Bootstrap Consistency for Quadratic Forms of Sample Averages with Increasing Dimension

Demian Pouzo *

UC Berkeley

November 17, 2014

Abstract

This paper establishes consistency of the weighted bootstrap for quadratic forms $\left(n^{-1/2}\sum_{i=1}^{n} Z_{i,n}\right)^{T}\left(n^{-1/2}\sum_{i=1}^{n} Z_{i,n}\right)$ where $(Z_{i,n})_{i=1}^{n}$ are mean zero, independent $\mathbb{R}^{d}$-valued random variables and $d = d(n)$ is allowed to grow with the sample size n, slower than $n^{1/4}$. The proof relies on an adaptation of Lindeberg interpolation technique whereby we simplify the original problem to a Gaussian approximation problem.

1 Introduction

Since its introduction by Efron (1979) the bootstrap has been widely used as a method for approximating the distribution of statistics. Many papers have extended the original idea in terms, both, of the applicability (see Horowitz (2001) and Hall (1986) for excellent reviews) and of its methodology; of particular interest for us are the bootstrap procedures: "wild bootstrap" (see Mammen (1993)) and more generally the "weighted bootstrap" (see Ma and Kosorok (2005)).

*Contact: Dept. of Economics at UC Berkeley. 530 Evans Hall # 3880. Berkeley, CA 94720. E-mail: dpouzo[at]econ.berkeley.edu. I would like to thank Xiaohong Chen, Noureddine El Karoui, Michael Jansson, Jim Powell and Elie Tamer for comments.

In this paper we attempt to expand the applicability of the weighted bootstrap procedure to quadratic forms with increasing dimensions. Namely, we study quadratic forms of the form

$$\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{i,n}\right)^T \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{i,n}\right) \tag{1}$$

where $(Z_{1,n}, ..., Z_{n,n})$ are independent (among each other) $\mathbb{R}^d$-valued random variables with mean zero and general covariance matrix Σ_n. We show that its distribution is well-approximated (under the Kolmogorov distance) by the distribution of

$$\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \omega_{i,n} Z_{i,n}\right)^T \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \omega_{i,n} Z_{i,n}\right) \tag{2}$$

where $(\omega_{1,n}, ..., \omega_{n,n})$ are independent *bootstrap weights*. The novelty in this paper is that we allow for $d = d(n)$ to increase with the sample size.

Studying the asymptotic behavior of quadratic forms, in particular establishing bootstrap consistency, is relevant since many statistics of interest can asymptotically be represented as quadratic forms of (scaled) sample averages. For instance, the likelihood ratio and Wald test statistics are asymptotically represented as quadratic forms of the scores; see Van der Vaart (2000) Ch. 16, and references therein. Portnoy (1988) establishes such representations for the likelihood ratio test statistics; there $d(n)$ is the dimension of the parameter of interest and is allowed to grow with n. Hjort *et al.* (2009) uses Portnoy's results to show a quadratic approximation result for Owen's (Owen (1990)) empirical likelihood, allowing for $d(n)^3/n \to 0$; see also Peng and Schick (2012). Therefore, by establishing the validity of the bootstrap for general quadratic forms, we propose an alternative method for inference for these statistics.

By letting d to increase with sample size we allow for different asymptotics, a "large-d and large-n" asymptotics, rather than the standard "fixed-d and large-n". The former type of asymptotics are more explicit about how the dimension, d, can affect the quality of the approximations. That is, even if the dimension of the parameters does not literally grow with n, if the model has a large number of parameters, doing "fixed-d large-n" asymptotics could be misleading, whereas doing "large-d large-n" asymptotics could depict a more accurate picture of the behavior for fixed samples; see Mammen (1989) for discussion. Our results can also be applied in cases where there is literally a growing number of parameters. For instance, Chen and Pouzo (2014) study the asymptotic behavior of the quasi-likelihood ratio and Wald test statistics in a semi-parametric conditional moment setup; in particular they show that the statistics are asymptotically equivalent to quadratic forms (1) under a null hypothesis of increasing dimensions (see Appendix A.4 in their paper); our results, in conjunction with theirs, could be applied to establish bootstrap-based inference for the quasi-likelihood ratio and Wald test statistics.

In order to establish our main result of bootstrap consistency, we use Lindeberg interpolation techniques (see Chatterjee (2006), Rollin (2013) and references therein) to approximate the quadratic forms of $n^{-1/2}\sum_{i=1}^{n}\omega_{i,n}Z_{i,n}$ and $n^{-1/2}\sum_{i=1}^{n}Z_{i,n}$ by the ones for Gaussian random variables with zero mean and variance $n^{-1}\sum_{i=1}^{n}Z_{i,n}Z_{i,n}^{T}$ and $E[Z_{1,n}Z_{1,n}^{T}]$, respectively.

By proceeding in this manner, we are able to reduce the original problem to a Gaussian approximation problem wherein we need to establish convergence of a Gaussian distribution with zero mean and variance $n^{-1}\sum_{i=1}^{n}Z_{i,n}Z_{i,n}^{T}$ to one with zero mean and variance $E[Z_{1,n}Z_{1,n}^{T}]$. We use Slepian interpolation (Slepian (1962), Rollin (2013), Chernozhukov *et al.* (2013a) and references therein) to accomplish this.

Due to the interpolation techniques used here, we need certain restrictions on the higher moments of the random variables. In particular, we impose growth restrictions on the higher moments of the bootstrap weights and the Euclidean norm of $Z_{1,n}$. These conditions essentially restrict the growth rate of $d(n)$. Although the precise growth rate depends on such conditions, the dimensions cannot grow faster $n^{1/4}$.

A number of papers develop large sample results allowing for increasing dimension. To name a few, Portnoy (1988) establishes the validity of the Wilks phenomenon for the likelihood ratio for exponential families when

$d(n)^{3/2}/n \to 0$. He and Shao (2000) derive the asymptotic distribution for M-estimators when the number of parameters is allowed to grow with the sample size. Recently, a few papers develop this type of results for quadratic forms of the form (1) allowing for increasing dimensions. In particular, Peng and Schick (2012) and Xu *et al.* (2014) develop a central limit theorems for quadratic forms of sample averages of vectors, allowing for the dimension to grow with n; both papers discuss several applications and examples. The results on our paper offer an alternative, bootstrap-based, method for inference for these cases.

Our paper also contributes to the growing literature of bootstrap results allowing for increasing dimensions. Mammen (1989) derives asymptotic expansion for M-estimators in linear models allowing for increasing dimension and use them to show consistency of a weighted bootstrap. In a different context, Radulovic (1998) uses Lindeberg interpolation methods allowing for increasing dimension to show that the functional bootstrap CLT holds under weaker conditions than equicontinuity; in his paper the restriction over the growth rate is $d(n)^6/n \to 0$. In Chernozhukov *et al.* (2013b), the authors derive a Gaussian weighted bootstrap approximation result for the *maximum* of the sum of high dimensional random vectors; in this specific setup the dimension is allowed to grow very fast, even at an exponential rate. Zhang and Cheng (2014) provide an extension of Chernozhukov *et al.* (2013b) to time series. In our paper the object of interest is the ℓ^2-norm of the sum of high dimensional random vectors (as opposed to the ℓ^∞-norm), so the results in these papers are not directly applicable. Finally, in a recent independent work, Spokoiny and Zhilova (2014) study the validity of the weighted bootstrap procedure for the likelihood ratio test statistics in finite samples and model misspecification; their results require $d(n)^3/n$ to be "small".

Organization of the Paper. In section 2 we define the problem and impose the required assumptions. Section 3 presents the main result and a discussion of its implications. Section 4 presents the proof of the main theorem. In order to keep the paper short, the proofs of intermediate results are gathered in the appendix.

Notation. For any vector $x \in \mathbb{R}^d$, we use $||x||_p^p$ to denote $\sum_{l=1}^d |x_l|^p$ and $x_{[l]}$ to denote the l-th coordinate of the vector; for $p = 2$ we use $||.||_e$. $tr\{A\}$ denotes the trace of matrix A. We use E_P to denote the expectation with respect to the probability measure P; for conditional distributions $P(\cdot|X)$ we use $E_{P(\cdot|X)}[\cdot]$ or sometimes directly $E_P[\cdot|X]$. We use $X_n \precsim Y_n$ to denote that $X_n \leq CY_n$ for some $C > 0$. We use $\partial^r f$ to denote the r-th derivative of f; for the cases of $r = 1$ and $r = 2$ we use the more standard f' and f'' notation.

2 Preliminaries

Let $\{Z_{i,n} \in \mathbb{R}^{d(n)} : i = 1, ..., n \text{ and } n \in \mathbb{N}\}$ with $(d(n))_{n\in\mathbb{N}}$ a non-decreasing real-valued sequence; $d(n)$ could diverge to infinity. For all $n \in \mathbb{N}$, let $Z^n \equiv (Z_{1,n}, ..., Z_{n,n})$ be independent among themselves with $Z_{i,n} \sim \mathbf{P}_n$ and $E_{\mathbf{P}_n}[(Z_{i,n})] = 0$ and $\Sigma_n \equiv E_{\mathbf{P}_n}[(Z_{i,n})(Z_{i,n})^T] \in \mathbb{R}^{d(n)\times d(n)}$ positive definite and finite.

Let $\mathbb{Z}_n \equiv n^{-1}\sum_{i=1}^n Z_{i,n}$, and

$$E_{\mathbf{P}_n}[(\sqrt{n}\mathbb{Z}_n)(\sqrt{n}\mathbb{Z}_n)^T] = n^{-1}\sum_{i=1}^n E_{\mathbf{P}_n}[(Z_{i,n})(Z_{i,n})^T] = \Sigma_n.$$

For a given matrix $A \in \mathbb{R}^{d\times d}$ we denote its Eigenvalues as $\{\lambda_1(A), ..., \lambda_d(A)\}$.

Assumption 2.1. *(i) There exists constants* $0 < c \leq C < \infty$ *such that* $c \leq \lambda_l(\Sigma_n) \leq C$ *for any* $l = 1, ..., d(n)$ *and* $n \in \mathbb{N}$, *and* $\frac{\max\{d(n)(E_{\mathbf{P}_n}[||Z_{1,n}||_e^3])^2, E_{\mathbf{P}_n}[||Z_{1,n}||_e^4], (d(n))^4\}}{n} = o(1)$*; (ii) there exists a* $\gamma > 0$ *such that* $\frac{(d(n))^{2+\gamma}}{n^\gamma} E_{\mathbf{P}_n}[||Z_{1,n}||_e^{4+2\gamma}] = o(1)$*; (iii) there exists a* $\kappa > 0$ *such that* $\frac{(\log(d(n)))^{\kappa/2} d(n)^{2+\kappa}}{n^{1+\kappa/2}} E_{\mathbf{P}_n}[||Z_{1,n}||_{2+\kappa}^{2(2+\kappa)}] = o(1)$.

2.1 Discussion of the assumption 2.1

The assumption that $c \leq \lambda_l(\Sigma_n) \leq C$ can be somewhat relaxed; for instance, it could be replaced by $\limsup_{n\to\infty} \frac{tr\{\Sigma_n^3\}}{(tr\{\Sigma_n^2\})^{3/2}} = 0$ and $\frac{tr\{\Sigma_n\}}{tr\{\Sigma_n^2\}} \leq C < \infty$. The rest of assumption 2.1 essentially imposed restrictions on the rate of growth of $d(n)$ relative to n. In order to shed more light on the implications of this part, and to provide sufficient conditions for it, is convenient to bound the quantities $E_{\mathbf{P}_n}[||Z_{1,n}||_e^{4+2\gamma}]$, etc in the assumption, in terms of $d(n)$.

Clearly, if $|Z_{[l],1,n}| \leq C < \infty$ a.s-$\mathbf{P}_n$ for all $l = 1, ..., d(n)$ and all $n \in \mathbb{N}$, then $E_{\mathbf{P}_n}[||Z_{1,n}||_e^{2q}] = O(d(n)^q)$ for any $q > 0$.[1] For example, such condition is imposed by Vershynin (2012a) in the context of estimation and approximation of covariance matrices of high dimensional distributions.

The next lemma shows that the result still holds if we impose the following (milder) restriction: $E_{\mathbf{P}_n}\left[e^{\lambda Z^2_{[l],1,n}}\right] \leq C < \infty$ for some $\lambda > 0$. For instance, if $(Z_{1,n,[l]})^2$ is a sub-Gamma random variable (Boucheron *et al.* (2013) p. 27), then the condition holds since $E_{\mathbf{P}_n}\left[e^{\lambda Z^2_{[l],1,n}}\right] \leq \exp\{\frac{\lambda^2 v}{2(1-c\lambda)}\}$ for any $\lambda \in (0, 1/c)$ and some $c > 0$. If $Z_{[l],1,n}$ is sub-Gaussian, then $(Z_{[l],1,n})^2$ is sub-exponential (see Vershynin (2012b) Lemma 5.14) and the condition holds by a similar argument.

An appealing feature of this result is that it only imposes restrictions on the marginal behavior of the components of the vector $Z_{1,n}$ and not its joint behavior.

Lemma 2.1. *Suppose that there exists a $C > 0$ and $\lambda > 0$ such $E_{\mathbf{P}_n}\left[e^{\lambda Z^2_{[l],1,n}}\right] \leq C$ for all $l = 1, ..., d(n)$ and all $n \in \mathbb{N}$. Then $E_{\mathbf{P}_n}[||Z_{1,n}||_e^{2q}] \precsim d(n)^q$ for any $q > 0$.*

Proof. Observe that

$$\begin{aligned} E_{\mathbf{P}_n}[(||Z_{1,n}||^2_e/d(n))^q] =& \int_0^\infty \mathbf{P}_n\left(||Z_{1,n}||^2_e/d(n) \geq t^{1/q}\right) dt \\ =& q\int_0^\infty u^{q-1}\mathbf{P}_n\left(||Z_{1,n}||^2_e/d(n) \geq u\right) du \end{aligned}$$

since $||Z_{1,n}||^2_e/d(n) = d(n)^{-1}\sum_{l=1}^{d(n)} |Z_{[l],1,n}|^2$, by the Markov inequality it follows that for any $\lambda > 0$

$$E_{\mathbf{P}_n}[(||Z_{1,n}||^2_e/d(n))^q] \leq \left(q\int_0^\infty u^{q-1}e^{-\lambda u}du\right) E_{\mathbf{P}_n}\left[e^{\lambda d(n)^{-1}\sum_{l=1}^{d(n)}|Z_{[l],1,n}|^2}\right].$$

[1] Recall that for a vector x, $x_{[l]}$ denotes the l-th component.

