SUPPLEMENTAL MATERIAL FOR "ESTIMATING SUFFICIENT REDUCTIONS OF THE PREDICTORS IN ABUNDANT HIGH-DIMENSIONAL REGRESSIONS"

By R. Dennis Cook[‡], Liliana Forzani and Adam J. Rothman[§]

University of Minnesota and Instituto de Matemática Aplicada del Litoral

APPENDIX A: INTRODUCTION

The referred equations in this appendix are labeled as (A.1), (A.2) and so on, whereas labels such as (1.1), (2.1), etc. refer to equations the main text. The sections in the appendix are labeled alphabetically. There are eight figures in the main article and so the first figure in supplement is number 9. The matrices $\mathbf{A}_{(\cdot)}$ and $\mathbf{T}_{(\cdot)}$ are used as transient notation to reduce the size of lengthy expressions. Results are arranged by section in the order that the appear in the main text and are ended with a $\square$. Preliminary results that will be used in our proofs are given at the beginning of the first section in which they are needed. Some of these (eg. Lemma F.2) may be of interest on their own. Proofs that seem straightforward or correspond to results taken from the literature are omitted.

APPENDIX B: DERIVATION OF THE ESTIMATORS

Recall from the definitions given in Section 3.1 that $\hat{\boldsymbol{\mu}} = \bar{\mathbf{X}}$, $\mathbb{Z} = \mathbb{X} - \mathbf{1}_n\hat{\boldsymbol{\mu}}^T$, $\widehat{\mathbf{B}} = \mathbb{Z}^T\mathbb{F}(\mathbb{F}^T\mathbb{F})^{-1} \in \mathbb{R}^{p\times r}$ and that the columns of $\widehat{\mathbf{V}}_d \in \mathbb{R}^{r\times d}$ are the first d eigenvectors of $\widehat{\mathbf{K}} = \boldsymbol{\Phi}_n^{1/2}\widehat{\mathbf{B}}^T\widehat{\mathbf{W}}\widehat{\mathbf{B}}\boldsymbol{\Phi}_n^{1/2} \in \mathbb{R}^{r\times r}$. Then we need to obtain $(\hat{\mathbf{b}}, \widehat{\boldsymbol{\Gamma}}) = \arg\min \operatorname{tr}\{\mathbf{J}(\mathbf{b}, \boldsymbol{\Gamma})\}$, where $\mathbf{J}(\mathbf{b}, \boldsymbol{\Gamma}) = (\mathbb{Z} - \mathbb{F}\mathbf{b}^T\boldsymbol{\Gamma}^T)\widehat{\mathbf{W}}(\mathbb{Z} - \mathbb{F}\mathbf{b}^T\boldsymbol{\Gamma}^T)^T$. Substituting the decomposition $\mathbb{Z} = \mathbf{P}_{\mathbb{F}}\mathbb{Z} + \mathbf{Q}_{\mathbb{F}}\mathbb{Z}$ into the objective function to be minimized, we have

$$\begin{aligned}\operatorname{tr}\{\mathbf{J}(\mathbf{b}, \boldsymbol{\Gamma})\} &= \operatorname{tr}[\{(\mathbf{P}_{\mathbb{F}}\mathbb{Z} - \mathbb{F}\mathbf{b}^T\boldsymbol{\Gamma}^T) + \mathbf{Q}_{\mathbb{F}}\mathbb{Z}\}\widehat{\mathbf{W}}\{(\mathbf{P}_{\mathbb{F}}\mathbb{Z} - \mathbb{F}\mathbf{b}^T\boldsymbol{\Gamma}^T) + \mathbf{Q}_{\mathbb{F}}\mathbb{Z}\}^T] \\ &= n\operatorname{tr}\{\boldsymbol{\Phi}_n(\widehat{\mathbf{B}}^T - \mathbf{b}^T\boldsymbol{\Gamma}^T)\widehat{\mathbf{W}}(\widehat{\mathbf{B}}^T - \mathbf{b}^T\boldsymbol{\Gamma}^T)^T\} + \operatorname{tr}\{\mathbf{Q}_{\mathbb{F}}\mathbb{Z}\widehat{\mathbf{W}}\mathbb{Z}^T\mathbf{Q}_{\mathbb{F}}\}.\end{aligned}$$

It follows that we can equivalently use $(\hat{\mathbf{b}}, \widehat{\boldsymbol{\Gamma}}) = \arg\min L(\mathbf{b}, \boldsymbol{\Gamma})$, where

$$\begin{aligned}L(\mathbf{b}, \boldsymbol{\Gamma}) &= n\operatorname{tr}\{\boldsymbol{\Phi}_n(\widehat{\mathbf{B}} - \boldsymbol{\Gamma}\mathbf{b})^T\widehat{\mathbf{W}}(\widehat{\mathbf{B}} - \boldsymbol{\Gamma}\mathbf{b})\} \\ &= n\operatorname{vec}^T(\widehat{\mathbf{B}} - \boldsymbol{\Gamma}\mathbf{b})(\boldsymbol{\Phi}_n \otimes \widehat{\mathbf{W}})\operatorname{vec}(\widehat{\mathbf{B}} - \boldsymbol{\Gamma}\mathbf{b}).\end{aligned}$$

[‡]Corresponding author. Supported by NSF Grants DMS-1007547 and DMS-1105650.
[§]Suppored by NSF Grant DMS-1105650.

If $\widehat{\mathbf{W}}$ is a consistent estimator of $\boldsymbol{\Delta}^{-1}$, $\widehat{\boldsymbol{\Gamma}}\hat{\mathbf{b}}$ will then be asymptotically efficient as $n \to \infty$ with p fixed (cf. Cook and Ni, 2005). However, as we will see later, the choice of a reasonable weight matrix can be much more intricate when p and n are both large.

Straightforward calculation will show that L is minimized for fixed $\mathbf{b}$ by the value $\boldsymbol{\Gamma} = \widehat{\mathbf{B}}\boldsymbol{\Phi}_n\mathbf{b}^T(\mathbf{b}\boldsymbol{\Phi}_n\mathbf{b}^T)^{-1} = \widehat{\mathbf{B}}\boldsymbol{\Phi}_n\mathbf{b}^T$, where the second equality follows from the constraint $\mathbf{b}\boldsymbol{\Phi}_n\mathbf{b}^T = \mathbf{I}_d$. It follows that $\boldsymbol{\Gamma}\mathbf{b}$ must satisfy the relationship $\boldsymbol{\Gamma}\mathbf{b} = \widehat{\mathbf{B}}\mathbf{P}^T_{\mathbf{b}^T(\boldsymbol{\Phi}_n)} = \widehat{\mathbf{B}}\boldsymbol{\Phi}_n^{1/2}\mathbf{P}_{\boldsymbol{\Phi}_n^{1/2}\mathbf{b}^T}\boldsymbol{\Phi}_n^{-1/2}$ and consequently $\boldsymbol{\Gamma}\mathbf{b}$ depends on $\mathbf{b}$ only through $\mathrm{span}(\boldsymbol{\Phi}_n^{1/2}\mathbf{b}^T)$.

Substituting $\boldsymbol{\Gamma}\mathbf{b} = \widehat{\mathbf{B}}\mathbf{P}^T_{\mathbf{b}^T(\boldsymbol{\Phi}_n)}$ into $L(\mathbf{b}, \boldsymbol{\Gamma})$, we next need to minimize

$$\begin{aligned} L_1(\mathbf{b}) &= n\,\mathrm{tr}\{\boldsymbol{\Phi}_n(\mathbf{I} - \mathbf{P}_{\mathbf{b}^T(\boldsymbol{\Phi}_n)})\widehat{\mathbf{B}}^T\widehat{\mathbf{W}}\widehat{\mathbf{B}}(\mathbf{I} - \mathbf{P}^T_{\mathbf{b}^T(\boldsymbol{\Phi}_n)})\} \\ &= n\,\mathrm{tr}\{\mathbf{Q}_{\boldsymbol{\Phi}_n^{1/2}\mathbf{b}^T}\boldsymbol{\Phi}_n^{1/2}\widehat{\mathbf{B}}^T\widehat{\mathbf{W}}\widehat{\mathbf{B}}\boldsymbol{\Phi}_n^{1/2}\mathbf{Q}_{\boldsymbol{\Phi}_n^{1/2}\mathbf{b}^T}\}. \end{aligned}$$

Then $\mathrm{span}(\widehat{\mathbf{V}}_d)$ minimizes L_1 over $\mathrm{span}(\boldsymbol{\Phi}_n^{1/2}\mathbf{b}^T)$ and the choice $\boldsymbol{\Phi}_n^{1/2}\hat{\mathbf{b}}^T = \widehat{\mathbf{V}}_d$ produces a $\widehat{\boldsymbol{\Gamma}}$ that satisfies the constraint on $\boldsymbol{\Gamma}$. This then leads to the estimators stated in Section 3.1.

APPENDIX C: ADDITIONAL SIMULATION RESULTS

C.1. Nonnormal errors and exponential correlations. To assess the performance of the reduction estimators when the errors are not Gaussian, we generated $\boldsymbol{\Delta}^{-1/2}\boldsymbol{\varepsilon}$ as independent observations from $N(0, 1)$, Uniform $(0, 1)$, T_5 and χ^2_5 distributions, each centered and scaled to have mean 0 and variance 1. Using these error distributions, which satisfy conditions W.1 and W.2 for all four reduction estimators, we repeated the simulation of Section 7.3.1 with $\theta = 0.5$. The results for $\widehat{\mathbf{R}}_{\widehat{\boldsymbol{\Delta}}_\lambda}$, $\widehat{\mathbf{R}}_{\widehat{\boldsymbol{\Delta}}}$ and $\widehat{\mathbf{R}}_{\mathrm{diag}}$ are shown in Figure 9a-c. Although each figure contains four curves, one for each distribution, they are not labeled because they are so close. The error distributions considered here are sub-Gaussian except for χ^2_5 and T_5, although this did not affect the performance of $\widehat{\mathbf{R}}_{\mathrm{diag}}$ and $\widehat{\mathbf{R}}_{\mathrm{spice}}$ for which our convergence bounds required sub-Gaussian errors.

