Electronic Journal of Statistics
Vol. 9 (2015) 3046–3097
ISSN: 1935-7524
DOI: 10.1214/15-EJS1090

Bootstrap consistency for quadratic forms of sample averages with increasing dimension

Demian Pouzo*

530 Evans Hall # 3880. Berkeley, CA 94720
e-mail: dpouzo@econ.berkeley.edu

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. We apply our bootstrap results to model-specification testing problems when the number of moments is allowed to grow with the sample size.

MSC 2010 subject classifications: 60F05, 60F17, 62E20, 62F40.
Keywords and phrases: Asymptotic theory, bootstrap, high dimensional statistics, quadratic forms, model specification test.

Received December 2014.

Contents

*I would like to thank the Editor, George Michailidis, an anonymous associate editor and two referees for their comments and suggestions. I would also like to thank Xiaohong Chen, Noureddine El Karoui, Bryan Graham, Michael Jansson, Jim Powell and Elie Tamer for comments.

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 Praestgaard (1990) and Praestgaard and Wellner (1993)).

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.

So as to further illustrate the applicability of our results, in Section 4 we study a concrete application motivated by the work of Donald et al. (2003) who consider model-specification tests for models defined by a *diverging* number of moment conditions (this quantity determines our $d(n)$). By applying our results, we establish bootstrap consistency results for the distribution of the model-specification test statistics of two ubiquitous estimators in econometrics and statistics: The generalized empirical likelihood (GEL; Smith (1997)) estimator and The generalized method of moments (GMM; Hansen (1982)) estimator. By employing our bootstrap result we are able to perform inference for *non-optimally weighted* GMM estimators. To our knowledge these results are new.

By letting d to increase with sample size in our general theory, 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 does not literally grow with n, if, for instance, the model has a large number of parameters (or moment conditions as in our application), doing "fixed-d and large-n" asymptotics could be misleading, whereas doing "large-d and 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 (2015) 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.[1]

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 covariance $n^{-1}\sum_{i=1}^{n}Z_{i,n}Z_{i,n}^{T}$ 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. (2013b) 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 than $n^{1/4}$.

[1]In Section 4 we provide more concrete examples of these two cases in the context of our application.

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 theorem 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. (2013a), 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. (2013a) 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 Theorem and a discussion of its implications. Section 4 presents an application to model-specification tests. Section 5 presents a numerical simulations. Section 6 presents the proof of the main Theorem. Section 7 presents some concluding remarks. 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. $tr\{A\}$ 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 universal $C > 0$. 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. $wpa1-P$ means "with probability approaching one under P".

2. Preliminaries

Let $\{Z_{i,n} \in \mathbb{R}^{d(n)} : i = 1,...,n \textit{ and } n \in \mathbb{N}\}$ with $(d(n))_{n\in\mathbb{N}}$ being a non-decreasing integer-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. Henceforth, we will typically omit the sub-index n in $Z_{i,n}$.

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

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

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 exist 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*

$$\frac{\max\{d(n)(E_{\mathbf{P}_n}[||Z_1||_2^3])^2, E_{\mathbf{P}_n}[||Z_1||_2^4], (d(n))^4\}}{n} = o(1);$$

(ii) there exists a $\gamma > 0$ such that $\frac{(d(n))^{2+\gamma}}{n^\gamma}E_{\mathbf{P}_n}[||Z_1||_2^{4+2\gamma}] = o(1)$; (iii) there exists a $\kappa \geq 0$ such that $\frac{(\log(d(n)))^{\kappa/2}d(n)^{2+\kappa}}{n^{1+\kappa/2}}E_{\mathbf{P}_n}[||Z_1||_{2+\kappa}^{2(2+\kappa)}] = o(1)$.

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 provide sufficient conditions for this part of Assumption 2.1, it is convenient to provide bounds in terms of $d(n)$ for the quantities $E_{\mathbf{P}_n}[||Z_1||_2^q]$ (for different q's) and $E_{\mathbf{P}_n}[||Z_1||_{2+\kappa}^{2(2+\kappa)}]$ in the assumption.

Clearly, if $|Z_{[l],1}| \leq C < \infty$ a.s-$\mathbf{P}_n$ for all $l = 1,...,d(n)$ and all $n \in \mathbb{N}$, then $E_{\mathbf{P}_n}[||Z_1||_2^{2q}] = O(d(n)^q)$ for any $q > 0$.[2] For example, such condition is imposed by Vershynin (2012b) 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: $E_{\mathbf{P}_n}\left[e^{\lambda Z_{[l],1}^2}\right] \leq C < \infty$ for some $\lambda > 0$. For instance, if $(Z_{[l],1})^2$ is a sub-Gamma random variable (Boucheron et al. (2013) p. 27), then the condition holds since $E_{\mathbf{P}_n}\left[e^{\lambda Z_{[l],1}^2}\right] \leq \exp\{\frac{\lambda^2 v}{2(1-c\lambda)}\}$ for any $\lambda \in (0,1/c)$ and some $c > 0$. If $Z_{[l],1}$ is sub-Gaussian, then $(Z_{[l],1})^2$ is sub-exponential (see Vershynin (2012a) Lemma 5.14) and the condition holds by the same 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 and not on its joint behavior.

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

Lemma 2.1. *Suppose that there exists a $C > 0$ and $\lambda > 0$ such $E_{\mathbf{P}_n}\left[e^{\lambda Z_{[l],1}^2}\right] \leq C$ for all $l = 1, ..., d(n)$ and all $n \in \mathbb{N}$. Then $E_{\mathbf{P}_n}[||Z_1||_2^{2q}] \precsim d(n)^q$ for any $q > 0$.*

Proof. Observe that

$$E_{\mathbf{P}_n}[(||Z_1||_2^2/d(n))^q] = \int_0^\infty \mathbf{P}_n\left(||Z_1||_2^2/d(n) \geq t^{1/q}\right) dt$$
$$= q\int_0^\infty u^{q-1}\mathbf{P}_n\left(||Z_1||_2^2/d(n) \geq u\right) du$$

since $||Z_1||_2^2/d(n) = d(n)^{-1}\sum_{l=1}^{d(n)} |Z_{[l],1}|^2$, by the Markov inequality it follows that for any $\lambda > 0$

$$E_{\mathbf{P}_n}[(||Z_1||_2^2/d(n))^q] \leq \left(q\int_0^\infty u^{q-1}e^{-\lambda u}du\right) E_{\mathbf{P}_n}\left[e^{\lambda d(n)^{-1}\sum_{l=1}^{d(n)} |Z_{[l],1}|^2}\right].$$

By Jensen inequality $E_{\mathbf{P}_n}\left[e^{\lambda d(n)^{-1}\sum_{l=1}^{d(n)} |Z_{[l],1}|^2}\right] \leq d(n)^{-1}\sum_{l=1}^{d(n)} E_{\mathbf{P}_n}\left[e^{\lambda |Z_{[l],1}|^2}\right]$ which is bounded by a constant 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 $q > 0$. □

Under the conditions in the lemma, 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 $\frac{d(n)^2}{n}E_{\mathbf{P}_n}[||Z_1||_2^4] \precsim \frac{d(n)^4}{n} \to 0$.

That is, under conditions that bound all (polynomial) moments of the individual components of Z_1, 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 $\{\omega_{in} \in \mathbb{R} : i = 1, ..., n \text{ and } n \in \mathbb{N}\}$ 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 $E_{\mathbf{P}_n^*(\cdot|Z^n)}\left[(\omega_{in})^2\right] = 1$; (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.[3]

[3]Of course, the technique of proof can be applied 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 $\mathbb{Z}_n^* \equiv n^{-1}\sum_{i=1}^n \omega_{in} Z_i$ be the bootstrap analog of $\mathbb{Z}_n$.

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

$$\sup_{t\in\mathbb{R}} \left|\mathbf{P}_n^*\left(||\sqrt{n}\mathbb{Z}_n^*||_2^2 \geq t \mid Z^n\right) - \mathbf{P}_n\left(||\sqrt{n}\mathbb{Z}_n||_2^2 \geq t\right)\right| = o_{\mathbf{P}_n}(1). \tag{3}$$

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 6; here we present an heuristic argument. The first step of the proof is to apply Lindeberg interpolation techniques (see Chatterjee (2006) and Rollin (2013) and references therein) to approximate $\sqrt{n}\mathbb{Z}_n^*$ by $\sqrt{n}\mathbb{U}_n$ and $\sqrt{n}\mathbb{Z}_n$ by $\sqrt{n}\mathbb{V}_n$, where $\mathbb{U}_n$ and $\mathbb{V}_n$ are Gaussian random variables with zero mean and covariances $n^{-1}\sum_{i=1}^n Z_i Z_i^T$ and $E[Z_{1,n}Z_{1,n}^T]$ respectively.

In order to do this, we first approximate the indicator function $x \mapsto 1\{||x||_2^2 \geq t\}$ by "smooth" functions $x \mapsto \mathcal{P}_{t,\delta,h}(||x||_2^2)$; 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 the "smaller" δ is, the closer the function $\mathcal{P}_{t,\delta,h}$ is to the indicator function; see Lemmas B.1, B.2 and B.3 in the Appendix B. It is worth to note that what we mean by δ to be "small" depends on how $||\sqrt{n}\mathbb{V}_n||_2^2$ concentrates mass. Lemma B.4 in the Appendix B establishes an anti-concentration result, wherein we obtain that this random variable puts very little mass in any given interval. Therefore δ could actually be quite large, of the order of $\sqrt{tr\{\Sigma_n^2\}}$.

Second, since $x \mapsto \mathcal{P}_{t,\delta,h}(||x||_2^2)$ belongs to a class of "smooth" functions, we show that it suffices to show consistency under the weak norm (as opposed to the norm implied in 3).[4] This is done in Lemmas 6.1 and 6.2. The relevant class of "smooth" functions is given by $\mathcal{C}_M$, which is the class of functions $f : \mathbb{R} \to \mathbb{R}$ that are three times continuously differentiable and $\sup_x |\partial^r f(x)| \leq (M)^r$ and $\sup_x |f(x)| \leq 1$.

The following Theorems formalize the aforementioned approximation of $\sqrt{n}\mathbb{Z}_n^*$ by $\sqrt{n}\mathbb{U}_n$ and $\sqrt{n}\mathbb{Z}_n$ by $\sqrt{n}\mathbb{V}_n$ and can be viewed of independent interest since they show that a "generalized invariance principle" holds in our setup. Henceforth, we use $\mathbf{\Phi}_n^*(\cdot|Z^n)$ and $\mathbf{\Phi}_n$ respectively, to denote their probability distributions.

[4] The formal definition of the norm is presented in Equation 6 in Section 6.

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

$$\sup_{f \in \mathcal{C}_{h^{-1}}} \left| E_{\mathbf{P}_n^*}\left[f\left(||\sqrt{n}\mathbb{Z}_n^*||_2^2 \right) | Z^n \right] - E_{\mathbf{\Phi}_n^*}\left[f\left(||\sqrt{n}\mathbb{U}_n||_2^2 \right) | Z^n \right] \right| = o_{\mathbf{P}_n}(h^{-2}).$$

Proof. See Appendix A. □

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

$$\sup_{f \in \mathcal{C}_{h^{-1}}} \left| E_{\mathbf{P}_n}\left[f\left(||\sqrt{n}\mathbb{Z}_n||_2^2 \right) \right] - E_{\mathbf{\Phi}_n}\left[f\left(||\sqrt{n}\mathbb{V}_n||_2^2 \right) \right] \right| = o(h^{-2}).$$

Proof. See Appendix A. □

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 $n^{-1}\sum_{i=1}^n Z_i Z_i^T$ to one with zero mean and variance $E[Z_1 Z_1^T]$. Lemma 6.3 in Section 6 — which is based in the Slepian interpolation (see Chernozhukov et al. (2013a), Chernozhukov et al. (2013b) and Rollin (2013) and references therein)— establishes that is enough to show that

$$d(n) \max_{1 \leq j,l \leq d(n)} \left| n^{-1} \sum_{i=1}^n Z_{[j],i} Z_{[l],i} - E_{\mathbf{P}_n}[Z_{[j],1} Z_{[l],1}] \right| = o_{\mathbf{P}_n}(1). \tag{4}$$

In Section 6, we show that, employing standard arguments, the expression 4 holds under our assumptions. A similar result is obtained by Chernozhukov et al. (2013a) without the scaling factor of $d(n)$; their setup, however, is different since the object of interest is $\max_{1 \leq j \leq d(n)} |n^{-1/2} \sum_{i=1}^n Z_{[j],i}|$ (as opposed to $||n^{-1/2}\sum_{i=1}^n Z_i||_2^2$). [5]

Asymptotic Distribution of $||\sqrt{n}\mathbb{Z}_n||_2^2$. An implication of the proof of Theorem 3.1 and Theorem 3.3 is that

$$\sup_{t \in \mathbb{R}} \left| \mathbf{P}_n \left(\frac{||\sqrt{n}\mathbb{Z}_n||_2^2 - d(n)}{\sqrt{d(n)}} \geq t \right) - \mathbf{\Phi}_n \left(\frac{||\sqrt{n}\mathbb{V}_n||_2^2 - d(n)}{\sqrt{d(n)}} \geq t \right) \right| = o(1). \tag{5}$$

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 $\frac{||\sqrt{n}\mathbb{Z}_n||_2^2 - d(n)}{\sqrt{2d(n)}} \Rightarrow N(0,1)$ or, informally, $||\sqrt{n}\mathbb{Z}_n||_2^2$ 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 $\frac{||\sqrt{n}\mathbb{Z}_n||_2^2 - tr\{\Sigma_n\}}{\sqrt{2tr\{\Sigma_n^2\}}}$ is

[5] An important consequence of this difference is that, as opposed to our case, Chernozhukov et al. (2013a) can use a "smooth maximum function" to approximate their quantity of interest; the approximation error is only of order $\log d$. This, allows them to obtain faster rates for the approximation of the indicator functions with smooth functions. This, in turn, translates into a faster overall rate of convergence — $d = o(\exp(n))$ in their case. See Wasserman (2014) for a discussion and a nice review of these results.

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 $||\sqrt{n}\mathbb{Z}_n||_2^2$ are "drifting" to infinity, the bootstrap still provides a good approximation since the moments of $||\sqrt{n}\mathbb{Z}_n^*||_2^2$ are mimicking this behavior.

On the Lindeberg Interpolation. Theorems 3.2 and 3.3 are based on the following Lindeberg interpolation for quadratic forms.[6]

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 $q>0$*

$$|E[f(||\sum_{i=1}^n A_i||_2^2)] - E[f(||\sum_{i=1}^n B_i||_2^2)]| \leq \mathbf{S}_n + L_2(f)\left(\frac{L_3(f)}{L_2(f)}\right)^q \mathbf{R}_n$$

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

$$\begin{aligned}
\mathbf{S}_{1,n} =& \sum_{i=1}^n |E\left[f''\left(||\mathbb{S}_{i:n}||_2^2\right)\right] E[||B_i||_2^4] - E[||A_i||_2^4]| \\
\mathbf{S}_{2,n} =& 4\sum_{i=1}^n |E\left[f''\left(||\mathbb{S}_{i:n}||_2^2\right)\mathbb{S}_{i:n}^T\right]\left(E[B_i||B_i||_2^2] - E[A_i||A_i||_2^2]\right)| \\
\mathbf{R}_n =& \sum_{i=1}^n E\left[\left(\mathbb{S}_{i:n}^T B_i + ||B_i||_2^2\right)^{2+q} + \left(\mathbb{S}_{i:n}^T A_i + ||A_i||_2^2\right)^{2+q}\right]
\end{aligned}$$

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

Proof. See Appendix A. □

It is worth pointing out that 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 $||\mathbb{S}_{i:n}+A_i||_2^2$ and $||\mathbb{S}_{i:n}+B_i||_2^2$, 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 a $d(n)$-dimensional 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 $g(x)\equiv f(||x||_2^2)$. As can be seen from the derivations in Chatterjee and Meckes (2008), the remainder term will also require bounds on higher

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

derivatives of g (and thus f), but of the form $\sup_{x\neq y}\frac{\|Hess(g)(x)-Hess(g)(y)\|_{op}}{||x-y||_2}$. [7] 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 $t_n(\alpha,Z^n)\equiv\inf\{t:\mathbf{P}_n^*\left(||\sqrt{n}\mathbb{Z}_n^*||_2^2\leq t\mid Z^n\right)\geq\alpha\}$. Due to the distribution consistency result proven in Theorem 3.1, we can approximate the α-th quantile of the distribution of $||\sqrt{n}\mathbb{Z}_n||_2^2$ by $t_n(\alpha,Z^n)$, in the sense that

$$\mathbf{P}_n\left(||\sqrt{n}\mathbb{Z}_n||_2^2\geq t_n(\alpha,Z^n)-\eta\right)\leq\alpha+o(1)$$

for any $\eta>0$. If $t_n(\alpha,Z^n)$ is a continuity point of $\mathbf{P}_n^*(\cdot|Z^n)$, then

$$\mathbf{P}_n^*\left(||\sqrt{n}\mathbb{Z}_n^*||_2^2\geq t_n(\alpha,Z^n)\mid Z^n\right)=\alpha,$$

and the first display becomes $\mathbf{P}_n\left(||\sqrt{n}\mathbb{Z}_n||_2^2\geq t_n(\alpha,Z^n)\right)=\alpha+o(1)$. Hence, Theorem 3.1 can be used to construct valid p-values based on the bootstrap.

4. An application to model specification tests for GEL and GMM estimators

In this section we apply our results to construct bootstrap-based specification tests for models with increasing number of moment restrictions. We do this for two estimators: generalized method of moment (GMM; see Hansen (1982)) estimator and generalized empirical likelihood (GEL; see Smith (1997)) estimator. Both estimators are widely used in econometrics and statistics and encompass a wide range of commonly used estimators such as Z-estimators (Van der Vaart (2000) Ch. 5), and empirical likelihood estimator (Owen (1988)), respectively.[8]

In models characterized by moment conditions, model-specification tests (MST) allow us to check whether the moment conditions match the data well or not. In this setup with increasing moment restrictions, MST has been studied by Donald et al. (2003) (DIN, henceforth); see also de Jong and Bierens (1994). They show that the MST statistic is asymptotically a quadratic form of scaled sample averages; however, they rely on inferential methods build on expressions akin to 5. Instead, by applying our Theorem 3.1, we can use the weighted bootstrap method to approximate the asymptotic distribution of MST statistics; thus complementing their results by providing an alternative way of constructing asymptotic p-values. Moreover, as explained below, by not relying on CLT-type results to approximate the limiting distribution, we are able to

[7] $Hess(g)$ 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.

[8] See Imbens (2002) for additional examples and a discussion. See also Hall (2005) for a review for GMM.

provide valid asymptotic inference for a larger class of GMM estimators than the one considered in DIN.

The setup closely follows that of DIN and is as follows. Suppose $(X_i)_{i=1}^n$ is an i.i.d. sample of real-valued random variables with $X_i \sim \mathbf{P}_n = \mathbf{P}$. The model we consider is one where the true parameter of interest, $\theta_0 \in Int(\Theta)$ — with Θ a compact subset of $\mathbb{R}^q$— is uniquely identified by the following set of moment conditions

$$E_{\mathbf{P}}[g(X, \theta_0)] = 0$$

where $g : \mathbb{R} \times \mathbb{R}^q \to \mathbb{R}^d$ is known to the researcher.