By Jensen inequality $E_{\mathbf{P}_n}\left[e^{\lambda d(n)^{-1}\sum_{l=1}^{d(n)}|Z_{[l],1,n}|^2}\right] \leq d(n)^{-1}\sum_{l=1}^{d(n)} E_{\mathbf{P}_n}\left[e^{\lambda|Z_{[l],1,n}|^2}\right] \leq C$. Thus, the desired result follows from the fact that $\left(q\int_0^\infty u^{q-1}e^{-\lambda u}du\right) = \left(q\lambda^{-q}\int_0^\infty w^{q-1}e^{-w}dw\right) = q\lambda^{-q}\Gamma(q) < \infty$ for any $q > 0$. $\square$

Therefore, assumption 2.1(i) boils down to $\frac{d(n)^4}{n} = o(1)$. For assumption 2.1(ii) is sufficient to impose $\frac{d(n)^{4+2\gamma}}{n^\gamma} = o(1)$; for $\gamma = 2$ it boils down to $\frac{d(n)^4}{n} = o(1)$ but for large γ it (roughly) becomes $\frac{d(n)^2}{n} = o(1)$. Finally, for, say $\kappa = 0$, assumption 2.1(iii) is reduced to $\frac{d(n)^2}{n}E_{\mathbf{P}_n}[||Z_{1,n}||_e^4] \precsim \frac{d(n)^4}{n} \to 0$.

That is, under conditions that bound all (polynomial) moments of the individual components of $Z_{1,n}$, the dimension is allowed to grow slower than the 4th-root of the sample size.

2.2 The Bootstrap Weights

The bootstrap weights are given by $\{\omega_{in} \in \mathbb{R} : i = 1,...,n \textit{ and } n \in \mathbb{N}\}$ where, for any $n \in \mathbb{N}$ and conditional on $Z^n = z^n$, $(\omega_{1n},...,\omega_{nn}) \sim \mathbf{P}_n^*(\cdot|z^n)$ for some $\mathbf{P}_n^*(\cdot|z^n)$.

Assumption 2.2. *For all $n \in \mathbb{N}$ and $i = 1,2,...,n$, (i) $(\omega_{1n},...,\omega_{nn})$ are independent and $E_{\mathbf{P}_n^*(\cdot|Z^n)}[\omega_{in}] = 0$ and $E_{\mathbf{P}_n^*(\cdot|Z^n)}\left[(\omega_{in}-1)^2\right] = 1$; (ii) there exists a $q \geq \max\{\gamma+2,4\}$, such that $E_{\mathbf{P}_n^*(\cdot|Z^n)}[|\omega_{in}|^q] \leq C_w < \infty$ for some constant $C_w > 0$.*

Part (i) is standard. Part (ii) is mild considering that the weights are chosen by the researcher. The technique of proof can easily be adapted to the case where the following (stronger) restriction is imposed: $E_{\mathbf{P}_n^*(\cdot|Z^n)}[\exp\{\omega_{in}\}] \leq C_w < \infty$.

3 The Main Result

We now present the main result of the paper. In what follows, for any measurable function $z^n \mapsto f(z^n)$ we use $|f(Z^n)| = o_{\mathbf{P}_n}(1)$ to denote: For any $\varepsilon > 0$, there exists a $N(\varepsilon)$ such that for all $n \geq N(\varepsilon)$, $\mathbf{P}_n(|f(Z^n)| \geq \varepsilon) < \varepsilon$.

Let $\mathbb{Z}_n^* \equiv n^{-1}\sum_{i=1}^n \omega_{i,n}Z_{i,n}$ be the bootstrap analog of $\mathbb{Z}_n$.

Theorem 3.1. *Suppose assumption 2.1 and 2.2 hold. Then*

$$\sup_{t\in\mathbb{R}} \left|\mathbf{P}_n^*\left(||\sqrt{n}\mathbb{Z}_n^*||_e^2 \geq t \mid Z^n\right) - \mathbf{P}_n\left(||\sqrt{n}\mathbb{Z}_n||_e^2 \geq t\right)\right| = o_{\mathbf{P}_n}(1).$$

3.1 Comments and discussion

We now present some remarks and discuss some implications of the preceding theorem.

Heuristics. We postpone the somewhat long proof of the theorem to section 4; here we present an heuristic argument. The first step in the proof is to approximate the indicator function $x \mapsto 1\{||x||_e^2 \geq t\}$ by "smooth" functions $x \mapsto \mathcal{P}_{t,\delta,h}(||x||_e^2)$; the exact expression for $\mathcal{P}_{t,\delta,h}$ is presented in lemma B.1 and follows from the suggestion by Pollard (2001) p. 247. The functions are indexed by (h,δ) where h is "small" compared to δ and as $\delta \to 0$ the function $\mathcal{P}_{t,\delta,h}$ converges to the indicator function

The second step uses the fact that $\mathcal{P}_{t,\delta,h}$ belongs to a class of "smooth" functions, and applies Lindeberg interpolation techniques (Chatterjee (2006) and Rollin (2013) among others) to approximate $\sqrt{n}\mathbb{Z}_n^*$ by $\sqrt{n}\mathbb{U}_n \equiv n^{-1/2}\sum_{i=1}^n U_{i,n}$ and $\sqrt{n}\mathbb{Z}_n$ by $\sqrt{n}\mathbb{V} \equiv n^{-1/2}\sum_{i=1}^n V_{i,n}$, where $(U_{i,n})_{i=1}^n$ are independent Gaussian with zero mean and variance $Z_{i,n}Z_{i,n}^T$ and $(V_{i,n})_{i=1}^n$ are independent Gaussian with zero mean and variance $E[Z_{1,n}Z_{1,n}^T]$. We use $\mathbf{\Phi}_n^*(\cdot|Z^n)$ and $\mathbf{\Phi}_n$ respectively, to denote their probability distributions. The following theorems formalize this, and can be viewed of independent interest since they show that a "generalized invariance principle" holds in our setup (formal proofs of these theorems are relegated to Appendix A).

Theorem 3.2. *Suppose assumption 2.1 and 2.2 hold. For any $h > 0$,*

$$\sup_{f\in\mathcal{C}_{h^{-1}}} \left|E_{\mathbf{P}_n^*}\left[f\left(||\sqrt{n}\mathbb{Z}_n^*||_e^2\right)|Z^n\right] - E_{\mathbf{\Phi}_n^*}\left[f\left(||\sqrt{n}\mathbb{U}_n||_e^2\right)|Z^n\right]\right| = o_{\mathbf{P}_n}(h^{-2})$$

where $\mathcal{C}_M$ be the class of functions $f:\mathbb{R}\to\mathbb{R}$ that are three times continuously differentiable and $\sup_x|\partial^r f(x)| \leq (M)^r$ and $\sup_x|f(x)| \leq 1$.

Theorem 3.3. *Suppose assumption 2.1 and 2.2 hold. For any $h > 0$,*

$$\sup_{f \in \mathcal{C}_{h^{-1}}} \left| E_{\mathbf{P}_n} \left[f \left(||\sqrt{n}\mathbb{Z}_n||_e^2 \right) \right] - E_{\mathbf{\Phi}_n} \left[f \left(||\sqrt{n}\mathbb{V}_n||_e^2 \right) \right] \right| = o(h^{-2}).$$

By using theorems 3.2 and 3.3 we have reduced the original problem to a Gaussian approximation problem. That is, we need to establish convergence (under the distance induced by $\mathcal{C}$) of a Gaussian distribution with zero mean and variance $n^{-1} \sum_{i=1}^n Z_{i,n} Z_{i,n}^T$ to one with zero mean and variance $E[Z_{1,n} Z_{1,n}^T]$. Lemma 4.3 in Section 4 — which is based in the Slepian interpolation (Chernozhukov *et al.* (2013b), Chernozhukov *et al.* (2013a) and Rollin (2013) and references therein)— establishes that is enough to show that

$$d(n) \max_{1 \leq j,l \leq d(n)} \left| n^{-1} \sum_{i=1}^n Z_{[j],i,n} Z_{[l],i,n} - E_{\mathbf{P}_n}[Z_{[j],1,n} Z_{[l],1,n}] \right| = o_{\mathbf{P}_n}(1). \quad (3)$$

A similar result is obtained by Chernozhukov *et al.* (2013b) without the scaling factor of $d(n)$; their setup, however, is different since the object of interest is $\max_{1 \leq j \leq d(n)} |n^{-1/2} \sum_{i=1}^n Z_{[j],i,n}|$ (as opposed to $||n^{-1/2} \sum_{i=1}^n Z_{i,n}||_e^2$). Below we show that, employing standard arguments, the expression 3 holds under our assumptions.

Asymptotic Distribution of $||\sqrt{n}\mathbb{Z}_n||_e^2$. An implication of the proof of Theorem 3.1 and Theorem 3.3 is that

$$\sup_{t \in \mathbb{R}} \left| \mathbf{P}_n \left(\frac{||\sqrt{n}\mathbb{Z}_n||_e^2 - d(n)}{\sqrt{d(n)}} \geq t \right) - \mathbf{\Phi}_n \left(\frac{||\sqrt{n}\mathbb{V}_n||_e^2 - d(n)}{\sqrt{d(n)}} \geq t \right) \right| = o(1).$$

That is, if $\Sigma_n = I_{d(n)}$ then this expression and a direct application of the CLT (when $d(n) \to \infty$) imply that $\frac{||\sqrt{n}\mathbb{Z}_n||_e^2 - d(n)}{\sqrt{2d(n)}} \Rightarrow N(0,1)$ or, informally, $||\sqrt{n}\mathbb{Z}_n||_e^2$ is approximately chi-square distributed with $d(n)$ degrees of freedom. When $\Sigma_n \neq I_{d(n)}$, the last claim is no longer true but it holds that $\frac{||\sqrt{n}\mathbb{Z}_n||_e^2 - tr\{\Sigma_n\}}{\sqrt{2tr\{\Sigma_n^2\}}}$ is approximately distributed as $\sum_{j=1}^{d(n)} \frac{\lambda_j(\Sigma_n)(\chi_j - 1)}{\sqrt{2\sum_{j=1}^{d(n)} \lambda_j^2(\Sigma_n)}}$ with χ_j^2 drawn from a chi-square with degree one; see Xu *et al.* (2014) and Peng and Schick (2012) for a discussion regarding these results.

We note that in Theorem 3.1 no scaling (by $-d(n)$ and $1/\sqrt{2d(n)}$ or $-tr\{\Sigma_n\}$ and $1/\sqrt{2tr\{\Sigma_n^2\}}$) is needed. That is, although the mean and variance of $||\sqrt{n}\mathbb{Z}_n||_e^2$ are "drifting" to infinity, the bootstrap still provides a good approximation since the moments of $||\sqrt{n}\mathbb{Z}_n^*||_e^2$ are mimicking this behavior.

On the Lindeberg Interpolation. Theorems 3.2 and 3.3 are based on the following Lindeberg interpolation for quadratic forms (formal proofs of this theorem are relegated to Appendix A).[2]

Theorem 3.4. *Let $(A_1, ..., A_n) \in \mathbb{R}^{d\times n}$ and $(B_1, ..., B_n) \in \mathbb{R}^{d\times n}$ be random matrices independent from each other. Suppose for each $1 \leq i \leq n$, A_i has finite second moments with $E[A_i] = 0$, $A_1, ..., A_n$ are independent, and B_i has finite second moments, with $E[B_i] = 0$ and $B_1, ..., B_n$ are independent. Suppose $E[A_iA_i^T] = E[B_iB_i^T] \equiv C_i$. Let $f : \mathbb{R} \to \mathbb{R}$ be three times differentiable and for $r = 1, 2, 3$, $|\partial^r f(\cdot)| \leq L_r(f)$. Then for any $\epsilon > 0$ and for any $q > 0$*

$$|E[f(||\sum_{i=1}^{n} A_i||_e^2)] - E[f(||\sum_{i=1}^{n} B_i||_e^2)]| \leq \mathbf{S}_n + L_2(f)\left(\frac{L_3(f)}{L_2(f)}\right)^q \mathbf{R}_n$$

where $\mathbf{S}_n = \mathbf{S}_{1,n} + \mathbf{S}_{2,n}$, with

$$\begin{aligned}
\mathbf{S}_{1,n} =& \sum_{i=1}^{n} |E\left[f''\left(||\mathbb{S}_{i:n}||_e^2\right)\right] E[||B_i||_e^4] - E[||A_i||_e^4]| \\
\mathbf{S}_{2,n} =& 4\sum_{i=1}^{n} |E\left[f''\left(||\mathbb{S}_{i:n}||_e^2\right)\mathbb{S}_{i:n}^T\right]\left(E[B_i||B_i||_e^2] - E[A_i||A_i||_e^2]\right)| \\
\mathbf{R}_n =& \sum_{i=1}^{n} E\left[\left(\mathbb{S}_{i:n}^T B_i + ||B_i||_e^2\right)^{2+q} + \left(\mathbb{S}_{i:n}^T A_i + ||A_i||_e^2\right)^{2+q}\right]
\end{aligned}$$

and $\mathbb{S}_{i:n} \equiv \sum_{j=1}^{i-1} A_j + 0 + \sum_{j=i+1}^{n} B_j$.