C.2. Constant correlations. We repeated the simulations of Sections 7.3.1, 7.3.2 and 7.4 using constant correlations instead of exponential correlations. In this setting, $\boldsymbol{\Delta}$ is generated as $\mathbf{D}^{1/2}\boldsymbol{\Theta}\mathbf{D}^{1/2}$ where $\mathbf{D}$ is a diagonal matrix with diagonal elements sampled from a Uniform $(1, 101)$ distribution and $\boldsymbol{\Theta} = (1-\theta)\mathbf{I}_p + \theta\mathbf{1}_p\mathbf{1}_p^T = \mathbf{Q}_{\mathbf{1}}(1-\theta) + (1+\theta(p-1))\mathbf{P}_{\mathbf{1}}$ where the parameter $0 < \theta < 1$ and $\mathbf{1}_p$ is a $p \times 1$ vector of 1's.

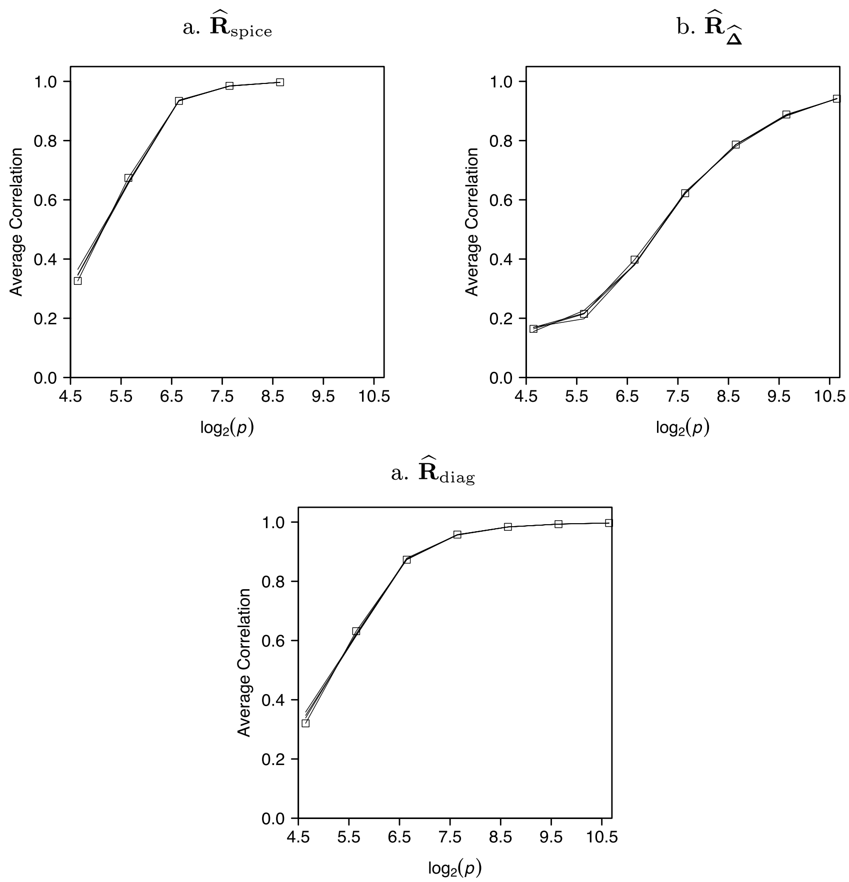

FIG 9. *Estimators of* $\mathbf{R}$ *when errors are Gaussian, Uniform,* T_5*, and* χ^2_5*, exponential error correlations:* $\theta_{ij} = 0.5^{|i-j|}$*, and* $n = p/2$*. Curve labels are omitted because they are indistinguishable.*

For all four reduction estimators, $h \asymp p$ implying $\kappa = n^{-1/2}$. As was the case for the exponential correlation model, as θ increases toward 1, the expected signal strength increases for $\widehat{\mathbf{R}}_{\text{diag}}, \widehat{\mathbf{R}}_{\widehat{\boldsymbol{\Delta}}}$, and $\widehat{\mathbf{R}}_{\text{spice}}$ since $\mathrm{E}_{\text{sim}}[\operatorname{tr}(\mathbf{W})] = \operatorname{tr}(\boldsymbol{\Theta}^{-1}) \log(101)/100$. More explicitly, a closed-form solution is available for $\boldsymbol{\Theta}^{-1}$ which allows us to compute,

$$\mathrm{E}_{\text{sim}}[\operatorname{tr}(\mathbf{W})] = p\left(\frac{1}{1-\theta} - \frac{\theta}{(1-\theta)[1+(p-1)\theta]}\right)\frac{\log 101}{100}.$$

Unlike the exponential correlation model, for constant correlations we have $\|\boldsymbol{\Delta}\| = O(p)$.

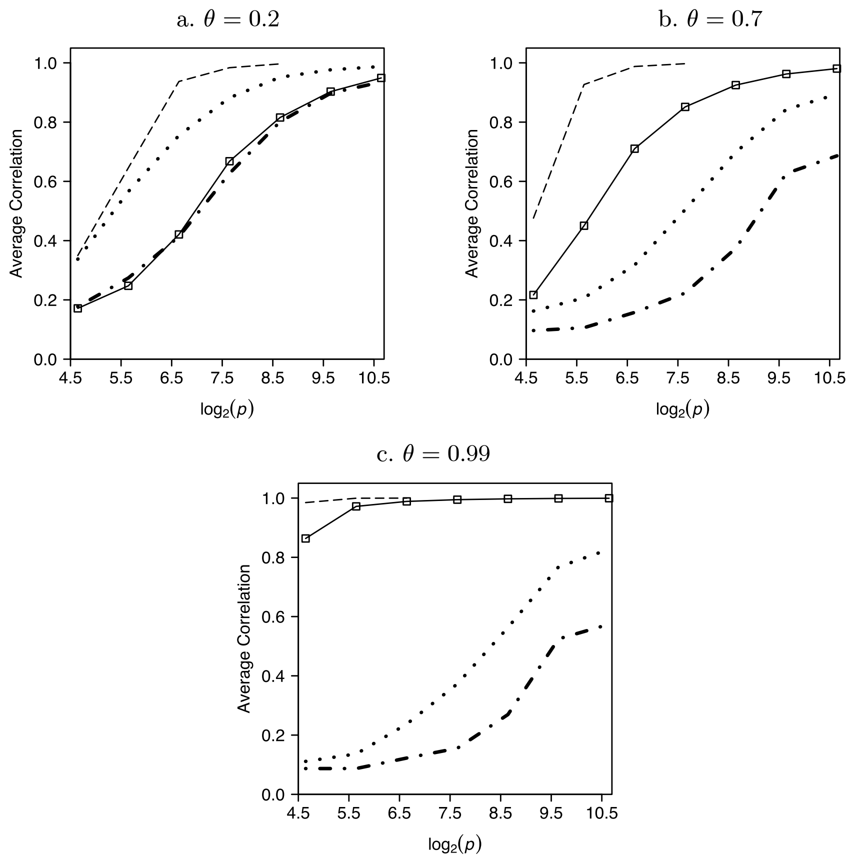

FIG 10. *Comparison of the four estimators of* $\mathbf{R}$*:* $\widehat{\mathbf{R}}_{\text{spice}}$ *(dashes),* $\widehat{\mathbf{R}}_{\widehat{\boldsymbol{\Delta}}}$ *(solid),* $\widehat{\mathbf{R}}_{\text{diag}}$ *(dots), and* $\widehat{\mathbf{R}}_{\mathbf{I}}$ *(dash dot), with constant error correlations, and* $n = p/2$*.*

C.2.1. $n = p/2$ *and* $p \to \infty$. In this setting, which corresponds to Section 7.3.1, n and p grow together, with $n = p/2$. The reduction estimators $\widehat{\mathbf{R}}_{\mathbf{I}}$ and $\widehat{\mathbf{R}}_{\text{diag}}$ are consistent. This follows since $\|\boldsymbol{\rho}\| = O(p)$ and $\|\boldsymbol{\rho}\|_F = O(p)$ implying that $\psi = \kappa = n^{-1/2}$ and as $n, p \to \infty$, $\boldsymbol{\nu} \to 0$ in probability. For example, for $\widehat{\mathbf{R}}_{\mathbf{I}}$ we have

$$
\begin{aligned}
\mathbf{G}_h^{1/2}\text{var}(\boldsymbol{\nu})\mathbf{G}_h^{1/2} &\leq p^{-1}\boldsymbol{\gamma}^T\boldsymbol{\rho}\boldsymbol{\gamma} \leq (101/p)\boldsymbol{\gamma}^T\boldsymbol{\Theta}\boldsymbol{\gamma} \\
&= (101 \times \theta/p)\boldsymbol{\gamma}^T\mathbf{P}_{\mathbf{1}}\boldsymbol{\gamma} + O(p^{-1}) \\
&= 101 \times \theta\frac{(\mathbf{1}_p^T\boldsymbol{\Gamma})^2}{p\boldsymbol{\Gamma}^T\boldsymbol{\Gamma}} + O(p^{-1}) = O(p^{-1})
\end{aligned}
$$

An analogous argument can be made for $\widehat{\mathbf{R}}_{\text{diag}}$.

For this model, our theory only guarantees consistency for $\widehat{\mathbf{R}}_{\text{spice}}$ if p

is bounded as $n \to \infty$. Since $\mathbf{\Delta}^{-1}$ is not sparse, $s \asymp p^2$ and $\omega_{\text{spice}} = n^{-1/2} p \log^{1/2} p$. In addition, if $p \to \infty$, the eigenvalues of $\mathbf{\Delta}$ are not uniformly bounded since $\|\mathbf{\Delta}\| = O(p)$, invalidating the bound on $\|\mathbf{S}\|$ for SPICE.