The main feature of this setup is that it allows $d \equiv d(n)$ to grow with the sample size. In many cases this departure from the standard theory is of relevance. For example, in many models the identifying condition is given by a conditional moment restriction, $E_{\mathbf{P}}[\rho(Y, \theta_0)|W]$ — where ρ maps into $\mathbb{R}^J$ with J fixed — and the researcher converts it to a series of unconditional moment restrictions $E_{\mathbf{P}}[\rho(Y, \theta_0) \otimes q^{K(n)}(W)]$ where $q^{K(n)}(w) = (q_1(w), ..., q_{K(n)}(w))$ are basis functions such as Fourier series, P-splines, etc; this is the case considered in DIN (see also de Jong and Bierens (1994) and references therein). For this case $x = (y, w)$, $d(n) = JK(n)$ and $g(x, \theta) = \rho(y, \theta) \otimes q^{K(n)}(w)$.

An alternative motivation to consider increasing d would be cases where although the number of moments is fixed, it could be large and thus treating it as a diverging sequence could deliver more accurate asymptotics. As pointed out by Koenker and Machado (1999) one example of this could be the panel data model in Arellano and Bond (1991) where $x = (y_1, ..., y_T)$ and the components of the vector $g(x, \theta)$ are given by $((y_t - y_{t-1}) - \theta(y_{t-1} - y_{t-2}))y_{t-s}$ for $s = 1, ..., t-1$ and $t = 3, ..., T$. Here, for a panel of length T, the number of instruments/moments is given by $d = (T-2)(T-1)/2$. [9]

The next assumptions impose some regularity conditions on g. These restrictions are standard in the literature and can be somewhat relaxed (e.g. see Donald et al. (2003) and references therein).

Assumption 4.1. $\Omega \equiv E_{\mathbf{P}}[g(X, \theta_0)g(X, \theta_0)^T]$ *exists with* $C^{-1} \leq \lambda_l(\Omega) \leq C$ *for all* $l = 1, ..., d$ *for some* $C \geq 1$.

For instance, for the case where $g = \rho \otimes q^{K(n)}$ (for simplicity, let $J = 1$) it suffices to assume that $E_{\mathbf{P}}[\rho(Y, \theta_0)^2|W]$ and the eigenvalues of $E_{\mathbf{P}}[q^{K(n)}(W, \theta_0)q^{K(n)}(W, \theta_0)^T]$ are both bounded bounded and bounded away from zero a.s.-$\mathbf{P}$. These assumptions are standard; see Donald et al. (2003) for a discussion. [10]

Let $\mathcal{N}$ be an open neighborhood of θ_0.

[9] For instance for $T = 4$, $d = d(n) = 3$ and for $T = 5$, $d = d(n) = 6$. In cases where $d(n) = o(n^{1/4})$, these values imply that, roughly speaking, the number of observations should be larger than 82 and 1300, resp. It is also worth to point out that in case where T is large, one can simply include fewer lags y_{t-s} in $((y_t - y_{t-1}) - \theta(y_{t-1} - y_{t-2}))y_{t-s}$ and thus reduce d.

[10] These assumptions are also standard in the context of series-based estimators; see Chen (2007).

Assumption 4.2. *For all n: (i) $E_{\mathbf{P}}\left[\sup_{\theta\in\mathcal{N}}||g(X,\theta)||_2^{2(2+\gamma)}\right] \precsim d(n)^{2+\gamma}$ for some $\gamma \geq 0$; (ii) $\theta \mapsto g(X,\theta)$ is continuously differentiable a.s.-$\mathbf{P}$; (iii) $E_{\mathbf{P}}[\sup_{\theta\in\mathcal{N}}||\nabla_\theta g(X,\theta)||_2^{2\beta}] \precsim d(n)^\beta$ for some $\beta \geq 1$; (iv) there exists a measurable $x \mapsto \delta_n(x)$ such that $||\nabla_\theta g(X,\theta) - \nabla_\theta g(X,\theta_0)||_2 \precsim \delta_n(X)||\theta-\theta_0||_2$ for all $\theta \in \mathcal{N}$ a.s.-$\mathbf{P}$, and $E_{\mathbf{P}}[\delta_n(X)^2] \precsim d(n)$.*[11]

For instance, for the case $g = \rho \otimes q^K$ for many basis functions such as splines and Fourier series it holds that $\sup_w ||q^K(w)||_2 \precsim \sqrt{K}$.[12] Thus, the previous assumption holds provided that $E[\sup_{\theta\in\mathcal{N}}||\rho(Y,\theta)||_2^{2(2+\gamma)}|W]$ and $E_{\mathbf{P}}[\sup_{\theta\in\mathcal{N}}||\nabla_\theta\rho(Y,\theta)||_2^{2\beta}|W]$ are bounded by a constant C, and $||\nabla_\theta\rho(Y,\theta) - \nabla_\theta\rho(Y,\theta_0)||_2 \precsim \delta(Y)||\theta-\theta_0||_2$ with $E_{\mathbf{P}}[\delta(Y)^2|W] \leq C$, a.s.-$\mathbf{P}$, for some $C > 0$.[13]

The GMM estimator is given by $\hat{\theta}_{GMM,n} = \arg\min_{\theta\in\Theta}\hat{Q}_{GMM,n}(\theta)$ where

$$\hat{Q}_{GMM,n}(\theta) \equiv n^{-1}\sum_{i=1}^n g(X_i,\theta)^T\hat{W}_n n^{-1}\sum_{i=1}^n g(X_i,\theta)$$

with $\hat{W}_n \in \mathbb{R}^{d\times d}$ is a (possibly random) positive definite matrix. The following mild condition is required

Assumption 4.3. *There exists a $W \in \mathbb{R}^{d(n)\times d(n)}$ positive definite and a $C \geq 1$ such that $||\hat{W}_n - W||_2 = o_{\mathbf{P}}(d(n)^{-1/2})$ and $C^{-1} \leq \lambda_l(W) \leq C$ for all $l = 1,...,d(n)$ and $n \in \mathbb{N}$.*

The bootstrap analog is given by $\hat{\theta}^*_{GMM,n} = \arg\min_{\theta\in\Theta}\hat{Q}^*_{GMM,n}(\theta)$ where

$$\hat{Q}^*_{GMM,n}(\theta) = n^{-1}\sum_{i=1}^n \omega_{i,n}g(X_i,\theta)^T\hat{W}_n n^{-1}\sum_{i=1}^n \omega_{i,n}g(X_i,\theta).$$

These formulas yield the MST statistic: $\hat{T}_{GMM,n} \equiv n\hat{Q}_{GMM,n}(\hat{\theta}_{GMM,n})$ and its bootstrap version $\hat{T}^*_{GMM,n} \equiv n\hat{Q}^*_{GMM,n}(\hat{\theta}^*_{GMM,n})$.

In order to simplify the exposition we directly impose that $(\omega_{i,n})_{i\leq n}$ satisfy Assumption 2.2 and also that they are uniformly bounded; this last assumption is not necessary for the results but imposing it greatly simplifies the technical derivations in our proofs.

It is worth to point out that DIN only considers GMM estimators with $W = \Omega^{-1}$ because they rely on CLT-type approximations for inference (e.g., see their Theorem 6.3). Since our result allow us to focus on bootstrap-based inference, the weighting matrix W does not need to coincide with Ω^{-1}; in fact it can simply be chosen as $\hat{W} = W = I$. That is, our results provide valid asymptotic inference for MST statistics for a larger class of GMM estimator, one with $W \neq \Omega^{-1}$.

---

[11]The notation $\nabla_\theta g(x,\theta)$ means the gradient with respect to θ of the function g; it is a $q \times d$ matrix. For any matrix, A, $||A||_2$ is defined as the operator norm.

[12]Other series like power series typically present $\sup_w ||q^K(w)||_2 \precsim K$, or more generally one can think of $\sup_w ||q^K(w)||_2 \precsim \zeta(K)$ for some function ζ. These cases can be accommodated in our theory, at the expense of further restricting the rate of growth of $d(n)$.

[13]These restrictions are analogous to Assumptions 4-6 in Donald et al. (2003).

The GEL estimator is given by

$$\hat{\theta}_{GEL,n} = \arg\min_{\theta\in\Theta} \hat{Q}_{GEL,n}(\theta),$$
$$\textit{where } \hat{Q}_{GEL,n}(\theta) \equiv \sup_{\lambda\in\Lambda(\theta)} \sum_{i=1}^{n} s(\lambda^T g(X_i,\theta))$$

where $s : \mathcal{V} \subseteq \mathbb{R} \mapsto \mathbb{R}$ is concave and twice continuously differentiable with Lipschitz second derivative, $\mathcal{V}$ includes a neighborhood of 0, and $\Lambda(\theta) \equiv \{\lambda \in \mathbb{R}^d : \lambda^T g(X,\theta) \in \mathcal{V},\ a.s. - \mathbf{P}\}$. The function s can be chosen to encompass several estimators of interest such as empirical likelihood ($s(\cdot) = \ln(1-\cdot)$), exponential tilting ($s(\cdot) = -\exp(\cdot)$; Imbens et al. (1998) and Kitamura and Stutzer (1997)) and continuously updating GMM ($s(\cdot) = -0.5(1+\cdot)^2$; Hansen et al. (1996)). Henceforth, to simplify the presentation we assume the following normalization $s'(0) = s''(0) = -1$.

Analogously to GMM, we have the following MST statistic for GEL: $\hat{T}_{GEL,n} \equiv 2\left\{\hat{Q}_{GEL,n}(\hat{\theta}_{GEL,n}) - ns(0)\right\}$ and its bootstrap version $\hat{T}^*_{GEL,n} \equiv 2\left\{\hat{Q}^*_{GEL,n}(\hat{\theta}^*_{GEL,n}) - ns(0)\right\}$, where $\hat{\theta}^*_{GEL,n} = \arg\min_{\theta\in\Theta} \hat{Q}^*_{GEL,n}(\theta)$ and $\hat{Q}^*_{GEL,n}$ is defined as $\hat{Q}_{GEL,n}$ but with $\omega_{i,n} g(x_i,\cdot)$ instead of $g(x_i,\cdot)$.[14]

The next assumption is a high level condition. Part (i) ensures existence of a minimizer for λ and part (ii) imposes convergence rates on the GMM and GEL estimators. Because our main goal is to establish the asymptotic behavior of the MST statistics, we directly impose this assumption to ease the exposition.

Assumption 4.4. *(i) $\hat{\lambda}^*_n = \arg\max_{\lambda\in\Lambda(\hat{\theta}^*_{GEL,n})} \sum_{i=1}^n s(\lambda^T \omega_{i,n} g(X_i, \hat{\theta}^*_{GEL,n}))$ exists wpa1-$\mathbf{P}$ and $||\hat{\lambda}^*_n||_2 = O_{\mathbf{P}^*_n(\cdot|Z^n)}(\sqrt{d(n)/n})$, wpa1-$\mathbf{P}$; (ii) $\hat{\theta}^*_{j,n} = \theta_0 + O_{\mathbf{P}^*_n(\cdot|Z^n)}(n^{-1/2})$ wpa1-$\mathbf{P}$ and $\hat{\theta}_{j,n} = \theta_0 + O_{\mathbf{P}}(n^{-1/2})$ for $j \in \{GEL, GMM\}$.*

The derivation of both parts of this assumption from more primitive conditions can be obtained from the results in DIN and references therein; in particular in Lemma A.10 and Theorems 5.4 and 5.6.

The following lemma establishes that the test statistics for both estimators are asymptotically equivalent to a quadratic form on sample averages of g.

Lemma 4.1. *Suppose Assumptions 4.1, 4.2, 4.3 and 4.4 hold. Also, suppose that $\frac{d(n)^{\max\{2+4/\gamma,4\}}}{n} = o(1)$. Then*

$$\hat{T}_{GMM,n} = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n g(X_i,\theta_0)\right)^T W \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n g(X_i,\theta_0)\right) + o_{\mathbf{P}}(\sqrt{d(n)})$$
$$\hat{T}_{GEL,n} = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n g(X_i,\theta_0)\right)^T \Omega^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n g(X_i,\theta_0)\right) + o_{\mathbf{P}}(\sqrt{d(n)})$$

[14]Abusing notation we still denote $\Lambda(\theta)$ as the set for the bootstrap case.

$$\hat{T}^*_{GMM,n} = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_{i,n} g(X_i,\theta_0)\right)^T W \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_{i,n} g(X_i,\theta_0)\right) + o_{\mathbf{P}^*_n(\cdot|Z^n)}(\sqrt{d(n)}),$$

$$\hat{T}^*_{GEL,n} = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_{i,n} g(X_i,\theta_0)\right)^T \Omega^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_{i,n} g(X_i,\theta_0)\right) + o_{\mathbf{P}^*_n(\cdot|Z^n)}(\sqrt{d(n)})$$

wpa1-$\mathbf{P}$.

Proof. See Appendix C. □

This Lemma establishes that the test statistics, asymptotically, behave as quadratic forms of (properly scaled) sample averages. Thus, our result in Theorem 3.1 can be applied to these cases with $Z_i \equiv g(X_i,\theta_0)W^{1/2}$ or $Z_i \equiv g(X_i,\theta_0)\Omega^{-1/2}$. The next Theorem formalizes this claim in this particular setting.

Theorem 4.1. *Suppose Assumptions 4.1, 4.2, 4.3 and 4.4 hold. Also, suppose that* $\frac{d(n)^{\max\{2+4/\gamma,4\}}}{n} = o(1)$. *Then*

$$\sup_{t\in\mathbb{R}} \left| \mathbf{P}^*_n\left(\frac{\hat{T}^*_{GMM,n}}{\sqrt{d(n)}} \geq t \mid Z^n\right) - \mathbf{P}\left(\frac{\hat{T}_{GMM,n}}{\sqrt{d(n)}} \geq t\right)\right| = o_{\mathbf{P}}(1),$$

and

$$\sup_{t\in\mathbb{R}} \left| \mathbf{P}^*_n\left(\frac{\hat{T}^*_{GEL,n}}{\sqrt{d(n)}} \geq t \mid Z^n\right) - \mathbf{P}\left(\frac{\hat{T}_{GEL,n}}{\sqrt{d(n)}} \geq t\right)\right| = o_{\mathbf{P}}(1).$$

Proof. See Appendix C. □

This result allow us to compute bootstrap-based p-values for the MST statistics for the general classes of GMM and GEL estimators, even when the number of moment restrictions increases with the sample size (but not too fast). In particular, for $\gamma \geq 2$, our condition on rate imposes that $d(n)^4/n = o(1)$ which is the one required in Theorem 6.4 in DIN, but at the cost of imposing restrictions on some higher moments of $||g(\cdot,\theta_0)||_2$ (see Assumption 4.2(i)).

5. Numerical simulations

In this section we present a Monte Carlo (MC) study to assess the finite sample behavior of our procedure. We perform 5000 MC repetitions and in each draw we perform 5000 bootstrap repetitions.

The design is as follows: In each MC repetition we draw $Z_i = V^{1/2}\sqrt{12}U_i$ with $U_i \sim U(-0.5, 0.5)$ for $i = 1, ..., n$, and V is a positive definite symmetric matrix specified below. Let

$$Q_n = n\left(n^{-1}\sum_{i=1}^{n} Z_i\right)^T\left(n^{-1}\sum_{i=1}^{n} Z_i\right)$$

and the associated bootstrapped version is given by

$$Q_n^* = n\left(n^{-1}\sum_{i=1}^{n} \omega_i Z_i\right)^T\left(n^{-1}\sum_{i=1}^{n} \omega_i Z_i\right).$$

Throughout the study we use $\omega \sim N(0,1)$.

We are interested in studying $\mathbb{K}_n^B = \sup_{a\in\mathbb{A}}\left|\mathbf{P}_n\left(Q_n \geq t_n^B(a, Z^n)\right) - (1-a)\right|$ and, for comparison, $\mathbb{K}_n = \sup_{a\in\mathbb{A}}\left|\mathbf{P}_n\left(Q_n \geq t_n(a)\right) - (1-a)\right|$, where $t_n^B(a, Z^n)$ is the a-th empirical percentile of Q_n^* and $t_n(a)$ is the a-th percentile of a chi-square with degrees of freedom $d(n)$.[15] [16] We let $\mathbb{A} = \{0.90, 0.95, 0.975, 0.99\}$. The typical application for our results is testing — like in the Section 4 —, and with this in mind $\mathbb{A}$ is designed to capture the relevant values of a for which we would like to assess the performance of the approximation.

Approximation Error. Figure 1 shows the $\log(\mathbb{K}_n/\mathbb{K}_n^B)$ for different values of the weighting matrix V and for $n = 500$ and $d(n) = 3$. When $V = I$ both, the chi-squared-based and boostrap-based procedures yield correct approximations of the limiting distribution, and thus $\log(\mathbb{K}_n/\mathbb{K}_n^B)$ is close to one. As expected, for cases where $V = (1+\epsilon/\sqrt{n})I$ with $\epsilon \neq 0$, the chi-squared-based approximation does not approximate the limiting distribution, whereas the bootstrap-based continues to do so. The simulations shows that even for

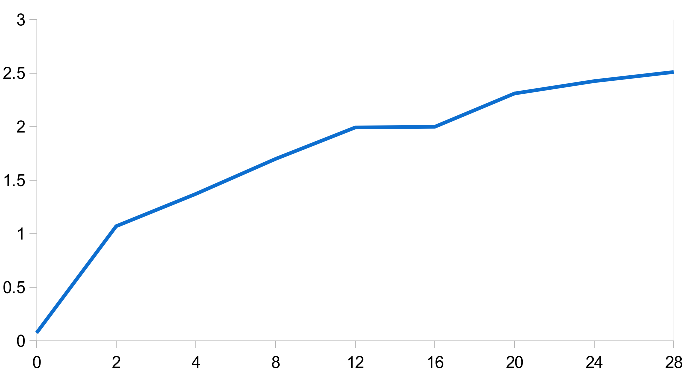

Fig 1. *Plot of* $\log(\mathbb{K}_n/\mathbb{K}_n^B)$ *for different values of* $V = (1+\epsilon/\sqrt{n})I$ *for* $\epsilon \in \{0, 2, ..., 28\}$.

[15] In both cases, we approximate $\mathbf{P}_n$ using the empirical cdf across MC repetitions.

[16] $\mathbb{K}_n^B$ is in fact the quantity of interest since, by construction, $1 - a$ coincides with the empirical quantile of Q_n^*, thus $\mathbb{K}_n^B$ approximates $\sup_{a\in\mathbb{A}}\left|\mathbf{P}_n\left(Q_n \geq t_n^B(a, Z^n)\right) - \mathbf{P}_n^*(Q_n^* \geq t_n^B(a, Z^n) \mid Z^n)\right|$. A similar observation holds for $\mathbb{K}_n$.

small values of $\epsilon/\sqrt{n}$, the difference is non-negligible. We also note that the deviations from $V = I$ we consider are "mild" and we expect that for more complex deviations the results will be even more stark.

Table 1 shows the value of $100 \times \left|\mathbf{P}_n\left(Q_n \geq t_n^B(a, Z^n)\right) - (1-a)\right|$ for each $a \in \mathbb{A}$. We can see that regardless of the value of V, the approximation error of our bootstrap procedure remains stable at low values, below 0.5%.