A few remarks regarding this theorem are in order. First, in lemma A.1

[2] This Lindeberg interpolation builds on the approach in Xu *et al.* (2014).

in the Appendix we provide bounds for $\mathbf{S}_n$ (and $\mathbf{R}_n$). These bounds only use restrictions imposed on the higher moments of the original data and the bootstrap weights (see assumptions 2.1(i)(ii) and 2.2). However, it is easy to see that if one would have additional information on the higher moments, one could obtain sharper bounds for $\mathbf{S}_n$. For instance, to show Theorem 3.2, we apply theorem 3.4 with $A_i = n^{-1/2}\omega_{i,n}Z_{i,n}$ and $B_i = n^{-1/2}u_i Z_{i,n}$ with $u_i \sim N(0,1)$. If we would have that $(\omega_{i,n})_{i=1}^n$ were such that $E[|\omega_{i,n}|^4] = E[(z)^4]$ with $z \sim N(0,1)$, then $\mathbf{S}_{1,n} = 0$ (a similar observation applies to $\mathbf{S}_{2,n}$). These bounds in $\mathbf{S}_n$, in turn, will translate to faster rates of the bootstrap approximation.

Second, the interpolation compares the quantities $\sum_{i=1}^n A_i$ with $\sum_{i=1}^n B_i$ by comparing "one component at a time". This comparison is essentially divided into two parts. First, we compare $||\mathbb{S}_{i:n} + A_i||_e^2$ and $||\mathbb{S}_{i:n} + B_i||_e^2$, which are real-valued quantities. Second, we exploit the smoothness of the *univariate* function f to bound its variation using Taylor's approximation. Loosely speaking, the first step reduces the problem to an univariate one. An alternative approach would be to consider interpolations for *multivariate* functions (e.g. Chatterjee and Meckes (2008)) of the form $g : \mathbb{R}^{d(n)} \to \mathbb{R}$ with $g(x) \equiv f(||x||_e^2)$. As can be seen from the derivations in Chatterjee and Meckes (2008), the reminder term will also require bounds on higher derivatives of g (and thus f), but of the form $\sup_{x\neq y} \frac{||Hess(g)(x) - Hess(g)(y)||_{op}}{||x-y||_e}$. [3] Which approach is better depends largely on what type of restrictions over the class of test functions are natural in the problem at hand. For us, $||\partial^r f||_{L^\infty} < \infty$ is a natural assumption, but in other applications it could be too strong.

More generally, this discussion illustrates the relationship between restrictions in the class of test functions ($\mathcal{C}$) and the bounds on higher order moments and ultimately the rate of growth of $d(n)$.

Bootstrap P-Value. For any $\alpha \in (0,1)$ and $Z^n \in \mathbb{R}^{d(n)}$, let $t_n(\alpha, Z^n) \equiv \inf\{t : \mathbf{P}_n^* \left(||\sqrt{n}\mathbb{Z}_n^*||_e^2 \leq t \mid Z^n\right) \geq \alpha\}$. Due to the distribution consistency

[3] $Hess(g)$ is the Hessian of the function and $||.||_{op}$ is the operator norm. Other type of bounds could be found in Raic (2004) based on Hilbert-Schmidt norm.

result proven in Theorem 3.1, we can approximate the α-th quantile of the distribution of $||\sqrt{n}\mathbb{Z}_n||_e^2$ by $t_n(\alpha, Z^n)$, in the sense that

$$\mathbf{P}_n\left(||\sqrt{n}\mathbb{Z}_n||_e^2 \geq t_n(\alpha, Z^n) - \eta\right) \leq \alpha + o(1)$$

for any $\eta > 0$. If $t_n(\alpha, Z^n)$ is a continuity point of $\mathbf{P}_n^*(\cdot|Z^n)$, then $\mathbf{P}_n^*\left(||\sqrt{n}\mathbb{Z}_n^*||_e^2 \geq t_n(\alpha, Z^n) \mid Z^n\right) = \alpha$, and the previous display becomes $\mathbf{P}_n\left(||\sqrt{n}\mathbb{Z}_n||_e^2 \geq t_n(\alpha, Z^n)\right) = \alpha + o(1)$. Hence, Theorem 3.1 can be used to construct valid p-values based on the bootstrap.

4 Proof of Theorem 3.1

Recall that $x \in \mathbb{R}^{d(n)} \mapsto ||x||_e^2 \equiv x^T x$ and that $\mathcal{C}_M$ is the class of functions $f : \mathbb{R} \to \mathbb{R}$ that are three times continuously differentiable and $\sup_x |\partial^r f(x)| \leq (M)^r$.

All the proofs of the lemmas in this section are relegated to Appendix B.

For any two probability measures Q and P, let

$$\Delta_M(P, Q) \equiv \sup_{f \in \mathcal{C}_M} |E_P[f(||X||_e^2)] - E_Q[f(||Y||_e^2)]|. \quad (4)$$

We want to establish the following: For any $\varepsilon' > 0$, there exists a $N(\varepsilon')$ such that

$$\mathbf{P}_n\left(\sup_{t\in\mathbb{R}}\left|\mathbf{P}_n^*\left(||\sqrt{n}\mathbb{Z}_n^*||_e^2 \geq t \mid Z^n\right) - \mathbf{P}_n\left(||\sqrt{n}\mathbb{Z}_n||_e^2 \geq t\right)\right| \geq \varepsilon'\right) < \varepsilon'$$

for all $n \geq N(\varepsilon)$. Observe that

$$\begin{aligned}
&\mathbf{P}_n\left(\sup_{t\in\mathbb{R}}\left|\mathbf{P}_n^*\left(||\sqrt{n}\mathbb{Z}_n^*||_e^2 \geq t \mid Z^n\right) - \mathbf{P}_n\left(||\sqrt{n}\mathbb{Z}_n||_e^2 \geq t\right)\right| \geq \varepsilon'\right) \\
\leq &\mathbf{P}_n\left(\{\sup_{t\in\mathbb{R}}\left|\mathbf{P}_n^*\left(||\sqrt{n}\mathbb{Z}_n^*||_e^2 \geq t \mid Z^n\right) - \mathbf{P}_n\left(||\sqrt{n}\mathbb{Z}_n||_e^2 \geq t\right)\right| \geq \varepsilon'\} \cap S_n\right) \\
&+ \mathbf{P}_n\left(S_n^C\right)
\end{aligned}$$

where $S_n \equiv \{Z^n : n^{-1}\sum_{i=1}^n ||Z_{i,n}||_e^2 \leq (0.5\varepsilon')^{-1} tr\{\Sigma_n\}\}$. By the Markov inequality $\mathbf{P}_n\left(S_n^C\right) \leq 0.5\varepsilon'$. Thus, it suffices to show that

$$\mathbf{P}_n\left(\left\{\sup_{t\in\mathbb{R}}\left|\mathbf{P}_n^*\left(||\sqrt{n}\mathbb{Z}_n^*||_e^2 \geq t \mid Z^n\right) - \mathbf{P}_n\left(||\sqrt{n}\mathbb{Z}_n||_e^2 \geq t\right)\right| \geq \varepsilon'\right\} \cap S_n\right) < 0.5\varepsilon' \tag{5}$$

By the triangle inequality, for all $t \in \mathbb{R}$ and Z^n

$$\begin{aligned}&|E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_e^2 \geq t\} \mid Z^n\right] - E_{\mathbf{P}_n}\left[1\{||\sqrt{n}\mathbb{Z}_n||_e^2 \geq t\}\right]| \\ \leq &|E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_e^2 \geq t\} \mid Z^n\right] - E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2 \geq t\}\right]| \\ &+ |E_{\mathbf{P}_n}\left[1\{||\sqrt{n}\mathbb{Z}_n||_e^2 \geq t\}\right] - E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2 \geq t\}\right]|\end{aligned}$$

where $\mathbb{V}_n \equiv n^{-1}\sum_{i=1}^n V_{i,n}$ with $V_{i,n} \sim i.i.d. - N(0,\Sigma_n)$. We use $\mathbf{\Phi}_n$ to denote the probability of $(V_{i,n})_{i=1}^n$.

Therefore, in order to obtain display 5, it suffices to bound

$$\mathbf{P}_n\left(\{\sup_{t\in\mathbb{R}}\left|E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_e^2 \geq t\} \mid Z^n\right] - E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2 \geq t\}\right]\right| \geq 0.5\varepsilon'\} \cap S_n\right) < 0.25\varepsilon' \tag{6}$$

and

$$\lim_{n\to\infty}\sup_{t\in\mathbb{R}}\left|E_{\mathbf{P}_n}\left[1\{||\sqrt{n}\mathbb{Z}_n||_e^2 \geq t\}\right] - E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2 \geq t\}\right]\right| = 0 \tag{7}$$

The next two lemmas allow us to "replace" the indicator functions by "smooth" functions.

Lemma 4.1. *Suppose assumption 2.1(i) holds. For any $\varepsilon > 0$, there exists a $\gamma(\varepsilon)$ and $N(\varepsilon)$ such that for all $n \geq N(\varepsilon)$ and all $h \leq h(\varepsilon, \sqrt{tr\{\Sigma_n^2\}}\gamma(\varepsilon))$*

$$\sup_{t\in\mathbb{R}}\left|E_{\mathbf{P}_n}\left[1\{||\sqrt{n}\mathbb{Z}_n||_e^2 \geq t\}\right] - E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2 \geq t\}\right]\right| \tag{8}$$

$$\leq \frac{\varepsilon}{1-\varepsilon} + 3\varepsilon + \Delta_{h^{-1}}(\mathbf{P}_n,\mathbf{\Phi}_n). \tag{9}$$

where, recall that, $\Delta_{h^{-1}}(\mathbf{P}_n,\mathbf{\Phi}_n) = \sup_{f\in\mathcal{C}_{h^{-1}}}\left|E_{\mathbf{P}_n}\left[f\left(||\sqrt{n}\mathbb{Z}_n||_e^2\right)\right] - E_{\mathbf{\Phi}_n}\left[f\left(||\sqrt{n}\mathbb{V}_n||_e^2\right)\right]\right|$.

And

Lemma 4.2. *Suppose assumption 2.1(i) holds. For any $\varepsilon > 0$, there exists a $\gamma(\varepsilon)$ and $N(\varepsilon)$ such that for all $n \geq N(\varepsilon)$ and all $h \leq h(\varepsilon, \sqrt{tr\{\Sigma_n^2\}}\gamma(\varepsilon))$*

$$\begin{aligned} &\sup_{t \in \mathbb{R}} \left| E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_e^2 \geq t\} | Z^n \right] - E_{Pr}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2 \geq t\} \right] \right| \\ \leq & \frac{\varepsilon}{1-\varepsilon} + 3\varepsilon + \Delta_{h^{-1}}(\mathbf{P}_n^*(\cdot|Z^n), \mathbf{\Phi}_n), \end{aligned} \tag{10}$$

for any $Z^n \in \mathbb{R}^{d(n)}$.

where, recall that,

$$\Delta_{h^{-1}}(\mathbf{P}_n^*(\cdot|Z^n), \mathbf{\Phi}_n) = \sup_{f \in \mathcal{C}_{h^{-1}}} \left| E_{\mathbf{P}_n^*}\left[f\left(||\sqrt{n}\mathbb{Z}_n^*||_e^2 \right) | Z^n \right] - E_{\mathbf{\Phi}_n}\left[f\left(||\sqrt{n}\mathbb{V}_n||_e^2 \right) \right] \right| \tag{11}$$

Remark 4.1. *The previous lemma holds for any h provided that is below $h \leq h(\varepsilon, \sqrt{tr\{\Sigma_n^2\}}\gamma(\varepsilon))$. The intuition from this restriction is as follows: h and $\delta_n \equiv \sqrt{tr\{\Sigma_n^2\}}\gamma(\varepsilon)$ index the "smooth" function we use to approximate $x \mapsto 1\{||x||_e^2 \geq t\}$; see lemma B.1 in the Appendix for a precise expression. It turns out that h has to be "small" relative to δ_n. Therefore, we need the bound $h(\varepsilon, \delta_n)$.*

It is worth to note that, for the "smooth" function to be a good approximation of $1\{||\cdot||_e^2 \geq t\}$, we need δ_n to be "small" (see the proof of lemma 4.2 in the Appendix). What we mean by δ_n to be "small" depends on how $||\sqrt{n}\mathbb{V}_n||_e^2$ concentrates mass. Lemma B.4 establishes an anti-concentration result, wherein we obtain that this random variable puts very little mass in any given interval. Therefore δ_n could actually be quite large, of the order of $\sqrt{tr\{\Sigma_n^2\}}$.

Therefore, by letting ε in the lemmas be such that $\frac{\varepsilon}{1-\varepsilon} + 3\varepsilon = 0.25\varepsilon'$ we obtain

$$\sup_{t \in \mathbb{R}} \left| E_{\mathbf{P}_n}\left[1\{||\sqrt{n}\mathbb{Z}_n||_e^2 \geq t\} \right] - E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2 \geq t\} \right] \right| \leq 0.25\varepsilon' + \Delta_{h^{-1}}(\mathbf{P}_n, \mathbf{\Phi}_n) \tag{12}$$

and

$$\mathbf{P}_n\left(\{\sup_{t\in\mathbb{R}}\left|E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_e^2\geq t\}\mid Z^n\right]-E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2\geq t\}\right]\right|\geq 0.5\varepsilon'\}\cap S_n\right)$$
$$\leq \mathbf{P}_n\left(\{\Delta_{h^{-1}}(\mathbf{P}_n^*(\cdot|Z^n),\mathbf{\Phi}_n)\geq 0.25\varepsilon'\}\cap S_n\right) \tag{13}$$

for all $n\geq N(\varepsilon)$ and all $h\leq h(\varepsilon,\delta_n)$.

By the triangle inequality and straightforward algebra, it follows that

$$\begin{aligned}
&\mathbf{P}_n\left(\{\Delta_{h^{-1}}(\mathbf{P}_n^*(\cdot|Z^n),\mathbf{\Phi}_n)\geq 0.25\varepsilon'\}\cap S_n\right)\\
&\leq \mathbf{P}_n\left(\{\Delta_{h^{-1}}(\mathbf{P}_n^*(\cdot|Z^n),\mathbf{\Phi}_n^*(\cdot|Z^n))\geq \frac{1}{8}\varepsilon'\}\cap S_n\right)\\
&\ + \mathbf{P}_n\left(\{\Delta_{h^{-1}}(\mathbf{\Phi}_n^*(\cdot|Z^n),\mathbf{\Phi}_n)\geq \frac{1}{8}\varepsilon'\}\cap S_n\right)
\end{aligned}$$

where $\mathbf{\Phi}_n^*(\cdot|Z^n)$ denotes the conditional probability (given the original data Z^n) associated to $\mathbb{U}_n\equiv n^{-1}\sum_{i=1}^n U_{i,n}$ with $U_{i,n}\sim i.i.d.-N(0,Z_{i,n}Z_{i,n}^T)$.

