities implies that

$$(\mathbb{E}[(Z'PZ)^4])^{1/4} \leq \mathbb{E}[Z'PZ] + D\|P\|_F \leq \|P\|_F^2 + D\|P\|_F^2 = O(\|P\|_F^2),$$

since $\mathbb{E}[Z'PZ] = \text{trace}(P) = \text{trace}(P^2) = \|P\|_F^2 = \text{rank } P$ is integer, and where $D > 0$ is an appropriate constant, not depending on m . The validity of Assumption (A1).(c) follows from the arguments in Section 1.4 in Srivastava and Vershynin [23].

For part (ii), simply note that under the factorization assumption in (C1) all the moments occurring in Conditions (A1).(d,e) are identical to those calculated under independence of the components of Z . Therefore, the result follows from part (i). □

Appendix B: Proofs of auxiliary results of Section 6.2

Proof of Lemma 6.1. For ease of notation we drop the subscript n that indexes the position of the matrix A_n in the array, i.e., we write $A = A_n$ and denote by a_{ij} the ij -th entry of that matrix. Similarly, we write $Z = (Z_1, \dots, Z_n)'$, where $Z_i = Z_{i,n}$. Now, expand

$$\begin{aligned} Z'AZ - \mathbb{E}[Z'AZ|\mathcal{G}_n] &= \sum_{i \neq j}^n Z_i Z_j a_{ij} + \sum_{j=1}^n a_{jj}(Z_j^2 - 1) \\ &= \sum_{j=1}^n 2Z_j \sum_{i=1}^{j-1} Z_i a_{ij} + \sum_{j=1}^n a_{jj}(Z_j^2 - 1) \\ &= \bar{T}_n + T_n^*, \end{aligned}$$

where we adopt the convention that empty sums are equal to zero. We show that $\bar{T}_n/(\sqrt{2}\|A\|_F) \xrightarrow{w} \mathcal{N}(0, 1)$ and $T_n^*/(\sqrt{2}\|A\|_F) \xrightarrow{i.p.} 0$, as $n \rightarrow \infty$.

The desired convergence of T_n^* follows from the straight forward calculation

$$\begin{aligned} \mathbb{E} \left[\left(\frac{T_n^*}{\sqrt{2}\|A\|_F} \right)^2 \middle| \mathcal{G}_n \right] &= \sum_{j=1}^n \frac{a_{jj}^2 \mathbb{E}[(Z_j^2 - 1)^2 | \mathcal{G}_n]}{2\|A\|_F^2} \\ &= \sum_{j=1}^n \frac{a_{jj}^2 (\mathbb{E}[Z_j^4 | \mathcal{G}_n] - 1)}{2\|A\|_F^2} \\ &\leq \frac{1}{2} \sum_{j=1}^n \frac{a_{jj}^2 \mathbb{E}[Z_j^4 | \mathcal{G}_n]}{\|A\|_F^2}, \end{aligned}$$

and by assumption.

To see the weak convergence of \bar{T}_n , for $j = 1, \dots, n$, define

$$V_{n,j} = \sqrt{2} Z_j \sum_{i=1}^{j-1} Z_i a_{ij} / \|A\|_F,$$

$\mathcal{F}_{n,0} = \mathcal{G}_n$ and $\mathcal{F}_{n,j} = \sigma(\mathcal{G}_n, Z_i : i \leq j)$, by which we mean the smallest sigma algebra for which Z_1, \dots, Z_j are measurable and which also contains \mathcal{G}_n . Note that for each $n, j \in \mathbb{N}$, $\mathcal{F}_{n,j-1} \subseteq \mathcal{F}_{n,j} \subseteq \mathcal{F}$, $\|A_n\|_F = \sqrt{\text{trace}(A^2)}$ is $\mathcal{F}_{n,0}$ measurable and $V_{n,j}$ is $\mathcal{F}_{n,j}$ measurable. Moreover, we have

$$\frac{\bar{T}_n}{\sqrt{2}\|A\|_F} = \sum_{j=1}^n V_{n,j}.$$

Now, by the central limit theorem for dependent random variables [see 10, 13, and notice the discussion in Helland [13] following eq. (2.7)], it remains to verify that

$$\sum_{j=1}^n \mathbb{E}[V_{n,j} | \mathcal{F}_{n,j-1}] \xrightarrow[n \rightarrow \infty]{i.p.} 0, \quad (\text{B.1})$$

$$\sum_{j=1}^n \text{Var}[V_{n,j} | \mathcal{F}_{n,j-1}] \xrightarrow[n \rightarrow \infty]{i.p.} 1, \quad \text{and} \quad (\text{B.2})$$

$$\sum_{j=1}^n \mathbb{E}[V_{n,j}^2 \mathbf{1}_{|V_{n,j}| > \delta} | \mathcal{F}_{n,j-1}] \xrightarrow[n \rightarrow \infty]{i.p.} 0 \quad \text{for all } \delta > 0, \quad (\text{B.3})$$

as in equations (2.5)–(2.7) in Helland [13]. The convergence in (B.1) is trivial, since $\mathbb{E}[V_{n,j} | \mathcal{F}_{n,j-1}] = 0$, in view of the conditional independence of the Z_i given \mathcal{G}_n .

For (B.2), abbreviate $T_n = \sum_{j=1}^n \text{Var}[V_{n,j} | \mathcal{F}_{n,j-1}] = \sum_{j=1}^n \mathbb{E}[V_{n,j}^2 | \mathcal{F}_{n,j-1}]$ and use conditional independence again to obtain

$$\mathbb{E}[V_{n,j}^2 | \mathcal{F}_{n,j-1}] = 2 \|A\|_F^{-2} \left(\sum_{i=1}^{j-1} Z_i a_{ij} \right)^2.$$

Expanding the squared sum gives

$$\left(\sum_{i=1}^{j-1} Z_i a_{ij} \right)^2 = \sum_{i,k}^{j-1} Z_i Z_k a_{ij} a_{kj} = \sum_{i=1}^{j-1} Z_i^2 a_{ij}^2 + 2 \sum_{i < k}^{j-1} Z_i Z_k a_{ij} a_{kj},$$

and therefore, the absolute difference $|T_n - 1|$ can be bounded as

$$\begin{aligned} |T_n - 1| &= \|A\|_F^{-2} \left| \sum_{j=1}^n 2 \left(\sum_{i=1}^{j-1} Z_i^2 a_{ij}^2 + 2 \sum_{i < k}^{j-1} Z_i Z_k a_{ij} a_{kj} \right) - \|A\|_F^2 \right| \\ &\leq \|A\|_F^{-2} \left(2 \left| \sum_{i < j}^n (Z_i^2 - 1) a_{ij}^2 \right| + 4 \left| \sum_{j=1}^n \sum_{i < k}^{j-1} Z_i Z_k a_{ij} a_{kj} \right| + \left| \sum_{j=1}^n a_{jj}^2 \right| \right). \end{aligned}$$

To establish the convergence in (B.2), it remains to show convergence to zero in probability of the terms in absolute values on the last line of the preceding display multiplied by $\|A\|_F^{-2}$.

First, note that $\|A\|_F^{-2} \sum_{j=1}^n a_{jj}^2$ converges to zero in probability by assumption and because of $\mathbb{E}[Z_j^4 | \mathcal{G}_n] \geq 1$. Now, write $T_{n,1} = \sum_{i < j}^n (Z_i^2 - 1) a_{ij}^2$ and $T_{n,2} = \sqrt{2} \sum_{j=1}^n \sum_{i < k}^{j-1} Z_i Z_k a_{ij} a_{kj}$ and observe that

$$\begin{aligned} \mathbb{E}[T_{n,1}^2 | \mathcal{G}_n] &= \sum_{i_1 < j_1}^n \sum_{i_2 < j_2}^n \mathbb{E}[(Z_{i_1}^2 - 1)(Z_{i_2}^2 - 1) | \mathcal{G}_n] a_{i_1 j_1}^2 a_{i_2 j_2}^2 \\ &= \sum_{j_1, j_2}^n \sum_{i=1}^{j_1 \wedge j_2 - 1} (\mathbb{E}[Z_i^4 | \mathcal{G}_n] - 1) a_{i j_1}^2 a_{i j_2}^2 \\ &\leq \left(\max_{j=1, \dots, n} \mathbb{E}[Z_j^4 | \mathcal{G}_n] \right) \sum_{j_1, j_2}^n \sum_{i=1}^{j_1 \wedge j_2 - 1} a_{i j_1}^2 a_{i j_2}^2, \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}[T_{n,2}^2 | \mathcal{G}_n] &= 2 \sum_{j_1, j_2}^n \sum_{i_1 < k_1}^{j_1-1} \sum_{i_2 < k_2}^{j_2-1} \mathbb{E}[Z_{i_1} Z_{k_1} Z_{i_2} Z_{k_2} | \mathcal{G}_n] a_{i_1 j_1} a_{k_1 j_1} a_{i_2 j_2} a_{k_2 j_2} \\ &= \sum_{j_1, j_2}^n 2 \sum_{i < k}^{j_1 \wedge j_2 - 1} a_{i j_1} a_{k j_1} a_{i j_2} a_{k j_2} \\ &\leq \left(\max_{j=1, \dots, n} \mathbb{E}[Z_j^4 | \mathcal{G}_n] \right) \sum_{j_1, j_2}^n \sum_{i \neq k}^{j_1 \wedge j_2 - 1} a_{i j_1} a_{k j_1} a_{i j_2} a_{k j_2}. \end{aligned}$$

Therefore, if we define the triangular truncation operator \tilde{A} of the symmetric matrix A by $\tilde{A} = \sum_{s>t} e_s a_{st} e'_t$, where $e_s \in \mathbb{R}^n$ is the s -th element of the standard basis in \mathbb{R}^n , we see that

$$\begin{aligned} \text{trace} \left((\tilde{A}' \tilde{A})^2 \right) &= \text{trace} \left(\left(\sum_{s_1 > t_1}^n \sum_{s_2 > t_2}^n e_{t_1} a_{s_1 t_1} e'_{s_1} e_{s_2} a_{s_2 t_2} e'_{t_2} \right)^2 \right) \\ &= \text{trace} \left(\left(\sum_{s=1}^n \sum_{t_1, t_2=1}^{s-1} e_{t_1} a_{s t_1} a_{s t_2} e'_{t_2} \right)^2 \right) \\ &= \sum_{s_1, s_2=1}^n \text{trace} \left(\sum_{t_1, t_2=1}^{s_1-1} e_{t_1} a_{s_1 t_1} a_{s_1 t_2} e'_{t_2} \sum_{u_1, u_2=1}^{s_2-1} e_{u_1} a_{s_2 u_1} a_{s_2 u_2} e'_{u_2} \right) \\ &= \sum_{s_1, s_2=1}^n \sum_{t_1, t_2=1}^{s_1 \wedge s_2 - 1} a_{s_1 t_1} a_{s_1 t_2} a_{s_2 t_2} a_{s_2 t_1}, \end{aligned} \quad (\text{B.4})$$

and, in turn, that

$$\begin{aligned} &\mathbb{E}[T_{n,1}^2 | \mathcal{G}_n] + \mathbb{E}[T_{n,2}^2 | \mathcal{G}_n] \\ &\leq \left(\max_{j=1, \dots, n} \mathbb{E}[Z_j^4 | \mathcal{G}_n] \right) \sum_{j_1, j_2}^n \sum_{i, k}^{j_1 \wedge j_2 - 1} a_{i j_1} a_{k j_1} a_{i j_2} a_{k j_2} \\ &= \left(\max_{j=1, \dots, n} \mathbb{E}[Z_j^4 | \mathcal{G}_n] \right) \text{trace} \left((\tilde{A}' \tilde{A})^2 \right). \end{aligned} \quad (\text{B.5})$$

Now, convergence to zero of $\|A\|_F^{-2} (|T_{n,1}| + |T_{n,2}|)$ in probability follows from the above considerations and Lemma 2.1 in Bhansali, Giraitis and Kokoszka [5], which yields the inequality

$$\begin{aligned} \mathbb{E} \left[\|A\|_F^{-4} (|T_{n,1}| + |T_{n,2}|)^2 | \mathcal{G}_n \right] &\leq 2 \left(\max_{j=1, \dots, n} \mathbb{E}[Z_j^4 | \mathcal{G}_n] \right) \|A\|_F^{-4} \|\tilde{A}' \tilde{A}\|_F^2 \\ &\leq 2C^2 \left(\max_{j=1, \dots, n} \mathbb{E}[Z_j^4 | \mathcal{G}_n] \right) \frac{\|A\|_S^2}{\|A\|_F^2}, \end{aligned}$$

where $C > 0$ is a global constant, not depending on n . Thus, by assumption, the bound on the far right-hand-side of the preceding display converges to zero, in probability, which establishes the convergence in (B.2).