The results for p and n growing with $n = p/2$ are shown in Figure 10a-c. The reduction estimators $\widehat{\mathbf{R}}_{\text{spice}}$ and $\widehat{\mathbf{R}}_{\widehat{\mathbf{\Delta}}}$ appear to be converging to the population reduction as n grows, even though our theory fails to guarantee their consistency in this setting. Although $p > n$, $\widehat{\mathbf{R}}_{\widehat{\mathbf{\Delta}}}$ outperforms $\widehat{\mathbf{R}}_{\text{diag}}$ and $\widehat{\mathbf{R}}_{\mathbf{I}}$ when $\theta = 0.7, 0.99$.

Results for $\widehat{\mathbf{R}}_{\text{spice}}$ were computed up to $p = 400$ when $\theta = 0.2, 0.7$ and up to $p = 100$ when $\theta = 0.99$ due to intractable computation time required for the glasso algorithm. Similar to the exponential correlations model, in the scenarios when $\widehat{\mathbf{R}}_{\text{spice}}$ was computed, it considerably outperformed the other reduction estimators, even though $\mathbf{\Delta}^{-1}$ is not sparse.

C.2.2. *$p = 100$ and $n \to \infty$.* In this setting, which corresponds to Section 7.3.2, we fix $p = 100$ and let n grow. Our theory guarantees that $\widehat{\mathbf{R}}_{\text{spice}}$ and $\widehat{\mathbf{R}}_{\widehat{\mathbf{\Delta}}}$ are $\sqrt{n}$-consistent. As in the exponential correlation model, $\widehat{\mathbf{R}}_{\mathbf{I}}$ and $\widehat{\mathbf{R}}_{\text{diag}}$ are inconsistent, since p is bounded and $\text{span}(\boldsymbol{\gamma})$ is not a reducing subspace of $\boldsymbol{\rho}$, implying $\boldsymbol{\nu}$ fails to vanish.

The results for $p = 100$ and n growing are illustrated in Figure 11a-c. As our theory suggests, $\widehat{\mathbf{R}}_{\text{spice}}$ and $\widehat{\mathbf{R}}_{\widehat{\mathbf{\Delta}}}$ both appear to converge to the population reduction as n grows. $\widehat{\mathbf{R}}_{\text{spice}}$ outperforms the other reduction estimators for the values of θ considered. As expected, the reduction estimator $\widehat{\mathbf{R}}_{\widehat{\mathbf{\Delta}}}$ outperforms $\widehat{\mathbf{R}}_{\text{diag}}$ and $\widehat{\mathbf{R}}_{\mathbf{I}}$ for all values of θ as n grew sufficiently large. This pattern also held when n was small and θ was large.

C.3. Results for $\boldsymbol{\xi} \neq \beta\mathbf{f}$. The data here were generated as described in Section 7.4, except we again used the constant error correlations instead of the exponential correlations. The results are shown in Figure 12. The conclusions are similar to those given in Section 7.4, but here we see that $\widehat{\mathbf{R}}_{\text{diag}}$ and $\widehat{\mathbf{R}}_{\mathbf{I}}$ are more influenced by the choice of $\mathbf{f}$.

APPENDIX D: PROOF OF LEMMA 4.1

Let $\boldsymbol{\gamma} = \mathbf{W}^{1/2}\mathbf{\Gamma}(\mathbf{\Gamma}^T\mathbf{W}\mathbf{\Gamma})^{-1/2}$, so $\boldsymbol{\gamma}$ is a semi-orthogonal matrix. Then

$$\begin{aligned}
\|\text{var}(\mathbf{R}(\mathbf{X}))\| &= \|(\mathbf{\Gamma}^T\mathbf{\Delta}^{-1}\mathbf{\Gamma})^{-1}\| = \|(\mathbf{\Gamma}^T\mathbf{W}^{1/2}\boldsymbol{\rho}^{-1}\mathbf{W}^{1/2}\mathbf{\Gamma})^{-1}\| \\
&= \|(\mathbf{\Gamma}^T\mathbf{W}\mathbf{\Gamma})^{-1/2}(\boldsymbol{\gamma}^T\boldsymbol{\rho}^{-1}\boldsymbol{\gamma})^{-1}(\mathbf{\Gamma}^T\mathbf{W}\mathbf{\Gamma})^{-1/2}\| \\
&\leq \|(\mathbf{\Gamma}^T\mathbf{W}\mathbf{\Gamma})^{-1}\| \times \|(\boldsymbol{\gamma}^T\boldsymbol{\rho}^{-1}\boldsymbol{\gamma})^{-1}\| \\
&\leq \|(\mathbf{\Gamma}^T\mathbf{W}\mathbf{\Gamma})^{-1}\| \times \|\boldsymbol{\rho}\| = \|\mathbf{G}_h^{-1}\| \times \|\boldsymbol{\rho}\|/h.
\end{aligned}$$

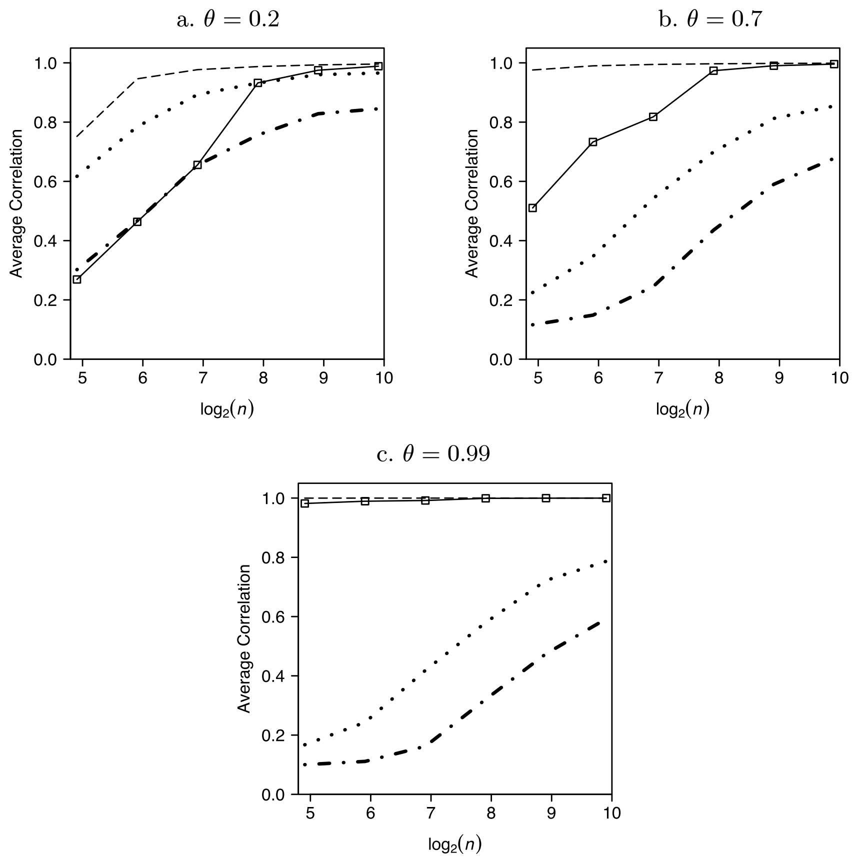

Fig 11. *Comparison of the four estimators of* $\mathbf{R}$*:* $\widehat{\mathbf{R}}_{\text{spice}}$ *(dashes),* $\widehat{\mathbf{R}}_{\widehat{\boldsymbol{\Delta}}}$ *(solid),* $\widehat{\mathbf{R}}_{\text{diag}}$ *(dots), and* $\widehat{\mathbf{R}}_{\mathbf{I}}$ *(dash dot), with constant error correlations, and* $p = 100$*.*

The conclusion follows because $\mathbf{G}_h$ converges. □

APPENDIX E: PROOF OF LEMMA 4.2

We prove this lemma by its parts.

(i). Magnus and Neudecker (1979, cor. 4.1) show that $\operatorname{var}(\boldsymbol{\varepsilon}^T \mathbf{W} \boldsymbol{\varepsilon}) = 2\operatorname{tr}(\boldsymbol{\rho}^2)$. The conclusion follows because $\operatorname{tr}(\boldsymbol{\rho}^2) \le \operatorname{tr}^2(\boldsymbol{\rho})$, which is $O(p^2)$ under condition W.1.

(ii). When $\mathbf{W} = \boldsymbol{\Delta}^{-1}$, $\operatorname{var}(\boldsymbol{\varepsilon}^T \mathbf{W} \boldsymbol{\varepsilon}) = \operatorname{var}(\boldsymbol{\epsilon}^T \boldsymbol{\epsilon}) = \sum_{i,j} \mathrm{E}(\epsilon_i^2 \epsilon_j^2) - p^2 \le (\phi - 1)p^2$.