Hence, by the previous display and equations 5, 6-7, 12 and 13, in order to show the desired result it suffices to show that: For all ε', there exists a $N(\varepsilon')$ such that

$$\mathbf{P}_n\left(\{\Delta_{h^{-1}}(\mathbf{P}_n^*(\cdot|Z^n),\mathbf{\Phi}_n^*(\cdot|Z^n))\geq\varepsilon'\}\cap S_n\right)<\varepsilon', \tag{14}$$
$$\mathbf{P}_n\left(\{\Delta_{h^{-1}}(\mathbf{\Phi}_n^*(\cdot|Z^n),\mathbf{\Phi}_n)\geq\varepsilon'\}\cap S_n\right)<\varepsilon', \tag{15}$$
$$\textit{and } \Delta_{h^{-1}}(\mathbf{P}_n,\mathbf{\Phi}_n)<\varepsilon' \tag{16}$$

for all $n\geq N(\varepsilon')$ and some $h\leq h(\varepsilon,\sqrt{tr\{\Sigma_n^2\}}\gamma(\varepsilon))$. Theorems 3.2 and 3.3 establish expressions 14 and 16.

Remark 4.2. *From lemma B.2, $h(\varepsilon,\sqrt{tr\{\Sigma_n^2\}}\gamma(\varepsilon))=\sqrt{tr\{\Sigma_n^2\}}\gamma(\varepsilon)/\Phi^{-1}(\varepsilon)$ and thus h can be taken to be proportional (up to a constant that depends on ε) to $\sqrt{tr\{\Sigma_n^2\}}$. Hence, under assumption 2.1(i), h can be taken to be such that $h^{-2}\precsim d(n)^{-1}$. Therefore, Theorems 3.2 and 3.3 actually imply a stronger result: $\Delta_{h^{-1}}(\mathbf{P}_n,\mathbf{\Phi}_n)=o(d(n)^{-1})$ and $\Delta_{h^{-1}}(\mathbf{P}_n^*(\cdot|Z^n),\mathbf{\Phi}_n^*(\cdot|Z^n))=o_{\mathbf{P}_n}(d(n)^{-1})$.*

We have thus reduced the original problem to a Gaussian approximation problem. That is, it remains to show that

$$\mathbf{P}_n\left(\{\Delta_{h^{-1}}(\mathbf{\Phi}_n^*(\cdot|Z^n),\mathbf{\Phi}_n)\geq\varepsilon'\}\cap S_n\right)<\varepsilon'. \tag{17}$$

Since $\sqrt{n}\mathbb{U}_n \sim N(0,\hat{\Sigma}_n)$ (with $\hat{\Sigma}_n = n^{-1}\sum_{i=1}^n Z_{i,n}Z_{i,n}^T$) and $\sqrt{n}\mathbb{V}_n \sim N(0,\Sigma_n)$, the previous display is equivalent to showing that

$$\mathbf{P}_n\left(\{\Delta_{h^{-1}}(N(0,\hat{\Sigma}_n),N(0,\Sigma_n))\geq\varepsilon'\}\cap S_n\right)<\varepsilon'.$$

Essentially, this expression follows by the fact that $\hat{\Sigma}_n$ converges in probability to Σ_n in a suitable norm. The following lemma formalizes this.

Lemma 4.3. *For any $h>0$ and any $n\in\mathbb{N}$*

$$\begin{aligned}\Delta_{h^{-1}}(\mathbf{\Phi}_n^*(\cdot|Z^n),\mathbf{\Phi}_n)\precsim\max_{j,l}&\left|\left\{n^{-1}\sum_{i=1}^n Z_{[j],i,n}Z_{[l],i,n}-\Sigma_{[l,j],n}\right\}\right|\\&\times h^{-1}d(n)\left(h^{-1}tr\{\Sigma_n\}+h^{-1}tr\{\hat{\Sigma}_n\}+2\right).\end{aligned}$$

Observe that for any $Z^n\in S_n=\{Z^n : n^{-1}\sum_{i=1}^n ||Z_{i,n}||_e^2\leq(0.5\varepsilon')^{-1}tr\{\Sigma_n\}\}$, the RHS of the expression in the Lemma is bounded above by $d(n)h^{-1}\{h^{-1}(\varepsilon')^{-1}tr\{\Sigma_n\}+2\}$.

Thus by lemma 4.3, in order to establish the desired result, it suffices to show that

$$\mathbf{P}_n\left(\max_{j,l}\left|n^{-1}\sum_{i=1}^n Z_{[l],i,n}Z_{[j],i,n}-\Sigma_{[j,l],n}\right|\geq\frac{(\varepsilon')^2}{d(n)h^{-2}tr\{\Sigma_n\}}\cap S_n\right)<\varepsilon' \tag{18}$$

for sufficiently large n. Henceforth, let $c_n\equiv\frac{(\varepsilon')^2}{d(n)h^{-2}tr\{\Sigma_n\}}$ and let $\mathbf{A}_{i,n}[j,l]\equiv Z_{[j],i,n}Z_{[l],i,n}$, observe that

$$E_{\mathbf{P}_n}[\mathbf{A}_{i,n}[j,l]]=E_{\mathbf{P}_n}[Z_{[j],i,n}Z_{[l],i,n}]=\Sigma_{[j,l],n}.$$

Let $\mathbf{A}_{i,n}[j,l]=\mathbf{A}_{i,n}^L[j,l]+\mathbf{A}_{i,n}^U[j,l]\equiv\mathbf{A}_{i,n}[j,l]1\{|\mathbf{A}_{i,n}[j,l]|\leq d_n\}+\mathbf{A}_{i,n}[j,l]1\{|\mathbf{A}_{i,n}[j,l]|\geq d_n\}$ where $(d_n)_n$ with $d_n>0$ is defined below.

Clearly, $\mathbf{A}_{i,n}^L[j,l] \leq d_n$. So, by Hoeffding inequality (see Boucheron *et al.* (2013) p. 34)

$$\begin{aligned}
&\mathbf{P}_n \left(\max_{j,l} |n^{-1} \sum_{i=1}^n \{\mathbf{A}_{i,n}^L[j,l] - E_{\mathbf{P}_n}[\mathbf{A}_{i,n}^L[j,l]]\}| \geq c_n \right) \\
\leq &\sum_{j,l} \mathbf{P}_n \left(|n^{-1} \sum_{i=1}^n \{\mathbf{A}_{i,n}^L[j,l] - E_{\mathbf{P}_n}[\mathbf{A}_{i,n}^L[j,l]]\}| \geq c_n \right) \\
\precsim &C \exp\left\{ 2\log(d(n)) - n\frac{c_n^2}{d_n^2} \right\}
\end{aligned}$$

Therefore, by setting $d_n = c_n\sqrt{\frac{n0.25}{\log(d(n))}}$, the previous display implies that

$$\mathbf{P}_n \left(|n^{-1} \sum_{i=1}^n \{\mathbf{A}_{i,n}^L[j,l] - E_{\mathbf{P}_n}[\mathbf{A}_{i,n}^L[j,l]]\}| \geq \varepsilon' \right) \leq \varepsilon',$$

for sufficiently large n.

Second, by the Markov inequality and the fact that

$$E_{\mathbf{P}_n}\left[\left(\{\mathbf{A}_{i,n}^U[j,l] - E_{\mathbf{P}_n}[\mathbf{A}_{i,n}^U[j,l]]\} \right) \left(\{\mathbf{A}_{k,n}^U[j,l] - E_{\mathbf{P}_n}[\mathbf{A}_{k,n}^U[j,l]]\} \right) \right] = 0 \tag{19}$$

for all $i \neq k$, it follows that

$$\begin{aligned}
&\mathbf{P}_n \left(\max_{j,l} |n^{-1} \sum_{i=1}^n \{\mathbf{A}_{i,n}^U[j,l] - E_{\mathbf{P}_n}[\mathbf{A}_{i,n}^U[j,l]]\}| \geq c_n \right) \\
\leq &\sum_{j,l} (c_n)^{-2} E_{\mathbf{P}_n} \left[\left(n^{-1} \sum_{i=1}^n \{\mathbf{A}_{i,n}^U[j,l] - E_{\mathbf{P}_n}[\mathbf{A}_{i,n}^U[j,l]]\} \right)^2 \right] \\
= &(c_n)^{-2} n^{-1} \sum_{j,l} E_{\mathbf{P}_n} \left[\left(\{\mathbf{A}_{1,n}^U[j,l] - E_{\mathbf{P}_n}[\mathbf{A}_{1,n}^U[j,l]]\} \right)^2 \right] \\
\leq &(c_n)^{-2} n^{-1} \sum_{j,l} E_{\mathbf{P}_n} \left[\left(\mathbf{A}_{1,n}^U[j,l] \right)^2 \right]
\end{aligned}$$

Therefore by the Markov inequality, for $p > 0$

$$\begin{aligned}
&\mathbf{P}_n\left(\max_{j,l} |n^{-1}\sum_{i=1}^{n}\{\mathbf{A}^U_{i,n}[j,l] - E_{\mathbf{P}_n}[\mathbf{A}^U_{i,n}[j,l]]\}| \geq c_n\right)\\
\leq&\frac{1}{c_n^2 n(d_n)^p}\sum_{j=1}^{d(n)}\sum_{l=1}^{d(n)} E_{\mathbf{P}_n}\left[\left(Z_{[j],1,n}Z_{[l],1,n}\right)^{2+p}\right]\\
=&\frac{1}{c_n^2 n(d_n)^p} E_{\mathbf{P}_n}\left[\left(\sum_{j=1}^{d(n)}(Z_{[j],1,n})^{2+p}\right)^2\right]
\end{aligned}$$

Since $d_n = c_n\sqrt{\frac{n0.25}{\log(d(n))}}$ and $c_n \equiv \frac{(\varepsilon')^2}{d(n)h^{-2}tr\{\Sigma_n\}}$, it follows that

$$\begin{aligned}
&\mathbf{P}_n\left(\max_{j,l} |n^{-1}\sum_{i=1}^{n}\{\mathbf{A}^U_{i,n}[j,l] - E_{\mathbf{P}_n}[\mathbf{A}^U_{i,n}[j,l]]\}| \geq c_n\right)\\
\precsim&\frac{\log(d(n))^{p/2}}{c_n^{2+p}n^{1+p/2}} E_{\mathbf{P}_n}\left[\left(\sum_{j=1}^{d(n)}(Z_{[j],1,n})^{2+p}\right)^2\right]\\
\precsim&\frac{(\log(d(n)))^{p/2}d(n)^{2+p}(tr\{\Sigma_n\})^{2+p}}{h^{4+2p}n^{1+p/2}} E_{\mathbf{P}_n}\left[\left(\sum_{j=1}^{d(n)}(Z_{[j],1,n})^{2+p}\right)^2\right]
\end{aligned}$$

Since we can set $h \asymp \sqrt{tr\{\Sigma_n^2\}}$, the RHS becomes $\frac{(\log(d(n)))^{p/2}d(n)^{2+p}}{n^{1+p/2}}\left(\frac{tr\{\Sigma_n\}}{tr\{\Sigma_n^2\}}\right)^{2+p} E_{\mathbf{P}_n}\left[\left(\sum_{j=1}^{d(n)}(Z_{[j],1,n})^{2+p}\right)^2\right]$. By choosing $p = \kappa$, by assumptions 2.1(i) and 2.1(iii), the term vanishes as $n \to \infty$.

Therefore, equation 18 is established and with that the proof of Theorem 3.1.

References

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration Inequalities*. Oxford Univ. Press.

Chatterjee, S. (2006). A generalization of the lindeberg principle. *The Annals of Probability*, **34**(6), 2061–2076.

Chatterjee, S. and Meckes, E. (2008). Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.*, **4**, 257–283.

Chen, X. and Pouzo, D. (2014). Sieve wald and qlr inferences on semi/nonparametric conditional moment models. *forthcoming in Econometrica*.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013a). Comparison and anti-concentration bounds for maxima of gaussian random vectors. *ArXiv 1301.4807*.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013b). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, **41**, 2786–2819.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, **7**(1), 1–26.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, volume II. Wiley, 2nd edition.

Hall, P. (1986). Methodology and theory for the bootstrap. In R. F. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 4, chapter 39, pages 2341–2381. Elsevier, 1 edition.

He, X. and Shao, Q.-M. (2000). On parameters of increasing dimensions. *Journal of Multivariate Analysis*, **73**, 120–135.

Hjort, N., McKeague, I., and Keilegom, I. V. (2009). Extending the scope of empirical likelihood. *The Annals of Statistics*, **37**(3), 1079–1111.

Horowitz, J. L. (2001). The bootstrap. In R. F. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 5, chapter 52, pages 3159–3228. Elsevier, 1 edition.

Johnson, W. B., Schechtman, G., and Zinn, J. (1985). Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *The Annals of Probability*, **13**(1), pp. 234–253.

Ma, S. and Kosorok, M. R. (2005). Robust semiparametric m-estimation and the weighted bootstrap. *Journal of Multivariate Analysis*, **96**(1), 190 – 217.

Mammen, E. (1989). Asymptotics with increasing dimension for robust regression with applications to the bootstrap. *The Annals of Statistics*, **17**(61), 382–400.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, **21**(1), 255–285.

Owen, A. (1990). *Empirical Likelihood*. Chapman and Hall/CRC.

Peng, H. and Schick, A. (2012). Asymptotic normality of quadratic forms with random vectors of increasing dimension. *Working Paper*.

Pollard, D. (2001). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.

Portnoy (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *The Annals of Statistics*, **16**(1), 356–366.

Radulovic, D. (1998). Can we bootstrap even if clt fails? *Journal of Theoretical Probability*, **11**(3), 813–830.