Finally, for (B.3) we abbreviate $m_n = \max_j \mathbb{E}[Z_j^4 | \mathcal{G}_n]$ and use the upper bound $V_{n,j}^2 \mathbf{1}_{|V_{n,j}| > \delta} \leq \delta^{-2} V_{n,j}^4$. Now,

$$\begin{aligned} \mathbb{E} \left[\sum_{j=1}^n \mathbb{E}[V_{n,j}^4 | \mathcal{F}_{n,j-1}] \middle| \mathcal{G}_n \right] &= 4 \|A\|_F^{-4} \sum_{j=1}^n \mathbb{E}[Z_j^4 | \mathcal{G}_n] \mathbb{E} \left[\left(\sum_{i=1}^{j-1} Z_i a_{ij} \right)^4 \middle| \mathcal{G}_n \right] \\ &= 4 \|A\|_F^{-4} \sum_{j=1}^n \mathbb{E}[Z_j^4 | \mathcal{G}_n] \left(3 \sum_{i_1 \neq i_2}^{j-1} a_{i_1 j}^2 a_{i_2 j}^2 + \sum_{i=1}^{j-1} \mathbb{E}[Z_i^4 | \mathcal{G}_n] a_{ij}^4 \right) \\ &\leq 4 m_n (m_n + 3) \|A\|_F^{-4} \sum_{j=1}^n \sum_{i_1, i_2=1}^{j-1} a_{i_1 j}^2 a_{i_2 j}^2, \end{aligned}$$

and furthermore

$$\begin{aligned} \|A\|_F^{-4} \sum_{j=1}^n \sum_{i_1, i_2=1}^{j-1} a_{i_1 j}^2 a_{i_2 j}^2 &= \|A\|_F^{-4} \sum_{j=1}^n \left(\sum_{i=1}^{j-1} a_{ij}^2 \right)^2 \\ &\leq \|A\|_F^{-4} \left(\max_j \sum_{i=1}^n a_{ij}^2 \right) \sum_{i,j=1}^n a_{ij}^2 \\ &= \frac{\max_j (A^2)_{jj}}{\|A\|_F^2}. \end{aligned}$$

Together with $m_n \geq 1$ and our assumption, this implies that the upper bound on the second-to-last display converges to zero in probability. \square

Proof of Lemma 6.3. For convenience, we drop the subscript n that indexes the position in the array whenever there is no risk of confusion. Let $w'_i = (1, x'_i)R$ denote the i -th row of the matrix W and define $\tilde{w}_i = \Omega_W^{-1/2} w_i$, $\tilde{W} = W \Omega_W^{-1/2}$ and $S_1 = \tilde{W}' \tilde{W} - \tilde{w}_1 \tilde{w}'_1 = \sum_{i=2}^n \tilde{w}_i \tilde{w}'_i = \Omega_W^{-1/2} R' U_{-1} U_{-1} R \Omega_W^{-1/2}$, where

$$\Omega_W = \mathbb{E}[w_1 w'_1] = R' \begin{bmatrix} 1 & \mu' \\ \mu & \Sigma + \mu \mu' \end{bmatrix} R$$

is positive definite and U_{-1} is defined as in Assumption (A1).(b). This assumption also entails that $W'W$, $\tilde{W}'\tilde{W}$ and S_1 are invertible with probability one, where we denote the corresponding null set by N . For convenience, we redefine these quantities in an arbitrary invertible and measurable way on N . Moreover, we must also have $p_n + 2 \leq n$ under (A1).(b).

Since $h_j = h_{j,n} = w'_j (W'W)^{-1} w_j$ on N^c , permuting the h_1, \dots, h_n is equivalent to a permutation of w_1, \dots, w_n , which are i.i.d., and therefore their joint distribution is invariant under permutation. Hence, the h_j are exchangeable random variables. In particular, the h_j are identically distributed and therefore

the fact that $\sum_{j=1}^n h_j = \text{trace}(P_W) = k_n$, on N^c , entails that $\mathbb{E}[h_1] = k_n/n$. We also note for later use that $\text{Var}[h_1] = \mathbb{E}[h_1^2] - \mathbb{E}[h_1]^2 \leq \mathbb{E}[h_1] - \mathbb{E}[h_1]^2 = (1 - k_n/n)k_n/n$, since $0 \leq h_1 \leq 1$. It only remains to show that the variance actually converges to zero.

As a preliminary consideration, we study h_1 in the case where $t_n := k_n/n \rightarrow t \in [0, 1]$. The general case of possibly non-converging t_n then follows from a standard subsequence argument (see the end of the proof). The case $t \in \{0, 1\}$ is immediate, because here $\text{Var}[h_1] \rightarrow 0$ as $n \rightarrow \infty$, and thus, $h_1 \rightarrow t$ in probability, by the arguments in the previous paragraph. Assume now that $t \in (0, 1)$. Note that $P_W = P_{\tilde{W}}$ and use the Sherman-Morrison formula to obtain

$$h_1 = \tilde{w}'_1 (S_1 + \tilde{w}_1 \tilde{w}'_1)^{-1} \tilde{w}_1 = \frac{\tilde{w}'_1 S_1^{-1} \tilde{w}_1}{1 + \tilde{w}'_1 S_1^{-1} \tilde{w}_1},$$

at least on N^c . For $\alpha \geq 0$, set $J_\alpha = (S_1 + (n - 1)\alpha I_{k_n})^{-1}$ and define the random function

$$\Psi_n(\alpha) = \frac{\tilde{w}'_1 J_\alpha \tilde{w}_1}{1 + \tilde{w}'_1 J_\alpha \tilde{w}_1},$$

which satisfies $\Psi_n(0) = h_1$, almost surely. Since $y \mapsto y/(1 + y)$ is non-decreasing on $[0, \infty)$, and $M \mapsto M^{-1}$ is non-increasing on invertible hermitian matrices [cf. 6, p. 114], the function Ψ_n is non-increasing on $[0, \infty)$. We establish the convergence in probability of $\Psi_n(0)$ by first analyzing the limiting behavior of $\Psi_n(\alpha)$ as $n \rightarrow \infty$, for every $\alpha > 0$. To this end, we consider the conditional mean and variance of $\tilde{w}'_1 J_\alpha \tilde{w}_1$ given S_1 .

Since \tilde{w}_1 and S_1 are independent and $\mathbb{E}[\tilde{w}_1 \tilde{w}'_1] = I_{k_n}$, one easily calculates $\mathbb{E}[\tilde{w}'_1 J_\alpha \tilde{w}_1 | S_1] = \text{trace}(J_\alpha)$. The conditional variance is slightly more involved. Abbreviate $\Sigma_W = \text{Var}[w_1]$, $\bar{\mu} = \mathbb{E}[\tilde{w}_1] = \Omega_W^{-1/2} R'(1, \mu)'$ and use Assumption (A1).(a) to write

$$\tilde{w}_1 = \bar{\mu} + \Omega_W^{-1/2} R' \begin{pmatrix} 0 & 0 \\ 0 & \Gamma \end{pmatrix} \begin{pmatrix} 0 \\ z_1 \end{pmatrix}.$$

Also notice that $\Sigma_W = \Omega_W - R'(1, \mu)'(1, \mu)R$, and hence $\Omega_W^{-1/2} \Sigma_W \Omega_W^{-1/2} = I_{k_n} - \bar{\mu} \bar{\mu}'$. Since $\Omega_W^{-1/2} \Sigma_W \Omega_W^{-1/2}$ is positive semidefinite, this also implies that $\|\bar{\mu}\| \leq 1$. Now, decompose the quantity of interest

$$\begin{aligned} \tilde{w}'_1 J_\alpha \tilde{w}_1 &= \bar{\mu}' J_\alpha \bar{\mu} + 2\bar{\mu}' J_\alpha \Omega_W^{-1/2} R' \begin{pmatrix} 0 & 0 \\ 0 & \Gamma \end{pmatrix} \begin{pmatrix} 0 \\ z_1 \end{pmatrix} \\ &+ (0, z'_1) \begin{pmatrix} 0 & 0 \\ 0 & \Gamma' \end{pmatrix} R \Omega_W^{-1/2} J_\alpha \Omega_W^{-1/2} R' \begin{pmatrix} 0 & 0 \\ 0 & \Gamma \end{pmatrix} \begin{pmatrix} 0 \\ z_1 \end{pmatrix}. \end{aligned} \quad (\text{B.6})$$

Conditional on S_1 , the variance of $\bar{\mu}' J_\alpha \bar{\mu}$ is zero, the variance of half of the mixed term is

$$\bar{\mu}' J_\alpha \Omega_W^{-1/2} \Sigma_W \Omega_W^{-1/2} J_\alpha \bar{\mu} = \bar{\mu}' J_\alpha^2 \bar{\mu} - (\bar{\mu}' J_\alpha \bar{\mu})^2$$

and the variance of the last term in (B.6) is

$$\begin{aligned} \text{Var}[(0, z'_1)M(0, z'_1)'|S_1] &= \text{Var}[z'_1M_{22}z_1|S_1] \\ &= O(\text{trace}(M_{22}^2)) + (\text{trace}(M_{22}))^2o(1) \\ &\leq O(\text{trace}(M^2)) + (\text{trace}(M))^2o(1), \end{aligned}$$

by assumption, and where we have abbreviated the symmetric positive semidefinite matrix in between the vectors $(0, z'_1)$ and $(0, z'_1)'$ in (B.6) by M and used the notation M_{22} to denote its bottom right sub matrix of order $m \times m$. Now, $\text{trace}(M) = \text{trace}(J_\alpha \Omega_W^{-1/2} \Sigma_W \Omega_W^{-1/2}) = \text{trace}(J_\alpha) - \bar{\mu}' J_\alpha \bar{\mu}$ and $\text{trace}(M^2) = \text{trace}(J_\alpha \Omega_W^{-1/2} \Sigma_W \Omega_W^{-1/2} J_\alpha \Omega_W^{-1/2} \Sigma_W \Omega_W^{-1/2}) = \text{trace}(J_\alpha^2) - 2\bar{\mu}' J_\alpha^2 \bar{\mu} + (\bar{\mu}' J_\alpha \bar{\mu})^2$. For $\alpha > 0$, $\|J_\alpha\|_S \leq [\alpha(n-1)]^{-1}$ and $\|\bar{\mu}\|^2 \leq 1$. Therefore, $\bar{\mu}' J_\alpha \bar{\mu}$ and $\bar{\mu}' J_\alpha^2 \bar{\mu}$ converge to zero, almost surely, for every $\alpha > 0$. Thus, in order to show that $\text{Var}[\tilde{w}'_1 J_\alpha \tilde{w}_1 | S_1]$ converges to zero, almost surely, for every $\alpha > 0$, it suffices to show that $\text{trace}(J_\alpha^2) \rightarrow 0$ as $n \rightarrow \infty$, almost surely, and that $\text{trace}(J_\alpha)$ is almost surely convergent. Moreover, if we can even show that $\text{trace}(J_\alpha) \rightarrow \psi_\alpha \in [0, \infty)$, almost surely, for every $\alpha > 0$, then we also have $\tilde{w}'_1 J_\alpha \tilde{w}_1 \rightarrow \psi_\alpha$, in probability (since $\mathbb{E}[\tilde{w}'_1 J_\alpha \tilde{w}_1 | S_1] = \text{trace}(J_\alpha)$), and thus $\Psi_n(\alpha) \rightarrow \Psi(\alpha) := \psi_\alpha / (1 + \psi_\alpha)$, in probability, for every $\alpha > 0$.