(iii). Under either symmetry or independence of the elements of $\boldsymbol{\epsilon}$, we have $\mathrm{E}(\epsilon_i \epsilon_j \epsilon_k \epsilon_l) = 0$ if at least one index is distinct (not equal to any of the

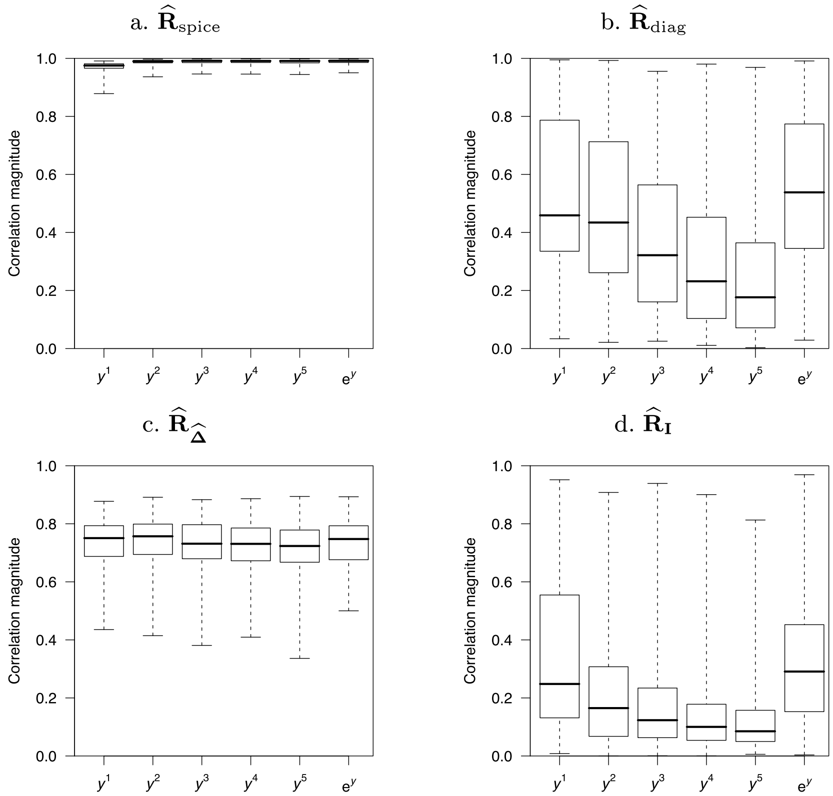

FIG 12. *Estimators of* $\mathbf{R}$ *when* $\mathbf{f}$ *is misspecified, using constant error correlations:* $\theta_{ij} = 0.7$, $n = 50$, *and* $p = 100$. *Boxplots are labeled by the highest order term in* $\mathbf{f}$.

other three indices). To reach the conclusion under symmetry it is sufficient to consider a pair (ϵ_i, ϵ_j). Letting g denote a generic density, $g(-\epsilon_i, \epsilon_j) = g(\epsilon_j | -\epsilon_i)g(-\epsilon_i) = g(\epsilon_j|\epsilon_i)g(\epsilon_i) = g(\epsilon_i, \epsilon_j)$.

If no indices are distinct then they must be equal in pairs, leading to the following possibilities (a) $\mathrm{E}(\epsilon_i\epsilon_i\epsilon_i\epsilon_i) = \mathrm{E}(\epsilon_i^4) \equiv \phi_{ii}$, and, for $i \neq j$, (b) $\mathrm{E}(\epsilon_i\epsilon_i\epsilon_j\epsilon_j) = \mathrm{E}(\epsilon_i\epsilon_j\epsilon_i\epsilon_j) = \mathrm{E}(\epsilon_i\epsilon_j\epsilon_j\epsilon_i) = \mathrm{E}(\epsilon_i^2\epsilon_j^2) = \phi_{ij}$. Let $\mathbf{u}_i \in \mathbb{R}^p$ have a 1 in position i and 0's elsewhere, and let $\mathbf{U}_{ij} = \mathbf{u}_i \mathbf{u}_j^T$. Then we have

$$\begin{aligned}
\mathrm{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T \otimes \boldsymbol{\epsilon}\boldsymbol{\epsilon}^T) &= \sum_{ijkl}^{p} \mathrm{E}(\epsilon_i \epsilon_j \epsilon_k \epsilon_l) \mathbf{U}_{ij} \otimes \mathbf{U}_{kl} \\
&\leq \phi \sum_{i \neq j}^{p} \{\mathbf{U}_{ii} \otimes \mathbf{U}_{jj} + \mathbf{U}_{ij} \otimes \mathbf{U}_{ij} + \mathbf{U}_{ij} \otimes \mathbf{U}_{ji}\} + \phi \sum_{i}^{p} \mathbf{U}_{ii} \otimes \mathbf{U}_{ii} \\
&\leq \phi \sum_{i,j}^{p} \{\mathbf{U}_{ii} \otimes \mathbf{U}_{jj} + \mathbf{U}_{ij} \otimes \mathbf{U}_{ij} + \mathbf{U}_{ij} \otimes \mathbf{U}_{ji}\} \\
&= \phi \{\mathbf{I}_{p^2} + \mathbf{C}_{p^2} + \mathrm{vec}(\mathbf{I}_p)\mathrm{vec}^T(\mathbf{I}_p)\}.
\end{aligned}$$

Letting $\mathbf{A} = \boldsymbol{\Delta}^{1/2} \mathbf{W} \boldsymbol{\Delta}^{1/2}$ and combining this bound with the previous results we have

$$\begin{aligned}
\mathrm{E}(\boldsymbol{\varepsilon}^T \mathbf{W} \boldsymbol{\varepsilon})^2 &= \mathrm{tr}\{(\mathbf{A} \otimes \mathbf{A}) \mathrm{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T \otimes \boldsymbol{\epsilon}\boldsymbol{\epsilon}^T)\} \\
&\leq \phi\, \mathrm{tr}\{(\mathbf{A} \otimes \mathbf{A})(\mathbf{I}_{p^2} + \mathbf{C}_{p^2} + \mathrm{vec}(\mathbf{I}_p)\mathrm{vec}^T(\mathbf{I}_p)\} \\
&= \phi\{\, \mathrm{tr}^2(\mathbf{A}) + 2\, \mathrm{tr}(\mathbf{A}^2)\} = \phi\{\, \mathrm{tr}^2(\boldsymbol{\rho}) + 2\, \mathrm{tr}(\boldsymbol{\rho}^2)\}.
\end{aligned}$$

□

APPENDIX F: PROOF OF PROPOSITION 5.1

F.1. Preliminary results. The following four lemmas will facilitate our proofs of Proposition 5.1. The first is well-known. The second, Lemma F.2, provides moments of linear and quadratic forms in $\boldsymbol{\varepsilon}$. The third shows how to go from the asymptotics when $\widehat{\mathbf{W}} = \mathbf{W}$ to the asymptotics for $\widehat{\mathbf{W}}$. The fourth gives order for terms that include $\mathbf{e}$ and $\widehat{\mathbf{W}}$, where $\mathbf{e} \in \mathbb{R}^{n \times p}$ was defined in (2.3).

LEMMA F.1. *(i) Let $\mathbf{A} \in \mathbb{R}^{p \times p}$ be a symmetric positive semi-definite matrix. Then, for any matrix $\mathbf{C} \in \mathbb{R}^{p \times q}$, $\mathbf{C}^T \mathbf{A} \mathbf{C} \leq \|\mathbf{A}\| \mathbf{C}^T \mathbf{C} \leq \mathrm{tr}(\mathbf{A}) \mathbf{C}^T \mathbf{C}$.*
(ii) Let $\mathbf{A} \in \mathbb{R}^{p \times p}$ be a symmetric matrix. Then for $\mathbf{a} \in \mathbb{R}^p$ and $\mathbf{c} \in \mathbb{R}^p$, $|\mathbf{a}^T \mathbf{A} \mathbf{c}|^2 \leq \|\mathbf{A}\|^2 \|\mathbf{a}\|^2 \|\mathbf{c}\|^2$.

LEMMA F.2. *Let the rows $\boldsymbol{\varepsilon}_j^T$ of $\mathbf{e} \in \mathbb{R}^{n \times p}$ be i.i.d. random vectors with mean 0, variance $\boldsymbol{\Delta}$ and finite fourth moments.*
(i) Let $\mathbf{L}_1 \in \mathbb{R}^{n \times r}$ and $\mathbf{L}_2 \in \mathbb{R}^{p \times r}$ be nonstochastic matrices. Then $\mathrm{var}\{\mathrm{vec}(\mathbf{L}_2^T \mathbf{e}^T \mathbf{L}_1)\} = \mathbf{L}_1^T \mathbf{L}_1 \otimes \mathbf{L}_2^T \boldsymbol{\Delta} \mathbf{L}_2$.
(ii) Let $\mathbf{L} \in \mathbb{R}^{n \times m}$ and $\mathbf{H} \in \mathbb{R}^{p \times p}$ be nonstochastic matrices of which $\mathbf{H} > 0$ is symmetric. Let $\mathbf{u}_i \in \mathbb{R}^n$ have a 1 in position i and 0's elsewhere,

and let $\mathbf{P}$ *be the projection matrix defined by* $\mathbf{P} = \sum_{i=1}^{n}(\mathbf{u}_i\mathbf{u}_i^T \otimes \mathbf{u}_i\mathbf{u}_i^T)$*. Then* $\mathrm{E}(\mathbf{L}^T\mathbf{e}\mathbf{H}\mathbf{e}^T\mathbf{L}) = \mathbf{L}^T\mathbf{L}\,\mathrm{tr}(\mathbf{H}\boldsymbol{\Delta})$ *and*

$$\begin{aligned}\mathrm{var}\{\mathrm{vec}(\mathbf{L}^T\mathbf{e}\mathbf{H}\mathbf{e}^T\mathbf{L})\} = {} & \mathrm{tr}(\mathbf{H}\boldsymbol{\Delta}\mathbf{H}\boldsymbol{\Delta})(\mathbf{I}_{m^2} + \mathbf{C}_{m^2})\mathbf{T}^T \otimes \mathbf{T} \\ & + \{\mathrm{var}(\boldsymbol{\epsilon}^T\boldsymbol{\Delta}^{1/2}\mathbf{H}\boldsymbol{\Delta}^{1/2}\boldsymbol{\epsilon}) - 2\,\mathrm{tr}(\mathbf{H}\boldsymbol{\Delta}\mathbf{H}\boldsymbol{\Delta})\}\mathbf{T}^T\mathbf{P}\mathbf{T}\end{aligned}$$

where $\mathbf{T} = \mathbf{L} \otimes \mathbf{L}$*,* $\mathbf{C}_{m^2} \in \mathbb{R}^{m^2 \times m^2}$ *is the commutation matrix, and* $\boldsymbol{\epsilon}$ *has mean 0, variance* $\mathbf{I}_p$ *and is distributed as a row of* $\mathbf{e}\boldsymbol{\Delta}^{-1/2}$*.*

PROOF. We prove this proposition by its parts.