TABLE 1
$100 \times \left|\mathbf{P}_n\left(Q_n \geq t_n^B(a, Z^n)\right) - (1-a)\right|$ *for all* $a \in \mathbb{A}$ *and different values of* ϵ*; for* $n = 500$ *and* $d(n) = 3$

ϵ	0	4	8	12	16	20	24	28
$a = 0.900$	0.06	0.50	0.40	0.30	0.20	0.02	0.20	0.10
$a = 0.950$	0.06	0.30	0.22	0.02	0.04	0.02	0.28	0.14
$a = 0.975$	0.30	0.20	0.08	0.02	0.14	0.26	0.06	0.06
$a = 0.990$	0.24	0.30	0.04	0.04	0.06	0.18	0.02	0.02

Table 2 shows $\mathbb{K}_n^B/\mathbb{K}_n$ for $d(n) = n^{1/5}$. We see that for all n under consideration the ratio is around one, and in almost all below one. These results suggest that, at least for the current design, the convergence rate of the bootstrap-based approximation is no worse than the one for the chi-squared-based.

TABLE 2
$\mathbb{K}_n^B/\mathbb{K}_n$ *for* $d(n) = n^{1/5}$ *and different values of* n

	$n = 250$	$n = 500$	$n = 1000$	$n = 2000$	$n = 3000$
$\mathbb{K}_n^B/\mathbb{K}_n$ for $d(n) = n^{1/5}$	0.937	0.950	0.750	0.892	1.030

Robustness to $d(n)$ and choice of weights. We now assess how robust our procedure is to the choice of $d(n)$. Recall that, for this specification, our theory predicts that is sufficient to have $d(n) = o(n^{1/4})$; for values higher than this our theory is silent about the validity of our bootstrap procedure. We are thus particularly interested on the performance of our procedure for the latter set of values. In this exercise, we set $V = I$ and consider different values of n and $d(n)$.

Table 3 columns 2–5 shows the value of $100 \times \mathbb{K}_n^B$ for different choices of $d(n)$ and n. For values of n less than 1000, the procedure seems to be quite robust to larger choices of $d(n)$ in the range of $n^{1/4}$ to $n^{1/2}$, but not higher. For values of n around 2000-3000, however, our procedure seems to deteriorate for values of $d(n)$ equal or larger than $n^{1/2}$.

TABLE 3
The value of $100 \times \mathbb{K}_n^B$ *for different values of* $(n, d(n))$

$n \setminus d(n)$	$n^{1/5}$	$n^{1/4}$	$n^{1/3}$	$n^{1/2}$	$n^{3/4}$
250	0.300	0.440	0.332	0.440	2.440
500	0.350	0.401	0.280	0.450	2.780
1000	0.340	0.240	0.540	0.500	2.080
2000	0.400	0.250	0.340	0.690	0.943
3000	0.200	0.201	0.341	0.601	1.463

We now assess the robustness of our procedure to different choices of weights. We compare the Gaussian weights with two other weights: $\omega_i \sim U(-0.5, 0.5)$ and $\omega_i \sim t - Student(3)$ (properly scaled to have unit variance). These choices are designed to study how different tail behavior of the weight's distribution affect the performance of our bootstrap procedure.

In order to ease the computational burden we lower the bootstrap repetitions to 2000 each. Table 4 presents the results. The overall pattern seems to suggest that the Gaussian and Uniform weights have comparable performances, and perform better than the t-Student weights. This pattern illustrates the discussion in Section 7 regarding desirable properties of weights.

TABLE 4
The value of $100 \times \mathbb{K}_n^B$ *for Gaussian, Uniform and t-Student weights with* $d(n) = n^{1/5}$

$n\backslash$ Weights	Gaussian	Uniform	t-Student
250	0.560	0.440	0.960
500	0.500	0.370	0.980
1000	0.139	0.319	0.400
2000	0.240	0.340	0.660
3000	0.180	0.200	0.400

Remarks. Overall, the simulations suggest that our procedure has a finite sample performance that is at least as good as, and in some cases better than, the "standard" chi-squared approach. Weights with "thin tails" such as Uniform and Gaussian seem to perform better than weights with heavier tails. Additionally, as also discussed in the context of our application in Section 4, our bootstrap-based approximation can be applied in situations that go beyond those covered by the chi-square approach.

6. Proof of Theorem 3.1

Recall that $x \in \mathbb{R}^{d(n)} \mapsto ||x||_2^2 \equiv x^T x$ 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

$$\Delta_M(P, Q) \equiv \sup_{f \in \mathcal{C}_M} |E_P[f(||X||_2^2)] - E_Q[f(||Y||_2^2)]|. \tag{6}$$

Remark 6.1. *Throughout the text we use this definition for* $X = n^{-1/2} \sum_{i=1}^n X_i$ *and* $Y = n^{-1/2} \sum_{i=1}^n Y_i$*, with* $(X_1, ..., X_n) \sim P$ *and* $(Y_1, ..., Y_n) \sim Q$*. For these cases, we abuse notation and use* $\Delta_M(P, Q)$ *to denote*

$$\sup_{f \in \mathcal{C}_M} |E_P[f(||n^{-1/2} \sum_{i=1}^n X_i||_2^2)] - E_Q[f(||n^{-1/2} \sum_{i=1}^n Y_i||_2^2)]|.$$

Also, in the cases where $X_i \sim i.i.d. - P$*, we abuse notation and still use* $\Delta_M(P, Q)$ *to denote the same quantity.*

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

$$\mathbf{P}_n \left(\sup_{t\in\mathbb{R}} \left| \mathbf{P}_n^* \left(||\sqrt{n}\mathbb{Z}_n^*||_2^2 \geq t \mid Z^n \right) - \mathbf{P}_n \left(||\sqrt{n}\mathbb{Z}_n||_2^2 \geq t \right) \right| \geq \varepsilon' \right) < \varepsilon'$$

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

$$\begin{aligned} &\mathbf{P}_n \left(\sup_{t\in\mathbb{R}} \left| \mathbf{P}_n^* \left(||\sqrt{n}\mathbb{Z}_n^*||_2^2 \geq t \mid Z^n \right) - \mathbf{P}_n \left(||\sqrt{n}\mathbb{Z}_n||_2^2 \geq t \right) \right| \geq \varepsilon' \right) \\ \leq &\mathbf{P}_n \left(\{ \sup_{t\in\mathbb{R}} \left| \mathbf{P}_n^* \left(||\sqrt{n}\mathbb{Z}_n^*||_2^2 \geq t \mid Z^n \right) - \mathbf{P}_n \left(||\sqrt{n}\mathbb{Z}_n||_2^2 \geq t \right) \right| \geq \varepsilon' \} \cap S_n \right) \\ &+ \mathbf{P}_n \left(S_n^C \right) \end{aligned}$$

where $S_n \equiv \{Z^n : n^{-1} \sum_{i=1}^n ||Z_i||_2^2 \leq (0.5\varepsilon')^{-1} tr\{\Sigma_n\}\}$. By the Markov inequality $\mathbf{P}_n \left(S_n^C \right) \leq 0.5\varepsilon'$. Thus, it suffices to show that

$$\mathbf{P}_n \left(\left\{ \sup_{t\in\mathbb{R}} \left| \mathbf{P}_n^* \left(||\sqrt{n}\mathbb{Z}_n^*||_2^2 \geq t \mid Z^n \right) - \mathbf{P}_n \left(||\sqrt{n}\mathbb{Z}_n||_2^2 \geq t \right) \right| \geq \varepsilon' \right\} \cap S_n \right) < 0.5\varepsilon'. \tag{7}$$

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

$$\begin{aligned} &|E_{\mathbf{P}_n^*} \left[1\{||\sqrt{n}\mathbb{Z}_n^*||_2^2 \geq t\} \mid Z^n \right] - E_{\mathbf{P}_n} \left[1\{||\sqrt{n}\mathbb{Z}_n||_2^2 \geq t\} \right] | \\ \leq &|E_{\mathbf{P}_n^*} \left[1\{||\sqrt{n}\mathbb{Z}_n^*||_2^2 \geq t\} \mid Z^n \right] - E_{\mathbf{\Phi}_n} \left[1\{||\sqrt{n}\mathbb{V}_n||_2^2 \geq t\} \right] | \\ &+ |E_{\mathbf{P}_n} \left[1\{||\sqrt{n}\mathbb{Z}_n||_2^2 \geq t\} \right] - E_{\mathbf{\Phi}_n} \left[1\{||\sqrt{n}\mathbb{V}_n||_2^2 \geq t\} \right] | \end{aligned}$$

where $\sqrt{n}\mathbb{V}_n \sim N(0, \Sigma_n)$. We use $\mathbf{\Phi}_n$ to denote this probability.

Therefore, in order to obtain display 7, it suffices to bound

$$\begin{aligned} &\mathbf{P}_n \left(\{ \sup_{t\in\mathbb{R}} \left| E_{\mathbf{P}_n^*} \left[1\{||\sqrt{n}\mathbb{Z}_n^*||_2^2 \geq t\} \mid Z^n \right] - E_{\mathbf{\Phi}_n} \left[1\{||\sqrt{n}\mathbb{V}_n||_2^2 \geq t\} \right] \right| \geq 0.5\varepsilon' \} \cap S_n \right) \\ &< 0.25\varepsilon' \end{aligned} \tag{8}$$

and

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}} \left| E_{\mathbf{P}_n} \left[1\{||\sqrt{n}\mathbb{Z}_n||_2^2 \geq t\} \right] - E_{\mathbf{\Phi}_n} \left[1\{||\sqrt{n}\mathbb{V}_n||_2^2 \geq t\} \right] \right| = 0. \tag{9}$$

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

Lemma 6.1. *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 $h \leq h(\sqrt{tr\{\Sigma_n^2\}}\gamma(\varepsilon), \varepsilon)$* [17]

$$\sup_{t\in\mathbb{R}} \left| E_{\mathbf{P}_n} \left[1\{||\sqrt{n}\mathbb{Z}_n||_2^2 \geq t\} \right] - E_{\mathbf{\Phi}_n} \left[1\{||\sqrt{n}\mathbb{V}_n||_2^2 \geq t\} \right] \right| \tag{10}$$

$$\leq \frac{\varepsilon}{1-\varepsilon} + 3\varepsilon + \Delta_{h^{-1}}(\mathbf{P}_n, \mathbf{\Phi}_n). \tag{11}$$

[17] The value $h(\sqrt{tr\{\Sigma_n^2\}}\gamma(\varepsilon), \varepsilon)$ is given by $\sqrt{tr\{\Sigma_n^2\}}\gamma(\varepsilon)/(-\Phi^{-1}(\varepsilon))$; see Lemma B.1.

(Recall $\Delta_{h^{-1}}(\mathbf{P}_n,\mathbf{\Phi}_n)=\sup_{f\in\mathcal{C}_{h^{-1}}}\left|E_{\mathbf{P}_n}\left[f\left(||\sqrt{n}\mathbb{Z}_n||_2^2\right)\right]-E_{\mathbf{\Phi}_n}\left[f\left(||\sqrt{n}\mathbb{V}_n||_2^2\right)\right]\right|$). And

Lemma 6.2. *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 $h\leq h(\sqrt{tr\{\Sigma_n^2\}}\gamma(\varepsilon),\varepsilon)$*

$$\sup_{t\in\mathbb{R}}\left|E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_2^2\geq t\}|Z^n\right]-E_{Pr}\left[1\{||\sqrt{n}\mathbb{V}_n||_2^2\geq t\}\right]\right|$$
$$\leq\frac{\varepsilon}{1-\varepsilon}+3\varepsilon+\Delta_{h^{-1}}(\mathbf{P}_n^*(\cdot|Z^n),\mathbf{\Phi}_n), \tag{12}$$

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

Remark 6.2. *The previous lemma holds for any h provided that is below $h\leq h(\sqrt{tr\{\Sigma_n^2\}}\gamma(\varepsilon),\varepsilon)$. 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 $x\mapsto 1\{||x||_2^2\geq t\}$; 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 $h(\delta_n,\varepsilon)$.*

It is worth to note that, for the "smooth" function to be a good approximation of $1\{||\cdot||_2^2\geq t\}$, we need δ_n to be "small" (see the proof of Lemma 6.2 in the Appendix). What we mean by δ_n to be "small" depends on how $||\sqrt{n}\mathbb{V}_n||_2^2$ 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 is 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

$$\sup_{t\in\mathbb{R}}\left|E_{\mathbf{P}_n}\left[1\{||\sqrt{n}\mathbb{Z}_n||_2^2\geq t\}\right]-E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_2^2\geq t\}\right]\right|\leq\frac{\varepsilon'}{4}+\Delta_{h^{-1}}(\mathbf{P}_n,\mathbf{\Phi}_n) \tag{13}$$

and

$$\mathbf{P}_n\left(\{\sup_{t\in\mathbb{R}}\left|E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_2^2\geq t\}\mid Z^n\right]-E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_2^2\geq t\}\right]\right|\geq\frac{\varepsilon'}{2}\}\cap S_n\right)$$
$$\leq\mathbf{P}_n\left(\{\Delta_{h^{-1}}(\mathbf{P}_n^*(\cdot|Z^n),\mathbf{\Phi}_n)\geq 0.25\varepsilon'\}\cap S_n\right) \tag{14}$$

for all $n\geq N(\varepsilon)$ and all $h\leq h(\delta_n,\varepsilon)$ (note that ε is a function of ε').

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) $\sqrt{n}\mathbb{U}_n\sim N(0,n^{-1}\sum_{i=1}^n Z_iZ_i^T)$

Hence, by the previous display and Equations 7, 8-9, 13 and 14, in order to show the desired result it suffices to show that: For all ε', there exists a $N(\varepsilon')$ such that

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

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

$$\text{and } \Delta_{h^{-1}}(\mathbf{P}_n, \mathbf{\Phi}_n) < \varepsilon' \tag{17}$$

for all $n \geq N(\varepsilon')$ and some $h \leq h(\sqrt{tr\{\Sigma^2_n\}}\gamma(\varepsilon), \varepsilon)$. Theorems 3.2 and 3.3 establish expressions 15 and 17.

Remark 6.3. *From Lemma B.2,* $h(\sqrt{tr\{\Sigma^2_n\}}\gamma(\varepsilon), \varepsilon) = \sqrt{tr\{\Sigma^2_n\}}\gamma(\varepsilon)/(-\Phi^{-1}(\varepsilon))$ *and thus* h *can be taken to be proportional (up to a constant that depends on* ε*) to* $\sqrt{tr\{\Sigma^2_n\}}$*. 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

$$\mathbf{P}_n\left(\{\Delta_{h^{-1}}(\mathbf{\Phi}^*_n(\cdot|Z^n), \mathbf{\Phi}_n) \geq \varepsilon'\} \cap S_n\right) < \varepsilon'. \tag{18}$$

Since $\sqrt{n}\mathbb{U}_n \sim N(0, \hat{\Sigma}_n)$ (with $\hat{\Sigma}_n = n^{-1}\sum_{i=1}^n Z_i Z_i^T$) 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 6.3. *For any* $h > 0$ *and any* $n \in \mathbb{N}$

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

Observe that for any $Z^n \in S_n = \{Z^n : n^{-1}\sum_{i=1}^n ||Z_i||^2_2 \leq (0.5\varepsilon')^{-1}tr\{\Sigma_n\}\}$, so the RHS of the expression in the Lemma is bounded above by

$$\frac{d(n)}{h}\{\frac{tr\{\Sigma_n\}}{h\varepsilon'} + 2\}.$$

Thus by Lemma 6.3, in order to establish the desired result, it suffices to show that

$$\mathbf{P}_n\left(\max_{j,l}\left|n^{-1}\sum_{i=1}^n Z_{[l],i}Z_{[j],i} - \Sigma_{[j,l]}\right| \geq \frac{(\varepsilon')^2}{d(n)h^{-2}tr\{\Sigma_n\}} \cap S_n\right) < \varepsilon' \tag{19}$$

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

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

Let $\mathbf{A}_{i,n}[j,l] = \mathbf{A}^L_{i,n}[j,l] + \mathbf{A}^U_{i,n}[j,l]$ with $\mathbf{A}^L_{i,n}[j,l] \equiv \mathbf{A}_{i,n}[j,l]1\{|\mathbf{A}_{i,n}[j,l]| \leq e_n\}$ and $\mathbf{A}^U_{i,n}[j,l] \equiv \mathbf{A}_{i,n}[j,l]1\{|\mathbf{A}_{i,n}[j,l]| \geq e_n\}$ where $(e_n)_n$ with $e_n > 0$ is defined below. Clearly, $\mathbf{A}^L_{i,n}[j,l] \leq e_n$. So, by Hoeffding inequality (see Boucheron et al. (2013) p. 34)

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

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

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

for sufficiently large n.

Second, by the Markov inequality and the fact that

$$E_{\mathbf{P}_n}\left[\left(\{\mathbf{A}^U_{i,n}[j,l] - E_{\mathbf{P}_n}[\mathbf{A}^U_{i,n}[j,l]]\}\right)\left(\{\mathbf{A}^U_{k,n}[j,l] - E_{\mathbf{P}_n}[\mathbf{A}^U_{k,n}[j,l]]\}\right)\right] = 0 \quad (20)$$

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

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

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

$$\mathbf{P}_n\left(\max_{j,l} |n^{-1}\sum_{i=1}^n \{\mathbf{A}^U_{i,n}[j,l] - E_{\mathbf{P}_n}[\mathbf{A}^U_{i,n}[j,l]]\}| \geq c_n\right)$$

$$\leq \frac{1}{c_n^2 n (e_n)^p} \sum_{j=1}^{d(n)} \sum_{l=1}^{d(n)} E_{\mathbf{P}_n} \left[\left(Z_{[j],1} Z_{[l],1} \right)^{2+p} \right]$$

$$= \frac{1}{c_n^2 n (e_n)^p} E_{\mathbf{P}_n} \left[\left(\sum_{j=1}^{d(n)} (Z_{[j],1})^{2+p} \right)^2 \right].$$

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

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

$$\precsim \frac{\log(d(n))^{p/2}}{c_n^{2+p} n^{1+p/2}} E_{\mathbf{P}_n} \left[\left(\sum_{j=1}^{d(n)} (Z_{[j],1})^{2+p} \right)^2 \right]$$

$$\precsim \frac{(\log(d(n)))^{p/2} d(n)^{2+p} (tr\{\Sigma_n\})^{2+p}}{h^{4+2p} n^{1+p/2}} E_{\mathbf{P}_n} \left[\left(\sum_{j=1}^{d(n)} (Z_{[j],1})^{2+p} \right)^2 \right].$$

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

Therefore, Equation 19 is established and with that the proof of Theorem 3.1.

7. Discussion

Applicability of our Results. The example developed in Section 4 illustrates a general feature present in several test statistics, namely that they behave asymptotically as quadratic forms of (properly scaled) sample averages. These are the main motivational examples to which we can apply our result in Theorem 3.1.

This remark is best illustrated in the Wald statistic case. To formalize this, consider i.i.d. data $(X_1, ..., X_n)$ drawn from $\mathbf{P}$ and parameter a $k \geq d = d(n)$ dimensional $\theta_{\mathbf{P}}$ and a "smooth" function $\theta \mapsto c(\theta) \in \mathbb{R}^d$ which represent the hypothesis we want to test; i.e., the null hypothesis is $c(\theta_{\mathbf{P}}) = 0$. [18] The fact that d grows with the sample size is of potential interest because in certain situations one could have that the dimension of the parameter, although fixed, is not "small" relative to n. Also, in some other situations, one could have a more explicit model of increasing dimensionality like in the cases discussed in Section 4 or in series or sieves estimators; see, for example, Chen and Pouzo (2015).

[18] The notation $\theta_{\mathbf{P}}$ stresses that the parameter is a (known) function of the probability distribution. Thus, an estimator can obtained by "plugging in" the empirical distribution P_n.