Raic, M. (2004). A multivariate clt for decomposable random vectors with finite second moments. *Journal of Theoretical Probability*, **17**(3), 573–603.

Rollin, A. (2013). Stein's method in high dimensions with applications. *ArXiv 1101.4454*.

Slepian, D. (1962). The one-sided barrier problem for gaussian noise. *Bell System Technical Journal*, **41**(2), 463–501.

Spokoiny, V. and Zhilova, M. (2014). Bootstrap confidence sets under a model misspecification. *arXiv:1410.0347v1*.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, **9**(6), 1135–1151.

Van der Vaart, A. (2000). *Asymptotic Statistics*. Cambridge University Press.

Vershynin, R. (2012a). How close is the sample covariance matrix to the actual covariance matrix? *Journal of Theoretical Probability*, **25**, 655–686.

Vershynin, R. (2012b). Introduction to the non-asymptotic analysis of random matrices. In *Compressed sensing*, pages 210–268. Cambridge Univ. Press.

Xu, M., Zhang, D., and Wu, W. B. (2014). l^2 asymptotics for high-dimensional data. *arXiv: 1405.7244*.

Zhang, X. and Cheng, G. (2014). Bootstrapping high dimensional time series. *arXiv:1406.1037v2*.

Appendix

A Proof of Theorems 3.4, 3.2 and 3.3

The next lemma provides a bound for $\mathbf{S}_n$ and $\mathbf{R}_n$ in theorem 3.4.

Lemma A.1. *Suppose the same conditions of Theorem 3.4. Then,*

$$\mathbf{S}_{1,n} \leq L_2(f) \sum_{i=1}^{n} E[||B_i||_e^4 + ||A_i||_e^4]$$
$$\mathbf{S}_{2,n} \leq L_2(f) \sqrt{\sum_{j=1}^{n} tr\{C_j\}} \sum_{i=1}^{n} \left(E\left[||B_i||_e^3\right] + E\left[||A_i||_e^3\right]\right).$$

And, for any $q > 0$

$$\mathbf{R}_n \precsim \sum_{i=1}^{n} \left(E\left[\left(\mathbb{S}_{i:n}^T B_i\right)^{2+q} + \left(\mathbb{S}_{i:n}^T A_i\right)^{2+q} \right] + E\left[||B_i||_e^{4+2q}\right] + E\left[||A_i||_e^{4+2q}\right] \right).$$

and

$$\sum_{i=1}^{n} E\left[\left(\mathbb{S}_{i:n}^T B_i\right)^{2+q} \right]$$
$$\precsim \sum_{i=1}^{n} E[||B_i||_e^{2+q}] \max \left\{ \left(\sum_{j=1}^{n} E\left[||S_j||_e^2\right] \right)^{1+0.5q}, \sum_{j=1}^{n} E\left[||S_j||_e^{2+q}\right] \right\}.$$

and an analogous expression holds for $\sum_{i=1}^{n} E\left[\left(\mathbb{S}_{i:n}^T A_i\right)^{2+q} \right]$.

Proof of Lemma A.1. $\mathbf{S}_{1,n}$ is trivially bounded by $L_2(f) \sum_{i=1}^{n} E[||B_i||_e^4 + ||A_i||_e^4]$. Regarding $\mathbf{S}_{2,n}$, observe that

$$\sum_{i=1}^{n} |E\left[f''\left(||\mathbb{S}_{i:n}||_e^2\right) \mathbb{S}_{i:n}^T\right] \left(E[B_i ||B_i||_e^2]\right)| \leq L_2(f) \sum_{i=1}^{n} E\left[||\mathbb{S}_{i:n}||_e ||B_i||_e^3\right]$$
$$\leq L_2(f) \sum_{i=1}^{n} \sqrt{E\left[||\mathbb{S}_{i:n}||_e^2\right]} E\left[||B_i||_e^3\right]$$

by independence of $\mathbb{S}_{i:n}$ and B_i and Cauchy-Swarchz. Also, $E[\mathbb{S}_{i:n}\mathbb{S}_{i:n}^T] = \sum_{j=1}^n E[S_jS_j^T]$, so $E\left[||\mathbb{S}_{i:n}||_e^2\right] = tr\{E[\mathbb{S}_{i:n}\mathbb{S}_{i:n}^T]\} = \sum_{j=1}^n tr\{C_j\}$. A similar results holds when B_i is replaced by A_i. Therefore

$$\mathbf{S}_{2,n} \leq L_2(f)\sqrt{\sum_{j=1}^n tr\{C_j\}}\sum_{i=1}^n \left(E\left[||B_i||_e^3\right] + E\left[||A_i||_e^3\right]\right).$$

Regarding $\mathbf{R}_n$. Note that

$$\sum_{i=1}^n E\left[\left(\mathbb{S}_{i:n}^T B_i + ||B_i||_e^2\right)^{2+q}\right] \precsim \left(\sum_{i=1}^n E\left[\left(\mathbb{S}_{i:n}^T B_i\right)^{2+q}\right] + \sum_{i=1}^n E\left[\left(||B_i||_e\right)^{4+2q}\right]\right).$$

Observe that $E\left[\left(\mathbb{S}_{i:n}^T B_i\right)^{2+q}\right] = E\left[E\left[\left(\sum_{j=1}^n S_j^T b_i\right)^{2+q} |B_i = b_i\right]\right]$. Since $(S_j)_j$ does not contain B_i, conditioning on $B_i = b_i$, $(S_j^T b_i)_j$ is an independent sequence.

Therefore, by Johnson *et al.* (1985), for any $q > 0$,

$$\begin{aligned} &E\left[\left(\mathbb{S}_{i:n}^T b_i\right)^{2+q}\right] \\ &\precsim \left(\max\left\{\sqrt{E\left[\left(\sum_{j=1}^n S_j^T b_i\right)^2\right]}, \left(\sum_{j=1}^n E\left[\left(S_j^T b_i\right)^{2+q}\right]\right)^{1/(2+q)}\right\}\right)^{2+q} \end{aligned}$$

(where the expectation is only with respect to $(S_j)_{j=1}^n$, not b_i). By independence, and the fact that $E[S_j^T b_i] = 0$,

$$E\left[\left(\sum_{j=1}^n S_j^T b_i\right)^2\right] = E\left[\sum_{j=1}^n \left(S_j^T b_i\right)^2\right] = \ tr\left\{E\left[\left(\sum_{j=1}^n S_j S_j^T\right)\right] b_i b_i^T\right\}.$$

Also, note that

$$\sum_{j=1}^n E\left[\left(S_j^T b_i\right)^{2+q}\right] \leq \sum_{j=1}^n E\left[\left(||S_j||_e ||b_i||_e\right)^{2+q}\right] = \left(||b_i||_e\right)^{2+q}\sum_{j=1}^n E\left[\left(||S_j||_e\right)^{2+q}\right].$$

Therefore, using these bounds and taken expectation with respect to B_i

and after straightforward algebra,

$$\sum_{i=1}^{n} E\left[\left(\mathbb{S}_{i:n}^{T} B_i\right)^{2+q}\right]$$
$$\lesssim \sum_{i=1}^{n} E[||B_i||_e^{2+q}] \max\left\{\left(\sum_{j=1}^{n} E\left[||S_j||_e^2\right]\right)^{1+0.5q}, \sum_{j=1}^{n} E\left[||S_j||_e^{2+q}\right]\right\}.$$

An analogous steps can be taken to show the same result replacing B_i by A_i; they will be omitted. □

Proof of Theorem 3.4 . Let $\mathbb{S}_{i:n} \equiv \sum_{j=1}^{i-1} A_j + 0 + \sum_{j=i+1}^{n} B_j \equiv \sum_{j=1}^{n} S_j$. Observe that $(S_i)_{i=1}^n$ are independent and $E[S_i] = 0$, also $E[S_i S_i^T] = E[B_i B_i^T] = C_i$. Also, note that $\mathbb{S}_{1:n} \equiv \sum_{i=1}^{n} B_i - B_1$ and $\mathbb{S}_{n:n} \equiv \sum_{i=1}^{n} A_i - A_n$. Moreover

$$\mathbb{S}_{i:n} + A_i = \left(\sum_{j=1}^{i} A_j + \sum_{j=i+1}^{n} B_j\right) = \mathbb{S}_{i+1:n} + B_{i+1}. \tag{20}$$

Therefore,

$$\sum_{i=1}^{n} E\left[f\left(||\mathbb{S}_{i:n} + B_i||_e^2\right) - f\left(||\mathbb{S}_{i:n} + A_i||_e^2\right)\right] = E\left[f\left(||\sum_{i=1}^{n} B_i||_e^2\right) - f\left(||\sum_{i=1}^{n} A_i||_e^2\right)\right].$$

Observe that $||\mathbb{S}_{i:n} + B_i||_e^2 = ||\mathbb{S}_{i:n}||_e^2 + ||B_i||_e^2 + 2\mathbb{S}_{i:n}^T B_i$. Therefore, by this fact and three times differentiability of f, it follows that

$$\begin{aligned} f\left(||\mathbb{S}_{i:n} + B_i||_e^2\right) - f\left(||\mathbb{S}_{i:n}||_e^2\right) =& f'\left(||\mathbb{S}_{i:n}||_e^2\right)\left(||B_i||_e^2 + 2\mathbb{S}_{i:n}^T B_i\right) \\ &+ 0.5 f''\left(||\mathbb{S}_{i:n}||_e^2\right)\left(||B_i||_e^2 + 2\mathbb{S}_{i:n}^T B_i\right)^2 \\ &+ R_{i,1,n} \end{aligned}$$

where $R_{i,1,n}$ is a reminder term which will be defined later. Similarly

$$\begin{aligned} f\left(||\mathbb{S}_{i:n} + A_i||_e^2\right) - f\left(||\mathbb{S}_{i:n}||_e^2\right) =& f'\left(||\mathbb{S}_{i:n}||_e^2\right)\left(||A_i||_e^2 + 2\mathbb{S}_{i:n}^T A_i\right) \\ &+ 0.5 f''\left(||\mathbb{S}_{i:n}||_e^2\right)\left(||A_i||_e^2 + 2\mathbb{S}_{i:n}^T A_i\right)^2 \\ &+ R_{i,2,n} \end{aligned}$$

Hence

$$\begin{aligned} & E\left[f\left(||\mathbb{S}_{i:n} + B_i||_e^2\right) - f\left(||\mathbb{S}_{i:n} + A_i||_e^2\right)\right] \\ =& E\left[f'\left(||\mathbb{S}_{i:n}||_e^2\right)\left(||A_i||_e^2 - ||B_i||_e^2 + 2\mathbb{S}_{i:n}^T (A_i - B_i)\right)\right] \\ &+ 0.5 E\left[f''\left(||\mathbb{S}_{i:n}||_e^2\right)\left\{\left(||B_i||_e^2 + 2\mathbb{S}_{i:n}^T B_i\right)^2 - \left(||A_i||_e^2 + 2\mathbb{S}_{i:n}^T A_i\right)^2\right\}\right] \\ &+ E\left[R_{i,1,n} - R_{i,2,n}\right] \\ \equiv& F_{i,n} + S_{i,n} + E\left[R_{i,1,n} - R_{i,2,n}\right] \end{aligned}$$

Therefore, it suffices to bound the *first order terms* $F_n \equiv \sum_{i=1}^n F_{i,n}$, *second order terms* $S_n \equiv \sum_{i=1}^n S_{i,n}$ and *the remainder terms* $E\left[R_{i,1,n} - R_{i,2,n}\right]$.

The First order terms, F_n. Since $\mathbb{S}_{i:n}$ is independent with A_i and B_i and $E[A_i] = E[B_i] = 0$ and $E[A_i A_i^T] = E[B_i B_i^T]$ it readily follows that

$$\sum_{i=1}^n E\left[f'\left(||\mathbb{S}_{i:n}||_e^2\right)\mathbb{S}_{i:n}^T\left(B_i - A_i\right)\right] = \sum_{i=1}^n E\left[f'\left(||\mathbb{S}_{i:n}||_e^2\right)\mathbb{S}_{i:n}^T\right] E\left[(B_i - A_i)\right] = 0$$

and

$$\sum_{i=1}^n E\left[f'\left(||\mathbb{S}_{i:n}||_e^2\right)\left(||B_i||_e^2 - ||A_i||_e^2\right)\right] = \sum_{i=1}^n E\left[f'\left(||\mathbb{S}_{i:n}||_e^2\right)\right] E\left[\left(||B_i||_e^2 - ||A_i||_e^2\right)\right] = 0.$$

The term Second order terms, S_n. For this term it suffices to study the following terms:

$$
\begin{aligned}
S_{1,n} \equiv & \sum_{i=1}^{n} E\left[f''\left(||\mathbb{S}_{i:n}||_e^2\right)\left(||B_i||_e^4 - ||A_i||_e^4\right)\right] \\
S_{2,n} \equiv & \sum_{i=1}^{n} E\left[f''\left(||\mathbb{S}_{i:n}||_e^2\right) 4\left((\mathbb{S}_{i:n}^T B_i)^2 - (\mathbb{S}_{i:n}^T A_i)^2\right)\right] \\
S_{3,n} \equiv & \sum_{i=1}^{n} E\left[f''\left(||\mathbb{S}_{i:n}||_e^2\right) 4\mathbb{S}_{i:n}^T\left(B_i||B_i||_e^2 - A_i||A_i||_e^2\right)\right].
\end{aligned}
$$

By independence of $\mathbb{S}_{i:n}$ with A_i and B_i, $S_{1,n} = \sum_{i=1}^{n} E\left[f''\left(||\mathbb{S}_{i:n}||_e^2\right)\right] E\left[||B_i||_e^4 - ||A_i||_e^4\right]$. Regarding $S_{2,n}$, because $\mathbb{S}_{i:n}$ is independent to A_i and B_i and $E[A_i A_i^T] = E[B_i B_i^T]$, it follows that $E\left[\mathbb{S}_{i:n}^T B_i B_i^T \mathbb{S}_{i:n}\right] = E\left[\mathbb{S}_{i:n}^T A_i A_i^T \mathbb{S}_{i:n}\right]$ and thus $S_{2,n} = 0$.