(i).

$$\begin{aligned}\mathrm{var}\{\mathrm{vec}(\mathbf{L}_2^T\mathbf{e}^T\mathbf{L}_1)\} &= \mathrm{var}\{(\mathbf{L}_1^T \otimes \mathbf{L}_2^T)\mathrm{vec}(\mathbf{e}^T)\} \\ &= (\mathbf{L}_1^T \otimes \mathbf{L}_2^T)(\mathbf{I}_n \otimes \boldsymbol{\Delta})(\mathbf{L}_1 \otimes \mathbf{L}_2) \\ &= \mathbf{L}_1^T\mathbf{L}_1 \otimes \mathbf{L}_2^T\boldsymbol{\Delta}\mathbf{L}_2.\end{aligned}$$

(ii). Write $\mathbf{e}\mathbf{H}\mathbf{e}^T = (\mathbf{e}\boldsymbol{\Delta}^{-1/2})(\boldsymbol{\Delta}^{1/2}\mathbf{H}\boldsymbol{\Delta}^{1/2})(\mathbf{e}\boldsymbol{\Delta}^{-1/2})^T = \boldsymbol{\epsilon}\mathbf{A}\boldsymbol{\epsilon}^T$, where the rows $\boldsymbol{\epsilon}_i^T$ of $\mathbf{e}\boldsymbol{\Delta}^{-1/2}$ are i.i.d. with mean 0 and variance $\mathbf{I}_p$, $i = 1, \ldots, n$, and $\mathbf{A} = \boldsymbol{\Delta}^{1/2}\mathbf{H}\boldsymbol{\Delta}^{1/2}$. Then the (i, j)-th element of $\mathbf{e}\mathbf{H}\mathbf{e}^T$ is $(\mathbf{e}\mathbf{H}\mathbf{e}^T)_{ij} = \boldsymbol{\epsilon}_i^T\mathbf{A}\boldsymbol{\epsilon}_j$. Consequently, $\mathrm{E}(\mathbf{e}\mathbf{H}\mathbf{e}^T)_{ij} = 0$ if $i \neq j$ and $\mathrm{E}(\mathbf{e}\mathbf{H}\mathbf{e}^T)_{ii} = \mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\boldsymbol{\Delta}\mathbf{H})$. It follows that $\mathrm{E}(\mathbf{L}^T\mathbf{e}\mathbf{H}\mathbf{e}^T\mathbf{L}) = \mathbf{L}^T\mathbf{L}\,\mathrm{tr}(\mathbf{H}\boldsymbol{\Delta})$.

Turning to the variance, let $\mathbf{V} = \mathrm{var}\{\mathrm{vec}(\boldsymbol{\epsilon}\mathbf{A}\boldsymbol{\epsilon}^T)\} \in \mathbb{R}^{n^2 \times n^2}$. Then element (ij, kl) of $\mathbf{V}$ is $\mathbf{V}_{ij,kl} = \mathrm{cov}(\boldsymbol{\epsilon}_i^T\mathbf{A}\boldsymbol{\epsilon}_j, \boldsymbol{\epsilon}_k^T\mathbf{A}\boldsymbol{\epsilon}_l)$. Thus, because $\boldsymbol{\epsilon}_i \perp\!\!\!\perp \boldsymbol{\epsilon}_j$ for $i \neq j$, $\mathbf{V}_{ij,kl} = 0$ if at least one of its four indices is distinct (not equal to any of the other three indices). If no indices are distinct then they must be equal in pairs, leaving four possibilities: (a) $\mathbf{V}_{ii,jj}$, (b) $\mathbf{V}_{ij,ij}$, (c) $\mathbf{V}_{ij,ji}$ and (d) $\mathbf{V}_{ii,ii}$, where $i \neq j$. Clearly, because $\boldsymbol{\epsilon}_i \perp\!\!\!\perp \boldsymbol{\epsilon}_j$, $\mathbf{V}_{ii,jj} = \mathrm{cov}(\boldsymbol{\epsilon}_i^T\mathbf{A}\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j^T\mathbf{A}\boldsymbol{\epsilon}_j) = 0$, leaving only possibilities (b)–(d). Let $\mathbf{U}_{ij} = \mathbf{u}_i \otimes \mathbf{u}_j$. We can now write

$$\begin{aligned}\mathbf{V} &= \sum_{ij,kl}^{n} \mathbf{V}_{ij,kl}\mathbf{U}_{ij}\mathbf{U}_{kl}^T \\ &= \sum_{ij}^{n} \mathbf{V}_{ij,ij}\mathbf{U}_{ij}\mathbf{U}_{ij}^T + \sum_{ij}^{n} \mathbf{V}_{ij,ji}\mathbf{U}_{ij}\mathbf{U}_{ji}^T - \sum_{i}^{n} \mathbf{V}_{ii,ii}\mathbf{U}_{ii}\mathbf{U}_{ii}^T.\end{aligned}$$

But, for $i \neq j$, $\mathbf{V}_{ij,ij} = \mathrm{cov}(\boldsymbol{\epsilon}_i^T\mathbf{A}\boldsymbol{\epsilon}_j, \boldsymbol{\epsilon}_i^T\mathbf{A}\boldsymbol{\epsilon}_j) = \mathrm{cov}(\boldsymbol{\epsilon}_i^T\mathbf{A}\boldsymbol{\epsilon}_j, \boldsymbol{\epsilon}_j^T\mathbf{A}\boldsymbol{\epsilon}_i) = \mathbf{V}_{ij,ji} = \mathbf{V}_{12,12}$, where the last equality follows because $\mathbf{V}_{ij,ij}$ is constant in $i \neq j$. Thus we can write, recalling that $\mathbf{C}_{n^2} = \sum_{ij}(\mathbf{u}_i\mathbf{u}_j \otimes \mathbf{u}_j\mathbf{u}_i)$ is a commutation matrix and that $\mathbf{P} = \sum_i^n (\mathbf{u}_i \otimes \mathbf{u}_i)(\mathbf{u}_i \otimes \mathbf{u}_i)^T$,

$$\begin{aligned}\mathbf{V} &= \mathbf{V}_{12,12} \sum_{ij}^{n} \{\mathbf{U}_{ij}\mathbf{U}_{ij}^T + \mathbf{U}_{ij}\mathbf{U}_{ji}^T\} + (\mathbf{V}_{11,11} - 2\mathbf{V}_{12,12})\mathbf{P} \\ &= \mathbf{V}_{12,12} \sum_{ij}^{n} \{\mathbf{U}_{ij}\mathbf{U}_{ij}^T + \mathbf{U}_{ij}\mathbf{U}_{ij}^T\mathbf{C}_{n^2}\} + (\mathbf{V}_{11,11} - 2\mathbf{V}_{12,12})\mathbf{P}.\end{aligned}$$

Because $\mathbf{I}_{n^2} = \sum_{ij}^n \mathbf{U}_{ij}\mathbf{U}_{ij}^T$ we next have

$$\begin{aligned}\mathbf{V} &= \mathbf{V}_{12,12}(\mathbf{I}_{n^2} + \mathbf{C}_{n^2}) + (\mathbf{V}_{11,11} - 2\mathbf{V}_{12,12})\mathbf{P} \\ &= \operatorname{tr}(\mathbf{A}^2)(\mathbf{I}_{n^2} + \mathbf{C}_{n^2}) + \{\operatorname{var}(\boldsymbol{\epsilon}^T\mathbf{A}\boldsymbol{\epsilon}) - 2\operatorname{tr}(\mathbf{A}^2)\}\mathbf{P},\end{aligned}$$

where $\mathbf{V}_{12,12} = \operatorname{var}(\boldsymbol{\epsilon}_1^T\mathbf{A}\boldsymbol{\epsilon}_2) = \mathrm{E}\{(\boldsymbol{\epsilon}_1^T\mathbf{A}\boldsymbol{\epsilon}_2)^2\} = \operatorname{tr}(\mathbf{A}^2)$. It now follows that

$$\begin{aligned}\operatorname{var}\{\operatorname{vec}(\mathbf{L}^T\mathbf{e}\mathbf{H}\mathbf{e}^T\mathbf{L})\} &= (\mathbf{L}^T \otimes \mathbf{L}^T)\operatorname{var}\{\operatorname{vec}(\mathbf{e}\mathbf{H}\mathbf{e}^T)\}(\mathbf{L} \otimes \mathbf{L}) \\ &= \operatorname{tr}(\mathbf{A}^2)(\mathbf{I}_{m^2} + \mathbf{C}_{m^2})(\mathbf{L}^T\mathbf{L} \otimes \mathbf{L}^T\mathbf{L}) \\ &\quad + \{\operatorname{var}(\boldsymbol{\epsilon}^T\mathbf{A}\boldsymbol{\epsilon}) - 2\operatorname{tr}(\mathbf{A}^2)\}(\mathbf{L}^T \otimes \mathbf{L}^T)\mathbf{P}(\mathbf{L} \otimes \mathbf{L}).\end{aligned}$$

□

In the following lemma $\mathbf{W}$, $\widehat{\mathbf{W}}$, $\mathbf{S}$ and ω are as defined at the outset of Section 5. The matrices $\widehat{\mathbf{A}}$, $\mathbf{B}$ and $\widehat{\mathbf{C}}$ are generic quantities that will assume various specific forms when applying the lemma. When $\widehat{\mathbf{A}} = \widehat{\mathbf{C}}$ the first two conditions – $h^{-1}\widehat{\mathbf{A}}^T\mathbf{W}\widehat{\mathbf{A}} = O_p(1)$ and $h^{-1}\widehat{\mathbf{C}}^T\mathbf{W}\widehat{\mathbf{C}} = O_p(1)$ – are not needed because they are then implied by the third condition $h^{-1}(\widehat{\mathbf{A}}^T\mathbf{W}\widehat{\mathbf{C}} - \mathbf{B}) = O_p(\tau)$.