Therefore, we need to study the limiting behavior of

$$\begin{aligned} \text{trace}((S_1 + (n-1)\alpha I_{k_n})^{-\ell}) &= \frac{k_n}{(n-1)^\ell} \frac{1}{k_n} \sum_{j=1}^{k_n} \frac{1}{(\lambda_j + \alpha)^\ell} \\ &= \frac{k_n}{(n-1)^\ell} \int_0^\infty (y + \alpha)^{-\ell} dF^{S_1/(n-1)}(y), \quad (\text{B.7}) \end{aligned}$$

for $\ell = 1, 2$, where $0 \leq \lambda_1 \leq \dots \leq \lambda_{k_n}$ are the ordered eigenvalues of $S_1/(n-1) = \sum_{j=2}^n \tilde{w}_j \tilde{w}'_j / (n-1)$ and $F^{S_1/(n-1)}$ denotes the corresponding empirical spectral distribution function. Now one easily verifies the assumptions of Theorem 1.1 in Bai and Zhou [3]. First note that the \tilde{w}_j are i.i.d. and $\mathbb{E}[\tilde{w}_1 \tilde{w}'_1] = I_{k_n}$. Second, by the same argument following (B.6) and for an arbitrary non-random $k_n \times k_n$ matrix B with bounded spectral norm, we have

$$\begin{aligned} \mathbb{E}[|\tilde{w}'_1 B \tilde{w}_1 - \text{trace}(B)|^2] &= \text{Var}[\tilde{w}'_1 (B/2 + B'/2) \tilde{w}_1] \\ &= O(\text{trace}((B/2 + B'/2)^2)) + (\text{trace}(B) + O(1))^2 o(1) + O(1) \\ &= O(k_n \|B\|_S^2) + k_n^2 \|B\|_S^2 o(1) + O(1) = o(n^2). \end{aligned}$$

Recall also that for now $t_n = k_n/n \rightarrow t \in (0, 1)$. Therefore, $F^{S_1/(n-1)}$ converges weakly, almost surely to the Marčenko-Pastur distribution with Lebesgue density $f^{MP}(y) = \sqrt{(y-a)(b-y)}/(2\pi ty)$ on $[a, b]$, where $a = (1 - \sqrt{t})^2$ and $b = (1 + \sqrt{t})^2$. Now we see that we can not use the same strategy to establish the convergence of (B.7) in the case where $\alpha = 0$, because the function $h_\alpha(y) = (y + \alpha)^{-1}$ is not bounded on $(0, \infty)$ in that case. However, for $\alpha > 0$,

h_α is bounded and continuous on $[0, \infty)$ and therefore the integral in (B.7) converges almost surely,

$$\int_0^\infty h_\alpha^\ell(y) dF^{S_1/(n-1)}(y) \xrightarrow{a.s.} \int_a^b h_\alpha^\ell(y) f^{MP}(y) dy \in (0, \infty).$$

Since $k_n/(n-1) \rightarrow t \in (0, 1)$ and $k_n/(n-1)^2 \rightarrow 0$, this means that $\text{trace}(J_\alpha) \rightarrow \psi_\alpha := t \int_a^b h_\alpha(y) f^{MP}(y) dy$ and $\text{trace}(J_\alpha^2) \rightarrow 0$, almost surely, for every $\alpha > 0$. As discussed at the end of the previous paragraph, this entails that $\Psi_n(\alpha) \rightarrow \Psi(\alpha)$, in probability, for every $\alpha > 0$, where the function Ψ is given by

$$\Psi(\alpha) = \frac{t \int_a^b h_\alpha(y) f^{MP}(y) dy}{1 + t \int_a^b h_\alpha(y) f^{MP}(y) dy}.$$

Now it is easy to see (e.g., by the dominated convergence theorem) that the function Ψ is continuous and non-increasing on $[0, \infty)$ (recall that here $t < 1$ and $a > 0$). Moreover, the limiting integral for $\alpha = 0$ can be evaluated as $\int_a^b h_0(y) f^{MP}(y) dy = 1/(1-t)$ [cf. 14, Lemma B.1] resulting in $\Psi(0) = t$.

Let us briefly recapitulate what we have found so far. First of all, we have seen that $h_1 = h_{1,n} \rightarrow t$ in probability, as $n \rightarrow \infty$, if $t_n \rightarrow t \in \{0, 1\}$. For $t_n \rightarrow t \in (0, 1)$, we know that $\mathbb{E}[h_{1,n}] = k_n/n \rightarrow t$ and $0 \leq \Psi_n(\alpha) \leq \Psi_n(0) = h_1 \leq 1$ almost surely. Moreover, $\Psi_n(\alpha) \rightarrow \Psi(\alpha)$ in probability, for every $\alpha > 0$, and $\Psi(\alpha) \rightarrow \Psi(0) = t$ as $\alpha \rightarrow 0$. Thus, Lemma C.2 applies and we obtain that $h_{1,n} \rightarrow t$ in probability, also in the case $t \in (0, 1)$.

Finally, consider $\Delta_n := |h_1 - k_n/n|$ with arbitrary $t_n = k_n/n \in [0, 1]$. Suppose that $c := \limsup \mathbb{E}[\Delta_n] > 0$. Then there exists a subsequence n' , such that $\mathbb{E}[\Delta_{n'}] \rightarrow c$, as $n' \rightarrow \infty$. By compactness, there exists a further subsequence n'' , such that $t_{n''} \rightarrow t \in [0, 1]$, as $n'' \rightarrow \infty$. But in this case, our previous arguments have shown that $\Delta_{n''} \rightarrow 0$ in probability, which also entails that $\mathbb{E}[\Delta_{n''}] \rightarrow 0$, by boundedness, contradicting the assertion that $\mathbb{E}[\Delta_{n'}] \rightarrow c > 0$. \square

Appendix C: Other technical results

Lemma C.1. *Under the model (2.1), suppose that $\frac{1}{n} \sum_{i=1}^n \mathbb{E}[(\varepsilon_i/\sigma_n)^4 | x_i] = O_{\mathbb{P}}(1)$ and $\text{rank}(U) = p_n + 1$, almost surely, for all $n \in \mathbb{N}$. If $\limsup_{n \rightarrow \infty} p_n/n < 1$, then $\sqrt{n} |\hat{\sigma}_n^2/\sigma_n^2 - 1| = O_{\mathbb{P}}(1)$. In particular, we have $\mathbb{P}(\hat{\sigma}_n^2 = 0) \rightarrow 0$ as $n \rightarrow \infty$.*

Remark. The assumption that $\frac{1}{n} \sum_{i=1}^n \mathbb{E}[(\varepsilon_i/\sigma_n)^4 | x_i] = O_{\mathbb{P}}(1)$ is clearly weaker than a uniform bound on $\mathbb{E}[(\varepsilon_1/\sigma_n)^4]$ or a uniform bound on $\mathbb{E}[\varepsilon_1^4]$ together with $\liminf_n \sigma_n^2 > 0$. Clearly, also Assumption (A2) implies $\frac{1}{n} \sum_{i=1}^n \mathbb{E}[(\varepsilon_i/\sigma_n)^4 | x_i] = O_{\mathbb{P}}(1)$.

Proof of Lemma C.1. Recall that $\hat{\sigma}_n^2 = \varepsilon' M \varepsilon$, almost surely, where $M := M(X) := (I_n - P_U)/(n - p - 1)$ is a function of the design matrix X . Note that y_1, \dots, y_n are conditionally independent given X , and hence, also $\varepsilon_i =$

$y_i - \mathbb{E}[y_i|x_i]$, for $i = 1, \dots, n$, are conditionally independent given X . Therefore, one easily obtains the almost sure identities

$$\begin{aligned} \mathbb{E}[\hat{\sigma}_n^2/\sigma_n^2|X] &= \text{trace}(M) \quad \text{and} \\ \text{Var}[\hat{\sigma}_n^2/\sigma_n^2|X] &= 2\text{trace}(M^2) + \sum_{i=1}^n (\mathbb{E}[(\varepsilon_i/\sigma_n)^4|x_i] - 3)M_{ii}^2. \end{aligned}$$

By our assumption on U , with probability one, $\text{trace}(M) = 1$ and $\text{trace}(M^2) = 1/(n-p-1)$, whereas $M_{ii}^2 \leq 1/(n-p-1)^2$ holds everywhere, since the diagonal entries of the projection matrix $I_n - P_U$ are always between 0 and 1. Taken together, we see that $\mathbb{E}[\hat{\sigma}_n^2/\sigma_n^2|X] = 1$ and $\text{Var}[\hat{\sigma}_n^2/\sigma_n^2|X] \leq 2/(n-p-1) + \sum_{i=1}^n \mathbb{E}[(\varepsilon_i/\sigma_n)^4|x_i]/(n-p-1)^2$. Now, the conditional Markov inequality yields

$$\begin{aligned} \mathbb{P}(\sqrt{n}|\hat{\sigma}_n^2/\sigma_n^2 - 1| > \delta) &= \mathbb{E}[\mathbb{P}(\sqrt{n}|\hat{\sigma}_n^2/\sigma_n^2 - 1| > \delta|X) \wedge 1] \\ &\leq \mathbb{E}\left[\left(\frac{n}{\delta^2} \text{Var}[\hat{\sigma}_n^2/\sigma_n^2|X]\right) \wedge 1\right] \\ &\leq \mathbb{P}(n \text{Var}[\hat{\sigma}_n^2/\sigma_n^2|X] > \delta) + \left(\frac{1}{\delta} \wedge 1\right). \end{aligned}$$

Since $n \text{Var}[\hat{\sigma}_n^2/\sigma_n^2|X] = O_{\mathbb{P}}(1)$, in view of the previous considerations and the assumptions $\limsup_{n \rightarrow \infty} p_n/n < 1$ and $\frac{1}{n} \sum_{i=1}^n \mathbb{E}[(\varepsilon_i/\sigma_n)^4|x_i] = O_{\mathbb{P}}(1)$, this finishes the proof of the first claim. The second assertion follows immediately, because of $\mathbb{P}(\hat{\sigma}_n^2 = 0) \leq \mathbb{P}(|\hat{\sigma}_n^2/\sigma_n^2 - 1| > 1/2)$. \square

Lemma C.2. For $n \in \mathbb{N}$ and $\alpha > 0$, let h_n and $\Psi_n(\alpha)$ be real random variables such that $0 \leq \Psi_n(\alpha) \leq h_n \leq 1$ almost surely, and, for $t \in [0, 1]$, let $\Psi : [0, \infty) \rightarrow [0, 1]$ be such that $\Psi(\alpha) \rightarrow t$ as $\alpha \rightarrow 0$. If for every $\alpha > 0$, $\Psi_n(\alpha) \rightarrow \Psi(\alpha)$ in probability, and $\mathbb{E}[h_n] \rightarrow t$ as $n \rightarrow \infty$, then $h_n \rightarrow t$ in probability, as $n \rightarrow \infty$.

Remark. Lemma C.2 is an asymptotic version of the well known fact that a random variable h that satisfies $h \geq t \in \mathbb{R}$ and $\mathbb{E}[h] = t$ must be equal to t , almost surely.