Finally, regarding $S_{3,n}$, observe that by independence of $\mathbb{S}_{i:n}$ and B_i and A_i

$$
|S_{3,n}| \leq 4 \sum_{i=1}^{n} |E\left[f''\left(||\mathbb{S}_{i:n}||_e^2\right)\mathbb{S}_{i:n}^T\right]\left(E[B_i||B_i||_e^2] - E[A_i||A_i||_e^2]\right)|
$$

The remainder terms, $R_{1,n}$ and $R_{2,n}$. By Taylor's theorem it follows that: For any $q > 0$

$$
\sum_{i=1}^{n} E\left[|R_{i,1,n}|\right] \precsim L_2(f)^{1-q} L_3(f)^q \sum_{i=1}^{n} E\left[\left(\mathbb{S}_{i:n}^T B_i + ||B_i||_e^2\right)^{2+q}\right]
$$

□

A.1 Proof of Theorem 3.2

Proof of Theorem 3.2. We first note that is enough to bound

$$
\mathbf{P}_n\left(\left\{\sup_{f \in \mathcal{C}_{h-1}} \left|E_{\mathbf{P}_n^*}\left[f\left(||\sqrt{n}\mathbb{Z}_n^*||_e^2\right)|Z^n\right] - E_{\mathbf{\Phi}_n^*}\left[f\left(||\sqrt{n}\mathbb{U}_n||_e^2\right)|Z^n\right]\right| \geq \varepsilon\right\} \cap K_n\right)
$$

where $K_n \equiv \{Z^n : n^{-1}\sum_{i=1}^{n} ||Z_{i,n}||_e^2 \leq (0.5\varepsilon')^{-1} tr\{\Sigma_n\} \equiv M_n\}$.

The strategy of proof consists of applying the results in Theorem 3.4 and Lemma A.1, with $A_i = n^{-1/2}\omega_{i,n} Z_{i,n}$ and $B_i = n^{-1/2} u_i Z_{i,n}$ where $u_i \sim N(0,1)$. Then use the Markov inequality and show that the expectation (under $\mathbf{P}_n$) of the terms in the RHS of the main expression in Theorem 3.4, $\mathbf{S}_n$ and $\mathbf{R}_n$, vanishes as $n \to \infty$.

The leading terms, $\mathbf{S}_n$. For this case $\sum_{i=1}^{n} E[(||B_i||_e)^4] \precsim n^{-2} \sum_{i=1}^{n} ||Z_{i,n}||_e^4$ and $\sum_{i=1}^{n} E[(||A_i||_e)^4] \precsim n^{-2} \sum_{i=1}^{n} ||Z_{i,n}||_e^4$, under assumption 2.2. Therefore, $\mathbf{S}_{1,n}$ in Theorem 3.4 is bounded above (up to a constant) by $n^{-1} \left(n^{-1} \sum_{i=1}^{n} ||Z_{i,n}||_e^4 \right)$.

Therefore, since $L_2(f) = h^{-2}$, $E_{\mathbf{P}_n}[\mathbf{S}_{1,n}] \precsim h^{-2} n^{-2} \sum_{i=1}^{n} E_{\mathbf{P}_n}[||Z_{i,n}||_e^4] = h^{-2} n^{-1} E_{\mathbf{P}_n}[||Z_{1,n}||_e^4]$ which is of order $o(h^{-2})$ by assumption 2.1(i).

Observe that in this case $E[S_i S_i^T] = n^{-1} Z_{i,n} Z_{i,n}^T$ and thus

$$\begin{aligned}
\mathbf{S}_{2,n} \precsim & h^{-2} \sqrt{n^{-1} \sum_{i=1}^{n} ||Z_{i,n}||_e^2} n^{-3/2} \sum_{i=1}^{n} E[|\omega_{i,n}|^3 + |u_{i,n}|^3] ||Z_{i,n}||_e^3 \\
\precsim & h^{-2} \sqrt{n^{-1} \sum_{i=1}^{n} ||Z_{i,n}||_e^2} n^{-3/2} \sum_{i=1}^{n} ||Z_{i,n}||_e^3.
\end{aligned}$$

For any $Z^n \in K_n$, $\mathbf{S}_{2,n} \precsim h^{-2} \sqrt{M_n} n^{-3/2} \sum_{i=1}^{n} ||Z_{i,n}||_e^3$. Therefore, $E_{\mathbf{P}_n}[\mathbf{S}_{2,n} 1\{K_n\}] \precsim h^{-2} \sqrt{M_n} n^{-3/2} \sum_{i=1}^{n} E_{\mathbf{P}_n}[||Z_{i,n}||_e^3] = h^{-2} \sqrt{M_n} n^{-1/2} E_{\mathbf{P}_n}[||Z_{1,n}||_e^3]$, which is of order $o(h^{-2})$ by assumption 2.1(i).

The remainder terms, $\mathbf{R}_n$. To bound the remainder term in the expression of Theorem 3.4 we use lemma A.1 and the fact that $L_2(f) = h^{-2}$. Observe that $\left(tr \left\{ \sum_{j=1}^{n} E\left[\left(S_j^T S_j \right) \right] \right\} \right)^{1+0.5q} = \left(tr \left\{ n^{-1} \sum_{j=1}^{n} Z_{j,n} Z_{j,n}^T \right\} \right)^{1+0.5q} = \left(n^{-1} \sum_{j=1}^{n} ||Z_{j,n}||_e^2 \right)^{1+0.5q}$. Also,

$$\sum_{i=1}^{n} E\left[(||B_i||_e)^{2+q} \right] = n^{-(1+0.5q)} \sum_{i=1}^{n} E\left[|u_{i,n}|^{2+q} \right] ||Z_i||_e^{2+q} \precsim n^{-(1+0.5q)} \sum_{i=1}^{n} ||Z_i||_e^{2+q}$$

because of the fact that $E[|u_{i,n}|^{2+q}] \leq C < \infty$ with $q = \gamma$. Similarly, under assumption 2.2,

$$\sum_{j=1}^{n} E\left[(||S_j||_e)^{2+q}\right] \precsim n^{-(1+0.5q)} \sum_{i=1}^{n} ||Z_i||_e^{2+q}.$$

Therefore,

$$\begin{aligned}
&\sum_{i=1}^{n} E\left[\left(\mathbb{S}_{i:n}^T B_i\right)^{2+q}\right] \\
\precsim & n^{-(1+0.5q)} \sum_{i=1}^{n} ||Z_{i,n}||_e^{2+q} \max\left\{\left(n^{-1}\sum_{j=1}^{n} ||Z_{j,n}||_e^2\right)^{1+0.5q}, n^{-(1+0.5q)} \sum_{i=1}^{n} ||Z_i||_e^{2+q}\right\} \\
\precsim & \max\left\{n^{-(1+0.5q)} \sum_{i=1}^{n} ||Z_{i,n}||_e^{2+q} \left(n^{-1}\sum_{j=1}^{n} ||Z_{j,n}||_e^2\right)^{1+0.5q}, n^{-(1+q)} \sum_{i=1}^{n} ||Z_i||_e^{4+2q}\right\}
\end{aligned}$$

where the last line follows from Jensen inequality. And, also note that $\sum_{i=1}^{n} E[(||B_i||_e)^{4+2q}] \precsim n^{-(2+q)} \sum_{i=1}^{n} ||Z_{i,n}||_e^{4+2q}$.

It is straightforward to check that analogous expressions hold for $\sum_{i=1}^{n} E\left[\left(\mathbb{S}_{i:n}^T A_i\right)^{2+q}\right]$ and $\sum_{i=1}^{n} E[(||A_i||_e)^{4+2q}]$.

Recall that $q = \gamma$. Thus, $E_{\mathbf{P}_n}[n^{-(2+q)} \sum_{i=1}^{n} ||Z_{i,n}||_e^{4+2q}] = n^{-(1+q)} E_{\mathbf{P}_n}[||Z_{1,n}||_e^{4+2q}]$ which vanishes as $n \to \infty$ under assumption 2.1(ii). Similarly, $E_{\mathbf{P}_n}\left[\sum_{i=1}^{n} E\left[\left(\mathbb{S}_{i:n}^T B_i\right)^{2+q}\right] 1\{Z^n \in K_n\}\right]$ (and $E_{\mathbf{P}_n}\left[\sum_{i=1}^{n} E\left[\left(\mathbb{S}_{i:n}^T A_i\right)^{2+q}\right] 1\{Z^n \in K_n\}\right]$) are bounded above (up to a constant) by $(M_n)^{1+0.5q} n^{-(0.5q)} E_{\mathbf{P}_n}[||Z_{1,n}||_e^{2+q}] + n^{-q} E_{\mathbf{P}_n}[||Z_{1,n}||_e^{4+q}]$; both terms vanish as $n \to \infty$ under assumption 2.1(ii) with $q = \gamma$.

The desired result follows by the Markov inequality, since we proven that $E_{\mathbf{P}_n}[\mathbf{S}_n 1\{K_n\}]$ and $E_{\mathbf{P}_n}[\mathbf{R}_n 1\{K_n\}]$ are of order $o(h^{-2})$. $\square$

A.2 Proof of Theorem 3.3

For the proof of Theorem 3.3 we need the following simple lemma.

Lemma A.2. *Let $d \geq 1$ and let $X \in \mathbb{R}^d$ such that $X \sim N(0, A)$ for some*

A positive definite. Then for any $q > 0$

$$E[||X||_e^{2q}] \leq C(q)(tr\{A\})^q$$

for some $C(q) \in (0, \infty)$.

Proof of Lemma A.2. It follows that we can cast X as $\Lambda^{1/2}\xi$ with $\xi \sim N(0, I_d)$ where Λ is a diagonal matrix of eigenvalues of A.

For any $q > 0$

$$E[||X||_e^{2q}] = tr\{A\}^q E\left[\left(\sum_{j=1}^d c_j(A)|\xi_j|^2\right)^q\right]$$

where $c_j(A) \equiv \frac{\lambda_j(A)}{\sum_{j=1}^d \lambda_j(A)}$. Since

$$\begin{aligned} E\left[\left(\sum_{j=1}^d c_j(A)|\xi_j|^2\right)^q\right] =& \int_0^\infty \Pr\left(\sum_{j=1}^d c_j(A)|\xi_j|^2 \geq t^{1/q}\right) dt \\ =& q\int_0^\infty u^{q-1} \Pr\left(\sum_{j=1}^d c_j(A)|\xi_j|^2 \geq u\right) du \\ =& q\int_0^\infty u^{q-1} e^{-0.25u} du E\left[e^{0.25\sum_{j=1}^d c_j(A)|\xi_j|^2}\right] \\ \leq& q\int_0^\infty u^{q-1} e^{-0.25u} du \sum_{j=1}^d c_j(A) E\left[e^{0.25|\xi_j|^2}\right] \end{aligned}$$

where the third line follows from the Markov inequality and the fourth from Jensen inequality. The result follows from the fact that $q\int_0^\infty u^{q-1}e^{-0.25u}du \leq C < \infty$ and $|\xi_j|^2 \sim \chi^2$ and $\sum_{j=1}^d c_j(A) = 1$. $\square$

Proof of Theorem 3.3. The strategy of proof consists of applying the results in Theorem 3.4 and Lemma A.1, with $A_i = n^{-1/2}Z_{i,n}$ and $B_i = n^{-1/2}V_{i,n}$. Observe that $E[A_iA_i^T] = E[B_iB_i^T] = \Sigma_n$.

THE TERM $\mathbf{S}_n$. For this case $\sum_{i=1}^n E[(||B_i||_e)^4] = n^{-2}\sum_{i=1}^n E[||V_{i,n}||_e^4] = n^{-1}E[||V_{1,n}||_e^4]$ and $\sum_{i=1}^n E[(||A_i||_e)^4] = n^{-2}\sum_{i=1}^n E[||Z_{i,n}||_e^4] = n^{-1}E[||Z_{1,n}||_e^4]$. Therefore, $\mathbf{S}_{1,n}$ in Theorem 3.4 is bounded above (up to a constant) by $h^{-2}n^{-1}\left(E[||Z_{1,n}||_e^4] + E[||V_{1,n}||_e^4]\right)$, and by Lemma A.2, this implies that

$$\mathbf{S}_{1,n} \precsim h^{-2}n^{-1}\left(E[||Z_{1,n}||_e^4] + (tr\{\Sigma_n\})^2\right).$$

both terms are of order $o(h^{-2})$ under assumption 2.1(ii).

Observe that in this case $E[S_j S_j^T] = n^{-1}\Sigma_n$ and thus

$$\begin{aligned}\mathbf{S}_{3,n} \precsim & h^{-2}\sqrt{tr\{\Sigma_n\}}n^{-3/2}\sum_{i=1}^n (E[||Z_{i,n}||_e^3] + E[||V_{i,n}||_e^3]) \\ = & h^{-2}\sqrt{tr\{\Sigma_n\}}n^{-1/2}(E[||Z_{1,n}||_e^3] + E[||V_{1,n}||_e^3]).\end{aligned}$$

By Lemma A.2, $E[||V_{1,n}||_e^3] = (tr\{\Sigma_n\})^{3/2}$. Thus, by assumption 2.1(i), $\mathbf{S}_{2,n}$ is of order $o(h^{-2})$.

We thus have established that $\mathbf{S}_n$ in Theorem 3.4 vanishes. We now establish that $\mathbf{R}_n$ also vanishes.