LEMMA F.3. *Let $\widehat{\mathbf{A}} \in \mathbb{R}^{p\times m}$ and $\widehat{\mathbf{C}} \in \mathbb{R}^{p\times l}$, with l and m fixed, be stochastic matrices so that, as $n, p \to \infty$, $h^{-1}(p)\widehat{\mathbf{A}}^T\mathbf{W}\widehat{\mathbf{A}} = O_p(1)$ and $h^{-1}(p)\widehat{\mathbf{C}}^T\mathbf{W}\widehat{\mathbf{C}} = O_p(1)$. Let $\mathbf{B} \in \mathbb{R}^{m\times l}$ be a nonstochastic matrix so that $h^{-1}(p)\mathbf{B}$ converges and $h^{-1}(p)(\widehat{\mathbf{A}}^T\mathbf{W}\widehat{\mathbf{C}} - \mathbf{B}) = O_p(\tau(n,p))$, as $\tau(n,p) \to 0$. Then $h^{-1}(p)(\widehat{\mathbf{A}}^T\widehat{\mathbf{W}}\widehat{\mathbf{C}} - \mathbf{B}) = O_p(\tau) + O_p(\omega)$.*

PROOF. Let $\widetilde{\mathbf{A}} = \widehat{\mathbf{A}}^T\mathbf{W}\widehat{\mathbf{C}}$. Then

$$\begin{aligned}h^{-1}(\widehat{\mathbf{A}}^T\widehat{\mathbf{W}}\widehat{\mathbf{C}} - \mathbf{B}) &= h^{-1}(\widehat{\mathbf{A}}^T\mathbf{W}\widehat{\mathbf{C}} - \mathbf{B}) + h^{-1}\widehat{\mathbf{A}}^T(\widehat{\mathbf{W}} - \mathbf{W})\widehat{\mathbf{C}} \\ &= h^{-1}(\widetilde{\mathbf{A}} - \mathbf{B}) + h^{-1}\widehat{\mathbf{A}}^T\mathbf{W}^{1/2}\mathbf{S}\mathbf{W}^{1/2}\widehat{\mathbf{C}}\end{aligned}$$

The first term on the right hand side is $O_p(\tau)$ by hypothesis. To deal with the second term we have by Lemma F.1(ii), for fixed vectors $\mathbf{a} \in \mathbb{R}^m$ and $\mathbf{c} \in \mathbb{R}^l$,

$$h^{-2}(\mathbf{a}^T\widehat{\mathbf{A}}^T\mathbf{W}^{1/2}\mathbf{S}\mathbf{W}^{1/2}\widehat{\mathbf{C}}\mathbf{c})^2 \le h^{-2}\left(\mathbf{a}^T\widehat{\mathbf{A}}^T\mathbf{W}\widehat{\mathbf{A}}\mathbf{a}\right)\left(\mathbf{c}^T\widehat{\mathbf{C}}^T\mathbf{W}\widehat{\mathbf{C}}\mathbf{c}\right)\|\mathbf{S}\|^2,$$

from which the conclusion follows since $\mathbf{a}$ and $\mathbf{c}$ are arbitrary. □

The proof of the next lemma and the proofs of other formal statements regarding orders are based on application of Chebyschev's inequality. Briefly, if $A_{n,p}$ is a stochastic sequence depending on n and p then $\Pr(|A_{n,p}| > \epsilon) \leq \epsilon^{-2}\mathrm{E}(A_{n,p}^2)$. If $\mathrm{E}(A_{n,p}^2) \asymp \alpha^2(n,p)$ as $n,p \to \infty$ in some relationship then $A_{n,p} = O_p(\alpha(n,p))$. When $\mathrm{E}(A_{n,p}) = 0$ it is necessary to compute only $\mathrm{var}(A_{n,p})$. When $\mathrm{E}(A_{n,p}) \neq 0$ we will find both $\mathrm{var}(A_{n,p})$ and $\mathrm{E}(A_{n,p})$, since then $\mathrm{E}(A_{n,p}^2) = \mathrm{var}(A_{n,p}) + \mathrm{E}^2(A_{n,p})$. This logic will typically be applied to a stochastic sequence of matrices $\mathbf{A}_{n,p} \in \mathbb{R}^{l \times m}$ of fixed dimension. If $A_{n,p} \equiv \mathbf{a}^T\mathbf{A}_{n,p}\mathbf{b} = O_p(\alpha(n,p))$ for all nonstochastic $\mathbf{a} \in \mathbb{R}^l$ and $\mathbf{b} \in \mathbb{R}^m$ then we write $\mathbf{A}_{n,p} = O_p(\alpha(n,p))$.

LEMMA F.4. *Assume the context and conditions of Proposition 5.1 and let $\boldsymbol{\alpha} \in \mathbb{R}^{n\times s}$ and $\boldsymbol{\gamma} \in \mathbb{R}^{d \times l}$ be nonstochastic matrices, where s, d and l are fixed, $\boldsymbol{\gamma} = O(1)$ and $\boldsymbol{\alpha}^T\boldsymbol{\alpha} = O(n)$. Then*
(i). $(h(p)n)^{-1}\boldsymbol{\alpha}^T\mathbf{e}\widehat{\mathbf{W}}\boldsymbol{\Gamma}\boldsymbol{\gamma} = O_p(\kappa) + O_p(\omega)$, and
(ii). $(h(p)n^2)^{-1}\boldsymbol{\alpha}^T\mathbf{e}\widehat{\mathbf{W}}\mathbf{e}^T\boldsymbol{\alpha} = O_p(\kappa^2) + O_p(\omega)$.

PROOF. We prove part (ii) first and then use it in the proof of part (i).
(ii). Letting $A = \{\mathrm{var}(\boldsymbol{\varepsilon}^T\mathbf{W}\boldsymbol{\varepsilon}) - 2\,\mathrm{tr}(\boldsymbol{\rho}^2)\}$, we have from Lemma F.2

$$
\begin{aligned}
(hn^2)^{-1}\mathrm{E}(\boldsymbol{\alpha}^T\mathbf{e}\mathbf{W}\mathbf{e}^T\boldsymbol{\alpha}) &= (hn)^{-1}(\boldsymbol{\alpha}^T\boldsymbol{\alpha}/n)\,\mathrm{tr}(\boldsymbol{\rho}) = \kappa^2(\boldsymbol{\alpha}^T\boldsymbol{\alpha}/n)\bar{\rho} \\
&= O(\kappa^2) \\
(hn^2)^{-2}\mathrm{var}\{\mathrm{vec}(\boldsymbol{\alpha}^T\mathbf{e}\mathbf{W}\mathbf{e}^T\boldsymbol{\alpha})\} &= (hn^2)^{-2}\,\mathrm{tr}(\boldsymbol{\rho}^2)(\mathbf{I}_{s^2} + \mathbf{C}_{s^2})(\boldsymbol{\alpha}^T\boldsymbol{\alpha} \otimes \boldsymbol{\alpha}^T\boldsymbol{\alpha}) \\
&\quad + (hn^2)^{-2}A(\boldsymbol{\alpha}^T \otimes \boldsymbol{\alpha}^T)\mathbf{P}(\boldsymbol{\alpha} \otimes \boldsymbol{\alpha}) \\
&= O(\kappa^4).
\end{aligned}
$$

The final equality follows by application of the hypothesis $\boldsymbol{\alpha}^T\boldsymbol{\alpha} = O(n)$, condition W.1, $\mathrm{tr}(\boldsymbol{\rho}^2/p^2) \leq \mathrm{tr}^2(\boldsymbol{\rho}/p) = O(1)$, and condition W.2, $A/p^2 = O(1)$. The conclusion of the lemma then follows from Lemma F.3.

(i). Clearly, $(nh)^{-1}\mathrm{E}\left(\boldsymbol{\gamma}^T\boldsymbol{\Gamma}^T\mathbf{W}\mathbf{e}^T\boldsymbol{\alpha}\right) = 0$, so we need to consider only its variance. Recalling that $\bar{\rho} = \mathrm{tr}(\boldsymbol{\rho})/p$ and applying Lemma F.2 with $\mathbf{L}_1 = \boldsymbol{\alpha}$ and $\mathbf{L}_2 = \mathbf{W}\boldsymbol{\Gamma}\boldsymbol{\gamma}$ we have

$$
\begin{aligned}
\mathrm{var}\left\{(nh)^{-1}\boldsymbol{\Gamma}^T\mathbf{W}\mathbf{e}^T\boldsymbol{\alpha}\right\} &= (nh)^{-2}\boldsymbol{\alpha}^T\boldsymbol{\alpha} \otimes \boldsymbol{\gamma}^T\boldsymbol{\Gamma}^T\mathbf{W}\boldsymbol{\Delta}\mathbf{W}\boldsymbol{\Gamma}\boldsymbol{\gamma} \\
&= n^{-1}h^{-2}(\boldsymbol{\alpha}^T\boldsymbol{\alpha}/n) \\
&\quad \otimes(\boldsymbol{\gamma}^T\boldsymbol{\Gamma}^T\mathbf{W}^{1/2})(\mathbf{W}^{1/2}\boldsymbol{\Delta}\mathbf{W}^{1/2})(\mathbf{W}^{1/2}\boldsymbol{\Gamma}\boldsymbol{\gamma}) \\
&\leq \kappa^2\bar{\rho}\{(\boldsymbol{\alpha}^T\boldsymbol{\alpha}/n) \otimes \boldsymbol{\gamma}^T\mathbf{G}_h\boldsymbol{\gamma}\} = O(\kappa^2).
\end{aligned}
$$

The order follows using that $\bar{\rho} = O(1)$ from condition W.1 and $\mathbf{G}_h = O(1)$. The conclusion now follows from Lemma F.3 since $h^{-1}\mathbf{\Gamma}^T\mathbf{W}\mathbf{\Gamma} = \mathbf{G}_h = O(1)$ and from part (ii) $(hn^2)^{-1}\boldsymbol{\alpha}^T\mathbf{e}\mathbf{W}\mathbf{e}^T\boldsymbol{\alpha} = O_p(\kappa^2) = O_p(1)$, the second equality following from universal condition (4.3). □