Proof of Lemma C.2. Fix $\delta > 0$, choose $\alpha = \alpha(\delta) > 0$ such that $|\Psi(\alpha) - t| < \delta/2$ and do the following standard bound,

$$\begin{aligned} \mathbb{P}(h_n < \mathbb{E}[h_n] - \delta) &\leq \mathbb{P}(|\Psi_n(\alpha) - \mathbb{E}[h_n]| > \delta) \\ &\quad + \mathbb{P}(h_n < \mathbb{E}[h_n] - \delta, |\Psi_n(\alpha) - \mathbb{E}[h_n]| \leq \delta). \end{aligned}$$

But $|\Psi_n(\alpha) - \mathbb{E}[h_n]| \leq |\Psi_n(\alpha) - \Psi(\alpha)| + |\Psi(\alpha) - t| + |t - \mathbb{E}[h_n]| \leq \delta/2 + o_{\mathbb{P}}(1)$, whereas $|\Psi_n(\alpha) - \mathbb{E}[h_n]| \leq \delta$ and $h_n < \mathbb{E}[h_n] - \delta$ together imply that $\Psi_n(\alpha) \geq \mathbb{E}[h_n] - \delta > h_n$, which, by assumption, happens only on a set of probability zero. Therefore, the upper bound in the previous display converges to zero. Now, by boundedness of h_n we have

$$\begin{aligned} \mathbb{E}[|h_n - \mathbb{E}[h_n]|] &\leq \mathbb{E}[|h_n - \mathbb{E}[h_n] + \delta|] + \delta \\ &\leq \mathbb{E}[(h_n - (\mathbb{E}[h_n] - \delta))\mathbf{1}_{\{h_n \geq \mathbb{E}[h_n] - \delta\}}] \end{aligned}$$

$$\begin{aligned}
& + \mathbb{P}(h_n < \mathbb{E}[h_n] - \delta) + \delta \\
\leq & \mathbb{E}[h_n \mathbf{1}_{\{h_n \geq \mathbb{E}[h_n] - \delta\}}] - \mathbb{E}[h_n] \mathbb{P}(h_n \geq \mathbb{E}[h_n] - \delta) \\
& + 2\delta + o(1),
\end{aligned}$$

and we also see that both $\mathbb{E}[h_n \mathbf{1}_{\{h_n \geq \mathbb{E}[h_n] - \delta\}}]$ and $\mathbb{E}[h_n] \mathbb{P}(h_n \geq \mathbb{E}[h_n] - \delta)$ converge to t . Since $\delta > 0$ was arbitrary, we must have $\limsup \mathbb{E}[|h_n - \mathbb{E}[h_n]|] = 0$ and thus, convergence in probability of h_n to t . \square

Lemma C.3. For every $n \in \mathbb{N}$, let $x_{1,n}, \dots, x_{n,n}$ be i.i.d. random p_n -vectors satisfying $x_{i,n} = \mu_n + \Gamma_n z_{i,n}$ as in Assumption (A1).(a) with positive semidefinite covariance matrix $\Sigma_n = \Gamma_n \Gamma_n'$. Set $X_n = [x_{1,n}, \dots, x_{n,n}]'$, $\hat{\Sigma}_n = X_n'(I_n - P_n)X_n/n$ and

$$S_n = \mathbb{E} \left[\begin{pmatrix} 1 \\ x_1 \end{pmatrix} (1 \quad x_1') \right] = \begin{pmatrix} 1 & \mu_n' \\ \mu_n & \Sigma_n + \mu_n \mu_n' \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_n \end{pmatrix} + \begin{pmatrix} 1 \\ \mu_n \end{pmatrix} (1 \quad \mu_n').$$

Moreover, let R_n be a $k_n \times (p_n + 1)$ matrix such that $R_n R_n' = I_{k_n}$ (i.e., $k_n \leq p_n + 1$) and set $\Omega_n = R_n S_n R_n'$.

- (i) Let $u_n \in \mathbb{R}^{p_n+1}$. If $\sup_{\|w\|=1} \mathbb{E}[|w' z_{1,n}|^\ell] = O(1)$ as $n \rightarrow \infty$, for some fixed $\ell \in \mathbb{N}$, not depending on n , then $\mathbb{E}[|u_n'(1, x_{1,n})'|^\ell] = O(|u_n' S_n u_n|^{\ell/2})$ as $n \rightarrow \infty$.
- (ii) Let $v_{n,1}, v_{n,2} \in \mathbb{R}^{p_n}$. If $\sup_{\|w\|=1} \mathbb{E}[|w' z_{1,n}|^4] = O(1)$ as $n \rightarrow \infty$, then $\text{Var}[\sqrt{n} v_{n,1}' \hat{\Sigma}_n v_{n,2}] = O(v_{n,1}' \Sigma_n v_{n,1} v_{n,2}' \Sigma_n v_{n,2})$.
- (iii) If Σ_n is positive definite, $z_{1,n}$ satisfies (A1).(c) and $\sup_{\|w\|=1} \mathbb{E}[|w' z_{1,n}|^4] = O(1)$, then the design matrix of the transformed data $W_n = [1, X] R_n'$ satisfies

$$\left\| \Omega_n^{-1/2} (W_n' W_n / n) \Omega_n^{-1/2} - I_{k_n} \right\|_S = \begin{cases} o_{\mathbb{P}}(1), & \text{if } k_n/n \rightarrow 0, \\ O_{\mathbb{P}}(1), & \text{if } k_n = O(n). \end{cases}$$

- (iv) If Σ_n is positive definite and $\text{Var}[z_{1,n}' M z_{1,n}] = O(\text{trace}(M^2)) + (\text{trace}(M))^2 o(1)$, as $n \rightarrow \infty$, for every symmetric matrix $M \in \mathbb{R}^{m_n \times m_n}$, then we have $\mathbb{E}[|(1, x_{1,n})' R_n' \Omega_n^{-1} R_n (1, x_{1,n})'|^2] = O(k_n^2)$.
- (v) Suppose that Σ_n is positive definite and that $\sup_{\|w\|=1} \mathbb{E}[|w' z_{1,n}|^8] = O(1)$ and $(\mathbb{E}[|z_{1,n}' P z_{1,n}|^4])^{1/4} = O(\|P\|_F)$, as $n \rightarrow \infty$, for every projection matrix P in \mathbb{R}^{m_n} , and partition $R_n = [t_1, T_1]$ with $t_1 \in \mathbb{R}^{k_n}$. If for every $n \in \mathbb{N}$ either one of (a) $\mu_n = 0$, or (b) $\text{rank } T_1 = k_n$ holds, then $\mathbb{E}[|(1, x_{1,n})' R_n' \Omega_n^{-1} R_n (1, x_{1,n})'|^4] = O(k_n^4)$.

Proof. For ease of notation we will drop the subscript n whenever there is no risk of confusion. A simple calculation involving the elementary inequality $|a + b|^\ell \leq 2^{\ell-1}(|a|^\ell + |b|^\ell)$ and the notation $u_n = u = (u_0, u_{-1})'$, with $u_{-1} \in \mathbb{R}^{p_n}$, yields

$$\begin{aligned}
\mathbb{E}[|u'(1, x_1')|^\ell] &= \mathbb{E}[|u'(1, \mu') + u'(0, z_1' \Gamma')|^\ell] \\
&\leq 2^{\ell-1} \mathbb{E}[|u'(1, \mu')|^\ell + |u'_{-1} \Gamma z_1|^\ell]
\end{aligned}$$

$$= 2^{\ell-1} \left(|u'(1, \mu')'(1, \mu')u|^{\ell/2} + |u'_{-1}\Sigma u_{-1}|^{\ell/2} \mathbb{E}[|w'z_1|^\ell] \right),$$

where $w = \Gamma'u_{-1}/\|\Gamma'u_{-1}\|$, if $\|\Gamma'u_{-1}\| > 0$ and $w = 0$, else. In the sum $u'(1, \mu')'(1, \mu')u + u'_{-1}\Sigma u_{-1} = u'S_n u$ both summands are non-negative and thus both summands are bounded by $u'S_n u$. Therefore, the upper bound in the previous display is itself bounded by a constant multiple of $|u'S_n u|^{\ell/2}$. This was the claim of part (i).

For part (ii), first note that because the distribution of $\hat{\Sigma}_n$ does not depend on μ , we may assume that $\mu = 0$, without loss of generality. By the same argument as above but with $\mu = 0$, $u_0 = 0$ and u_{-1} is either $v_{n,1}$ or $v_{n,2}$, we see that $\mathbb{E}[|v'_{n,s}x_1|^4] = O(|v'_{n,s}\Sigma v_{n,s}|^2)$, $s = 1, 2$. Now

$$\begin{aligned} \text{Var}[\sqrt{n}v'_{n,1}\hat{\Sigma}_n v_{n,2}] &= n \text{Var}[v'_{n,1}X'Xv_{n,2}/n - v'_{n,1}X'\mu'Xv_{n,2}/n^2] \\ &\leq 2n \left(\frac{1}{n^2} \sum_{i=1}^n \text{Var}[v'_{n,1}x_i v'_{n,2}x_i] + \frac{1}{n^4} \text{Var} \left[\sum_{i,j=1}^n v'_{n,1}x_i v'_{n,2}x_j \right] \right) \\ &\leq 2\sqrt{\mathbb{E}[|v'_{n,1}x_1|^4]\mathbb{E}[|v'_{n,2}x_1|^4]} + \frac{2}{n^3} \sum_{i,j,k,l=1}^n \mathbb{E}[v'_{n,1}x_i v'_{n,2}x_j v'_{n,1}x_k v'_{n,2}x_l] \\ &= O(v'_{n,1}\Sigma v_{n,1}v'_{n,2}\Sigma v_{n,2}) + \frac{2}{n^2} \mathbb{E}[|v'_{n,1}x_1|^2 |v'_{n,2}x_1|^2] \\ &\quad + \frac{4}{n^3} \sum_{i \neq j} \mathbb{E}[v'_{n,1}x_i v'_{n,2}x_i] \mathbb{E}[v'_{n,1}x_j v'_{n,2}x_j] \\ &\quad + \frac{2}{n^3} \sum_{i \neq j} \mathbb{E}[|v'_{n,1}x_i|^2] \mathbb{E}[|v'_{n,2}x_j|^2], \end{aligned}$$

which is of order $O(v'_{n,1}\Sigma v_{n,1}v'_{n,2}\Sigma v_{n,2})$ because $\mathbb{E}[|v'_{n,s}x_1|^2] = v'_{n,s}\Sigma v_{n,s}$ and $\mathbb{E}[v'_{n,1}x_1 v'_{n,2}x_1] \leq \sqrt{\mathbb{E}[|v'_{n,1}x_1|^2]\mathbb{E}[|v'_{n,2}x_1|^2]}$.

For parts (iii), (iv) and (v) we make the following preliminary considerations. First, note that in all three of these statements Σ is assumed to be positive definite and thus Ω is regular. Abbreviate $\bar{\mu}' := (1, \mu')R'\Omega^{-1/2}$ and

$$\Sigma_W = R \begin{pmatrix} 0 & 0 \\ 0 & \Sigma \end{pmatrix} R'.$$

Since $\Sigma_W = \Omega - R(1, \mu')'(1, \mu')R'$ we have

$$\Omega^{-1/2}\Sigma_W\Omega^{-1/2} = I_{k_n} - \bar{\mu}\bar{\mu}' = A \begin{pmatrix} 1 - \|\bar{\mu}\|^2 & 0 \\ 0 & I_{k_n-1} \end{pmatrix} A', \quad (\text{C.1})$$

for some orthogonal matrix A whose first column is $\bar{\mu}/\|\bar{\mu}\|$ if $\|\bar{\mu}\| > 0$, and $A = I_{k_n}$ if $\bar{\mu} = 0$. Here, quantities of dimension $k_n - 1$ have to be removed in case $k_n = 1$. The matrix $\Omega^{-1/2}\Sigma_W\Omega^{-1/2}$ in the previous display is positive semidefinite, which means that $0 \leq \|\bar{\mu}\| \leq 1$. For later use, we partition the

matrix $B := \Omega^{-1/2}A$ as $B = [b_1, B_1]$ where $b_1 \in \mathbb{R}^{k_n}$ and note that B satisfies

$$B'\Sigma_W B = \begin{pmatrix} 1 - \|\bar{\mu}\|^2 & 0 \\ 0 & I_{k_n-1} \end{pmatrix}, \quad (\text{C.2})$$

and $I_{k_n} = B'\Omega B = B'\Sigma_W B + B'R(1, \mu')'(1, \mu')R'B$, which entails that

$$B'R(1, \mu')'(1, \mu')R'B = \begin{pmatrix} \|\bar{\mu}\|^2 & 0 \\ 0 & 0 \end{pmatrix}. \quad (\text{C.3})$$

This finishes the preliminary considerations.