THE REMAINDER TERMS, $\mathbf{R}_n$. To bound the remainder term in the expression of Theorem 3.4 we use lemma A.1, $L_2(f) = h^{-2}$ and also set $q = \gamma$. Observe that $\left(tr\left\{\sum_{j=1}^n E\left[\left(S_j^T S_j\right)\right]\right\}\right)^{1+0.5q} = (tr\{\Sigma_n\})^{1+0.5q}$. Also,

$$\sum_{i=1}^n E\left[(||B_i||_e)^{2+q}\right] = n^{-0.5q}E[||V_1||_e^{2+q}] \precsim n^{-0.5q}(tr\{\Sigma_n\})^{1+0.5q}$$

by lemma A.2. Therefore,

$$\sum_{i=1}^n E\left[\left(\mathbb{S}_{i:n}^T B_i\right)^{2+q}\right] \precsim n^{-0.5q}(tr\{\Sigma_n\})^{1+0.5q}\max\left\{(tr\{\Sigma_n\})^{1+0.5q}, \sum_{j=1}^n E[||S_j||_e^{2+q}]\right\}.$$

Observe that $\sum_{j=1}^n E\left[(||S_j||_e)^{2+q}\right] \precsim n^{-(1+0.5q)}\left(\sum_{j=1}^{i-1} E\left[(||Z_{j,n}||_e)^{2+q}\right] + (n-i)tr\{\Sigma_n\}^{1+0.5q}\right)$

by lemma A.2. Under assumption 2.1(ii),

$$\sum_{j=1}^{n} E\left[(||S_j||_e)^{2+q}\right] \precsim n^{-(1+0.5q)}\left(iE\left[(||Z_{1,n}||_e)^{2+q}\right] + (n-i)tr\{\Sigma_n\}^{1+0.5q}\right)$$
$$\leq n^{-(0.5q)}\left(E\left[(||Z_{1,n}||_e)^{2+q}\right] + tr\{\Sigma_n\}^{1+0.5q}\right) \to 0,\ as\ n \to 0$$

because, $n^{-(0.5q)}tr\{\Sigma_n\}^{1+0.5q} = \left(n^{-1/2}tr\{\Sigma_n\}^{0.5+1/q}\right)^q$ and with $q = \gamma > 2$ is implied by 2.1(ii); and due to Jensen inequality $n^{-(0.5q)}E\left[(||Z_{1,n}||_e)^{2+q}\right] \leq \sqrt{n^{-q}E\left[(||Z_{1,n}||_e)^{4+2q}\right]}$ which vanishes for $q = \gamma$.

Also, by assumption 2.1(ii), $n^{-(0.5q)}(tr\{\Sigma_n\})^{2+q} \to 0$ as $n \to \infty$. Finally, note that $\sum_{i=1}^n E[(||B_i||_e)^{4+2q}] \precsim n^{-(2+q)} \sum_{i=1}^n E[||V_{i,n}||_e^{4+2q}] = n^{-(1+q)}E[||V_{1,n}||_e^{4+2q}] \precsim n^{-(1+q)}(tr\{\Sigma_n\})^{2+q}$ by lemma A.2. By assumption 2.1(ii) and the previous calculations, $n^{-(1+q)}(tr\{\Sigma_n\})^{2+q} = o(1)$. Similarly, $\sum_{i=1}^n E[(||A_i||_e)^{4+2q}] \precsim n^{-(2+q)} \sum_{i=1}^n E[||Z_{i,n}||_e^{4+2q}] = n^{-(1+q)}E[||Z_{1,n}||_e^{4+2q}] = o(1)$ by assumption 2.1(ii).

We have established that the remainder term $\mathbf{R}_n$ in Theorem 3.4 vanishes, and thus the desired result follows. □

B Proofs of Lemmas in Section 4

In order to prove the lemmas in section 4 we need the following lemmas.

B.1 Supplementary Lemmas

Let for any $t \in \mathbb{R}$, $\delta > 0$, $n \in \mathbb{N}$, and $h > 0$

$$\mathcal{P}_{t,\delta,h}(||x||_e^2) = \int p_{t,\delta}(||x||_e^2 + hz)\phi(z)dz,\ \forall x \in \mathbb{R}^{d(n)}$$

where $\mathbb{R} \ni u \mapsto p_{t,\delta}(u) = 1\{u \geq t\} + \frac{u-t+\delta}{\delta}1\{u \in (t-\delta, t)\}$ and ϕ is the standard Gaussian pdf.

Lemma B.1. *For any $\varepsilon \in (0,1)$, $\delta > 0$ and $n \in \mathbb{N}$, there exists $h(\delta,\varepsilon) = \frac{\delta}{\Phi^{-1}(\varepsilon)}$ such that for all $h \leq h(\delta,\varepsilon)$:*

(i)

$$E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_e^2 \geq t\}|Z^n\right] \leq \frac{1}{1-\varepsilon}E_{\mathbf{P}_n^*}\left[\mathcal{P}_{t-\delta,\delta,h}(||\sqrt{n}\mathbb{Z}_n^*||_e^2)|Z^n\right] \tag{21}$$

(ii)

$$E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_e^2 \geq t\}|Z^n\right] \geq \frac{1}{1-\varepsilon}E_{\mathbf{P}_n^*}\left[\mathcal{P}_{t+2\delta,\delta,h}(||\sqrt{n}\mathbb{Z}_n^*||_e^2)|Z^n\right] - \frac{\varepsilon}{1-\varepsilon} \tag{22}$$

Lemma B.2. *For any $\varepsilon \in (0,1)$, $\delta > 0$ and $n \in \mathbb{N}$, there exists $h(\delta,\varepsilon) = \frac{\delta}{\Phi^{-1}(\varepsilon)}$ such that for all $h \leq h(\delta,\varepsilon)$:*

(i)

$$E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2 \geq t\}\right] \leq \frac{1}{1-\varepsilon}E_{\mathbf{\Phi}_n}\left[\mathcal{P}_{t-\delta,\delta,h}(||\sqrt{n}\mathbb{V}_n||_e^2)\right] \tag{23}$$

(ii)

$$E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2 \geq t\}\right] \geq \frac{1}{1-\varepsilon}E_{\mathbf{\Phi}_n}\left[\mathcal{P}_{t+2\delta,\delta,h}(||\sqrt{n}\mathbb{V}_n||_e^2)\right] - \frac{\varepsilon}{1-\varepsilon} \tag{24}$$

Lemma B.3. *For any $\varepsilon \in (0,1)$, $\delta > 0$ and $n \in \mathbb{N}$, there exists $h(\delta,\varepsilon) = \frac{\delta}{\Phi^{-1}(\varepsilon)}$ such that for all $h \leq h(\delta,\varepsilon)$:*

(i)

$$E_{\mathbf{P}_n}\left[1\{||\sqrt{n}\mathbb{Z}_n||_e^2 \geq t\}\right] \leq \frac{1}{1-\varepsilon}E_{\mathbf{P}_n}\left[\mathcal{P}_{t-\delta,\delta,h}(||\sqrt{n}\mathbb{Z}_n||_e^2)\right] \tag{25}$$

(ii)

$$E_{\mathbf{P}_n}\left[1\{||\sqrt{n}\mathbb{Z}_n||_e^2 \geq t\}\right] \geq \frac{1}{1-\varepsilon}E_{\mathbf{P}_n}\left[\mathcal{P}_{t+2\delta,\delta,h}(||\sqrt{n}\mathbb{Z}_n||_e^2)\right] - \frac{\varepsilon}{1-\varepsilon} \tag{26}$$

Lemma B.4. *Suppose assumption 2.1(i) holds. For any $\varepsilon > 0$, there exists* *a $N(\varepsilon)$ and $\gamma(\varepsilon)$ such that for all $\gamma \leq \gamma(\varepsilon)$ and all $n \geq N(\varepsilon)$:*

$$\sup_t \mathbf{\Phi}_n \left(||| \sqrt{n} \mathbb{V}_n ||_e^2 - t| \leq \sqrt{tr\{\Sigma_n^2\}} \gamma \right) \leq \varepsilon \tag{27}$$

Remark B.1. *It is easy to see that from this lemma it follows that: For any $\varepsilon > 0$, there exists a $N(\varepsilon)$ and $\gamma(\varepsilon)$ such that for all $\gamma \leq \gamma(\varepsilon)$ and all $n \geq N(\varepsilon)$:*

$$\mathbf{\Phi}_n \left(|| \sqrt{n} \mathbb{V}_n ||_e^2 \geq t \right) \leq \varepsilon + \mathbf{\Phi}_n \left(|| \sqrt{n} \mathbb{V}_n ||_e^2 \geq t + \sqrt{tr\{\Sigma_n^2\}} \gamma \right) \tag{28}$$

for all $t \geq 0$.

Proof of Lemma B.1. **Part (i)** By definition of $\mathcal{P}_{t,\delta,h}$, for any $||x||_e^2 \geq t + \delta$

$$\mathcal{P}_{t,\delta,h}(||x||_e^2) \geq \int 1\{z : ||x||_e^2 + hz \geq t\} \phi(z) dz \geq \int 1\{z : hz \geq -\delta\} \phi(z) dz = 1 - \Phi(-\delta/h).$$

Thus, for any $h \leq \frac{\delta}{\Phi^{-1}(\varepsilon)} \equiv h(\delta, \varepsilon)$, $\mathcal{P}_{t,\delta,h}(||x||_e^2) \geq (1 - \varepsilon) 1\{||x||_e^2 \geq t + \delta\}$. Thus

$$E_{\mathbf{P}_n^*} \left[1\{|| \sqrt{n} \mathbb{Z}_n^* ||_e^2 \geq t\} | Z^n \right] \leq \frac{1}{1 - \varepsilon} E_{\mathbf{P}_n^*} \left[\mathcal{P}_{t-\delta,\delta,h}(|| \sqrt{n} \mathbb{Z}_n^* ||_e^2) | Z^n \right]$$

for any $h \leq h(\delta, \varepsilon)$.

Part (ii) Observe that for any $x : ||x||_e^2 \leq t - 2\delta$,

$$\mathcal{P}_{t,\delta,h}(||x||_e^2) \leq \int 1\{z : ||x||_e^2 + hz) \geq t - \delta\} \phi(z) dz \leq \int 1\{z : hz \geq \delta\} \phi(z) dz.$$

Thus $\mathcal{P}_{t,\delta,h}(||x||_e^2) \leq \varepsilon$ for any $x : ||x||_e^2 \leq t - 2\delta$ and $h \leq h(\delta, \varepsilon)$. Thus, for all $x \in \mathbb{R}^d$, $\mathcal{P}_{t,\delta,h}(||x||_e^2) \leq (1 - \varepsilon) 1\{||x||_e^2 \geq t - 2\delta\} + \varepsilon$. The result follows by taken expectations at both sides. □

Proof of Lemma B.2. The proof is identical to that of Lemma B.1 and will be omitted. □

Proof of Lemma B.3. The proof is identical to that of Lemma B.1 and will be omitted. $\square$

Proof of Lemma B.4. Observe that $\xi_n \equiv \sqrt{n}\mathbb{V}_n \sim N(0, \Sigma_n)$ (recall $\Sigma_n = E[Z_{1,n}Z_{1,n}^T]$). Note that

$$\begin{aligned}\xi_n^T \xi_n = (\Sigma_n^{-1/2}\xi_n)^T \Sigma_n (\Sigma_n^{-1/2}\xi_n) =& (U_n \Sigma_n^{-1/2}\xi_n)^T \Lambda_n (U_n \Sigma_n^{-1/2}\xi_n) \\ \equiv& (\zeta_n)^T \Lambda_n (\zeta_n) = \sum_{l=1}^{d(n)} \lambda_l \zeta_{l,n}^2\end{aligned}$$

where the third inequality follows from the diagonalization of Σ_n, where Λ_n is a diagonal matrix of eigenvalues and U_n is an unitary matrix. Observe that $\zeta_n = U_n \Sigma_n^{-1/2}\xi_n \sim N(0, I_{d(n)})$ and thus its components are iid standard Gaussian, so $\zeta_l^2 \sim \chi_1^2$ and $\lambda_l \zeta_l^2 \sim \Gamma(1/2, 2\lambda_l)$. Moreover, it is easy to see that

$$E[\lambda_l \zeta_{l,n}^2] = \lambda_l \textit{ and } Var(\lambda_l \zeta_{l,n}^2) = 2\lambda_l^2$$

which implies that $Var(\sum_{l=1}^{d(n)} \lambda_l \zeta_{l,n}^2) = 2tr\{\Sigma_n^2\}$. Also, $E[|\lambda_l \zeta_{l,n}^2|^3] = \lambda_l^3 E[|\zeta_{l,n}|^6] \leq C\left(\lambda_{max}(\Sigma_n)\right)^3$ where $\lambda_{max}(A)$ is the largest eigen value of a matrix A.

If $d(n) \leq d < \infty$, the proof follows from the fact that $\Gamma(1/2, 2\lambda_l)$ does not have mass points and is straight forward to show that the statement holds for any n.

Suppose that $d(n) \to \infty$ as $n \to \infty$. [4] Therefore,

$$\begin{aligned}
\sup_t \mathbf{\Phi}_n \left(|||\sqrt{n}\mathbb{V}_n||_e^2 - t| \leq \sqrt{tr\{\Sigma_n^2\}}\gamma \right) =& \sup_t \mathbf{\Phi}_n \left(|\frac{||\xi_n||_e^2}{\sqrt{2tr\{\Sigma_n^2\}}} - \frac{t}{\sqrt{2tr\{\Sigma_n^2\}}}| \leq \gamma/\sqrt{2} \right) \\
=& \sup_{t'} \mathbf{\Phi}_n \left(|\frac{||\xi_n||_e^2}{\sqrt{2tr\{\Sigma_n^2\}}} - t'| \leq \gamma/\sqrt{2} \right) \\
=& \sup_{t'} \mathbf{\Phi}_n \left(|\frac{\sum_{l=1}^{d(n)} \lambda_l(\zeta_{l,n}^2 - 1)}{\sqrt{2tr\{\Sigma_n^2\}}} - t' + tr\{\Sigma_n\}| \leq \gamma/\sqrt{2} \right) \\
=& \sup_{t''} \mathbf{\Phi}_n \left(|\frac{\sum_{l=1}^{d(n)} \lambda_l(\zeta_{l,n}^2 - 1)}{\sqrt{2tr\{\Sigma_n^2\}}} - t''| \leq \gamma/\sqrt{2} \right)
\end{aligned}$$

where the second line and fourth line follow from the fact that if $t \in \mathbb{R}$, then $\frac{t}{\sqrt{2tr\{\Sigma_n^2\}}} \in \mathbb{R}$.