Suppose there exists an estimator θ_{P_n} (P_n is the empirical distribution), then the Wald statistic is given by

$$\mathbb{W}_n(P_n,\mathbf{P}) = n\left(c(\theta_{P_n}) - c(\theta_{\mathbf{P}})\right)^T V_n \left(c(\theta_{P_n}) - c(\theta_{\mathbf{P}})\right)$$

where $V_n \in \mathbb{R}^{d\times d}$ is some (possibly random) matrix to be determined later.

Suppose θ_{P_n} admits an asymptotic linear representation (ALR) of the form [19]

$$\left\|\sqrt{n}(c(\theta_{P_n}) - c(\theta_{\mathbf{P}})) - \sqrt{n}E_{P_n}[\psi(X,\theta_{\mathbf{P}})]\right\|_2 = o_{\mathbf{P}}(1+\sqrt{n}||c(\theta_{P_n}) - c(\theta_{\mathbf{P}})||_2) \tag{21}$$

with $E_{\mathbf{P}}[\psi(X,\theta_{\mathbf{P}})] = 0$ and finite second moment. [20]

The bootstrap analog of the Wald statistic are of the "plug-in" type, i.e., [21]

$$\mathbb{W}_n(P_n^*,P_n) = n\left(c(\theta_{P_n^*}) - c(\theta_{P_n})\right)^T V_n \left(c(\theta_{P_n^*}) - c(\theta_{P_n})\right)$$

where P_n^* is given by $n^{-1}\sum_{i=1}^n \omega_{i,n}\delta_{X_i}$ (the dependence of P_n^* on X^n is omitted to ease the notational burden). The bootstrap ALR (B-ALR) is given by

$$\begin{aligned}&\left\|\sqrt{n}(c(\theta_{P_n^*}) - c(\theta_{P_n})) - n^{-1/2}\sum_{i=1}^n \omega_{i,n}\psi(X_i,\theta_{\mathbf{P}})\right\|_2 \\ &= o_{\mathbf{P}_n^*}(1+\sqrt{n}||c(\theta_{P_n^*}) - c(\theta_{P_n})||_2),\ wpa1-\mathbf{P}.\end{aligned} \tag{22}$$

Given the asymptotic linear representations, we can show that the Wald and Bootstrapped Wald statistics can be represented asymptotically as quadratic forms, and thus fall in the framework studied in this paper. The following proposition formalizes such representation, and thereby allow us to apply our Theorem 3.1 with $Z_i \equiv \psi(X_i,\theta_{\mathbf{P}})V^{1/2}$ to approximate the limiting distribution of $\mathbb{W}_n(P_n,\mathbf{P})$.

Proposition 7.1. *Let V be a matrix such that there exists a $c \geq 1$ such that $c^{-1} \leq \lambda_l(V) \leq c$ for all $l = 1,...,d$ and $||V_n - V||_2 = o_{\mathbf{P}}(1)$. Then, under the null hypothesis, ALR and B-ALR yield*

$$\begin{aligned}\mathbb{W}_n(P_n,\mathbf{P})(1+o_{\mathbf{P}}(1)) = &\left(n^{-1/2}\sum_{i=1}^n \psi(X_i,\theta_{\mathbf{P}})\right)^T V \left(n^{-1/2}\sum_{i=1}^n \psi(X_i,\theta_{\mathbf{P}})\right) \\ &+ o_{\mathbf{P}}(\sqrt{d(n)}),\end{aligned}$$

and

$$\begin{aligned}\mathbb{W}_n(P_n^*,P_n)(1+o_{\mathbf{P}_n^*}(1)) = &\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_{i,n}\psi(X_i,\theta_{\mathbf{P}})\right)^T V \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \omega_{i,n}\psi(X_i,\theta_{\mathbf{P}})\right) \\ &+ o_{\mathbf{P}_n^*}(\sqrt{d(n)})\ wpa1-\mathbf{P}.\end{aligned}$$

[19] See Van der Vaart (2000) and references therein for a discussion regarding ALR and sufficient conditions for it. Here we follow Murphy and der Vaart (2000).

[20] Note that $E_{P_n}[\psi(X,\theta_{\mathbf{P}})] = n^{-1}\sum_{i=1}^n \psi(X_i,\theta_{\mathbf{P}})$.

[21] In principle, one could also "replace" V_n — which typically is a function of P_n, V_{P_n} — by $V_{P_n^*}$. Our results could be extended to this case too.

Proof. See Appendix D. □

A few remarks are in order. First, and more importantly, we note that, our results can be applied to other test statistic provided that are *asymptotically* equivalent (up to $o(\sqrt{d(n)})$) to $\mathbb{W}(P_n, \mathbf{P})$ or to a quadratic form as in the proposition. Typically this is the case for the Likelihood ratio and Lagrange Multiplier (or Score) test statistics; see Newey and McFadden (1994) Section 9.

Second, for the Chi-square-based approximation to be valid, V must coincide with $(E_{\mathbf{P}}[\psi(X,\theta_0)\psi(X,\theta_0)^T])^{-1}$. The bootstrap-based approximation, however, does not require this assumption. This situation may arise, for instance, in Likelihood ratio tests under model misspecification.

Choice of Weights. We now provide some heuristic discussion regarding the weights.

The bootstrap procedure studied in this paper uses independent weights. Such restriction has also been used in several papers; e.g. Chernozhukov et al. (2013a) and Ma and Kosorok (2005). This choice is largely due to the fact that the independent behavior of weights makes many of the proofs easier. It would be of interest still to extend our results to non-iid weights such as Multinomial weights — which yield non-parametric and m-out-n bootstrap procedures. While such an extension is beyond the scope of the paper, we point out that one possible way to extend our results to this case is to "remove" the dependence by employing Poissonization results; see e.g. Chapter 3.5-3.6 of Van der Vaart and Wellner (1996). The key step would be to ensure the validity of Theorem 3.4 and Lemma A.1 (in the Appendix).[22]

Even within the class of independent weights, one could wonder what properties are desirable for the weights to have. Clearly, as indicated by our Assumption 2.2, restrictions on the tail behavior of the weights are important for our results. We now present a discussion, which expands on the quantitative explorations in Section 5, about what other properties might be desirable to have.

Heuristically, the Lindeberg interpolation result — Theorem 3.4 — relies on "matching" the first and second moments. By choosing the weights to match higher moment one could expect to improve the approximation rates.[23] More precisely, the bounds for $\mathbf{S}_n$ (and $\mathbf{R}_n$) obtained in Lemma A.1 in the Appendix only use restrictions in the original data and the bootstrap weights present in 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 $A_i = n^{-1/2}\omega_{i,n}Z_i$ and $B_i = n^{-1/2}u_iZ_i$ 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 $\mathbf{S}_{1,n} = 0$.

[22] Another possibility is to directly extend the Theorem and Lemma to allow for non-independent data; at least for one sequence, either the A's or B's in the Theorem. Independence is, due to the technique of proof, particularly important for establishing Lemma A.1.

[23] These observations are related to the four moment Theorem of Tao and Vu in the context of random matrices; see Tao and Vu (2011).

A similar observation applies to $\mathbf{S}_{2,n}$ but in this case the relevant moments are $E[\omega_{i,n}^3]$ and $E[Z^3]$.[24]

Finally, another extension that is linked to the previous discussion, is that of refinements (or lack thereof) of certain choices of bootstrap weights. We leave this for future research.

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

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

$$\mathbf{S}_{1,n} \leq L_2(f) \sum_{i=1}^{n} E[||B_i||_2^4 + ||A_i||_2^4]$$
$$\mathbf{S}_{2,n} \leq L_2(f) \sqrt{\sum_{j=1}^{n} tr\{C_j\}} \sum_{i=1}^{n} \left(E\left[||B_i||_2^3\right] + E\left[||A_i||_2^3\right]\right).$$

And, for any $q > 0$

$$\mathbf{R}_n \precsim \sum_{i=1}^{n} \left(E\left[\left(\mathbb{S}_{i:n}^T B_i\right)^{2+q} + \left(\mathbb{S}_{i:n}^T A_i\right)^{2+q} \right] + E\left[||B_i||_2^{4+2q}\right] + E\left[||A_i||_2^{4+2q}\right]\right).$$

And

$$\sum_{i=1}^{n} E\left[\left(\mathbb{S}_{i:n}^T B_i\right)^{2+q}\right]$$
$$\precsim \sum_{i=1}^{n} E[||B_i||_2^{2+q}] \max\left\{ \left(\sum_{j=1}^{n} E\left[||S_j||_2^2\right]\right)^{1+0.5q}, \sum_{j=1}^{n} E\left[||S_j||_2^{2+q}\right]\right\}.$$

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 $L_2(f)\sum_{i=1}^{n} E[||B_i||_2^4 + ||A_i||_2^4]$. Regarding $\mathbf{S}_{2,n}$, observe that

$$\sum_{i=1}^{n} |E\left[f''\left(||\mathbb{S}_{i:n}||_2^2\right)\mathbb{S}_{i:n}^T\right]\left(E[B_i||B_i||_2^2]\right)| \leq L_2(f)\sum_{i=1}^{n} E\left[||\mathbb{S}_{i:n}||_2\right] E\left[||B_i||_2^3\right]$$
$$\leq L_2(f) \sum_{i=1}^{n} \sqrt{E\left[||\mathbb{S}_{i:n}||_2^2\right]} E\left[||B_i||_2^3\right]$$

[24]The rate of convergence of the term $\mathbf{R}_n$ is regulated by q, which is linked to the bound on higher moments of the data and weights (see Assumptions 2.2 and 2.1).

by independence of $\mathbb{S}_{i:n}$ and B_i and Cauchy-Schwarz. Also, $E[\mathbb{S}_{i:n}\mathbb{S}_{i:n}^T] = \sum_{j=1}^n E[S_jS_j^T]$, so $E\left[||\mathbb{S}_{i:n}||_2^2\right] = tr\{E[\mathbb{S}_{i:n}\mathbb{S}_{i:n}^T]\} = \sum_{j=1}^n tr\{C_j\}$. A similar results holds when B_i is replaced by A_i. Therefore

$$\mathbf{S}_{2,n} \leq L_2(f)\sqrt{\sum_{j=1}^n tr\{C_j\}}\sum_{i=1}^n \left(E\left[||B_i||_2^3\right] + E\left[||A_i||_2^3\right]\right).$$

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

$$\sum_{i=1}^n E\left[\left(\mathbb{S}_{i:n}^T B_i + ||B_i||_2^2\right)^{2+q}\right] \precsim \left(\sum_{i=1}^n E\left[\left(\mathbb{S}_{i:n}^T B_i\right)^{2+q}\right] + \sum_{i=1}^n E\left[\left(||B_i||_2\right)^{4+2q}\right]\right).$$

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

$$\sum_{j=1}^n E\left[\left(S_j^T b_i\right)^{2+q}\right] \leq \sum_{j=1}^n E\left[\left(||S_j||_2||b_i||_2\right)^{2+q}\right] = \left(||b_i||_2\right)^{2+q}\sum_{j=1}^n E\left[\left(||S_j||_2\right)^{2+q}\right].$$

Therefore, using these bounds and taking expectation with respect to B_i and after straightforward algebra,

$$\begin{aligned} &\sum_{i=1}^n E\left[\left(\mathbb{S}_{i:n}^T B_i\right)^{2+q}\right] \\ &\precsim \sum_{i=1}^n E[||B_i||_2^{2+q}] \max\left\{\left(\sum_{j=1}^n E\left[||S_j||_2^2\right]\right)^{1+0.5q}, \sum_{j=1}^n E\left[||S_j||_2^{2+q}\right]\right\}. \end{aligned}$$

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 . 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

$$\mathbb{S}_{i:n} + A_i = \left(\sum_{j=1}^{i} A_j + \sum_{j=i+1}^{n} B_j \right) = \mathbb{S}_{i+1:n} + B_{i+1}. \tag{23}$$

Therefore,

$$\begin{aligned} &\sum_{i=1}^n E\left[f\left(||\mathbb{S}_{i:n} + B_i||_2^2 \right) - f\left(||\mathbb{S}_{i:n} + A_i||_2^2 \right) \right] \\ =&E\left[f\left(||\sum_{i=1}^n B_i||_2^2 \right) - f\left(||\sum_{i=1}^n A_i||_2^2 \right) \right]. \end{aligned}$$

Observe that $||\mathbb{S}_{i:n} + B_i||_2^2 = ||\mathbb{S}_{i:n}||_2^2 + ||B_i||_2^2 + 2\mathbb{S}_{i:n}^T B_i$. Therefore, by this fact and three times differentiability of f, it follows that

$$\begin{aligned} f\left(||\mathbb{S}_{i:n} + B_i||_2^2 \right) - f\left(||\mathbb{S}_{i:n}||_2^2 \right) =& f'\left(||\mathbb{S}_{i:n}||_2^2 \right) \left(||B_i||_2^2 + 2\mathbb{S}_{i:n}^T B_i \right) \\ &+ 0.5 f''\left(||\mathbb{S}_{i:n}||_2^2 \right) \left(||B_i||_2^2 + 2\mathbb{S}_{i:n}^T B_i \right)^2 \\ &+ R_{i,1,n} \end{aligned}$$

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

$$\begin{aligned} f\left(||\mathbb{S}_{i:n} + A_i||_2^2 \right) - f\left(||\mathbb{S}_{i:n}||_2^2 \right) =& f'\left(||\mathbb{S}_{i:n}||_2^2 \right) \left(||A_i||_2^2 + 2\mathbb{S}_{i:n}^T A_i \right) \\ &+ 0.5 f''\left(||\mathbb{S}_{i:n}||_2^2 \right) \left(||A_i||_2^2 + 2\mathbb{S}_{i:n}^T A_i \right)^2 \\ &+ R_{i,2,n}. \end{aligned}$$

Hence

$$\begin{aligned} &E\left[f\left(||\mathbb{S}_{i:n} + B_i||_2^2 \right) - f\left(||\mathbb{S}_{i:n} + A_i||_2^2 \right) \right] \\ =&E\left[f'\left(||\mathbb{S}_{i:n}||_2^2 \right) \left(||A_i||_2^2 - ||B_i||_2^2 + 2\mathbb{S}_{i:n}^T (A_i - B_i) \right) \right] \\ &+ 0.5 E\left[f''\left(||\mathbb{S}_{i:n}||_2^2 \right) \left\{ \left(||B_i||_2^2 + 2\mathbb{S}_{i:n}^T B_i \right)^2 - \left(||A_i||_2^2 + 2\mathbb{S}_{i:n}^T A_i \right)^2 \right\} \right] \\ &+ E\left[R_{i,1,n} - R_{i,2,n} \right] \\ \equiv& F_{i,n} + S_{i,n} + E\left[R_{i,1,n} - R_{i,2,n} \right]. \end{aligned}$$

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

$$\sum_{i=1}^n E\left[f'\left(||\mathbb{S}_{i:n}||_2^2 \right) \mathbb{S}_{i:n}^T (B_i - A_i) \right] = \sum_{i=1}^n E\left[f'\left(||\mathbb{S}_{i:n}||_2^2 \right) \mathbb{S}_{i:n}^T \right] E\left[(B_i - A_i) \right] = 0$$

and

$$\sum_{i=1}^{n} E\left[f'\left(||\mathbb{S}_{i:n}||_2^2\right)\left(||B_i||_2^2 - ||A_i||_2^2\right)\right] = \sum_{i=1}^{n} E\left[f'\left(||\mathbb{S}_{i:n}||_2^2\right)\right] E\left[\left(||B_i||_2^2 - ||A_i||_2^2\right)\right] \\ = 0.$$

THE TERM SECOND ORDER TERMS, S_n. For this term it suffices to study the following terms:

$$\begin{aligned} S_{1,n} &\equiv \sum_{i=1}^{n} E\left[f''\left(||\mathbb{S}_{i:n}||_2^2\right)\left(||B_i||_2^4 - ||A_i||_2^4\right)\right] \\ S_{2,n} &\equiv \sum_{i=1}^{n} E\left[f''\left(||\mathbb{S}_{i:n}||_2^2\right) 4\left((\mathbb{S}_{i:n}^T B_i)^2 - (\mathbb{S}_{i:n}^T A_i)^2\right)\right] \\ S_{3,n} &\equiv \sum_{i=1}^{n} E\left[f''\left(||\mathbb{S}_{i:n}||_2^2\right) 4\mathbb{S}_{i:n}^T\left(B_i||B_i||_2^2 - A_i||A_i||_2^2\right)\right]. \end{aligned}$$

By independence of $\mathbb{S}_{i:n}$ with A_i and B_i, it follows that

$$S_{1,n} = \sum_{i=1}^{n} E\left[f''\left(||\mathbb{S}_{i:n}||_2^2\right)\right] E\left[||B_i||_2^4 - ||A_i||_2^4\right].$$

Regarding $S_{2,n}$, because $\mathbb{S}_{i:n}$ is independent to A_i and B_i and $E[A_iA_i^T] = E[B_iB_i^T]$, 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

$$|S_{3,n}| \leq 4\sum_{i=1}^{n} |E\left[f''\left(||\mathbb{S}_{i:n}||_2^2\right)\mathbb{S}_{i:n}^T\right]\left(E[B_i||B_i||_2^2] - E[A_i||A_i||_2^2]\right)|.$$

THE REMAINDER TERMS, $R_{1,n}$ AND $R_{2,n}$. By Taylor's Theorem it follows that: For any $q > 0$

$$\sum_{i=1}^{n} E\left[|R_{i,1,n}|\right] \precsim L_2(f)^{1-q} L_3(f)^q \sum_{i=1}^{n} E\left[\left(\mathbb{S}_{i:n}^T B_i + ||B_i||_2^2\right)^{2+q}\right]. \qquad \square$$

A.1. Proof of Theorem 3.2

Proof of Theorem 3.2. We first note that is enough to show that for all $\varepsilon > 0$, there exists a $N(\varepsilon)$ such that

$$\mathbf{P}_n\left(\left\{\sup_{f\in\mathcal{C}_{h^{-1}}} h^2\left|E_{\mathbf{P}_n^*}\left[f\left(||\sqrt{n}\mathbb{Z}_n^*||_2^2\right)|Z^n\right] - E_{\mathbf{\Phi}_n^*}\left[f\left(||\sqrt{n}\mathbb{U}_n||_2^2\right)|Z^n\right]\right| \geq \varepsilon\right\} \cap K_n\right) \\ < \varepsilon$$

for all $n \geq N(\varepsilon)$, where $K_n \equiv \{Z^n : n^{-1}\sum_{i=1}^{n} ||Z_i||_2^2 \leq (0.5\varepsilon)^{-1} tr\{\Sigma_n\} \equiv M_n\}$.

The strategy of proof consists of applying the results in Theorem 3.4 and Lemma A.1, with $A_i = n^{-1/2}\omega_{in} Z_i$ and $B_i = n^{-1/2} u_i Z_i$ where $u_i \sim N(0,1)$.[25] Then uses the Markov inequality and shows 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 $\mathbf{R}_n$, are such that the first one vanishes at rate h^{-2} as $n \to \infty$, and the second one vanishes as $n \to \infty$.[26]

The leading terms, $\mathbf{S}_n$. For this case $\sum_{i=1}^n E[(||B_i||_2)^4] \precsim n^{-2}\sum_{i=1}^n ||Z_i||_2^4$ and $\sum_{i=1}^n E[(||A_i||_2)^4] \precsim n^{-2}\sum_{i=1}^n ||Z_i||_2^4$, under Assumption 2.2. Therefore, $\mathbf{S}_{1,n}$ in Theorem 3.4 is bounded above (up to a constant) by $n^{-1}\left(n^{-1}\sum_{i=1}^n ||Z_i||_2^4\right)$.