Now, for the proof of part (iii), write the quantity of interest as

$$\begin{aligned} & \left\| \Omega^{-1/2}(W'W/n)\Omega^{-1/2} - I_{k_n} \right\|_S = \left\| A'\Omega^{-1/2}(W'W/n)\Omega^{-1/2}A - A'A \right\|_S \\ & = \left\| B'(W'W/n)B - I_{k_n} \right\|_S \leq \left\| B'\hat{\Sigma}_W B - \begin{pmatrix} 1 - \|\bar{\mu}\|^2 & 0 \\ 0 & I_{k_n-1} \end{pmatrix} \right\|_S \\ & \quad + \left\| B'R(1, \hat{\mu}')'(1, \hat{\mu}')R'B - \begin{pmatrix} \|\bar{\mu}\|^2 & 0 \\ 0 & 0 \end{pmatrix} \right\|_S \\ & = \left\| \begin{pmatrix} b_1'\hat{\Sigma}_W b_1 - (1 - \|\bar{\mu}\|^2) & b_1'\hat{\Sigma}_W B_1 \\ B_1'\hat{\Sigma}_W b_1 & B_1'\hat{\Sigma}_W B_1 - I_{k_n-1} \end{pmatrix} \right\|_S \end{aligned} \quad (\text{C.4})$$

$$+ \left\| \begin{pmatrix} b_1'R(1, \hat{\mu}')'(1, \hat{\mu}')R'b_1 - \|\bar{\mu}\|^2 & b_1'R(1, \hat{\mu}')'(1, \hat{\mu}')R'B_1 \\ B_1'R(1, \hat{\mu}')'(1, \hat{\mu}')R'b_1 & B_1'R(1, \hat{\mu}')'(1, \hat{\mu}')R'B_1 \end{pmatrix} \right\|_S. \quad (\text{C.5})$$

where

$$\hat{\Sigma}_W = R \begin{pmatrix} 0 & 0 \\ 0 & \hat{\Sigma}_n \end{pmatrix} R' \quad \text{and} \quad \hat{\mu} = X'\iota/n.$$

For a partitioned matrix as above we have

$$\begin{aligned} & \left\| \begin{pmatrix} c_{11} & c'_{12} \\ c_{21} & C_{22} \end{pmatrix} \right\|_S^2 = \sup_{\|w\|=1} \left\| \begin{pmatrix} c_{11}w_1 + c_{12}w_{-1} \\ c_{21}w_1 + C_{22}w_{-1} \end{pmatrix} \right\|^2 \\ & \leq (\|c_{11}\| + \|c_{12}\|)^2 + (\|c_{21}\| + \|C_{22}\|_S)^2. \end{aligned}$$

Therefore, it suffices to show that the norms of the respective blocks are $O_{\mathbb{P}}(1)$, if $k_n = O(n)$, and converge to zero in probability, if $k_n/n \rightarrow 0$.

We begin with the terms involving $\hat{\mu}$ in (C.5). First,

$$\begin{aligned} & \mathbb{E} [b_1'R(1, \hat{\mu}')'(1, \hat{\mu}')R'b_1] - \|\bar{\mu}\|^2 \\ & = b_1'R \left[\begin{pmatrix} 0 & 0 \\ 0 & \Sigma/n \end{pmatrix} + \begin{pmatrix} 1 \\ \mu \end{pmatrix} (1 \quad \mu') \right] R'b_1 - \|\bar{\mu}\|^2 \\ & = (1 - \|\bar{\mu}\|^2)/n \xrightarrow[n \rightarrow \infty]{} 0, \end{aligned}$$

in view of $\mathbb{E}[\hat{\mu}\hat{\mu}'] = \Sigma/n + \mu\mu'$, (C.2) and (C.3). Moreover, the variance satisfies

$$\text{Var} \left[(b_1'R(1, \hat{\mu}')')^2 \right] = \frac{1}{n^4} \text{Var} \left[\sum_{i,j=1}^n b_1'R(1, x'_i)' b_1'R(1, x'_j)' \right]$$

$$\begin{aligned}
 &= \frac{1}{n^4} \text{Var} \left[\sum_{i \neq j}^n b'_1 R[(1, x'_i)'(1, x'_j) - (1, \mu')'(1, \mu')] R' b_1 \right. \\
 &\qquad \qquad \qquad \left. + \sum_{i=1}^n b'_1 R[(1, x'_i)'(1, x'_i) - S] R' b_1 \right] \\
 &\leq \frac{2}{n^4} \left(\sum_{i \neq j}^n \sum_{r \neq s}^n \mathbb{E}[b'_1 R[(1, x'_i)'(1, x'_j) - (1, \mu')'(1, \mu')] R' b_1 \times \right. \\
 &\qquad \qquad \qquad b'_1 R[(1, x'_r)'(1, x'_s) - (1, \mu')'(1, \mu')] R' b_1] \\
 &\qquad \qquad \qquad \left. + \sum_{i=1}^n \mathbb{E}[(b'_1 R(1, x'_i)'(1, x'_i) R' b_1)^2] \right).
 \end{aligned}$$

To work out the combinatorics of the quadruple sum above, abbreviate $F_{ij} = b'_1 R[(1, x'_i)'(1, x'_j) - (1, \mu')'(1, \mu')] R' b_1$ and note that $\mathbb{E}[F_{ij}] = 0$ if $i \neq j$ and $\mathbb{E}[F_{ij} F_{rs}] = 0$ if all four indices are distinct. Moreover, there are only $O(n^3)$ summands in which not all four indices are distinct, i.e., there are only $O(n^3)$ non-zero summands. Moreover, the non-zero summands can always be bounded by

$$\begin{aligned}
 |E[F_{ij} F_{rs}]| &\leq \sqrt{\mathbb{E}[F_{ij}^2] \mathbb{E}[F_{rs}^2]} = \mathbb{E}[F_{ij}^2] = \text{Var}[F_{ij}] = \text{Var}[b'_1 R(1, x'_i)'(1, x'_j) R' b_1] \\
 &\leq \mathbb{E}[(1, x'_i) R' b_1)^2 (1, x'_j) R' b_1)^2] = (\mathbb{E}[(1, x'_1) R' b_1)^2])^2,
 \end{aligned}$$

if $i \neq j$ and $r \neq s$. Since $\mathbb{E}[(b'_1 R(1, x'_1)'(1, x'_1) R' b_1)^2] = b'_1 R S R' b_1 = b'_1 \Omega b_1 = 1$, by definition of B , we see that the quadruple sum in the second-to-last display is of order $O(n^3)$. The remaining sum in the same display is of order $O(n)$, since $\mathbb{E}[(b'_1 R(1, x'_1)'(1, x'_1) R' b_1)^4] = O(1)$, by part (i) and the assumption $\sup_{\|w\|=1} \mathbb{E}[|w' z_1|^4] = O(1)$. Thus, we have shown that $b'_1 R(1, \hat{\mu}')'(1, \hat{\mu}') R' b_1 - \|\hat{\mu}\|^2 \rightarrow 0$, in probability.

Next, consider $\|b'_1 R(1, \hat{\mu}')'(1, \hat{\mu}') R' B_1\|^2 \leq |b'_1 R(1, \hat{\mu}')'|^2 \|B'_1 R(1, \hat{\mu}')'\|^2$. The first factor in the upper bound was just shown to be $O_{\mathbb{P}}(1)$. For the second factor note that $\mathbb{E}[\|B'_1 R(1, \hat{\mu}')'\|^2] = \text{trace}((B'_1 \Sigma_W B_1/n + B'_1 R(1, \mu')'(1, \mu') R' B_1)) = (k_n - 1)/n$, by (C.2) and (C.3). Since $\|B'_1 R(1, \hat{\mu}')'(1, \hat{\mu}') R' B_1\|_S = \|B'_1 R(1, \hat{\mu}')'\|^2$, we see that the spectral norm in (C.5) is $O_{\mathbb{P}}(1)$ if $k_n = O(n)$, and converges to zero in probability, if $k_n/n \rightarrow 0$.

For the spectral norm in (C.4), we may restrict to $\mu = 0$. First, write $R = [t_1, T_1]$ with $t_1 \in \mathbb{R}^{k_n}$ and use (C.2) to see that $\mathbb{E}[b'_1 \hat{\Sigma}_W b_1] - (1 - \|\hat{\mu}\|^2) = b'_1 \Sigma_W b_1 (n - 1)/n - (1 - \|\hat{\mu}\|^2) = (1 - \|\hat{\mu}\|^2)/n \rightarrow 0$, whereas $\text{Var}[b'_1 \hat{\Sigma}_W b_1] = \text{Var}[b'_1 T_1 \hat{\Sigma}_n T_1' b_1] \rightarrow 0$, in view of the result in part (ii) with $v_n = v_{n,1} = v_{n,2} = T_1' b_1/n^{1/4}$, which satisfies $v_n' \Sigma v_n = b'_1 \Sigma_W b_1/\sqrt{n} = (1 - \|\hat{\mu}\|^2)/\sqrt{n} \rightarrow 0$. For the off-diagonal block $B'_1 \hat{\Sigma}_W b_1$, note that it has mean zero in view of (C.2). Therefore, $\mathbb{E}[\|B'_1 \hat{\Sigma}_W b_1\|^2] = \sum_{j=1}^{k_n-1} \mathbb{E}[(e_j' B'_1 \hat{\Sigma}_W b_1)^2] = \sum_{j=1}^{k_n-1} \text{Var}[e_j' B'_1 \hat{\Sigma}_W b_1]$, where e_1, \dots, e_{k_n-1} is the standard basis in \mathbb{R}^{k_n-1} . Now, $\text{Var}[e_j' B'_1 \hat{\Sigma}_W b_1] = \text{Var}[e_j' B'_1 T_1 \hat{\Sigma}_n T_1' b_1]$, and part (ii) applies with $v_{n,1} = T_1' B_1 e_j/n^{1/4}$ and $v_{n,2} = T_1' b_1/n^{1/4}$, which satisfy $v_{n,1}' \Sigma v_{n,1} = e_j' B'_1 \Sigma_W B_1 e_j/\sqrt{n} = 1/\sqrt{n}$ and $v_{n,2}' \Sigma v_{n,2} =$

$b_1' \Sigma_W b_1 / \sqrt{n} = (1 - \|\bar{\mu}\|^2) / \sqrt{n}$, in view of (C.2). Therefore, $\mathbb{E}[\|B_1' \hat{\Sigma}_W b_1\|^2] = \sum_{j=1}^{k_n-1} O(1/n) = O(k_n/n)$. Hence, the only remaining term is $\|B_1' \hat{\Sigma}_W B_1 - I_{k_n-1}\|_S \leq \|\frac{1}{n} \sum_{i=1}^n B_1' T_1 x_i x_i' T_1' B_1 - I_{k_n-1}\|_S + \|B_1' T_1 \hat{\mu} \hat{\mu}' T_1' B_1\|_S$. For the second term in the upper bound, one easily finds its expected value to be $(k_n - 1)/n$, as in the previous paragraph. For the spectral norm of the remaining covariance term we verify the strong regularity (SR) condition of Srivastava and Vershynin [23, Theorem 1.1] for the random $(k_n - 1)$ -vectors $\bar{x}_i = B_1' T_1 x_i = B_1' T_1 \Gamma z_i$. First, note that the \bar{x}_i are independent and isotropic, since $\mu = 0$ and $\mathbb{E}[\bar{x}_i \bar{x}_i'] = B_1' T_1 \Sigma T_1' B_1 = B_1' \Sigma_W B_1 = I_{k_n-1}$. Fix a projection matrix P in \mathbb{R}^{k_n-1} and note that $\Gamma' T_1' B_1 P B_1' T_1 \Gamma$ is a projection matrix in \mathbb{R}^{m_n} of the same rank as P . Since the z_i satisfy Assumption (A1).(c) and $\|P \bar{x}_1\|^2 = \|P B_1' T_1 \Gamma z_1\|^2 = z_1' \Gamma' T_1' B_1 P B_1' T_1 \Gamma z_1 = \|\Gamma' T_1' B_1 P B_1' T_1 \Gamma z_1\|^2$, we see that the (SR) condition holds for \bar{x}_1 and with the same constants c, C as in (A1).(c). Therefore, Corollary 1.4 of Srivastava and Vershynin [23] shows that $\|\frac{1}{n} \sum_{i=1}^n B_1' T_1 x_i x_i' T_1' B_1 - I_{k_n-1}\|_S$ is $O_{\mathbb{P}}(1)$ if $k_n = O(n)$, and converges to zero, in probability, if $k_n/n \rightarrow 0$. This finishes part (iii).