Then, by Berry-Essen bound (Theorem 2, p. 544 feller Feller (1971)).

$$\sup_t \left| \mathbf{\Phi}_n \left(\frac{\sum_{l=1}^{d(n)} \lambda_l(\zeta_{l,n}^2 - 1)}{\sqrt{2tr\{\Sigma_n^2\}}} \leq t' \right) - \Phi(t') \right| \leq 6C \frac{\sum_{l=1}^{d(n)} \lambda_l^3}{(2tr\{\Sigma_n^2\})^{3/2}}$$

where Φ is the standard Gaussian cdf. Since $\frac{\sum_{l=1}^{d(n)} \lambda_l^3}{(2tr\{\Sigma_n^2\})^{3/2}} = \frac{tr\{\Sigma_n^3\}}{(2tr\{\Sigma_n^2\})^{3/2}}$, by assumption 2.1(i), for any $\varepsilon > 0$, there exists a $N(\varepsilon)$ such that $\frac{tr\{\Sigma_n^3\}}{(tr\{\Sigma_n^2\})^{3/2}} < 0.5\varepsilon$ for all $n \geq N(\varepsilon)$. Thus,

$$\begin{aligned}
\sup_{t\in\mathbb{R}} \mathbf{\Phi}_n \left(|||\xi_n||_e^2 - t| \leq \sqrt{tr\{\Sigma_n^2\}}\gamma \right) =& \sup_{t\in\mathbb{R}} \mathbf{\Phi}_n \left(\sqrt{tr\{\Sigma_n^2\}}\gamma - t \leq ||\xi_n||_e^2 \leq t + \sqrt{tr\{\Sigma_n^2\}}\gamma \right) \\
\leq& \sup_{t\in\mathbb{R}} \left| \Phi \left(t + \gamma/\sqrt{2} \right) - \Phi \left(t - \gamma/\sqrt{2} \right) \right| + 0.5\varepsilon.
\end{aligned}$$

Since for any $\varepsilon > 0$, there exists a $\gamma(\varepsilon)$ such that $\left| \Phi \left(t + \gamma/\sqrt{2} \right) - \Phi \left(t - \gamma/\sqrt{2} \right) \right| < 0.5\varepsilon$, the desired result follows. $\square$

[4] The relevant cases for us are: (i) $d(n) \leq d < \infty$ or (ii) $d(n) \uparrow \infty$, that is why we implicitly assume the limit of $(d(n))_n$ exist.

B.2 Proofs of Lemmas in Section 4

Proof of Lemma 4.1. The proof is analogous to that of Lemma 4.2 and will not be repeated here. $\square$

Proof of Lemma 4.2. Throughout the proof, let $\delta_n \equiv \sqrt{tr\{\Sigma_n^2\}}\gamma(\varepsilon)$, where $\gamma(\varepsilon)$ as in lemma B.4. By remark B.1 (applied thrice),

$$E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2 \geq t\}\right] \geq E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2 \geq t - 3\delta_n\}\right] - 3\varepsilon \tag{29}$$

for all $n \geq N(\varepsilon)$. By lemma B.2(ii),

$$E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2 \geq t\}\right] \geq \frac{1}{1-\varepsilon}E_{\mathbf{\Phi}_n}\left[\mathcal{P}_{t-\delta_n,\delta_n,h}(||\sqrt{n}\mathbb{V}_n||_e^2)\right] - \frac{\varepsilon}{1-\varepsilon} - 3\varepsilon \tag{30}$$

for all $h \leq h(\varepsilon, \delta_n)$ and all $n \geq N(\varepsilon)$. By lemma B.1(i), for all $h \leq h(\varepsilon, \delta_n)$

$$E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_e^2 \geq t\}|Z^n\right] \leq \frac{1}{1-\varepsilon}E_{\mathbf{P}_n^*}\left[\mathcal{P}_{t-\delta_n,\delta_n,h}(||\sqrt{n}\mathbb{Z}_n^*||_e^2)|Z^n\right]. \tag{31}$$

Hence, for all $h \leq h(\varepsilon, \delta_n)$ and all $n \geq N(\varepsilon)$,

$$\begin{aligned}&E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_e^2 \geq t\}|Z^n\right] - E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2 \geq t\}\right]\\ \leq&\frac{1}{1-\varepsilon}\left(E_{\mathbf{P}_n^*}\left[\mathcal{P}_{t-\delta_n,\delta_n,h}(||\sqrt{n}\mathbb{Z}_n^*||_e^2)|Z^n\right] - E_{\mathbf{\Phi}_n}\left[\mathcal{P}_{t-\delta_n,\delta_n,h}(||\sqrt{n}\mathbb{V}_n||_e^2)\right]\right)\\ &+\frac{\varepsilon}{1-\varepsilon} + 3\varepsilon.\end{aligned} \tag{32}$$

Similarly, by lemma B.1(ii), for all $h \leq h(\varepsilon, \delta_n)$

$$E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_e^2 \geq t\}|Z^n\right] \geq \frac{1}{1-\varepsilon}E_{\mathbf{P}_n^*}\left[\mathcal{P}_{t+2\delta_n,\delta_n,h}(||\sqrt{n}\mathbb{Z}_n^*||_e^2)|Z^n\right] - \frac{\varepsilon}{1-\varepsilon}. \tag{33}$$

By remark B.1 (applied thrice),

$$E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2 \geq t\}\right] \leq E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2 \geq t + 3\delta_n\}\right] + 3\varepsilon \tag{34}$$

for all $n \geq N(\varepsilon)$. By lemma B.2(ii),

$$E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2 \geq t\}\right] \leq \frac{1}{1-\varepsilon} E_{\mathbf{\Phi}_n}\left[\mathcal{P}_{t+2\delta_n,\delta_n,h}(||\sqrt{n}\mathbb{V}_n||_e^2)\right] + 3\varepsilon \tag{35}$$

for all $h \leq h(\varepsilon, \delta_n)$ and all $n \geq N(\varepsilon)$.

Hence,

$$\begin{aligned}
&E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_e^2 \geq t\}|Z^n\right] - E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_e^2 \geq t\}\right] \\
\geq &\frac{1}{1-\varepsilon}\left(E_{\mathbf{P}_n^*}\left[\mathcal{P}_{t+2\delta_n,\delta_n,h}(||\sqrt{n}\mathbb{Z}_n^*||_e^2)|Z^n\right] - E_{\mathbf{\Phi}_n}\left[\mathcal{P}_{t+2\delta_n,\delta_n,h}(||\sqrt{n}\mathbb{V}_n||_e^2)\right]\right) \\
&- \frac{\varepsilon}{1-\varepsilon} - 3\varepsilon.
\end{aligned} \tag{36}$$

By displays 32 and 36, in order to obtain the desired result it suffices to verify that $a \in \mathbb{R} \mapsto \mathcal{P}_{t,\delta,h}(a) \in \mathcal{C}_{h^{-1}}$. It is straight forward to check that $\mathcal{P}_{t,\delta,h}$ is three times continuously differentiable. Moreover, for any $a \in \mathbb{R}$,

$$|\partial \mathcal{P}_{t,\delta,h}(a)| \leq h^{-1}.$$

To show this expression, observe that by the Dominated Convergence Theorem, for any $a \in \mathbb{R}$,

$$\begin{aligned}
|\partial \mathcal{P}_{t,\delta,h}(a)| =& h^{-1}\left|\int p_{t,\delta}(u)(u-a)h^{-2}\phi((u-a)h^{-1})du\right| \\
=& h^{-1}\int |u-a|\, h^{-2}\phi((u-a)h^{-1})du \\
\leq& h^{-2}\sqrt{\int |u-a|^2\, h^{-1}\phi((u-a)h^{-1})du} \\
=& h^{-1}
\end{aligned}$$

where the second line follows from the fact that $0 \leq p_{t,\delta}(u) \leq 1$. Similarly calculations yield

$$|\partial^r \mathcal{P}_{t,\delta,h}(a)| \leq h^{-r}$$

which holds uniformly in $a \in \mathbb{R}$, δ, and t. $\square$

Proof of Lemma 4.3. Establishing the result is analogous to establishing a bound for $\Delta_{h^{-1}}(\mathbf{Q}_n^*(\cdot|Z^n), \mathbf{Q}_n)$ where $\mathbf{Q}_n^*(\cdot|Z^n)$ is $N(0, \hat{\Sigma}_n)$ and $\mathbf{Q}_n$ is $N(0, \Sigma_n)$. Let $\tilde{\xi}_n \sim \mathbf{Q}_n^*(\cdot|Z^n)$ and $\xi_n \sim \mathbf{Q}_n$.

For any $x \in \mathbb{R}^d$, let $f(x) \equiv g(||x||_e^2)$. Observe that for any $g \in \mathcal{C}_{h^{-1}}$, $\partial_i f(x) = g'(||x||_e^2) 2x_i$ and $\partial_{ij} f(x) = g''(||x||_e^2) 4x_i x_j + 2g'(||x||_e^2) 1\{i = j\}$.

By the Slepian interpolation (Rollin (2013) p. 4 — there the construction itself is slightly different, using $\sqrt{t}$ instead of $\cos(t)$ —),

$$E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[f\left(\tilde{\xi}_n\right) - f\left(\xi_n\right)\right] = \sum_{j=1}^{d(n)} \int_0^{\pi/2} E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[\partial_j f\left(\xi_n(t)\right) \dot{\xi}_{[j],n}(t)\right] dt$$

where $\xi_n(t) = \cos(t)\xi_n + \sin(t)\tilde{\xi}_n$ and $\dot{\xi}_{[j],n}(t)$ denotes the j-th coordinate of $\dot{\xi}_n(t)$ (the same holds for ξ_n, etc). Observe that $\dot{\xi}_{[j],n}(t) = -\sin(t)\xi_{[j],n} + \cos(t)\tilde{\xi}_{[j],n}$. Hence $(\dot{\xi}_{[j],n}(t), \xi_n(t))$ are jointly Gaussian with mean 0 a.s.-$\mathbf{P}_n$, for any t. Hence, by Stein's Identity (Stein (1981) and Chernozhukov *et al.* (2013b) Lemma H.2),

$$\begin{aligned} & E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[\partial_j f\left(\xi_n(t)\right) \dot{\xi}_{[j],n}(t)\right] \\ = & \sum_{l=1}^{d(k(n))} E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[\partial_{jl} f\left(\xi_n(t)\right)\right] E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[\xi_{[l],n}(t)\dot{\xi}_{[j],n}(t)\right]. \end{aligned}$$

It follows that

$$E\left[\xi_{[l],n}(t)\dot{\xi}_{[j],n}(t)\right] = E\left[(\tilde{\xi}_{[l],n}\tilde{\xi}_{[j],n} - \xi_{[l],n}\xi_{[j],n})\right] \sin(t)\cos(t).$$

Therefore,

$$\begin{aligned} & E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[f\left(\tilde{\xi}_n\right) - f\left(\xi_n\right)\right] = \sum_{j=1}^{d(n)}\sum_{l=1}^{d(n)} E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[(\tilde{\xi}_{[l],n}\tilde{\xi}_{[j],n} - \xi_{[l],n}\xi_{[j],n})\right] \\ & \times \int_0^{\pi/2} E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[\partial_{jl} f\left(\xi_n(t)\right)\right] \sin(t)\cos(t) dt \\ = & \sum_{j=1}^{d(n)}\sum_{l=1}^{d(n)} \left\{ n^{-1}\sum_{i=1}^{n} Z_{[l],i,n}Z_{[j],i,n} - \Sigma_{[j,l],n} \right\} \times \int_0^{\pi/2} E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[\partial_{jl} f\left(\xi_n(t)\right)\right] \sin(t)\cos(t) dt \end{aligned}$$

where the second line follows from the fact that $\tilde{\xi}_n \sim N(0, n^{-1}\sum_{i=1}^{n} Z_{i,n} Z_{i,n}^T)$, under $\mathbf{Q}_n^*(\cdot|Z^n)$.

Therefore,

$$
\begin{aligned}
& E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[f\left(\tilde{\xi}_n\right) - f\left(\xi_n\right)\right] \\
\leq & \max_{j,l}\left| n^{-1}\sum_{i=1}^{n} Z_{[l],i,n} Z_{[j],i,n} - \Sigma_{[j,l],n}\right| \times \sum_{j=1}^{d(n)}\sum_{l=1}^{d(n)}\int_0^{\pi/2} E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[|\partial_{jl} f\left(\xi_n(t)\right)|\right] |\sin(t)\cos(t)| dt.
\end{aligned}
$$

Observe that, by Cauchy-Swarchz inequality and the fact that $\partial_{ij} f(x) = g''(||x||_e^2) 4 x_i x_j + 2 g'(||x||_e^2) 1\{i = j\}$

$$
\begin{aligned}
\sum_{j=1}^{d(n)}\sum_{l=1}^{d(n)} E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[|\partial_{jl} f\left(\xi_n(t)\right)|\right] \leq & 4h^{-2}\sum_{j=1}^{d(n)}\sum_{l=1}^{d(n)} E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[|\xi_{[j],n}(t)||\xi_{[l],n}(t)|\right] + 2h^{-1} d(n) \\
\leq & 4h^{-2}\left(\sum_{j=1}^{d(n)}\sqrt{E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[|\xi_{[j],n}(t)|^2\right]}\right)^2 + 2h^{-1} d(n) \\
\leq & 4h^{-2} d(n) E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[||\xi_n(t)||_e^2\right] + 2h^{-1} d(n)
\end{aligned}
$$

Therefore, since $||\xi_n(t)||_e^2 \precsim \{||\xi_n||_e^2 + ||\tilde{\xi}_n||_e^2\}$,

$$
\begin{aligned}
\sum_{j=1}^{d(n)}\sum_{l=1}^{d(n)} E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[|\partial_{jl} f\left(\xi_n(t)\right)|\right] \precsim & d(n) h^{-1}\{h^{-1} E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[||\xi_n||_e^2 + ||\tilde{\xi}_n||_e^2\right] + 2\} \\
= & d(n) h^{-1}\{h^{-1}\left(tr\{\Sigma_n\} + tr\{\hat{\Sigma}_n\}\right) + 2\}.
\end{aligned}
$$

The desired result from the fact that $\int_0^{\pi/2} |\sin(t)\cos(t)| dt < \infty$. $\square$