Therefore, since $L_2(f) = h^{-2}$, $E_{\mathbf{P}_n}[\mathbf{S}_{1,n}] \precsim h^{-2}n^{-2}\sum_{i=1}^n E_{\mathbf{P}_n}[||Z_i||_2^4] = h^{-2}n^{-1}E_{\mathbf{P}_n}[||Z_1||_2^4]$ which is of order $o(h^{-2})$ by Assumption 2.1(i).

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

$$\begin{aligned}\mathbf{S}_{2,n} \precsim & h^{-2}\sqrt{n^{-1}\sum_{i=1}^n ||Z_i||_2^2}\, n^{-3/2}\sum_{i=1}^n E[|\omega_{in}|^3 + |u_{i,n}|^3]||Z_i||_2^3 \\ \precsim & h^{-2}\sqrt{n^{-1}\sum_{i=1}^n ||Z_i||_2^2}\, n^{-3/2}\sum_{i=1}^n ||Z_i||_2^3.\end{aligned}$$

For any $Z^n \in K_n$, $\mathbf{S}_{2,n} \precsim h^{-2}\sqrt{M_n}n^{-3/2}\sum_{i=1}^n ||Z_i||_2^3$. Therefore,

$$\begin{aligned}E_{\mathbf{P}_n}[\mathbf{S}_{2,n}1\{K_n\}] \precsim & \frac{\sqrt{M_n}}{h^2 n^{3/2}}\sum_{i=1}^n E_{\mathbf{P}_n}[||Z_i||_2^3] \\ & = h^{-2}\sqrt{M_n}n^{-1/2}E_{\mathbf{P}_n}[||Z_1||_2^3],\end{aligned}$$

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. Observe that $(tr\{\sum_{j=1}^n E[(S_j^T S_j)]\})^{1+0.5q} = (tr\{n^{-1}\sum_{j=1}^n Z_j Z_j^T\})^{1+0.5q} = (n^{-1}\sum_{j=1}^n ||Z_j||_2^2)^{1+0.5q}$. Also,

$$\sum_{i=1}^n E\left[(||B_i||_2)^{2+q}\right] = \frac{1}{n^{1+0.5q}}\sum_{i=1}^n E\left[|u_{i,n}|^{2+q}\right]||Z_i||_2^{2+q} \precsim \frac{1}{n^{1+0.5q}}\sum_{i=1}^n ||Z_i||_2^{2+q}$$

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

$$\sum_{j=1}^n E\left[(||S_j||_2)^{2+q}\right] \precsim n^{-(1+0.5q)}\sum_{i=1}^n ||Z_i||_2^{2+q}.$$

[25] Note that $\mathbb{U}_n$ can be cast as $n^{-1}\sum_{i=1}^n u_i Z_i$.

[26] For the last one is enough to show that it simply vanishes because in Theorem 3.4, $\mathbf{R}_n$ is already scaled by $L_2(f)(L_3(f)/L_2(f))^q = h^{-2}h^{-q}$. For any $h > 0$, the RHS is trivially $O(h^{-2})$; this is true — in fact is $o(h^{-2})$ — even in the case h diverges to infinity. It will not be the case, however, if h is taken to converge to 0, but we never consider such case.

Therefore,

$$\sum_{i=1}^{n} E\left[\left(\mathbb{S}_{i:n}^{T} B_i\right)^{2+q}\right]$$
$$\precsim \frac{1}{n^{1+0.5q}} \sum_{i=1}^{n} ||Z_i||_2^{2+q} \max\left\{ \left(n^{-1} \sum_{j=1}^{n} ||Z_j||_2^2 \right)^{1+0.5q}, n^{-(1+0.5q)} \sum_{i=1}^{n} ||Z_i||_2^{2+q} \right\}$$
$$\precsim \max\left\{ n^{-(1+0.5q)} \sum_{i=1}^{n} ||Z_i||_2^{2+q} \left(n^{-1} \sum_{j=1}^{n} ||Z_j||_2^2 \right)^{1+0.5q}, n^{-(1+q)} \sum_{i=1}^{n} ||Z_i||_2^{4+2q} \right\}$$

where the last line follows from Jensen inequality. And, $\sum_{i=1}^{n} E[(||B_i||_2)^{4+2q}] \precsim n^{-(2+q)} \sum_{i=1}^{n} ||Z_i||_2^{4+2q}$.

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 $\sum_{i=1}^{n} E[(||A_i||_2)^{4+2q}]$.

Recall that $q = \gamma$. Thus, $E_{\mathbf{P}_n}[\frac{1}{n^{2+q}} \sum_{i=1}^{n} ||Z_i||_2^{4+2q}] = \frac{1}{n^{1+q}} E_{\mathbf{P}_n}[||Z_1||_2^{4+2q}]$ 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 $(M_n)^{1+0.5q} n^{-(0.5q)} E_{\mathbf{P}_n}[||Z_1||_2^{2+q}] + n^{-q} E_{\mathbf{P}_n}[||Z_1||_2^{4+q}]$; both terms vanish as $n \to \infty$ under Assumption 2.1(ii) with $q = \gamma$.

The desired result follows by the Markov inequality, Theorem 3.4 and the fact that $L^l(f) = h^{-l}$ with $h > 0$ and $l = 2, 3$; given that we have proven that $E_{\mathbf{P}_n}[\mathbf{S}_n 1\{K_n\}] = o(h^{-2})$ and $E_{\mathbf{P}_n}[\mathbf{R}_n 1\{K_n\}] = o(1)$.[27] □

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 $q > 0$*

$$E[||X||_2^{2q}] \leq C(q)(tr\{A\})^q$$

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

[27] As pointed out above, $L_2(f)(L_3(f)/L_2(f))^q = h^{-2}h^{-q}$ and thus for any $h > 0$, the RHS is trivially $O(h^{-2})$; this is true — in fact is $o(h^{-2})$ — even in the case h diverges to infinity. It will not be the case, however, if h is taken to converge to 0, but we never consider such case.

Proof of Lemma A.2. Let $U \sim N(0, I_d)$ and let Λ be the diagonal matrix of eigenvalues of A and V the eigenvector matrix. For any $q > 0$

$$\begin{aligned} E[||X||_2^{2q}] &= E[(X^T X)^q] = E[(U^T AU)^q] \\ &= E[(\xi^T \Lambda \xi)^q], \ where\ \xi = V^T U \\ &= tr\{A\}^q E\left[\left(\sum_{j=1}^{d} c_j(A)|\xi_j|^2\right)^q\right] \end{aligned}$$

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

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

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 $\sum_{j=1}^{d} c_j(A) = 1$. □

Proof of Theorem 3.3. Firs note that we can always write $\mathbb{V}_n \equiv n^{-1}\sum_{i=1}^{n} V_{i,n}$ with $V_{i,n} \sim i.i.d. - N(0, \Sigma_n)$.

The strategy of proof follows the one for Theorem 3.2, which consists of applying the results in Theorem 3.4 and Lemma A.1, with $A_i = n^{-1/2}Z_i$ and $B_i = n^{-1/2}V_{i,n}$, and then bounding the terms $\mathbf{S}_n$ and $\mathbf{R}_n$ as in the proof of Theorem 3.2; we refer the reader to that proof for details and discussion.

Observe that $E[A_i A_i^T] = E[B_i B_i^T] = \Sigma_n$.

The term $\mathbf{S}_n$. For this case $\sum_{i=1}^{n} E[(||B_i||_2)^4] = n^{-2}\sum_{i=1}^{n} E[||V_{i,n}||_2^4] = n^{-1}E[||V_{1,n}||_2^4]$ and $\sum_{i=1}^{n} E[(||A_i||_2)^4] = n^{-2}\sum_{i=1}^{n} E[||Z_i||_2^4] = n^{-1}E[||Z_1||_2^4]$. Therefore, $\mathbf{S}_{1,n}$ in Theorem 3.4 is bounded above (up to a constant) by $h^{-2}n^{-1}\left(E[||Z_1||_2^4] + E[||V_{1,n}||_2^4]\right)$, and by Lemma A.2, this implies that

$$\mathbf{S}_{1,n} \precsim h^{-2}n^{-1}\left(E[||Z_1||_2^4] + (tr\{\Sigma_n\})^2\right)$$

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

$$\mathbf{S}_{3,n} \precsim h^{-2}\sqrt{tr\{\Sigma_n\}} n^{-3/2}\sum_{i=1}^{n}(E[||Z_i||_2^3] + E[||V_{i,n}||_2^3])$$

$$=h^{-2}\sqrt{tr\{\Sigma_n\}}n^{-1/2}(E[||Z_1||_2^3]+E[||V_{1,n}||_2^3]).$$

By Lemma A.2, $E[||V_{1,n}||_2^3]=(tr\{\Sigma_n\})^{3/2}$. 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 (at rate h^{-2}). 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 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,

$$\sum_{i=1}^n E\left[(||B_i||_2)^{2+q}\right]=n^{-0.5q}E[||V_1||_2^{2+q}]\precsim n^{-0.5q}(tr\{\Sigma_n\})^{1+0.5q}$$

by Lemma A.2. Therefore,

$$\sum_{i=1}^n E\left[\left(\mathbb{S}_{i:n}^T B_i\right)^{2+q}\right]\precsim\frac{(tr\{\Sigma_n\})^{1+0.5q}}{n^{0.5q}}\max\left\{(tr\{\Sigma_n\})^{1+0.5q},\sum_{j=1}^n E[||S_j||_2^{2+q}]\right\}.$$

Observe that

$$\sum_{j=1}^n E\left[(||S_j||_2)^{2+q}\right]\precsim n^{-(1+0.5q)}\left(\sum_{j=1}^{i-1}E\left[(||Z_j||_2)^{2+q}\right]+(n-i)tr\{\Sigma_n\}^{1+0.5q}\right)$$

by Lemma A.2. Under Assumption 2.1(ii),

$$\begin{aligned}\sum_{j=1}^n E\left[(||S_j||_2)^{2+q}\right]\precsim&n^{-(1+0.5q)}\left(iE\left[(||Z_1||_2)^{2+q}\right]+(n-i)tr\{\Sigma_n\}^{1+0.5q}\right)\\ \leq&n^{-(0.5q)}\left(E\left[(||Z_1||_2)^{2+q}\right]+tr\{\Sigma_n\}^{1+0.5q}\right)\to 0,\ as\ n\to 0\end{aligned}$$

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 Assumption 2.1(ii); and due to Jensen inequality $n^{-(0.5q)}E\left[(||Z_1||_2)^{2+q}\right]\leq\sqrt{n^{-q}E\left[(||Z_1||_2)^{4+2q}\right]}$ 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, by Lemma A.2, $\sum_{i=1}^n E[(||B_i||_2)^{4+2q}]\precsim n^{-(2+q)}\sum_{i=1}^n E[||V_{i,n}||_2^{4+2q}]\precsim n^{-(1+q)}(tr\{\Sigma_n\})^{2+q}$. By Assumption 2.1(ii) and the previous calculations, $n^{-(1+q)}(tr\{\Sigma_n\})^{2+q}=o(1)$. Similarly,

$$\begin{aligned}\sum_{i=1}^n E[(||A_i||_2)^{4+2q}]\precsim n^{-(2+q)}\sum_{i=1}^n E[||Z_i||_2^{4+2q}]\\ =n^{-(1+q)}E[||Z_1||_2^{4+2q}]=o(1)\end{aligned}$$

by Assumption 2.1(ii).

We thus have established that the remainder term $\mathbf{R}_n$ in Theorem 3.4 vanishes.

The desired result therefore follows by the same arguments as those in the proof of Theorem 3.2. □

Appendix B: Proofs of lemmas in Section 6

In order to prove the lemmas in Section 6 we need the following lemmas.

B.1. Supplementary lemmas

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

$$\mathcal{P}_{t,\delta,h}(||x||_2^2) = \int p_{t,\delta}(||x||_2^2 + hz)\phi(z)dz, \ \forall x \in \mathbb{R}^{d(n)}$$

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.

The next three lemmas show that we can use $\mathcal{P}_{t,\delta,h}(\cdot)$ to approximate the indicator function $1\{\cdot \geq t\}$ in expectation for the variables $||\sqrt{n}\mathbb{Z}_n^*||_2$, $||\sqrt{n}\mathbb{V}_n||_2$ and $||\sqrt{n}\mathbb{Z}_n||_2$, respectively.

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)

$$E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_2^2 \geq t\}|Z^n\right] \leq \frac{1}{1-\varepsilon} E_{\mathbf{P}_n^*}\left[\mathcal{P}_{t-\delta,\delta,h}(||\sqrt{n}\mathbb{Z}_n^*||_2^2)|Z^n\right]. \qquad (24)$$

(ii)

$$E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_2^2 \geq t\}|Z^n\right] \geq \frac{1}{1-\varepsilon} E_{\mathbf{P}_n^*}\left[\mathcal{P}_{t+2\delta,\delta,h}(||\sqrt{n}\mathbb{Z}_n^*||_2^2)|Z^n\right] - \frac{\varepsilon}{1-\varepsilon}. \qquad (25)$$

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)

$$E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_2^2 \geq t\}\right] \leq \frac{1}{1-\varepsilon} E_{\mathbf{\Phi}_n}\left[\mathcal{P}_{t-\delta,\delta,h}(||\sqrt{n}\mathbb{V}_n||_2^2)\right]. \qquad (26)$$

(ii)

$$E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_2^2 \geq t\}\right] \geq \frac{1}{1-\varepsilon} E_{\mathbf{\Phi}_n}\left[\mathcal{P}_{t+2\delta,\delta,h}(||\sqrt{n}\mathbb{V}_n||_2^2)\right] - \frac{\varepsilon}{1-\varepsilon}. \qquad (27)$$

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)

$$E_{\mathbf{P}_n}\left[1\{||\sqrt{n}\mathbb{Z}_n||_2^2 \geq t\}\right] \leq \frac{1}{1-\varepsilon} E_{\mathbf{P}_n}\left[\mathcal{P}_{t-\delta,\delta,h}(||\sqrt{n}\mathbb{Z}_n||_2^2)\right]. \tag{28}$$

(ii)

$$E_{\mathbf{P}_n}\left[1\{||\sqrt{n}\mathbb{Z}_n||_2^2 \geq t\}\right] \geq \frac{1}{1-\varepsilon} E_{\mathbf{P}_n}\left[\mathcal{P}_{t+2\delta,\delta,h}(||\sqrt{n}\mathbb{Z}_n||_2^2)\right] - \frac{\varepsilon}{1-\varepsilon}. \tag{29}$$

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 $n \geq N(\varepsilon)$:*

$$\sup_t \mathbf{\Phi}_n\left(|||\sqrt{n}\mathbb{V}_n||_2^2 - t| \leq \sqrt{tr\{\Sigma_n^2\}}\gamma\right) \leq \varepsilon. \tag{30}$$

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 $n \geq N(\varepsilon)$:*

$$\mathbf{\Phi}_n\left(||\sqrt{n}\mathbb{V}_n||_2^2 \geq t\right) \leq \varepsilon + \mathbf{\Phi}_n\left(||\sqrt{n}\mathbb{V}_n||_2^2 \geq t + \sqrt{tr\{\Sigma_n^2\}}\gamma\right) \tag{31}$$

for all $t \geq 0$.

Proof of Lemma B.1. **Part (i)** By definition of $\mathcal{P}_{t,\delta,h}$, for any $||x||_2^2 \geq t+\delta$

$$\begin{aligned}\mathcal{P}_{t,\delta,h}(||x||_2^2) &\geq \int 1\{z : ||x||_2^2 + hz \geq t\}\phi(z)dz \geq \int 1\{z : hz \geq -\delta\}\phi(z)dz \\ &= 1 - \Phi(-\delta/h).\end{aligned}$$

Thus, for any $h \leq \frac{\delta}{-\Phi^{-1}(\varepsilon)} \equiv h(\delta,\varepsilon)$, $\mathcal{P}_{t,\delta,h}(||x||_2^2) \geq (1-\varepsilon)1\{||x||_2^2 \geq t+\delta\}$. Thus

$$E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_2^2 \geq t\}|Z^n\right] \leq \frac{1}{1-\varepsilon} E_{\mathbf{P}_n^*}\left[\mathcal{P}_{t-\delta,\delta,h}(||\sqrt{n}\mathbb{Z}_n^*||_2^2)|Z^n\right]$$

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

Part (ii) Observe that for any $x : ||x||_2^2 \leq t - 2\delta$,

$$\mathcal{P}_{t,\delta,h}(||x||_2^2) \leq \int 1\{z : ||x||_2^2 + hz \geq t-\delta\}\phi(z)dz \leq \int 1\{z : hz \geq \delta\}\phi(z)dz.$$

Thus $\mathcal{P}_{t,\delta,h}(||x||_2^2) \leq 1-\Phi(\delta/h)$. Since $h \leq \frac{\delta}{-\Phi^{-1}(\varepsilon)}$, $\mathcal{P}_{t,\delta,h}(||x||_2^2) \leq \varepsilon$. Therefore, $\mathcal{P}_{t,\delta,h}(||x||_2^2) \leq \varepsilon$ for any $x : ||x||_2^2 \leq t - 2\delta$ and $h \leq h(\delta,\varepsilon)$. Thus, for all $x \in \mathbb{R}^d$, $\mathcal{P}_{t,\delta,h}(||x||_2^2) \leq (1-\varepsilon)1\{||x||_2^2 \geq t-2\delta\} + \varepsilon$. The result follows by taken expectation, $E_{\mathbf{P}_n^*}[.]$, 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. □

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

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

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, $E[|\lambda_l\zeta_{l,n}^2|^3] = \lambda_l^3E[|\zeta_{l,n}|^6] \leq C\left(\lambda_{max}(\Sigma_n)\right)^3$ 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 $n \to \infty$. [28] Therefore,

$$\begin{aligned}
&\sup_t \mathbf{\Phi}_n\left(|||\sqrt{n}\mathbb{V}_n||_2^2 - t| \leq \sqrt{tr\{\Sigma_n^2\}}\gamma\right)\\
=&\sup_t \mathbf{\Phi}_n\left(|\frac{||\xi_n||_2^2}{\sqrt{2tr\{\Sigma_n^2\}}} - \frac{t}{\sqrt{2tr\{\Sigma_n^2\}}}| \leq \gamma/\sqrt{2}\right)\\
=&\sup_{t'} \mathbf{\Phi}_n\left(|\frac{||\xi_n||_2^2}{\sqrt{2tr\{\Sigma_n^2\}}} - t'| \leq \gamma/\sqrt{2}\right)\\
=&\sup_{t'} \mathbf{\Phi}_n\left(|\frac{\sum_{l=1}^{d(n)}\lambda_l(\zeta_{l,n}^2-1)}{\sqrt{2tr\{\Sigma_n^2\}}} - t' + tr\{\Sigma_n\}| \leq \gamma/\sqrt{2}\right)\\
=&\sup_{t''} \mathbf{\Phi}_n\left(|\frac{\sum_{l=1}^{d(n)}\lambda_l(\zeta_{l,n}^2-1)}{\sqrt{2tr\{\Sigma_n^2\}}} - t''| \leq \gamma/\sqrt{2}\right).
\end{aligned}$$

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

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

[28] 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.