For the proof of parts (iv) and (v), take $\ell \in \mathbb{N}$ and consider the elementary bound

$$\begin{aligned} \mathbb{E}[|(1, x_1') R' \Omega^{-1} R(1, x_1')'|^\ell] &= \mathbb{E}[|(1, \mu') R' \Omega^{-1} R(1, \mu')' \\ &\quad + 2(1, \mu') R' \Omega^{-1} R(0, z_1' \Gamma')' + (0, z_1' \Gamma') R' \Omega^{-1} R(0, z_1' \Gamma')'|^\ell] \\ &\leq 2^{\ell-1} \left(\|\bar{\mu}\|^{2\ell} + 2^{\ell-1} \left\{ 2^\ell \mathbb{E} \left[|\bar{\mu}' \Omega^{-1/2} R(0, z_1' \Gamma')'|^\ell \right] \right. \right. \\ &\quad \left. \left. + \mathbb{E} \left[|(0, z_1' \Gamma') R' \Omega^{-1} R(0, z_1' \Gamma')'|^\ell \right] \right\} \right). \end{aligned} \quad (\text{C.6})$$

Partition $R = [t_1, T_1]$ as above and abbreviate $M = \Gamma' T_1' \Omega^{-1} T_1 \Gamma$, so that the expectation on the last line of the previous display can be written as $\mathbb{E}[|z_1' M z_1|^\ell]$. Now, if $\ell = 2$, this can be evaluated as $\mathbb{E}[|z_1' M z_1|^2] = \text{Var}[z_1' M z_1] + (\mathbb{E}[z_1' M z_1])^2 = O(\text{trace}(M^2)) + (\text{trace}(M))^2 o(1) + (\text{trace}(M))^2$ under the assumption of part (iv). Since $\text{trace}(M) = \text{trace}(\Omega^{-1/2} \Sigma_W \Omega^{-1/2}) = k_n - \|\bar{\mu}\|^2$ and $\text{trace}(M^2) = k_n - 1 + (1 - \|\bar{\mu}\|^2)$, by (C.1), we see that $\mathbb{E}[|z_1' M z_1|^2] = O(k_n^2)$. Furthermore, $\mathbb{E}[|\bar{\mu}' \Omega^{-1/2} R(0, z_1' \Gamma')|^2] = \bar{\mu}' \Omega^{-1/2} \Sigma_W \Omega^{-1/2} \bar{\mu} = \|\bar{\mu}\|^2 - \|\bar{\mu}\|^4 \leq 1/4$, which finishes part (iv).

For part (v) we begin with the expectation in (C.6) with $\ell = 4$, which can be written as

$$\begin{aligned} \mathbb{E}[|\bar{\mu}' \Omega^{-1/2} T_1 \Gamma z_1|^4] &\leq \|\bar{\mu}' \Omega^{-1/2} T_1 \Gamma\|^4 \sup_{\|w\|=1} \mathbb{E}[|w' z_1|^4] \\ &= O(\|\bar{\mu}' \Omega^{-1/2} \Sigma_W \Omega^{-1/2} \bar{\mu}\|^2) = O(1). \end{aligned}$$

For $\mathbb{E}[|z_1' M z_1|^4]$, we begin with case (a) $\mu = 0$. Then

$$\mathbb{E}[|z_1' M z_1|^4] = \mathbb{E} \left[\left| (0, z_1') \begin{pmatrix} 1 & 0 \\ 0 & \Gamma' \end{pmatrix} R' \left(R \begin{pmatrix} 1 & 0 \\ 0 & \Sigma \end{pmatrix} R' \right)^{-1} R \begin{pmatrix} 1 & 0 \\ 0 & \Gamma \end{pmatrix} (0, z_1')' \right|^4 \right],$$

and we denote the matrix corresponding to the quadratic form in the vector $(0, z_1)'$ on the right-hand-side of this display by P . Clearly, P is a projection matrix which we partition as

$$P = \begin{pmatrix} p_{11} & p'_{21} \\ p_{21} & P_{22} \end{pmatrix},$$

with $p_{11} \in [0, 1]$. Exploiting the idempotency and symmetry of P , one can show that the generalized Schur complement of p_{11} in P , i.e, the matrix $P_{22} - p_{21}p_{11}^\dagger p'_{21}$, is again a projection matrix [cf. 4, Corollary 2.1], where $p_{11}^\dagger = p_{11}^{-1}$, if $p_{11} \neq 0$, and $p_{11}^\dagger = 0$, else.¹⁴ Moreover, since $||P_{22}||_S - ||p_{21}p_{11}^\dagger p'_{21}||_S \leq ||P_{22} - p_{21}p_{11}^\dagger p'_{21}||_S \leq 1$ and $||P_{22}||_S \leq ||P||_S = 1$, we see that $||p_{21}(p_{11}^\dagger)^{1/2}||^2 = ||p_{21}p_{11}^\dagger p'_{21}||_S \leq 2$ and that the Frobenius norm of the generalized Schur complement satisfies $||P_{22} - p_{21}p_{11}^\dagger p'_{21}||_F^2 = \text{trace}((P_{22} - p_{21}p_{11}^\dagger p'_{21})) \leq \text{trace}(P_{22}) \leq \text{trace}(P) = k_n$. Therefore, using our assumptions, we calculate

$$\begin{aligned} \mathbb{E}[|z_1' M z_1|^4] &= \mathbb{E}[|z_1' P_{22} z_1|^4] \\ &\leq 2^3 \left(\mathbb{E}[|z_1' (P_{22} - p_{21}p_{11}^\dagger p'_{21}) z_1|^4] + \mathbb{E}[|z_1' p_{21}p_{11}^\dagger p'_{21} z_1|^4] \right) \\ &\leq 2^3 \left(O(k_n^4) + \mathbb{E}[|(p_{11}^\dagger)^{1/2} p'_{21} z_1|^8] \right) = O(k_n^4). \end{aligned}$$

Finally, in the case (b), where $\text{rank } T_1 = k_n$, the matrix $T_1 \Sigma T_1'$ in the representation $\Omega = R S R' = T_1 \Sigma T_1' + R(1, \mu')'(1, \mu')R'$, is regular and thus we can invert Ω by the Sherman-Morrison formula to get

$$\begin{aligned} M &= \Gamma' T_1' \Omega^{-1} T_1 \Gamma \\ &= \Gamma' T_1' (T_1 \Sigma T_1')^{-1} T_1 \Gamma - \frac{\Gamma' T_1' (T_1 \Sigma T_1')^{-1} R(1, \mu')'(1, \mu')' R' (T_1 \Sigma T_1')^{-1} T_1 \Gamma}{1 + (1, \mu')' R' (T_1 \Sigma T_1')^{-1} R(1, \mu')} \end{aligned}$$

Therefore, we make use of the abbreviations $P = \Gamma' T_1' (T_1 \Sigma T_1')^{-1} T_1 \Gamma$ and $v = \Gamma' T_1' (T_1 \Sigma T_1')^{-1} R(1, \mu')'$ to bound the fourth moment of the quadratic form $z_1' M z_1$ by

$$\begin{aligned} \mathbb{E}[|z_1' M z_1|^4] &\leq 2^3 \left(\mathbb{E}[|z_1' P z_1|^4] + \mathbb{E} \left[\left(\frac{|v' z_1|^2}{1 + ||v||^2} \right)^4 \right] \right) \\ &\leq 2^3 \left(O(||P||_F^8) + \left(\frac{||v||^2}{1 + ||v||^2} \right)^4 \sup_{||w||=1} \mathbb{E}[|w' z_1|^8] \right). \end{aligned}$$

Since $||P||_F^8 = (\text{trace}(P))^4 = k_n^4$, the upper bound is of order $O(k_n^4)$, which finishes the proof of part (v). □

¹⁴Baksalary, Baksalary and Szulc [4] actually prove a more general result. The special case we are interested in here can also be easily derived by direct calculation.

Lemma C.4. *Let $1 \leq q \leq p$ be positive integers. If T is a $(p + 1) \times (p + 1)$ orthogonal matrix that is partitioned as*

$$T = \begin{bmatrix} t_0 & T_0 \\ t_1 & T_1 \end{bmatrix},$$

where $t_0 \in \mathbb{R}^q$, $t_1 \in \mathbb{R}^{p+1-q}$, $T_0 \in \mathbb{R}^{q \times p}$ and $T_1 \in \mathbb{R}^{(p+1-q) \times p}$, then $\|t_0\| > 0$ if and only if, $\text{rank } T_1 = p + 1 - q$.

Proof. By orthogonality,

$$I_{p+1} = TT' = \begin{bmatrix} t_0 t_0' + T_0 T_0' & t_0 t_1' + T_0 T_1' \\ t_1 t_0' + T_1 T_0' & t_1 t_1' + T_1 T_1' \end{bmatrix},$$

and $\|t_0\|^2 + \|t_1\|^2 = 1$. Hence,

$$T_1 T_1' = I_{p+1-q} - t_1 t_1'$$

has eigenvalues 1, with multiplicity $p - q$, and a single eigenvalue $1 - \|t_1\|^2 = \|t_0\|^2$, which is strictly positive if and only if, $\text{rank } T_1 = p + 1 - q$. \square

Lemma C.5. *If the $n \times p$ random matrix X has i.i.d. rows following the $\mathcal{N}(0, \Sigma)$ -distribution with positive definite Σ , $v \in \mathbb{R}^p$, $v \neq 0$ and $T \in \mathbb{R}^{q \times p}$ has orthonormal rows, then $v'(T\hat{\Sigma}_n^{-1}T')^{-1}v \sim v'(T\Sigma^{-1}T')^{-1}v \chi_{n-1-(p-q)}^2/n$, where $\hat{\Sigma}_n = X'(I_n - P_t)X/n$ is the sample covariance matrix.*

Proof. It is well known that $n\hat{\Sigma}_n \sim \mathcal{W}_p(\Sigma, n-1)$ has a Wishart distribution with scale matrix Σ and $n-1$ degrees of freedom [e.g., 19, Theorem 3.4.4.(c)]. If $q = p$, then T is orthogonal and $nv'(T\hat{\Sigma}_n^{-1}T')^{-1}v = v'Tn\hat{\Sigma}_nT'v \sim v'T\Sigma T'v \chi_{n-1}^2 = v'(T\Sigma^{-1}T')^{-1}v \chi_{n-1}^2$ [cf. 19, Theorem 3.4.2]. So assume that $q < p$. Let $S \in \mathbb{R}^{(p-q) \times p}$ be such that $R = [S', T']'$ is an orthogonal matrix. Then, by block matrix inversion of

$$R\hat{\Sigma}_nR' = \begin{pmatrix} S\hat{\Sigma}_nS' & S\hat{\Sigma}_nT' \\ T\hat{\Sigma}_nS' & T\hat{\Sigma}_nT' \end{pmatrix} \sim \frac{1}{n}\mathcal{W}_p(R\Sigma R', n-1),$$

we see that the matrix $(T\hat{\Sigma}_n^{-1}T')^{-1} = ([0, I_q](R\hat{\Sigma}_nR')^{-1}[0, I_q]')^{-1} = T\hat{\Sigma}_nT' - T\hat{\Sigma}_nS'(S\hat{\Sigma}_nS')^{-1}S\hat{\Sigma}_nT'$ is the Schur complement of $S\hat{\Sigma}_nS'$ in $R\hat{\Sigma}_nR'$, which follows the $\mathcal{W}_q(\Omega_{22.1}, n-1-(p-q))$ -distribution divided by n , where $\Omega_{22.1} = T\Sigma T' - T\Sigma S'(S\Sigma S')^{-1}S\Sigma T'$ [cf. 19, Theorem 3.4.6.(a)]. Therefore, $nv'(T\hat{\Sigma}_n^{-1}T')^{-1}v \sim v'\Omega_{22.1}v \chi_{n-1-(p-q)}^2$, and $\Omega_{22.1} = (T\Sigma^{-1}T')^{-1}$. \square