F.2. Proof of Proposition 5.1 parts (i) and (ii). Recall that the model is $\mathbb{X} = \mathbf{1}_n\boldsymbol{\mu}^T + \mathbf{\Xi}\mathbf{\Gamma}^T + \mathbf{e}$ where the rows of $\mathbf{e}$ are i.i.d. random variables with mean 0 and variance $\mathbf{\Delta}$ satisfying conditions W.1 and W.2. Recall also that $\mathbb{Z} = \mathbb{X} - \mathbf{1}_n\hat{\boldsymbol{\mu}}$, so $\mathbb{F}^T\mathbb{Z} = \mathbb{F}^T\mathbf{\Xi}\mathbf{\Gamma}^T + \mathbb{F}^T\mathbf{e} = n\mathbf{\Phi}_n\boldsymbol{\beta}^T\mathbf{\Gamma}^T + \mathbb{F}^T\mathbf{e}$, where

$$\boldsymbol{\beta} = n^{-1}\mathbf{\Xi}^T\mathbb{F}\mathbf{\Phi}_n^{-1} \in \mathbb{R}^{d\times r} \tag{F.1}$$

is the coefficient matrix from the multivariate linear regression of $\boldsymbol{\xi}_i$ on $\mathbf{f}_i$. Using the rank condition (3.3) and the definition of the scaling matrix in Section 4.1, we see that $\boldsymbol{\beta}$ has the following two properties

$$\boldsymbol{\beta} = O(1) \text{ as } n \to \infty, \text{ and } \boldsymbol{\beta}\mathbf{\Phi}_n\boldsymbol{\beta}^T = \mathbf{I}_d. \tag{F.2}$$

To prove the first two parts of the proposition it is enough to consider $h^{-1}\mathbf{\Phi}_n^{1/2}\widehat{\mathbf{K}}\mathbf{\Phi}_n^{1/2} = (hn^2)^{-1}\mathbb{F}^T\mathbb{Z}\widehat{\mathbf{W}}\mathbb{Z}^T\mathbb{F}$. Substituting for $\mathbb{F}^T\mathbb{Z}$, it is sufficient to study the following three terms. $\mathbf{T}_1 = (n^2h)^{-1}\mathbb{F}^T\mathbf{\Xi}\mathbf{\Gamma}^T\widehat{\mathbf{W}}\mathbf{\Gamma}\mathbf{\Xi}^T\mathbb{F}$, $\mathbf{T}_2 = (n^2h)^{-1}\mathbb{F}^T\mathbf{e}\widehat{\mathbf{W}}\mathbf{\Gamma}\mathbf{\Xi}^T\mathbb{F}$ and $\mathbf{T}_3 = (hn^2)^{-1}\mathbb{F}^T\mathbf{e}\widehat{\mathbf{W}}\mathbf{e}^T\mathbb{F}$. The first conclusion of the proposition follows by computing the expectation of these terms with $\widehat{\mathbf{W}} = \mathbf{W}$, aided by Lemma F.2(ii) with $\mathbf{L} = \mathbb{F}$ and $\mathbf{H} = \mathbf{W}$: $h^{-1}\mathbf{\Phi}_n^{1/2}\mathrm{E}(\widehat{\mathbf{K}})\mathbf{\Phi}_n^{1/2} = \mathbf{T}_1 + \mathrm{E}(\mathbf{T}_3) = \mathbf{T}_1 + (hn^2)^{-1}\mathbb{F}^T\mathbb{F}\,\mathrm{tr}(\mathbf{W}\mathbf{\Delta})$.

For the second conclusion we use Lemma F.4(i) with $\boldsymbol{\alpha} = \mathbb{F}$ and $\boldsymbol{\gamma} = \mathbf{\Xi}^T\mathbb{F}/n$ to get that $\mathbf{T}_2 = O_p(\kappa) + O_p(\omega)$. Using Lemma F.4(ii) with $\boldsymbol{\alpha} = \mathbb{F}$ we get that $\mathbf{T}_3 = O_p(\kappa^2) + O_p(\omega)$. Now, using Lemma F.3,

$$\mathbf{T}_1 - h^{-1}\mathbf{\Phi}_n^{1/2}\mathbf{K}\mathbf{\Phi}_n^{1/2} = (n^2h)^{-1}\mathbb{F}^T\mathbf{\Xi}\mathbf{\Gamma}^T(\widehat{\mathbf{W}} - \mathbf{W})\mathbf{\Gamma}\mathbf{\Xi}^T\mathbb{F} = O_p(\omega).$$

□

F.3. Proof of Proposition 5.1(iii). The following proposition summarizes the orders of various terms that we will encounter in our study of $\widehat{\mathbf{R}}_{\widehat{\mathbf{W}}}(\mathbf{X}_N) - \mathbf{R}(\mathbf{X}_N)$. For notational convenience, let $\widehat{\mathbf{G}}_h = \widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}\widehat{\mathbf{\Gamma}}/h \in \mathbb{R}^{d\times d}$ and recall that $\mathbf{G}_h = \mathbf{\Gamma}^T\mathbf{W}\mathbf{\Gamma}/h$.

Proposition F.1. *Assume the condition of Proposition 5.1. Then*

(i). $\widehat{\mathbf{V}}_d = \mathbf{\Phi}_n^{1/2}\boldsymbol{\beta}^T + O_p(\kappa) + O_p(\omega)$, *where the columns* $\widehat{\mathbf{V}}_d$ *are the first* d *eigenvectors of* $\widehat{\mathbf{K}}$.

(ii). $\widehat{\mathbf{G}}_h = O_p(1)$.
(iii). $h^{-1}(p)\mathbf{\Gamma}^T(\widehat{\mathbf{W}} - \mathbf{W})\mathbf{\Gamma} = O_p(\omega)$.
(iv) $\widehat{\mathbf{G}}_h - \mathbf{G}_h = O_p(\kappa) + O_p(\omega)$.
(v). $\widehat{\mathbf{G}}_h - h^{-1}(p)\widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}\mathbf{\Gamma} = O_p(\kappa) + O_p(\omega)$.

PROOF. We prove this proposition by its parts.

(i). It follows from Proposition 5.1(ii) that the projection over span of the first d eigenvectors of $h^{-1}\widehat{\mathbf{K}}$ minus projection over span of the first d eigenvectors of $h^{-1}\mathbf{K}$ is of order $O_p(\kappa)+O_p(\omega)$ (See Tyler 1981, Lemma 4.1, for further results along these lines). Using the constraints (F.2), the first d eigenvectors of $\mathbf{K}$ are then $n^{-1}\mathbf{\Phi}_n^{-1/2}\mathbb{F}^T\mathbf{\Xi} = \mathbf{\Phi}_n^{1/2}\boldsymbol{\beta}^T$ with eigenvalues given by the diagonal elements of the diagonal matrix $\mathbf{G}_h$:

$$\begin{aligned} n^{-1}\mathbf{K}\mathbf{\Phi}_n^{-1/2}\mathbb{F}^T\mathbf{\Xi} &= n^{-3}\mathbf{\Phi}_n^{-1/2}\mathbb{F}^T\mathbf{\Xi}\mathbf{G}_h\mathbf{\Xi}^T\mathbb{F}\mathbf{\Phi}_n^{-1/2} \times \mathbf{\Phi}_n^{-1/2}\mathbb{F}^T\mathbf{\Xi} \\ &= n^{-2}\mathbf{\Phi}_n^{-1/2}\mathbb{F}^T\mathbf{\Xi}\mathbf{G}_h\mathbf{\Xi}^T\mathbf{P}_{\mathbb{F}}\mathbf{\Xi} \\ &= n^{-1}\mathbf{\Phi}_n^{-1/2}\mathbb{F}^T\mathbf{\Xi}\mathbf{G}_h. \end{aligned}$$

(ii). $\widehat{\mathbf{G}}_h$ is the diagonal matrix with the largest d eigenvalues of $h^{-1}\widehat{\mathbf{K}}$, $h^{-1}(\widehat{\mathbf{K}} - \mathbf{K}) = O_p(\kappa) + O_p(\omega)$ and the eigenvalues of $h^{-1}\mathbf{K}$ are bounded.

(iii). This follows immediately from (4.1) and Lemma F.3.

(iv). Let $\mathbf{O}_{\kappa\omega} = O_p(\kappa) + O_p(\omega)$ be a generic $d \times r$ matrix of the indicated order. Then $\widehat{\mathbf{\Gamma}} = \widehat{\mathbf{B}}\mathbf{\Phi}_n^{1/2}\widehat{\mathbf{V}}_d = n^{-1}(\mathbb{Z}^T\mathbb{F})\mathbf{\Phi}_n^{-1/2}\widehat{\mathbf{V}}_d = (\mathbf{\Gamma}\boldsymbol{\beta}\mathbf{\Phi}_n + n^{-1}\mathbf{e}^T\mathbb{F})\mathbf{\Phi}_n^{-1/2}\widehat{\mathbf{V}}_d$ and, using part (i), $\widehat{\mathbf{V}}_d^T\mathbf{\Phi}_n^{-1/2} = \boldsymbol{\beta}+\mathbf{O}_{\kappa\omega}\mathbf{\Phi}_n^{-1/2} = \boldsymbol{\beta}+\mathbf{O}_{\kappa\omega}$ and therefore

$$\begin{aligned} \widehat{\mathbf{\Gamma}}^T &= (\boldsymbol{\beta} + \mathbf{O}_{\kappa\omega})(\mathbf{\Phi}_n\boldsymbol{\beta}^T\mathbf{\Gamma}^T + n^{-1}\mathbb{F}^T\mathbf{e}) \\ &= \boldsymbol{\beta}\mathbf{\Phi}_n\boldsymbol{\beta}^T\mathbf{\Gamma}^T + \mathbf{O}_{\kappa\omega}\boldsymbol{\beta}^T\mathbf{\Gamma}^T + n^{-1}(\boldsymbol{\beta} + \mathbf{O}_{\kappa\omega})\mathbb{F}^T\mathbf{e} \\ &= \mathbf{\Gamma}^T + \mathbf{O}_{\kappa\omega}\boldsymbol{\beta}^T\mathbf{\Gamma}^T + n^{-1}(\boldsymbol{\beta} + \mathbf{O}_{\kappa\omega})\mathbb{F}^T\mathbf{e} \qquad \text{(F.3)} \end{aligned}$$

where we use (F.2) in the last equality. For notational simplicity, let $\mathbf{A}_1 = \mathbf{\Gamma}\boldsymbol{\beta}(\boldsymbol{\beta}\mathbf{\Phi}_n+\mathbf{O}_{\kappa\omega})^T$, $\mathbf{A}_2 = \mathbf{e}^T\mathbb{F}(\boldsymbol{\beta}+\mathbf{O}_{\kappa\omega})^T$, $\mathbf{T}_1 = \mathbf{A}_1^T\widehat{\mathbf{W}}\mathbf{A}_1$, $\mathbf{T}_2 = n^{-1}\mathbf{A}_1^T\widehat{\mathbf{W}}\mathbf{A}_2$ and $\mathbf{T}_3 = n^{-2}\mathbf{A}_2^T\widehat{\mathbf{W}}\mathbf{A}_2$. Then

$$\widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}\widehat{\mathbf{\Gamma}} = \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_2^T + \mathbf{T}_3. \qquad \text{(F.4)}$$