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 $6C\frac{tr\{\Sigma_n^3\}}{(2tr\{\Sigma_n^2\})^{3/2}} < 0.5\varepsilon$ for all $n \geq N(\varepsilon)$. Thus,

$$\begin{aligned}
&\sup_{t\in\mathbb{R}} \mathbf{\Phi}_n \left(|||\xi_n||_2^2 - t| \leq \sqrt{tr\{\Sigma_n^2\}}\gamma\right) \\
=&\sup_{t\in\mathbb{R}} \mathbf{\Phi}_n \left(\sqrt{tr\{\Sigma_n^2\}}\gamma - t \leq ||\xi_n||_2^2 \leq t + \sqrt{tr\{\Sigma_n^2\}}\gamma\right) \\
\leq&\sup_{t\in\mathbb{R}} \left|\Phi\left(t+\gamma/\sqrt{2}\right) - \Phi\left(t-\gamma/\sqrt{2}\right)\right| + 0.5\varepsilon.
\end{aligned}$$

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. □

B.2. Proofs of lemmas in Section 6

Proof of Lemma 6.1. The proof is analogous to that of Lemma 6.2 and will not be repeated here. □

Proof of Lemma 6.2. 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),

$$E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_2^2 \geq t\}\right] \geq E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_2^2 \geq t - 3\delta_n\}\right] - 3\varepsilon \tag{32}$$

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

$$E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_2^2 \geq t\}\right] \geq \frac{1}{1-\varepsilon}E_{\mathbf{\Phi}_n}\left[\mathcal{P}_{t-\delta_n,\delta_n,h}(||\sqrt{n}\mathbb{V}_n||_2^2)\right] - \frac{\varepsilon}{1-\varepsilon} - 3\varepsilon \tag{33}$$

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

$$E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_2^2 \geq t\}|Z^n\right] \leq \frac{1}{1-\varepsilon}E_{\mathbf{P}_n^*}\left[\mathcal{P}_{t-\delta_n,\delta_n,h}(||\sqrt{n}\mathbb{Z}_n^*||_2^2)|Z^n\right]. \tag{34}$$

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

$$\begin{aligned}
&E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_2^2 \geq t\}|Z^n\right] - E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_2^2 \geq t\}\right] \\
\leq&\frac{1}{1-\varepsilon}\left(E_{\mathbf{P}_n^*}\left[\mathcal{P}_{t-\delta_n,\delta_n,h}(||\sqrt{n}\mathbb{Z}_n^*||_2^2)|Z^n\right] - E_{\mathbf{\Phi}_n}\left[\mathcal{P}_{t-\delta_n,\delta_n,h}(||\sqrt{n}\mathbb{V}_n||_2^2)\right]\right) \\
&+ \frac{\varepsilon}{1-\varepsilon} + 3\varepsilon.
\end{aligned} \tag{35}$$

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

$$E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_2^2 \geq t\}|Z^n\right] \geq \frac{1}{1-\varepsilon}E_{\mathbf{P}_n^*}\left[\mathcal{P}_{t+2\delta_n,\delta_n,h}(||\sqrt{n}\mathbb{Z}_n^*||_2^2)|Z^n\right] - \frac{\varepsilon}{1-\varepsilon}. \tag{36}$$

By Remark B.1 (applied thrice),

$$E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_2^2 \geq t\}\right] \leq E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_2^2 \geq t+3\delta_n\}\right] + 3\varepsilon \tag{37}$$

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

$$E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_2^2 \geq t\}\right] \leq \frac{1}{1-\varepsilon} E_{\mathbf{\Phi}_n}\left[\mathcal{P}_{t+2\delta_n,\delta_n,h}(||\sqrt{n}\mathbb{V}_n||_2^2)\right] + 3\varepsilon \tag{38}$$

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

Hence,

$$\begin{aligned}
&E_{\mathbf{P}_n^*}\left[1\{||\sqrt{n}\mathbb{Z}_n^*||_2^2 \geq t\}|Z^n\right] - E_{\mathbf{\Phi}_n}\left[1\{||\sqrt{n}\mathbb{V}_n||_2^2 \geq t\}\right] \\
\geq &\frac{1}{1-\varepsilon}\left(E_{\mathbf{P}_n^*}\left[\mathcal{P}_{t+2\delta_n,\delta_n,h}(||\sqrt{n}\mathbb{Z}_n^*||_2^2)|Z^n\right] - E_{\mathbf{\Phi}_n}\left[\mathcal{P}_{t+2\delta_n,\delta_n,h}(||\sqrt{n}\mathbb{V}_n||_2^2)\right]\right) \\
&- \frac{\varepsilon}{1-\varepsilon} - 3\varepsilon.
\end{aligned} \tag{39}$$

By displays 35 and 39, in order to obtain the desired result it suffices to verify that $a \in \mathbb{R} \mapsto \mathcal{P}_{t,\delta,h}(a)$ belong to $\mathcal{C}_{h^{-1}}$. 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 $0 \leq p_{t,\delta}(u) \leq 1$. Similar calculations yield

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

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

Proof of Lemma 6.3. 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 $f(x) \equiv g(||x||_2^2)$. Observe that for any $g \in \mathcal{C}_{h^{-1}}$, $\partial_i f(x) = g'(||x||_2^2)2x_i$ and $\partial_{ij} f(x) = g''(||x||_2^2)4x_i x_j + 2g'(||x||_2^2)1\{i=j\}$.

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. (2013a) 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 $\tilde{\xi}_n\sim N(0,n^{-1}\sum_{i=1}^n Z_iZ_i^T)$, under $\mathbf{Q}_n^*(\cdot|Z^n)$.

Therefore,

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

Observe that, by Cauchy-Schwarz inequality and the fact that $\partial_{ij} f(x) = g''(||x||_2^2)4x_i x_j + 2g'(||x||_2^2)1\{i = j\}$

$$\begin{aligned}\sum_{j=1}^{d(n)}\sum_{l=1}^{d(n)} E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[|\partial_{jl} f(\xi_n(t))|\right] \leq & \frac{4}{h^2}\sum_{j=1}^{d(n)}\sum_{l=1}^{d(n)} E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[|\xi_{[j],n}(t)||\xi_{[l],n}(t)|\right] \\ & + 2h^{-1}d(n) \\ \leq & 4h^{-2}\left(\sum_{j=1}^{d(n)}\sqrt{E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[|\xi_{[j],n}(t)|^2\right]}\right)^2 \\ & + 2h^{-1}d(n) \\ \leq & \frac{4d(n)}{h^2} E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[||\xi_n(t)||_2^2\right] + \frac{2d(n)}{h}.\end{aligned}$$

Therefore, since $||\xi_n(t)||_2^2 \precsim \{||\xi_n||_2^2 + ||\tilde{\xi}_n||_2^2\}$,

$$\begin{aligned}\sum_{j=1}^{d(n)}\sum_{l=1}^{d(n)} E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[|\partial_{jl} f(\xi_n(t))|\right] \precsim & \frac{d(n)}{h^2} E_{\mathbf{Q}_n^*(\cdot|Z^n)\cdot\mathbf{Q}_n}\left[||\xi_n||_2^2 + ||\tilde{\xi}_n||_2^2\right] \\ & + \frac{d(n)}{h}2 \\ = & d(n)h^{-1}\{h^{-1}\left(tr\{\Sigma_n\} + tr\{\hat{\Sigma}_n\}\right) + 2\}.\end{aligned}$$

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

Appendix C: Proofs for Section 4

We first introduce some notation and lemmas needed in the proofs of the results in Section 4 (the proofs of these lemmas are relegated to the end of the section). Let $\hat{g}_n^* \equiv n^{-1}\sum_{i=1}^n \omega_{in} g(X_i, \hat{\theta}^*_{GMM,n})$ and $\bar{g}_n^* \equiv n^{-1}\sum_{i=1}^n \omega_{in} g(X_i, \theta_0)$. Let $\bar{G}_n^*(\theta) = n^{-1}\sum_{i=1}^n \omega_{in}\nabla_\theta g(X_i, \theta) \in \mathbb{R}^{d(n)\times q}$.

Lemma C.1. *Suppose Assumption 4.2(ii)(iii)(iv) holds. Then:*

(1) $\sqrt{n}||\bar{g}_n^*||_2 = O_{\mathbf{P}_n^*(\cdot|Z^n)}(\sqrt{n^{-1}\sum_{i=1}^n ||g(X_i, \theta_0)||_2^2})$.

(2) Uniformly over $\theta \in \{\theta \in \Theta : ||\theta - \theta_0||_2 \precsim \Delta_n\}$ *with* $\Delta_n = o(1)$,

$$||\bar{G}_n^*(\theta)||_2 = O_{\mathbf{P}_n^*(\cdot|Z^n)}(\sqrt{d(n)}(n^{-1/2} + \Delta_n)),\ wpa1 - \mathbf{P}.$$

Let

$$\begin{aligned}& R_n^*(\theta, \lambda) \\ & \equiv \lambda^T\left(n^{-1}\sum_{i=1}^n \int_0^1 s''(t\lambda^T\omega_{in} g(X_i, \theta))dt\omega_{in}^2 g(X_i, \theta)g(X_i, \theta)^T - s''(0)\Omega\right)\lambda.\end{aligned}$$

Lemma C.2. *Suppose Assumption 4.2(i)(ii)(iii) holds and* $d(n)^4/n = o(1)$. *Then:*

(1) For all $\theta \in \mathcal{N}$, $\{\lambda : ||\lambda||_2 \precsim \sqrt{d(n)/n}\} \subseteq \Lambda(\theta)$, *wpa1-*$\mathbf{P}$.[29]

(2) Uniformly over $\lambda \in \{\lambda \in \Lambda(\hat{\theta}^*_{GEL,n}) : ||\lambda||_2 \precsim \sqrt{d(n)/n}\}$ *and* $||\theta - \theta_0||_2 \precsim \Delta_n$ *with* $\Delta_n = o(1)$,

$$nR^*_n(\theta, \lambda) = O_{\mathbf{P}^*_n(\cdot|Z^n)}\left(\sqrt{d(n)}(o(1) + d(n)^{3/2}\Delta_n)\right), \ wpa1 - \mathbf{P}.$$

The following lemma is a general result that provides a relationship between $O_{\mathbf{P}^*_n(\cdot|Z^n)}$ (and $o_{\mathbf{P}^*_n(\cdot|Z^n)}$) and $O_{\mathbf{P}}$ variables that we use throughout.

Lemma C.3. *Let* $(W_i)_i$ *and* $(X_i)_i$ *be sequences of random variables such that* W_n *is* $(\omega_{in}, Z_i)_{i\leq n}$ *measurable and* X_n *is* $(Z_i)_{i\leq n}$ *measurable and* $X_n \neq 0$ *a.s.-*$\mathbf{P}$. *Let* $(c_n)_n$ *be a sequence of positive real numbers. Then:*

(1) If $W_n = O_{\mathbf{P}^*_n(\cdot|Z^n)}(|X_n|)$ *and* $X_n = O_{\mathbf{P}}(c_n)$, *then* $W_n = O_{\mathbf{P}^*_n(\cdot|Z^n)}(c_n)$ *wpa1-*$\mathbf{P}$.

(2) If $W_n = O_{\mathbf{P}^*_n(\cdot|Z^n)}(|X_n|)$ *and* $X_n = o_{\mathbf{P}}(c_n)$, *then* $W_n = o_{\mathbf{P}^*_n(\cdot|Z^n)}(c_n)$ *wpa1-*$\mathbf{P}$.

Proof of Lemma 4.1. The proof for $\hat{T}_{GMM,n}$ is in Lemma 6.1 in DIN and also analogous to that of $\hat{T}^*_{GMM,n}$, so it will be omitted.

We now establish the result for $\hat{T}^*_{GMM,n}$. It follows that $n|(\bar{g}^*_n)^T\hat{W}_n\bar{g}^*_n - (\bar{g}^*_n)^T W_n \bar{g}^*_n| \leq ||\hat{W}_n - W_n||_2 \times ||\sqrt{n}\bar{g}^*_n||^2_2$. By Lemma C.1(1),

$$n|(\bar{g}^*_n)^T\hat{W}_n\bar{g}^*_n - (\bar{g}^*_n)^T W_n \bar{g}^*_n| = O_{\mathbf{P}^*_n(\cdot|Z^n)}\left(n^{-1}\sum_{i=1}^n ||g(X_i,\theta_0)||^2_2||\hat{W}_n - W_n||_2\right).$$

Under Assumption 4.3 and since $E_{\mathbf{P}}[n^{-1}\sum_{i=1}^n ||g(X_i,\theta_0)||^2_2] = tr\{\Omega\} = O(d(n))$, it follows by the Markov inequality that $n^{-1}\sum_{i=1}^n ||g(X_i,\theta_0)||^2_2||\hat{W}_n - W_n||_2 = o_{\mathbf{P}}(\sqrt{d(n)})$. Thus, by Lemma C.3, $n|(\bar{g}^*_n)^T\hat{W}_n\bar{g}^*_n - (\bar{g}^*_n)^T W_n \bar{g}^*_n| = o_{\mathbf{P}^*_n(\cdot|Z^n)}(\sqrt{d(n)})$ wpa1-$\mathbf{P}$.

Given this, it suffices to show that
$n|(\hat{g}^*_n)^T\hat{W}_n\hat{g}^*_n - (\bar{g}^*_n)^T\hat{W}_n\bar{g}^*_n| = o_{\mathbf{P}^*_n(\cdot|Z^n)}(\sqrt{d(n)})$ wpa1-$\mathbf{P}$. Note that

$$\begin{aligned} n|(\hat{g}^*_n)^T\hat{W}_n\hat{g}^*_n - (\bar{g}^*_n)^T\hat{W}_n\bar{g}^*_n| \leq & 2n|(\hat{g}^*_n - \bar{g}^*_n)^T\hat{W}_n\bar{g}^*_n| + n|(\hat{g}^*_n - \bar{g}^*_n)^T\hat{W}_n(\hat{g}^*_n - \bar{g}^*_n)| \\ = & 2n|(\hat{\theta}^*_{GMM,n} - \theta_0)^T(\Gamma^*_n)^T\hat{W}_n\bar{g}^*_n| \\ & + n|(\hat{\theta}^*_{GMM,n} - \theta_0)^T(\Gamma^*_n)^T\hat{W}_n(\Gamma^*_n)(\hat{\theta}^*_{GMM,n} - \theta_0)| \\ \equiv & T^*_{1,n} + T^*_{2,n} \end{aligned}$$

where the second line follows Assumption 4.2(i) and the mean value Theorem; here $\Gamma^*_n \equiv \int_0^1 \bar{G}^*_n(\hat{\theta}^*_n(t))dt$ with $\hat{\theta}^*_n(t) \equiv \theta_0 + t(\hat{\theta}^*_{GMM,n} - \theta_0)$. The desired result

[29] The set $\mathcal{N}$ is the one in Assumption 4.2.

follows by establishing that $T^*_{i,n} = o_{\mathbf{P}^*_n(\cdot|Z^n)}(\sqrt{d(n)})$ wpa1-**P** for $i = 1, 2$. We do this next.

We note that,

$$T^*_{1,n} \precsim \sqrt{n}||\hat{\theta}^*_{GMM,n} - \theta_0||_2||(\Gamma^*_n)^T\hat{W}_n\sqrt{n}\bar{g}^*_n||_2$$

wpa1-**P**.

By assumption $\sqrt{n}||\hat{\theta}^*_{GMM,n} - \theta_0||_2 = O_{\mathbf{P}^*_n(\cdot|Z^n)}(\sqrt{d(n)})$ wpa1-**P**. Moreover, under Assumption 4.3, $\lambda_{max}(\hat{W}_n) \leq C$ wpa1-**P** and thus $||(\Gamma^*_n)^T\hat{W}_n\sqrt{n}\bar{g}^*_n||_2 \precsim ||\Gamma^*_n||_2||\sqrt{n}\bar{g}^*_n||_2$. We can apply Lemma C.1(2) with $\Delta_n = n^{-1/2}\sqrt{d(n)}$ and obtain $||\Gamma^*_n||_2 = O_{\mathbf{P}^*_n(\cdot|Z^n)}(d(n)/\sqrt{n})$ wpa1-**P**. By Lemma C.1, and since $E_{\mathbf{P}}[n^{-1}\sum_{i=1}^n ||g(X_i,\theta_0)||_2^2] = tr\{\Omega\} = O(d(n))$, it follows by Lemma C.3 that $||\sqrt{n}\bar{g}^*_n||_2 = O_{\mathbf{P}^*_n(\cdot|Z^n)}(\sqrt{(d(n)})$. Thus $T^*_{1,n} = O_{\mathbf{P}^*_n(\cdot|Z^n)}((\sqrt{d(n)}d(n)/\sqrt{n})\sqrt{d(n)}) = O_{\mathbf{P}^*_n(\cdot|Z^n)}(\sqrt{d(n)}\frac{d(n)^{3/2}}{\sqrt{n}})$ wpa1-**P** since $\frac{d(n)^3}{n} \to 0$ the result follows.

Finally, by our assumption $||\hat{\theta}^*_{GMM,n} - \theta_0||_2 = O_{\mathbf{P}^*_n(\cdot|Z^n)}(\sqrt{d(n)}n^{-1/2})$, Lemma C.1, and Assumption 4.3, it follows that $T_{2,n} = O_{\mathbf{P}^*_n(\cdot|Z^n)}(d(n)^{1/2}\frac{d(n)^{5/2}}{n}) = o_{\mathbf{P}^*_n(\cdot|Z^n)}(d(n)^{1/2})$ wpa1-**P** because $d(n)^3/n \to 0$.

Therefore, we conclude that

$$n|(\hat{g}^*_n)^T\hat{W}_n\hat{g}^*_n - (\bar{g}^*_n)^T W_n\bar{g}^*_n| = o_{\mathbf{P}^*_n(\cdot|Z^n)}(\sqrt{d(n)})$$

wpa1-**P**.

We now establish the result for $\hat{T}^*_{GEL,n}$. The proof for $\hat{T}_{GEL,n}$ is completely analogous and therefore omitted. Abusing notation, we denote $\hat{g}^*_n \equiv n^{-1}\sum_{i=1}^n \omega_{in}g(X_i,\hat{\theta}^*_{GEL,n})$. Define the following function

$$\lambda \mapsto F^*_n(\lambda) = s'(0)\lambda^T\bar{g}^*_n + 0.5s''(0)\lambda^T\Omega\lambda.$$

Since $s''(0) < 0$, the maximum of this function is achieved at $\lambda_0 = -\frac{s'(0)}{s''(0)}\Omega^{-1}\bar{g}^*_n$ and $F^*_n(\lambda_0) = 0.5\frac{(s'(0))^2}{s''(0)}(\bar{g}^*_n)^T\Omega^{-1}\bar{g}^*_n$. By Lemma C.1(1) and the fact that Ω has eigenvalues uniformly bounded away from zero (Assumption 4.1), $||\lambda_0||_2 = O_{\mathbf{P}^*_n(\cdot|Z^n)}(\sqrt{d(n)/n})$ wpa1-**P**. Hence, $\lambda_0 \in \Lambda(\hat{\theta}^*_{GEL,n})$ wpa1-**P** by Lemma C.2(1).

By definition of $\hat{T}^*_{GEL,n}$ and the mean value Theorem

$$\hat{T}^*_{GEL,n} \geq 2\sum_{i=1}^n \left(s(\lambda^T\omega_{in}g(X_i,\hat{\theta}^*_{GEL,n})) - s(0)\right) = 2nF^*_n(\lambda) + nR^*_n(\hat{\theta}^*_{GEL,n},\lambda)$$

for all $\lambda \in \Lambda(\hat{\theta}^*_{GEL,n})$ with R^*_n defined in Lemma C.2.