Lemma C.6. *Let $\mu \in \mathbb{R}^p$ and Σ be a symmetric, positive definite $p \times p$ matrix. Let $T = [R'_0, R'_1]'$ be a $(p+1) \times (p+1)$ orthogonal matrix such that $R_0 \in \mathbb{R}^{q \times (p+1)}$ and set*

$$S = \begin{pmatrix} 1 & \mu' \\ \mu & \Sigma + \mu\mu' \end{pmatrix}, \quad \Omega = TST' = \begin{pmatrix} R_0SR'_0 & R_0SR'_1 \\ R_1SR'_0 & R_1SR'_1 \end{pmatrix} = \begin{pmatrix} \Omega_{00} & \Omega_{01} \\ \Omega_{10} & \Omega_{11} \end{pmatrix},$$

$$\Sigma_T = T \begin{pmatrix} 0 & 0 \\ 0 & \Sigma \end{pmatrix} T' = \begin{pmatrix} R_0 \\ R_1 \end{pmatrix} [0, I_p]' \Sigma [0, I_p] \begin{pmatrix} R'_0 & R'_1 \end{pmatrix} = \begin{pmatrix} \Sigma_{00} & \Sigma_{01} \\ \Sigma_{10} & \Sigma_{11} \end{pmatrix}.$$

If $\Sigma_{11} := R_1[0, I_p]' \Sigma [0, I_p] R_1'$ is regular, then the Schur complement of Ω_{11} in Ω is related to the Schur complement of Σ_{11} in Σ_T by $\Omega_{00} - \Omega_{01} \Omega_{11}^{-1} \Omega_{10} = \Sigma_{00} - \Sigma_{01} \Sigma_{11}^{-1} \Sigma_{10} + \tilde{\mu} \tilde{\mu}' / (1 + \nu)$, where $\tilde{\mu} = (R_0 - \Sigma_{01} \Sigma_{11}^{-1} R_1)(1, \mu)'$ and $\nu = (1, \mu') R_1' \Sigma_{11}^{-1} R_1 (1, \mu)'$.

Proof. First note that $\Omega_{ij} = \Sigma_{ij} + R_i(1, \mu')'(1, \mu)R_j$, for $i, j \in \{0, 1\}$. Abbreviate $\tilde{\mu}_i = R_i(1, \mu)'$, for $i = 0, 1$ and $\nu = \tilde{\mu}_1' \Sigma_{11}^{-1} \tilde{\mu}_1$ and use the Sherman-Morrison formula to write

$$\begin{aligned} \Omega_{00} - \Omega_{01} \Omega_{11}^{-1} \Omega_{10} &= \Sigma_{00} + \tilde{\mu}_0 \tilde{\mu}_0' - (\Sigma_{01} + \tilde{\mu}_0 \tilde{\mu}_1)(\Sigma_{11} + \tilde{\mu}_1 \tilde{\mu}_1')^{-1} (\Sigma_{10} + \tilde{\mu}_1 \tilde{\mu}_0') \\ &= \Sigma_{00} + \tilde{\mu}_0 \tilde{\mu}_0' - (\Sigma_{01} + \tilde{\mu}_0 \tilde{\mu}_1') \left(\Sigma_{11}^{-1} - \frac{\Sigma_{11}^{-1} \tilde{\mu}_1 \tilde{\mu}_1' \Sigma_{11}^{-1}}{1 + \nu} \right) (\Sigma_{10} + \tilde{\mu}_1 \tilde{\mu}_0') \\ &= \Sigma_{00} - \Sigma_{01} \Sigma_{11}^{-1} \Sigma_{10} \\ &\quad + \tilde{\mu}_0 \tilde{\mu}_0' - \tilde{\mu}_0 \tilde{\mu}_1' \Sigma_{11}^{-1} \Sigma_{10} - \Sigma_{01} \Sigma_{11}^{-1} \tilde{\mu}_1 \tilde{\mu}_0' - \tilde{\mu}_0 \tilde{\mu}_1' \Sigma_{11}^{-1} \tilde{\mu}_1 \tilde{\mu}_0' \\ &\quad + \Sigma_{01} \frac{\Sigma_{11}^{-1} \tilde{\mu}_1 \tilde{\mu}_1' \Sigma_{11}^{-1}}{1 + \nu} \Sigma_{10} + \tilde{\mu}_0 \tilde{\mu}_1' \frac{\Sigma_{11}^{-1} \tilde{\mu}_1 \tilde{\mu}_1' \Sigma_{11}^{-1}}{1 + \nu} \Sigma_{10} + \Sigma_{01} \frac{\Sigma_{11}^{-1} \tilde{\mu}_1 \tilde{\mu}_1' \Sigma_{11}^{-1}}{1 + \nu} \tilde{\mu}_1 \tilde{\mu}_0' \\ &\quad + \tilde{\mu}_0 \tilde{\mu}_1' \frac{\Sigma_{11}^{-1} \tilde{\mu}_1 \tilde{\mu}_1' \Sigma_{11}^{-1}}{1 + \nu} \tilde{\mu}_1 \tilde{\mu}_0' \\ &= \Sigma_{00} - \Sigma_{01} \Sigma_{11}^{-1} \Sigma_{10} + \tilde{\mu}_0 (1 - \nu) \tilde{\mu}_0' + \tilde{\mu}_0 \frac{\nu^2}{1 + \nu} \tilde{\mu}_0' \\ &\quad + \tilde{\mu}_0 \left(\frac{\nu}{1 + \nu} - 1 \right) \tilde{\mu}_1' \Sigma_{11}^{-1} \Sigma_{10} + \Sigma_{01} \Sigma_{11}^{-1} \tilde{\mu}_1 \left(\frac{\nu}{1 + \nu} - 1 \right) \tilde{\mu}_0' \\ &\quad + \Sigma_{01} \frac{\Sigma_{11}^{-1} \tilde{\mu}_1 \tilde{\mu}_1' \Sigma_{11}^{-1}}{1 + \nu} \Sigma_{10} \\ &= \Sigma_{00} - \Sigma_{01} \Sigma_{11}^{-1} \Sigma_{10} \\ &\quad + (\tilde{\mu}_0 \tilde{\mu}_0' - \tilde{\mu}_0 \tilde{\mu}_1' \Sigma_{11}^{-1} \Sigma_{10} - \Sigma_{01} \Sigma_{11}^{-1} \tilde{\mu}_1 \tilde{\mu}_0' + \Sigma_{01} \Sigma_{11}^{-1} \tilde{\mu}_1 \tilde{\mu}_1' \Sigma_{11}^{-1} \Sigma_{10}) / (1 + \nu) \\ &= \Sigma_{00} - \Sigma_{01} \Sigma_{11}^{-1} \Sigma_{10} + \frac{\tilde{\mu} \tilde{\mu}'}{1 + \nu}, \end{aligned}$$

where $\tilde{\mu} = \tilde{\mu}_0 - \Sigma_{01} \Sigma_{11}^{-1} \tilde{\mu}_1$. \square

Lemma C.7. Let k, n be positive integers such that $k < n - 1$. If X is a random $n \times k$ matrix whose rows are i.i.d. distributed according to $\mathcal{N}(\mu, \Sigma)$, where $\mu \in \mathbb{R}^k$ and Σ is positive definite, then

$$\frac{1}{n} \iota'(I_n - P_X) \iota \sim \frac{\xi}{\xi + \zeta},$$

where ξ and ζ are independent and distributed according to $\xi \sim \chi_{n-k}^2$ and $\zeta \sim \chi_k^2(\lambda_n)$, with non-centrality parameter $\lambda_n = n\mu'\Sigma^{-1}\mu$.

Remark. Lemma C.7 is a slight variation of Lemma A.2 in Leeb [17].

Proof. Note that $P_X = P_{\bar{X}}$, where $\bar{X} = X\Sigma^{-1/2}$ has i.i.d. rows following the $\mathcal{N}(\Sigma^{-1/2}\mu, I_n)$ distribution. Writing $\hat{\mu}_n = \bar{X}'\iota/n$ and $\hat{\Sigma}_n = \bar{X}'(I_n - P_\iota)\bar{X}/n =$

$\bar{X}'\bar{X}/n - \hat{\mu}_n\hat{\mu}_n'$, for the sample mean and sample covariance matrix of the transformed data, we have, at least on an event of probability one,

$$\begin{aligned} \frac{1}{n}\iota'(I_n - P_X)\iota &= 1 - \hat{\mu}_n'(\hat{\Sigma}_n + \hat{\mu}_n\hat{\mu}_n')^{-1}\hat{\mu}_n \\ &= 1 - \left[\hat{\mu}_n'\hat{\Sigma}_n^{-1}\hat{\mu}_n - \frac{(\hat{\mu}_n'\hat{\Sigma}_n^{-1}\hat{\mu}_n)^2}{1 + \hat{\mu}_n'\hat{\Sigma}_n^{-1}\hat{\mu}_n} \right] \\ &= \frac{1}{1 + \hat{\mu}_n'\hat{\Sigma}_n^{-1}\hat{\mu}_n}. \end{aligned}$$

Since $n\hat{\Sigma}_n$ has a standard Wishart distribution with $n-1$ degrees of freedom and is independent of $\hat{\mu}_n$, we get from Mardia, Kent and Bibby [19, Theorem 3.4.7] that, conditional on $\hat{\mu}_n$, the quantity $\hat{\mu}_n'\hat{\Sigma}_n^{-1}\hat{\mu}_n = n\|\hat{\mu}_n\|^2(\hat{\mu}_n/\|\hat{\mu}_n\|)'(n\hat{\Sigma}_n)^{-1} \times (\hat{\mu}_n/\|\hat{\mu}_n\|)$ has the same distribution as $n\|\hat{\mu}_n\|^2/\xi$, where $\xi \sim \chi_{n-k}^2$ is independent of $\hat{\mu}_n$. The proof is finished upon noting that $\zeta := n\|\hat{\mu}_n\|^2 = \|\bar{X}'\iota/\sqrt{n}\|^2 \sim \chi_k^2(n\mu'\Sigma^{-1}\mu)$ and that $1/(1 + \zeta/\xi) = \xi/(\xi + \zeta)$. \square

Lemma C.8. *Let $q_n \leq p_n + 1 < n$ be positive integer sequences such that $\limsup_n p_n/n < 1$, and let Δ_n be a non-negative real sequence of order $o(q_n/n)$. Moreover, let s_n and b_n be as in Theorem 2.1. Then a non-central F-distributed random variable $F_{q_n, n-p_n-1}(\lambda_n)$ with q_n and $n-p_n-1$ degrees of freedom and non-centrality parameter $\lambda_n = \Delta_n(n-p_n-1+q_n)$ satisfies*

$$\begin{aligned} s_n^{-1/2} (F_{q_n, n-p_n-1}(\lambda_n) - 1) - \sqrt{n}\Delta_n b_n \\ \xrightarrow[n \rightarrow \infty]{w} \begin{cases} \mathcal{N}(0, 1), & \text{if } q_n \rightarrow \infty, \\ (2q)^{-1/2}\chi_q^2 - \sqrt{q/2}, & \text{if } q_n = q, \quad \forall n \in \mathbb{N}. \end{cases} \end{aligned}$$

Remark. This result is elementary and follows from basic properties of the non-central χ^2 distribution [cf. 16, Chapter 29.5]. For the convenience of the reader we include a proof nonetheless. The proof also nicely resembles the main steps of the much more involved argument needed to treat the non-Gaussian cases of Theorem 2.1 and Theorem 3.1.