By Lemma C.2(2) with $\Delta_n = n^{-1/2}\sqrt{d(n)}$, it follows that $nR^*_n(\hat{\theta}^*_{GEL,n},\lambda_0) = o_{\mathbf{P}^*_n(\cdot|Z^n)}(d(n)^{1/2})$ wpa1-**P** since $d(n)^4/n = o(1)$ by assumption. Moreover, $\lambda_0 \in \Lambda(\hat{\theta}^*_{GEL,n})$, so $\hat{T}^*_{GEL,n} \geq 2nF^*_n(\lambda_0) + o_{\mathbf{P}^*_n(\cdot|Z^n)}(d(n)^{1/2})$ wpa1-**P**.

By definition of λ_0 it also follows that $F^*_n(\lambda_0) \geq F^*_n(\hat{\lambda}^*_n)$ (recall that $\hat{\lambda}^*_n$ is the maximizer of $\sum_{i=1}^n s(\lambda^T \omega_{in} g(X_i, \hat{\theta}^*_{GEL,n}))$; see Assumption 4.4).

Therefore,

$$2nF^*_n(\lambda_0) \geq 2nF^*_n(\hat{\lambda}^*_n) = \hat{T}^*_{GEL,n} - nR^*_n(\hat{\theta}^*_{GEL,n}, \hat{\lambda}^*_n).$$

Observe that, since $||\hat{\lambda}^*_n||_2 = O_{\mathbf{P}^*_n(\cdot|Z^n)}(\sqrt{d(n)/n})$ (by Assumption 4.4), by Lemma C.2(2) with $\Delta_n = n^{-1/2}\sqrt{d(n)}$, $nR^*_n(\hat{\theta}^*_{GEL,n}, \hat{\lambda}^*_n) = o_{\mathbf{P}^*_n(\cdot|Z^n)}(\sqrt{d(n)})$ wpa1-**P**, and thus $2nF^*_n(\lambda_0) \geq 2nF^*_n(\hat{\lambda}^*_n) \geq \hat{T}^*_{GEL,n} + o_{\mathbf{P}^*_n(\cdot|Z^n)}(\sqrt{d(n)})$ wpa1-**P**.

Therefore, it follows that

$$\begin{aligned}\hat{T}^*_{GEL,n} =& 2nF^*_n(\lambda_0) + o_{\mathbf{P}^*_n(\cdot|Z^n)}(d(n)^{1/2}) \\ =& \frac{(s'(0))^2}{s''(0)} n(\bar{g}^*_n)^T \Omega^{-1} \bar{g}^*_n + o_{\mathbf{P}^*_n(\cdot|Z^n)}(\sqrt{d(n)})\end{aligned}$$

wpa1-**P**. □

Throughout the proof, for any matrix M, let $||t||^2_M \equiv t'Mt$.

Proof of Theorem 4.1. We only establish the result for the GMM estimator; the one for the GEL estimator is completely analogous. We divide the proof into several steps.

STEP 1. By Lemma 4.1, for any $\varepsilon > 0$,

$$\begin{aligned}&\mathbf{P}^*_n\left(\frac{\hat{T}^*_{GMM,n}}{\sqrt{d(n)}} \geq t \mid Z^n\right) \\ &\leq (\geq)\mathbf{P}^*_n\left(\frac{\left\|n^{-1/2}\sum_{i=1}^n \omega_{in} g(X_i, \theta_0)\right\|^2_W}{\sqrt{d(n)}} \geq t - (+)\varepsilon \mid Z^n\right) + o_{\mathbf{P}}(1)\end{aligned}$$

and similarly,

$$\mathbf{P}\left(\frac{\hat{T}_{GMM,n}}{\sqrt{d(n)}} \geq t\right) \geq (\leq)\mathbf{P}\left(\frac{\left\|n^{-1/2}\sum_{i=1}^n g(X_i, \theta_0)\right\|^2_W}{\sqrt{d(n)}} \geq t + (-)\varepsilon\right) - o(1).$$

STEP 2. We now verify Assumptions 2.2 and 2.1 for $Z_i \equiv W^{1/2} g(X_i, \theta_0)$. The former is directly imposed, so we only need to verify the latter.

Note that $\Sigma_n = W^{1/2} E_{\mathbf{P}}[g(X, \theta_0) g(X, \theta_0)^T] W^{1/2} = W^{1/2} \Omega W^{1/2}$. Thus, Assumption 2.1(i), the first part, follows by the fact that $C^{-1} \leq \lambda_l(\Omega) \leq C$ for all $l = 1, ..., d(n)$ (Assumption 4.1) and Assumption 4.3. Regarding the second part of Assumption 2.1(i), note that under Assumption 4.2(i), for any $l \leq 2(2 + \gamma)$,

$$E_{\mathbf{P}_n}[||Z_1||^l_2] = E_{\mathbf{P}}[||W^{1/2} g(X, \theta_0)||^l_2] \precsim E_{\mathbf{P}}[||g(X, \theta_0)||^l_2] \text{ (by Assumption 4.3)}$$

and by Assumption 4.2(i), $E_{\mathbf{P}_n}[||Z_1||^l_2] \precsim d(n)^{l/2}$. Thus $E_{\mathbf{P}_n}[||Z_1||^4_2] \precsim d(n)^2$ and $(E_{\mathbf{P}_n}[||Z_1||^3_2])^2 \precsim d(n)^3$. Hence, the expression in the second part of Assumption 2.1(i) is of order $d(n)^4/n$ which is $o(1)$ by assumption.

Assumption 2.1(ii) follows because $E_{\mathbf{P}}[||g(X,\theta_0)||_2^{2(2+\gamma)}] \precsim d(n)^{2+\gamma}$ and $\frac{d(n)^{4+2\gamma}}{n^\gamma} = \left(\frac{d(n)^{2+4/\gamma}}{n}\right)^\gamma = o(1)$ by assumption. Finally, part (iii) of the Assumption 2.1 follows with $\kappa = 0$ and $\frac{d(n)^4}{n} = o(1)$.

Step 3. We now show that: For any $\varepsilon > 0$ and $t \in \mathbb{R}$, there exists a $N(\varepsilon)$ such that

$$\mathbf{P}\left(\frac{\left\|n^{-1/2}\sum_{i=1}^n g(X_i,\theta_0)\right\|_W^2}{\sqrt{d(n)}} \geq t+\varepsilon\right) \geq \mathbf{P}_n\left(\frac{\left\|n^{-1/2}\sum_{i=1}^n g(X_i,\theta_0)\right\|_W^2}{\sqrt{d(n)}} \geq t-\varepsilon\right) - \varepsilon - o(1)$$

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

By the Expression 5, uniformly over t,

$$\mathbf{P}\left(\frac{\left\|n^{-1/2}\sum_{i=1}^n g(X_i,\theta_0)\right\|_W^2}{\sqrt{d(n)}} \geq t+\varepsilon\right) \geq \mathbf{P}\left(\frac{\left\|W^{1/2}\sqrt{n}\mathbb{V}_n\right\|_2^2}{\sqrt{d(n)}} \geq t+\varepsilon\right) - o(1)$$

where $\sqrt{n}\mathbb{V}_n \sim N(0,\Omega)$. Under our assumptions $d(n) \asymp tr\{(W^{1/2}\Omega W^{1/2})^2\}$ and thus by Lemma B.4 (and its Remark B.1), it follows that for sufficiently small ε, $\mathbf{P}\left(\frac{\|W^{1/2}\sqrt{n}\mathbb{V}_n\|_2^2}{\sqrt{d(n)}} \geq t+\varepsilon\right) \geq \mathbf{P}\left(\frac{\|W^{1/2}\sqrt{n}\mathbb{V}_n\|_2^2}{\sqrt{d(n)}} \geq t-\varepsilon\right) - 0.5\varepsilon$ for any $n \geq N(\varepsilon)$. Invoking again Expression 5, the desired result follows.

Step 4. For any $t \in \mathbb{R}$ and all $n \geq N(\varepsilon)$,

$$\begin{aligned}
&\mathbf{P}_n^*\left(\frac{\hat{T}^*_{GMM,n}}{\sqrt{d(n)}} \geq t \mid Z^n\right) - \mathbf{P}\left(\frac{\hat{T}_{GMM,n}}{\sqrt{d(n)}} \geq t\right)\\
\leq&\mathbf{P}_n^*\left(\frac{\left\|n^{-1/2}\sum_{i=1}^n \omega_{i,n} g(X_i,\theta_0)\right\|_W^2}{\sqrt{d(n)}} \geq t-\varepsilon \mid Z^n\right)\\
&- \mathbf{P}\left(\frac{\left\|n^{-1/2}\sum_{i=1}^n g(X_i,\theta_0)\right\|_W^2}{\sqrt{d(n)}} \geq t+\varepsilon\right) + o_{\mathbf{P}}(1),\ (\textit{by Step } 1)\\
\leq&\mathbf{P}_n^*\left(\frac{\left\|n^{-1/2}\sum_{i=1}^n \omega_{i,n} g(X_i,\theta_0)\right\|_W^2}{\sqrt{d(n)}} \geq t-\varepsilon \mid Z^n\right)\\
&- \mathbf{P}\left(\frac{\left\|n^{-1/2}\sum_{i=1}^n g(X_i,\theta_0)\right\|_W^2}{\sqrt{d(n)}} \geq t-\varepsilon\right) - \varepsilon + o_{\mathbf{P}}(1),\ (\textit{by Step } 3).
\end{aligned}$$

An analogous result holds for $-\left(\mathbf{P}_n^*\left(\frac{\hat{T}^*_{GMM,n}}{\sqrt{d(n)}} \geq t \mid Z^n\right) - \mathbf{P}\left(\frac{\hat{T}_{GMM,n}}{\sqrt{d(n)}} \geq t\right)\right)$.

Therefore, for any $\varepsilon > 0$ and $n \geq N(\varepsilon)$,

$$\begin{aligned}
&\sup_{t\in\mathbb{R}} \left| \mathbf{P}_n^* \left(\frac{\hat{T}^*_{GMM,n}}{\sqrt{d(n)}} \geq t \mid Z^n \right) - \mathbf{P}\left(\frac{\hat{T}_{GMM,n}}{\sqrt{d(n)}} \geq t \right) \right| \\
\leq & \sup_{t\in\mathbb{R}} \left| \mathbf{P}_n^* \left(\frac{\left\| n^{-1/2} \sum_{i=1}^n \omega_{i,n} g(X_i, \theta_0) \right\|_W^2}{\sqrt{d(n)}} \geq t \mid Z^n \right) \right. \\
& \left. - \mathbf{P}\left(\frac{\left\| n^{-1/2} \sum_{i=1}^n g(X_i, \theta_0) \right\|_W^2}{\sqrt{d(n)}} \geq t \right) \right| + \varepsilon + o_{\mathbf{P}}(1) \\
= & \sup_{t\in\mathbb{R}} \left| \mathbf{P}_n^* \left(\left\| n^{-1/2} \sum_{i=1}^n \omega_{i,n} g(X_i, \theta_0) \right\|_W^2 \geq t \mid Z^n \right) - \mathbf{P}\left(\left\| n^{-1/2} \sum_{i=1}^n g(X_i, \theta_0) \right\|_W^2 \geq t \right) \right| \\
& + \varepsilon + o_{\mathbf{P}}(1)
\end{aligned}$$

where the last line follows from the fact that $\sqrt{d(n)}t \in \mathbb{R}$ for any $t \in \mathbb{R}$. The desired result thus follows from Theorem 3.1 with $Z_i \equiv W^{1/2} g(X_i, \theta_0)$ for all $i = 1, ..., n$. □

C.1. Proofs of Lemmas C.1, C.2 and C.3

Proof of Lemma C.1. (1) Note that

$$\begin{aligned}
& E_{\mathbf{P}_n^*(\cdot|Z^n)}[||\sqrt{n}\bar{g}_n^*||_2^2] \\
= & tr\{E_{\mathbf{P}_n^*(\cdot|Z^n)}[(n^{-1/2} \sum_{i=1}^n \omega_{in} g(X_i, \theta_0))(n^{-1/2} \sum_{i=1}^n \omega_{in} g(X_i, \theta_0))^T]\} \\
= & tr\{n^{-1} \sum_{i=1}^n E_{\mathbf{P}_n^*(\cdot|Z^n)}[\omega_{in}^2] g(X_i, \theta_0) g(X_i, \theta_0)^T\}
\end{aligned}$$

because under Assumption 2.2, the weights are centered and independent. Thus $E_{\mathbf{P}_n^*(\cdot|Z^n)}[||\sqrt{n}\bar{g}_n^*||_2^2] = n^{-1} \sum_{i=1}^n ||g(X_i, \theta_0)||_2^2$, and the desired result follows by the Markov inequality.

(2) By the triangle inequality

$$\begin{aligned}
||\bar{G}_n^*(\theta)||_2 \leq & ||n^{-1} \sum_{i=1}^n \omega_{in} \{\nabla_\theta g(X_i, \theta) - \nabla_\theta g(X_i, \theta_0)\}||_2 \\
& + n^{-1/2} ||n^{-1/2} \sum_{i=1}^n \omega_{in} \nabla_\theta g(X_i, \theta_0)||_2 \\
\equiv & T_{1,n} + T_{2,n}
\end{aligned}$$

where $\nabla_\theta g(X,\theta_0) \in \mathbb{R}^{d(n)\times q}$. Recall that for matrices, $||A||_2$ is the spectral norm. Let $||A|| \equiv tr\{A^T A\}$; it is clear that $||A||_2 \leq ||A||$. Moreover,

$$\begin{aligned}
&||n^{-1/2}\sum_{i=1}^{n}\omega_{in}\nabla_\theta g(X_i,\theta_0)||^2\\
=&tr\left\{\left(n^{-1/2}\sum_{i=1}^{n}\omega_{in}\nabla_\theta g(X_i,\theta_0)\right)^T\left(n^{-1/2}\sum_{i=1}^{n}\omega_{in}\nabla_\theta g(X_i,\theta_0)\right)\right\}\\
=&tr\left\{n^{-1}\sum_{i=1}^{n}\omega_{in}^2\left(\nabla_\theta g(X_i,\theta_0)\right)^T\nabla_\theta g(X_i,\theta_0)\right\}\\
&+tr\left\{n^{-1}\sum_{i\neq j}\omega_{in}\omega_{jn}\left(\nabla_\theta g(X_i,\theta_0)\right)^T\nabla_\theta g(X_j,\theta_0)\right\}.
\end{aligned}$$

Applying $E_{\mathbf{P}_n^*(\cdot|Z^n)}$ the second term in the RHS vanishes because of independence of the weights and zero mean. Thus, since $E_{\mathbf{P}_n^*(\cdot|Z^n)}[\omega_{in}^2]=1$, it follows by the Markov inequality that

$$T_{2,n}=O_{\mathbf{P}_n^*(\cdot|Z^n)}\left(n^{-1/2}\sqrt{n^{-1}\sum_{i=1}^{n}||\nabla_\theta g(X_i,\theta_0)||^2}\right).$$

Also, note that $n^{-1}\sum_{i=1}^{n}||\nabla_\theta g(X_i,\theta_0)||^2=O_{\mathbf{P}}(d(n))$ by the Markov inequality, Assumption 4.2(iii), and the fact that $||A||\leq\sqrt{q}||A||_2$. Therefore by Lemma C.3, $T_{2,n}=O_{\mathbf{P}_n^*(\cdot|Z^n)}\left(\sqrt{d(n)/n}\right)$ wpa1-$\mathbf{P}$.

Regarding $T_{1,n}$, note that $T_{1,n} \leq n^{-1}\sum_{i=1}^{n}|\omega_{in}| \times ||\nabla_\theta g(X_i,\theta)-\nabla_\theta g(X_i,\theta_0)||_2$. Under Assumption 4.2(iv),

$$T_{1,n}\leq n^{-1}\sum_{i=1}^{n}|\omega_{in}|\delta_n(X_i)||\theta-\theta_0||_2\leq n^{-1}\sum_{i=1}^{n}|\omega_{in}|\delta_n(X_i)\Delta_n.$$

Since weights are uniformly bounded, $T_{1,n}\precsim\Delta_n n^{-1}\sum_{i=1}^{n}\delta_n(X_i)$ a.s-$\mathbf{P}_n^*$. Thus under Assumption 4.2(iv), the Markov inequality and Lemma C.3, $T_{1,n}=O_{\mathbf{P}_n^*(\cdot|Z^n)}(\Delta_n\sqrt{d(n)})$ wpa1-$\mathbf{P}$. □

Proof of Lemma C.2. (1) Observe that $|\lambda^T\omega_{in}g(X_i,\theta)|\leq||\lambda||_2|\omega_{in}|||g(X_i,\theta)||_2\precsim\sqrt{d(n)/n}|\omega_{in}|||g(X_i,\theta)||_2$. It suffices to show that

$$\sqrt{d(n)/n}\max_{i\leq n}|\omega_{in}|||g(X_i,\theta)||_2=o_{\mathbf{P}_n^*(\cdot|Z^n)}(1)$$

wpa1-$\mathbf{P}$, uniformly in $\theta\in\mathcal{N}$. Since weights are uniformly bounded, it suffices to show that $\sqrt{d(n)/n}\max_{i\leq n}\sup_{\theta\in\mathcal{N}}||g(X_i,\theta)||_2=o_{\mathbf{P}}(1)$. By the Markov inequality

$$\mathbf{P}(\max_{i\leq n}\sup_{\theta\in\mathcal{N}}||g(X_i,\theta)||_2\geq K_n)\leq\frac{n}{K_n^{2\alpha}}E_{\mathbf{P}}[\sup_{\theta\in\mathcal{N}}||g(X_i,\theta)||_2^{2\alpha}].$$

Thus by Assumption 4.2(i) and $K_n = n^{1/(2\alpha)}\sqrt{d(n)}$ it follows that

$$\sqrt{d(n)/n}\max_{i\leq n}\sup_{\theta\in\mathcal{N}}||g(X_i,\theta)||_2 \precsim \frac{d(n)}{n^{0.5(1-1/\alpha)}}$$

since $d(n)^4/n = o(1)$ and $\alpha \geq 2$ this implies the desired result.

(2) It follows that

$$\begin{aligned}
&R_n^*(\theta,\lambda)\\
&\leq ||\lambda||_2^2 \left\| \int_0^1 n^{-1}\sum_{i=1}^n s''(t\lambda^T\omega_{in}g(X_i,\theta))\omega_{in}^2 g(X_i,\theta)g(X_i,\theta)^T dt - s''(0)\Omega \right\|_2\\
&\leq ||\lambda||_2^2 s''(0)\left\| n^{-1}\sum_{i=1}^n \omega_{in}^2 g(X_i,\theta)g(X_i,\theta)^T - \Omega\right\|_2\\
&+ ||\lambda||_2^2\left\| n^{-1}\sum_{i=1}^n \int_0^1 (s''(t\lambda^T\omega_{in}g(X_i,\theta)) - s''(0))dt\omega_{in}^2 g(X_i,\theta)g(X_i,\theta)^T\right\|_2\\
&\leq ||\lambda||_2^2 s''(0)\left\| n^{-1}\sum_{i=1}^n \omega_{in}^2\{g(X_i,\theta)g(X_i,\theta)^T - g(X_i,\theta_0)g(X_i,\theta_0)^T\}\right\|_2\\
&+ ||\lambda||_2^2 s''(0)\left\| n^{-1}\sum_{i=1}^n(\omega_{in}^2 - 1)g(X_i,\theta_0)g(X_i,\theta_0)^T\right\|_2\\
&+ ||\lambda||_2^2 s''(0)\left\| n^{-1}\sum_{i=1}^n g(X_i,\theta_0)g(X_i,\theta_0)^T - \Omega\right\|_2\\
&+ ||\lambda||_2^2\left\| n^{-1}\sum_{i=1}^n \int_0^1 (s''(t\lambda^T\omega_{in}g(X_i,\theta)) - s''(0))dt\omega_{in}^2 g(X_i,\theta)g(X_i,\theta)^T\right\|_2\\
&\equiv ||\lambda||_2^2\{s''(0)(T_{1,n} + T_{2,n} + T_{3,n}) + T_{4,n}(\lambda)\}.
\end{aligned}$$

Regarding $T_{1,n}$, it is easy to see that

$$T_{1,n} = O_{\mathbf{P}_n^*(\cdot|Z^n)}\left(n^{-1}\sum_{i=1}^n \left\|g(X_i,\theta)g(X_i,\theta)^T - g(X_i,\theta_0)g(X_i,\theta_0)^T\right\|_2\right).$$

Hence, by Lemma C.3 and after some algebra it follows that it suffices to show that $n^{-1}\sum_{i=1}^n ||g(X_i,\theta) - g(X_i,\theta_0)||_2^2 = O_{\mathbf{P}}(\Delta_n^2 d(n))$ and

$$\begin{aligned}
n^{-1}\sum_{i=1}^n ||g(X_i,\theta) - g(X_i,\theta_0)||_2||g(X_i,\theta_0)||_2 \leq& \sqrt{n^{-1}\sum_{i=1}^n ||g(X_i,\theta) - g(X_i,\theta_0)||_2^2}\\
&\times\sqrt{n^{-1}\sum_{i=1}^n ||g(X_i,\theta_0)||_2^2}\\
=&O_{\mathbf{P}}(\Delta_n d(n)).
\end{aligned}$$

These two results follow because under Assumption 4.2(ii),

$$||g(X_i,\theta)-g(X_i,\theta_0)||_2 \leq \int_0^1 ||\nabla_\theta g(X_i,\theta_0+t(\theta-\theta_0))||_2 dt||\theta-\theta_0||_2$$
$$\leq \sup_{\theta\in\mathcal{N}} ||\nabla_\theta g(X_i,\theta)||_2 \Delta_n.$$

And under Assumption 4.2(iii) and the Markov inequality,

$$n^{-1}\sum_{i=1}^n \sup_{\theta\in\mathcal{N}} ||\nabla_\theta g(X_i,\theta)||_2 = O_{\mathbf{P}}(d(n)^{1/2}).$$

Finally, under Assumption 4.2(i) and the Markov inequality,

$$n^{-1}\sum_{i=1}^n ||g(X_i,\theta_0)||_2^2 = O_{\mathbf{P}}(d(n)).$$

Therefore $n||\lambda||_2^2 T_{1,n} = O_{\mathbf{P}_n^*(\cdot|Z^n)}\left(d(n)^2\Delta_n\right)$ wpa1-$\mathbf{P}$.

Regarding $T_{2,n}$ and $T_{3,n}$ it can be shown that are $O_{\mathbf{P}_n^*(\cdot|Z^n)}(d(n)/\sqrt{n})$ wpa1-$\mathbf{P}$ and $O_{\mathbf{P}}(d(n)/\sqrt{n})$ resp.; the calculations are analogous to those in the proof of Lemma A.6 in DIN and thus omitted. It thus follows, $n||\lambda||_2^2(T_{2,n}+T_{3,n}) = O_{\mathbf{P}_n^*(\cdot|Z^n)}(\frac{d(n)^2}{\sqrt{n}}) = o_{\mathbf{P}_n^*(\cdot|Z^n)}(\sqrt{d(n)})$ wpa1-$\mathbf{P}$, since $(d(n))^{3/2}/\sqrt{n} = o(1)$ by assumption.

Regarding the term $T_{4,n}$, since s'' is Lipschitz at 0, it follows that

$$|\int_0^1 (s''(t\lambda^T\omega_{in}g(X_i,\theta)) - s''(0))dt| \precsim |\lambda^T\omega_{in}g(X_i,\theta)|$$

for all $t\in[0,1]$. Therefore,

$$\begin{aligned} T_{4,n}(\lambda) \leq & \left\| n^{-1}\sum_{i=1}^n |\omega_{in}|^3 |\lambda^T g(X_i,\theta)| g(X_i,\theta)g(X_i,\theta)^T \right\|_2 \\ \precsim & \left\| n^{-1}\sum_{i=1}^n |\lambda^T g(X_i,\theta)| g(X_i,\theta)g(X_i,\theta)^T \right\|_2 \\ \leq & n^{-1}\sum_{i=1}^n |\lambda^T g(X_i,\theta)| \left\| g(X_i,\theta)g(X_i,\theta)^T \right\|_2 \\ \leq & \sqrt{\lambda^T n^{-1}\sum_{i=1}^n g(X_i,\theta)g(X_i,\theta)^T\lambda}\sqrt{n^{-1}\sum_{i=1}^n \|g(X_i,\theta)g(X_i,\theta)^T\|_2^2} \end{aligned}$$

where the second line follows from the weights being uniformly bounded. By analogous arguments to those in Lemma A.6 in DIN it can be shown that $\lambda_{max}(n^{-1}\sum_{i=1}^n g(X_i,\theta)g(X_i,\theta)^T) \leq C < \infty$ wpa1-$\mathbf{P}$ and thus $\sqrt{\lambda^T n^{-1}\sum_{i=1}^n g(X_i,\theta)g(X_i,\theta)^T\lambda} \precsim \sqrt{\frac{d(n)}{n}}$ wpa1-$\mathbf{P}$. It follows that

$n^{-1}\sum_{i=1}^{n}\left\|g(X_i,\theta)g(X_i,\theta)^T\right\|_2^2 \leq n^{-1}\sum_{i=1}^{n}\|g(X_i,\theta)\|_2^4 = O_{\mathbf{P}}(d(n)^2)$ by Assumption 4.2(i) (observe that $\theta \in \mathcal{N}$ eventually). Therefore, by Lemma C.3, $n||\lambda||_2^2 T_{4,n}(\lambda) = O_{\mathbf{P}_n^*(\cdot|Z^n)}(d(n)\sqrt{\frac{d(n)}{n}}d(n)) = O_{\mathbf{P}_n^*(\cdot|Z^n)}(\sqrt{d(n)}\frac{d(n)^2}{\sqrt{n}})$, and since $\frac{d(n)^2}{\sqrt{n}} = o(1)$ by assumption, equals $= o_{\mathbf{P}_n^*(\cdot|Z^n)}(\sqrt{d(n)})$. □

Proof of Lemma C.3. (1) We want to establish that for any $\epsilon > 0$, there exists a $M = M(\epsilon)$ and $N(\epsilon)$ such that

$$\mathbf{P}\left(\mathbf{P}_n^*\left(|W_n| \geq c_n M \mid Z^n\right) \leq \epsilon\right) \geq 1-\epsilon,\ \forall n \geq N(\epsilon).$$

This is equivalent to establishing that $\mathbf{P}\left(\mathbf{P}_n^*\left(|W_n| \geq c_n M \mid Z^n\right) \geq \epsilon\right) \leq \epsilon$. Let $A_n \equiv \{Z^n : \mathbf{P}_n^*\left(|W_n| \geq c_n M \mid Z^n\right) \geq \epsilon\}$ and $B_n \equiv \{X_n : |X_n| \leq \sqrt{M}c_n\}$. Given $X_n \in B_n$, then $\{W_n : |W_n| \geq c_n M\} \subseteq \{W_n : |W_n| \geq |X_n|\sqrt{M}\}$, therefore

$$\begin{aligned}\mathbf{P}(A_n) \leq \mathbf{P}_n(A_n \cap B_n) + \mathbf{P}(B_n^C) \leq & \mathbf{P}\left(\mathbf{P}_n^*\left(|W_n| \geq |X_n|\sqrt{M} \mid Z^n\right) \geq \epsilon\right)\\ & + \mathbf{P}(\{X_n : |X_n| \geq \sqrt{M}c_n\}).\end{aligned}$$

Since $W_n = O_{\mathbf{P}_n^*(\cdot|Z^n)}(|X_n|)$, the first term in the RHS can be made less than ϵ for sufficiently large M; similarly since $X_n = O_{\mathbf{P}}(c_n)$ the second term can also be made arbitrary small.

(2) The proof for this result is analogous to (1) and thus omitted. □

Appendix D: Proof of Proposition 7.1

Proof of Proposition 7.1. By assumption over V_n, it follows that $\mathbb{W}_n(P_n,\mathbf{P}) = n\left\|c(\theta_{P_n}) - c(\theta_{\mathbf{P}})\right\|_V^2(1+o_{\mathbf{P}}(1))$. Henceforth in the proof, we abuse notation and use $\mathbb{W}_n(P_n,\mathbf{P})$ to denote $n\left\|c(\theta_{P_n}) - c(\theta_{\mathbf{P}})\right\|_V^2$.

Let $a_n \equiv 1 + \sqrt{n}||c(\theta_{P_n}) - c(\theta_{\mathbf{P}})||_2$. Under the null $c(\theta_{\mathbf{P}}) = 0$, the representation 21 and our assumption over eigenvalues of V, it follows that $a_n \precsim 1 + \sqrt{\mathbb{W}_n(P_n,\mathbf{P})}$ and

$$\left\|\sqrt{n}(c(\theta_{P_n}) - c(\theta_{\mathbf{P}})) - \sqrt{n}E_{P_n}[\psi(X,\theta_{\mathbf{P}})]\right\|_V = o_{\mathbf{P}}(1 + \sqrt{\mathbb{W}_n(P_n,\mathbf{P})}).$$

Note that for any x and y, $||x-y|| = o(1+||y||)$, implies $|||x|| - ||y||| \leq o(1+||y||)$ and thus $||y||(1+o(1)) \leq ||x|| + o(1)$ and $||x|| \leq ||y||(1+o(1)) + o(1)$. Thus applying this to $x = \sqrt{n}E_{P_n}[\psi(X,\theta_{\mathbf{P}})]$ and $y = \sqrt{n}(c(\theta_{P_n}) - c(\theta_{\mathbf{P}}))$, it follows that

$$\sqrt{\mathbb{W}_n(P_n,\mathbf{P})}(1+o_{\mathbf{P}}(1)) \leq \left\|\sqrt{n}E_{P_n}[\psi(X,\theta_{\mathbf{P}})]\right\|_V + o_{\mathbf{P}}(1)$$

and

$$\sqrt{\mathbb{W}_n(P_n,\mathbf{P})}(1+o_{\mathbf{P}}(1)) \geq \left\|\sqrt{n}E_{P_n}[\psi(X,\theta_{\mathbf{P}})]\right\|_V - o_{\mathbf{P}}(1).$$

Therefore, after simple algebra and squaring at both sides,

$$\mathbb{W}_n(P_n,\mathbf{P})(1+o_{\mathbf{P}}(1)) = \left\|\sqrt{n}E_{P_n}[\psi(X,\theta_{\mathbf{P}})]\right\|_V^2 + o_{\mathbf{P}}(1 + \left\|\sqrt{n}E_{P_n}[\psi(X,\theta_{\mathbf{P}})]\right\|_V).$$

It thus remains to show that $\|\sqrt{n}E_{P_n}[\psi(X,\theta_{\mathbf{P}})]\|_V \precsim \|\sqrt{n}E_{P_n}[\psi(X,\theta_{\mathbf{P}})]\|_2 = O_{\mathbf{P}}(\sqrt{d(n)})$. Note that $E[\|E_{P_n}[\psi(X,\theta_{\mathbf{P}})]\|_2^2] = \sum_{j=1}^{d(n)} E\left[\left(E_{P_n}[\psi_{[j]}(X,\theta_{\mathbf{P}})]\right)^2\right]$, and

$$\begin{aligned} E\left[\left(n^{-1}\sum_{i=1}^{n}\psi_{[j]}(X_i,\theta_{\mathbf{P}})\right)^2\right] =& E\left[n^{-2}\sum_{i=1}^{n}\left(\psi_{[j]}(X_i,\theta_{\mathbf{P}})\right)^2\right] \\ \leq& n^{-1}E\left[|\psi_{[j]}(X,\theta_{\mathbf{P}})|^2\right] \\ \precsim& n^{-1} \end{aligned}$$

where the first equality follows because for $i \neq l$

$$E\left[\psi_{[j]}(X_i,\theta_{\mathbf{P}})\psi_{[j]}(X_l,\theta_{\mathbf{P}})\right] = E\left[\psi_{[j]}(X_i,\theta_{\mathbf{P}})\right]E\left[\psi_{[j]}(X_l,\theta_{\mathbf{P}})\right] = 0.$$

Thus, by the Markov inequality the result follows.

The proof of the representation for $\mathbb{W}_n(P_n^*, P_n)$ is analogous and omitted; it is worth pointing out, however, that for this the null hypothesis is not imposed. ☐

References

Arellano, M. and Bond, S. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, 58(2):277–297, 1991.

Boucheron, S., Lugosi, G., and Massart, P. *Concentration Inequalities.* Oxford Univ. Press, 2013. MR3185193

Chatterjee, S. A generalization of the Lindeberg principle. *The Annals of Probability*, 34(6):2061–2076, 2006. MR2294976

Chatterjee, S. and Meckes, E. Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.*, 4:257–283, 2008. MR2453473

Chen, X. and Pouzo, D. Sieve Wald and QLR inferences on semi/nonparametric conditional moment models. *Econometrica*, 83(3):1013–1079, May 2015. MR3357484

Chen, X. Large sample sieve estimation of semi-nonparametric models. Part B of *Handbook of Econometrics*, Vol. 6, chapter 76, pages 5549–5632. Elsevier, 2007.

Chernozhukov, V., Chetverikov, D., and Kato, K. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41:2786–2819, 2013a. MR3161448

Chernozhukov, V., Chetverikov, D., and Kato, K. Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *arXiv:1301.4807*, 2013b. MR3350040

de Jong, R. M. and Bierens, H. J. On the limit behavior of a Chi-Square type test if the number of conditional moments tested approaches infinity. *Econometric Theory*, 10(01):70–90, 1994. MR1279691

Donald, S., Imbens, G., and Newey, W. Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics*, 117:55–93, 2003. MR2002302

Efron, B. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):1–26, 1979. MR0515681

Feller, W. *An Introduction to Probability Theory and its Applications*, Vol. II. Wiley, 2nd edition, 1971. MR0270403

Hall, A. *Generalized Method of Moments*. Advanced Texts in Econometrics Series. Oxford University Press, 2005. MR2135106

Hall, P. Methodology and theory for the bootstrap. In Engle, R. F. and McFadden, D., editors, *Handbook of Econometrics*, Vol. 4, chapter 39, pages 2341–2381. Elsevier, 1 edition, 1986. MR1315974

Hansen, L. P. Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054, 1982. MR0666123

Hansen, L. P., Heaton, J., and Yaron, A. Finite-sample properties of some alternative GMM estimators. *Journal of Business and Economic Statistics*, 14(3):262–280, 1996.

He, X. and Shao, Q.-M. On parameters of increasing dimensions. *Journal of Multivariate Analysis*, 73:120–135, 2000. MR1766124

Hjort, N., McKeague, I., and Van Keilegom, I. Extending the scope of empirical likelihood. *The Annals of Statistics*, 37(3):1079–1111, 2009. MR2509068

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

Imbens, G. W. Generalized method of moments and empirical likelihood. *Journal of Business and Economic Statistics*, 20(4):493–506, 2002. MR1973800

Imbens, G. W., Spady, R. H., and Johnson, P. Information-theoretic approaches to inference in moment condition models. *Econometrica*, 66(2):333–358, 1998. MR1612246

Johnson, W. B., Schechtman, G., and Zinn, J. Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *The Annals of Probability*, 13(1):234–253, 1985. MR0770640

Kitamura, Y. and Stutzer, M. An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, 65(4):861–874, 1997. MR1458431

Koenker, R. and Machado, J. A. F. GMM inference when the number of moment conditions is large. *Journal of Econometrics*, 93(2):327–344, 1999. MR1721103

Ma, S. and Kosorok, M. R. Robust semiparametric m-estimation and the weighted bootstrap. *Journal of Multivariate Analysis*, 96(1):190–217, 2005. MR2202406

Mammen, E. Asymptotics with increasing dimension for robust regression with applications to the bootstrap. *The Annals of Statistics*, 17(61):382–400, 1989. MR0981457

Mammen, E. Bootstrap and Wild bootstrap for high dimensional linear models. *The Annals of Statistics*, 21(1):255–285, 1993. MR1212176

Murphy, S. and Van der Vaart, A. On profile likelihood. 95:449–485, 2000. MR1803168

Newey, W. K. and McFadden, D. arge sample estimation and hypothesis testing. *Handbook of Econometrics*, Vol. 4, chapter 36, pages 2111–2245. Elsevier, 1994. MR1315971

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

Owen, A. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249, 1988. MR0946049

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

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

Portnoy, S. Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *The Annals of Statistics*, 16(1):356–366, 1988. MR0924876

Praestgaard, J. Bootstrap with general weights and multiplier central limit theorems. *Technical Report 195, Dept. Statistics, Univ. Washington.*, 1990.

Praestgaard, J. and Wellner, J. A. Exchangeably weighted bootstraps of the general empirical process. *The Annals of Probability*, 21(4):2053–2086, 1993. MR1245301

Radulovic, D. Can we bootstrap even if CLT fails? *Journal of Theoretical Probability*, 11(3):813–830, 1998. MR1633407

Raic, M. A multivariate CLT for decomposable random vectors with finite second moments. *Journal of Theoretical Probability*, 17(3):573–603, 2004. MR2091552

Rollin, A. Stein's method in high dimensions with applications. *arXiv:1101.4454*, 2013. MR3088380

Slepian, D. The one-sided barrier problem for Gaussian noise. *Bell System Technical Journal*, 41(2):463–501, 1962. MR0133183

Smith, R. J. Alternative semi-parametric likelihood approaches to generalized method of moments estimation. *The Economic Journal*, 107(441):503–519, 1997.

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

Stein, C. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 9(6):1135–1151, 1981. MR0630098

Tao, T. and Vu, V. Random matrices: The Four Moment Theorem for Wigner ensembles. *arXiv:1112.1976*, 2011. MR3380700

Van der Vaart, A. and Wellner, J. *Weak Convergence and Empirical Processes*. Springer, 1996. MR1385671

Van der Vaart, A. *Asymptotic Statistics*. Cambridge University Press, 2000. MR1652247

Vershynin, R. Introduction to the non-asymptotic analysis of random matrices. In *Compressed sensing*, pages 210–268. Cambridge Univ. Press, 2012a. MR2963170

VERSHYNIN, R. How close is the sample covariance matrix to the actual covariance matrix? *Journal of Theoretical Probability*, 25:655–686, 2012b. MR2956207

WASSERMAN, L. Stein's method and the bootstrap in low and high dimensions: A tutorial. *Working Paper*, 2014.

XU, M., ZHANG, D., and WU, W. B. L^2 asymptotics for high-dimensional data. *arXiv:1405.7244*, 2014.

ZHANG, X. and CHENG, G. Bootstrapping high dimensional time series. *arXiv:1406.1037v2*, 2014.