Proof. Let $Y_1, \dots, Y_{q_n}, X_1, \dots, X_{n-p_n-1}$ be i.i.d. standard normal and let $\mu = (\mu_1, \dots, \mu_{q_n})' \in \mathbb{R}^{q_n}$ such that $\mu'\mu = \lambda_n$. Then $F_{q_n, n-p_n-1}(\lambda_n)$ can be represented as

$$F_{q_n, n-p_n-1}(\lambda_n) \sim \frac{\sum_{i=1}^{q_n} (Y_i + \mu_i)^2 / q_n}{\sum_{j=1}^{n-p_n-1} X_j^2 / (n-p_n-1)}.$$

Therefore,

$$\begin{aligned} s_n^{-1/2} (F_{q_n, n-p_n-1}(\lambda_n) - 1) - \sqrt{n}\Delta_n b_n \\ \sim s_n^{-1/2} \frac{\frac{1}{q_n} \sum_{i=1}^{q_n} Y_i^2 - \frac{1}{n-p_n-1} \sum_{j=1}^{n-p_n-1} X_j^2 + \frac{2}{q_n} \sum_{i=1}^{q_n} Y_i \mu_i + \frac{\mu'\mu}{q_n}}{\frac{1}{n-p_n-1} \sum_{j=1}^{n-p_n-1} X_j^2} - \sqrt{n}\Delta_n b_n \end{aligned}$$

$$\begin{aligned}
 &= s_n^{-1/2} \left(\frac{1}{q_n} \sum_{i=1}^{q_n} Y_i^2 - \frac{1}{n-p_n-1} \sum_{j=1}^{n-p_n-1} X_j^2 \right) \left(1 + O_{\mathbb{P}}((n-p_n)^{-1/2}) \right) \\
 &\quad + \left(1 + O_{\mathbb{P}}((n-p_n)^{-1/2}) \right) \frac{2}{q_n \sqrt{s_n}} \sum_{i=1}^{q_n} Y_i \mu_i \\
 &\quad + \frac{\lambda_n}{q_n \sqrt{s_n}} \left(1 + O_{\mathbb{P}}((n-p_n)^{-1/2}) \right) - \sqrt{n} \Delta_n b_n.
 \end{aligned}$$

For the mixed term $\frac{2}{q_n \sqrt{s_n}} \sum_{i=1}^{q_n} Y_i \mu_i$ we note that it has mean zero and variance equal to $4(q_n^2 s_n)^{-1} \lambda_n = 4(q_n s_n)^{-1} \Delta_n (n/q_n)(n-p_n-1+q_n)/n = o(1)$, since $\Delta_n = o(q_n/n)$, by assumption, and $q_n s_n = 2(1+q_n/(n-p_n-1)) \geq 2$. So the mixed term converges to zero in probability. For the non-centrality term $\frac{\lambda_n}{q_n \sqrt{s_n}} (1 + O_{\mathbb{P}}((n-p_n)^{-1/2})) - \sqrt{n} \Delta_n b_n$ we first note that

$$\sqrt{n} \Delta_n b_n = \frac{n \Delta_n}{q_n \sqrt{s_n}} \frac{n-p_n-1+q_n}{n} = \frac{\lambda_n}{q_n \sqrt{s_n}},$$

so that

$$\begin{aligned}
 &\frac{\lambda_n}{q_n \sqrt{s_n}} \left(1 + O_{\mathbb{P}}((n-p_n)^{-1/2}) \right) - \sqrt{n} \Delta_n b_n = \frac{\lambda_n}{q_n \sqrt{s_n}} O_{\mathbb{P}}((n-p_n)^{-1/2}) \\
 &= \frac{n}{q_n} \Delta_n \frac{n-p_n-1+q_n}{n} (q_n s_n)^{-1/2} O_{\mathbb{P}}(\sqrt{q_n/(n-p_n)}) = o_{\mathbb{P}}(1),
 \end{aligned}$$

because $q_n/(n-p_n) \leq n/(n-p_n) = (1-p_n/n)^{-1}$ is bounded in view of our assumption that $\limsup_n p_n/n < 1$. The same assumption also guarantees that $n-p_n \rightarrow \infty$ as $n \rightarrow \infty$, so it remains to establish the appropriate convergence of

$$\begin{aligned}
 &s_n^{-1/2} \left(\frac{1}{q_n} \sum_{i=1}^{q_n} Y_i^2 - \frac{1}{n-p_n-1} \sum_{j=1}^{n-p_n-1} X_j^2 \right) \\
 &= s_n^{-1/2} \left(\frac{1}{q_n} \sum_{i=1}^{q_n} (Y_i^2 - 1) - \frac{1}{n-p_n-1} \sum_{j=1}^{n-p_n-1} (X_j^2 - 1) \right) \\
 &= \sqrt{\frac{n-p_n-1}{n-p_n-1+q_n}} \frac{1}{\sqrt{2q_n}} \sum_{i=1}^{q_n} (Y_i^2 - 1) \tag{C.7}
 \end{aligned}$$

$$- \sqrt{\frac{q_n}{n-p_n-1+q_n}} \frac{1}{\sqrt{2(n-p_n-1)}} \sum_{j=1}^{n-p_n-1} (X_j^2 - 1). \tag{C.8}$$

First, we consider the case where $q_n \rightarrow \infty$. Define the vectors

$$Z_n := \left(\begin{array}{c} \frac{1}{\sqrt{2q_n}} \sum_{i=1}^{q_n} (Y_i^2 - 1) \\ \frac{1}{\sqrt{2(n-p_n-1)}} \sum_{j=1}^{n-p_n-1} (X_j^2 - 1) \end{array} \right) \quad \text{and} \quad v_n := \left(\begin{array}{c} \sqrt{\frac{n-p_n-1}{n-p_n-1+q_n}} \\ -\sqrt{\frac{q_n}{n-p_n-1+q_n}} \end{array} \right).$$

By independence and the CLT, Z_n converges weakly to a bivariate standard normal distribution as $n \rightarrow \infty$ and $\|v_n\|_2 = 1$. Thus, $v_n'Z_n \rightarrow \mathcal{N}(0, 1)$, weakly, as $n \rightarrow \infty$, by compactness of the unit circle. If $q_n = q \in \mathbb{N}$ does not depend on n , then the expression in (C.8) converges to zero in probability and the expression in (C.7) is distributed as $(1 + o(1))[(2q)^{-1/2}\chi_q^2 - \sqrt{q/2}]$. \square

References

- [1] BAI, Z. D. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. [MR1399305](#)
- [2] BAI, Z. D. and YIN, Y. Q. (1993). Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Probab.* **21** 1275–1294. [MR1235416](#)
- [3] BAI, Z. and ZHOU, W. (2008). Large sample covariance matrices without independence structures in columns. *Statist. Sinica* **18** 425–442. [MR2411613](#)
- [4] BAKSALARY, J. K., BAKSALARY, O. M. and SZULC, T. (2004). Properties of Schur complements in partitioned idempotent matrices. *Linear Algebra Appl.* **379** 303–318. [MR2039745](#)
- [5] BHANSALI, R. J., GIRAITIS, L. and KOKOSZKA, P. S. (2007). Convergence of quadratic forms with nonvanishing diagonal. *Statist. Probab. Lett.* **77** 726–734. [MR2356512](#)
- [6] BHATIA, R. (1997). *Matrix Analysis. Graduate Texts in Mathematics* **169**. Springer, New York. [MR1477662](#)
- [7] BRYC, W. (1995). *The Normal Distribution: Characterizations with Applications*. Springer, New York. [MR1335228](#)
- [8] CAMBANIS, S., HUANG, S. and SIMONS, G. (1981). On the theory of elliptically contoured distributions. *J. Multivariate Anal.* **11** 368–385. [MR0629795](#)
- [9] DÜMBGEN, L. and CONTE-ZERIAL, P. (2013). On low-dimensional projections of high-dimensional distributions. In *From Probability to Statistics and Back: High-Dimensional Models and Processes. A Festschrift in Honor of Jon Wellner*, (M. Banerjee, F. Bunea, J. Huang, V. Koltchinskii and M. H. Maathius, eds.). *IMS Collections* **9** 91–104. IMS, Beachwood, Ohio. [MR3186751](#)
- [10] DVORETZKY, A. (1972). Asymptotic normality for sums of dependent random variables. In *Proc. 6th Berkeley Symp. Math. Statist. Probab.* **2** 513–535. University of California Press. [MR0415728](#)
- [11] EDELMAN, A. (1991). The distribution and moments of the smallest eigenvalue of a random matrix of Wishart type. *Linear Algebra Appl.* **159** 55–80. [MR1133335](#)
- [12] GHOSH, B. K. (1973). Some monotonicity theorems for χ^2 , F and t distributions with applications. *J. Roy. Statist. Soc. Ser. B* **35** 480–492. [MR0336872](#)
- [13] HELLAND, I. S. (1982). Central limit theorems for martingales with discrete or continuous time. *Scand. J. Statist.* **9** 79–94. [MR0668684](#)

- [14] HUBER, N. and LEEB, H. (2013). Shrinkage estimators for prediction out-of-sample: Conditional performance. *Commun. Statist. – Theory Methods* **42** 1246–1264. [MR3031279](#)
- [15] JOHNSON, L. N., KOTZ, S. and BALAKRISHNAN, N. (1995a). *Continuous Univariate Distributions 1*. Wiley, New York.
- [16] JOHNSON, L. N., KOTZ, S. and BALAKRISHNAN, N. (1995b). *Continuous Univariate Distributions 2*. Wiley, New York.
- [17] LEEB, H. (2009). Conditional predictive inference post model selection. *Ann. Statist.* **37** 2838–2876. [MR2541449](#)
- [18] LUGOSI, G. (2009). Concentration of measure inequalities. Lecture notes available online.
- [19] MARDIA, K. V., KENT, J. T. and BIBBY, J. M. (1995). *Multivariate Analysis*. Academic Press, London.
- [20] PORTNOY, S. (1985). On the central limit theorem in R^p when $p \rightarrow \infty$. *Probab. Theory Relat. Fields* **73** 571–583. [MR0863546](#)
- [21] PORTNOY, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16** 356–366. [MR0924876](#)
- [22] RAO, C. R. and TOUTENBURG, H. (1995). *Linear Models: Least Squares and Alternatives*. *Springer Series in Statistics*. Springer, New York. [MR1354840](#)
- [23] SRIVASTAVA, N. and VERSHYNIN, R. (2013). Covariance estimation for distributions with $2 + \varepsilon$ moments. *Ann. Probab.* **41** 3081–3111. [MR3127875](#)
- [24] STEINBERGER, L. and LEEB, H. (2014). On conditional moments of high-dimensional random vectors given lower-dimensional projections. *arXiv preprint arXiv:1405.2183*.
- [25] R DEVELOPMENT CORE TEAM (2011). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria ISBN 3-900051-07-0.
- [26] VAN AUBEL, A. and GAWRONSKI, W. (2003). Analytic properties of non-central distributions. *Appl. Math. Comput.* **141** 3–12. [MR1984225](#)
- [27] WANG, S. and CUI, H. (2013). Generalized F test for high dimensional linear regression coefficients. *J. Multivariate Anal.* **117** 134–149. [MR3053539](#)
- [28] WHITTLE, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Theory Probab. Appl.* **5** 302–305. [MR0133849](#)
- [29] 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](#)