We next study each of these terms separately. Expanding $\mathbf{T}_1$ we have

$$\mathbf{T}_1 = \mathbf{\Gamma}^T\widehat{\mathbf{W}}\mathbf{\Gamma} + \mathbf{O}_{\kappa\omega}\boldsymbol{\beta}^T\mathbf{\Gamma}^T\widehat{\mathbf{W}}\mathbf{\Gamma} + \mathbf{\Gamma}^T\widehat{\mathbf{W}}\mathbf{\Gamma}\boldsymbol{\beta}\mathbf{O}_{\kappa\omega}^T + \mathbf{O}_{\kappa\omega}\boldsymbol{\beta}^T\mathbf{\Gamma}^T\widehat{\mathbf{W}}\mathbf{\Gamma}\boldsymbol{\beta}\mathbf{O}_{\kappa\omega}^T$$

and thus $h^{-1}(\mathbf{T}_1 - \mathbf{\Gamma}^T\mathbf{W}\mathbf{\Gamma}) = O_p(\omega) + O_p(\kappa)$ from (iii). Next, expanding $\mathbf{T}_2$ and using Lemma F.4(i) with $\boldsymbol{\alpha} = \mathbb{F}\boldsymbol{\beta}^T$ and $\boldsymbol{\gamma} = \mathbf{I}_d$,

$$\begin{aligned} \mathbf{T}_2 &= n^{-1}\boldsymbol{\beta}\mathbb{F}^T\mathbf{e}\widehat{\mathbf{W}}\mathbf{\Gamma} + n^{-1}\mathbf{O}_{\kappa\omega}\mathbb{F}^T\mathbf{e}\widehat{\mathbf{W}}\mathbf{\Gamma} + n^{-1}\boldsymbol{\beta}\mathbb{F}^T\mathbf{e}\widehat{\mathbf{W}}\mathbf{\Gamma}\boldsymbol{\beta}\mathbf{O}_{\kappa\omega}^T \\ &\quad + n^{-1}\mathbf{O}_{\kappa\omega}\mathbb{F}^T\mathbf{e}\widehat{\mathbf{W}}\mathbf{\Gamma}\boldsymbol{\beta}\mathbf{O}_{\kappa\omega}^T \\ h^{-1}\mathbf{T}_2 &= O_p(\kappa) + O_p(\omega). \end{aligned}$$

Finally, expanding $\mathbf{T}_3$ and of Lemma F.4(ii) with $\boldsymbol{\alpha} = \mathbb{F}\boldsymbol{\beta}^T$

$$\begin{aligned} \mathbf{T}_3 &= n^{-2}\boldsymbol{\beta}\mathbb{F}^T\mathbf{e}\widehat{\mathbf{W}}\mathbf{e}^T\mathbb{F}\boldsymbol{\beta}^T + n^{-2}\boldsymbol{\beta}\mathbb{F}^T\mathbf{e}\widehat{\mathbf{W}}\mathbf{e}^T\mathbb{F}\mathbf{O}_{\kappa\omega}^T + n^{-2}\mathbf{O}_{\kappa\omega}\mathbb{F}^T\mathbf{e}\widehat{\mathbf{W}}\mathbf{e}^T\mathbb{F}\boldsymbol{\beta}^T \\ &\quad + n^{-2}\mathbf{O}_{\kappa\omega}\mathbb{F}^T\mathbf{e}\widehat{\mathbf{W}}\mathbf{e}^T\mathbb{F}\mathbf{O}_{\kappa\omega}^T \\ h^{-1}\mathbf{T}_3 &= O_p(\kappa^2) + O_p(\omega). \end{aligned}$$

(v). This follows from part (iv), the relationship $h^{-1}\left(\widehat{\mathbf{\Gamma}}^T\mathbf{W}\mathbf{\Gamma} - \mathbf{\Gamma}^T\mathbf{W}\mathbf{\Gamma}\right) = O_p(\kappa)+O_p(\omega)$, which can be shown following the proof of (iv), and Lemma F.3. □

Turning to the proof of Proposition 5.1(iii), let $\mathbf{X}_N$ be a new observation on $\mathbf{X}$ from model (2.2) with coordinate vector $\boldsymbol{\xi}_N$ and error $\boldsymbol{\varepsilon}_N$. Then we need to find the limiting behavior of

$$\begin{aligned} \widehat{\mathbf{R}}_{\widehat{\mathbf{W}}}(\mathbf{X}_N) - \mathbf{R}(\mathbf{X}_N) &= \left((\widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}\widehat{\mathbf{\Gamma}})^{-1}\widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}\mathbf{\Gamma} - \mathbf{I}_d\right)\boldsymbol{\xi}_N \\ &\quad + (\widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}\widehat{\mathbf{\Gamma}})^{-1}\,\widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}(\boldsymbol{\mu} - \bar{\mathbf{X}}) + (\widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}\widehat{\mathbf{\Gamma}})^{-1}\widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}\boldsymbol{\varepsilon}_N \\ &\quad - (\mathbf{\Gamma}^T\mathbf{\Delta}^{-1}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^T\mathbf{\Delta}^{-1}\boldsymbol{\varepsilon}_N \\ &= (\widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}\widehat{\mathbf{\Gamma}})^{-1}\left(\widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}\mathbf{\Gamma} - \widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}\widehat{\mathbf{\Gamma}}\right)\boldsymbol{\xi}_N \\ &\quad + (\widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}\widehat{\mathbf{\Gamma}})^{-1}\,\widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}(\boldsymbol{\mu} - \bar{\mathbf{X}}) + (\widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}\widehat{\mathbf{\Gamma}})^{-1}\widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}\boldsymbol{\varepsilon}_N \\ &\quad - (\mathbf{\Gamma}^T\mathbf{\Delta}^{-1}\mathbf{\Gamma})^{-1}\mathbf{\Gamma}^T\mathbf{\Delta}^{-1}\boldsymbol{\varepsilon}_N \\ &= \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3 - \mathbf{T}_4 \end{aligned}$$

Term $\mathbf{T}_1$*.* It follows from Proposition F.1(ii) and (v) that $\mathbf{T}_1 = O_p(\kappa) + O_p(\omega)$.

Term $\mathbf{T}_2$*.* For this term we have $\mathbf{T}_2 = \widehat{\mathbf{G}}_h^{-1}\{\widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}(\boldsymbol{\mu}-\bar{\mathbf{X}})/h\}$. From Proposition F.1(ii), $\widehat{\mathbf{G}}_h = O_p(1)$. To find the order of the second factor we use $h^{-1}\left(\widehat{\mathbf{\Gamma}}^T\widehat{\mathbf{W}}(\boldsymbol{\mu} - \bar{\mathbf{X}})\right) = \widehat{\mathbf{V}}_d^T\mathbf{\Phi}_n^{-1/2}(\mathbb{F}^T\mathbb{Z}/nh)\widehat{\mathbf{W}}(\boldsymbol{\mu} - \bar{\mathbf{X}})$, and

$$\begin{aligned} (nh)^{-1}\mathbb{F}^T\mathbb{Z}\widehat{\mathbf{W}}(\boldsymbol{\mu} - \bar{\mathbf{X}}) &= -(n^2h)^{-1}\mathbb{F}^T\mathbb{X}\widehat{\mathbf{W}}\mathbf{e}^T\mathbf{1}_n \\ &= -(nh)^{-1}(\mathbb{F}^T\mathbf{\Xi}/n)\mathbf{\Gamma}^T\widehat{\mathbf{W}}\mathbf{e}^T\mathbf{1}_n + (n^2h)^{-1}\mathbb{F}^T\mathbf{e}\widehat{\mathbf{W}}\mathbf{e}^T\mathbf{1}_n \\ &= \mathbf{T}_{21} + \mathbf{T}_{22}. \end{aligned}$$

From Lemma F.4(i) with $\boldsymbol{\alpha} = \mathbf{1}_n$ and $\boldsymbol{\gamma} = \boldsymbol{\Xi}^T\mathbb{F}/n$ (cf. (3.3)) it follows that $\mathbf{T}_{21} = O_p(\kappa) + O_p(\omega)$.

For $\mathbf{T}_{22}$ we get using Lemma F.4(ii) with $\boldsymbol{\alpha} = \mathbb{F}$ and with $\boldsymbol{\alpha} = \mathbf{1}_n$,

$$|\mathbf{T}_{22}| \leq \left(\frac{\mathbb{F}^T\mathbf{e}\widehat{\mathbf{W}}\mathbf{e}^T\mathbb{F}}{n^2h}\right)^{1/2}\left(\frac{\mathbf{1}_n^T\mathbf{e}\widehat{\mathbf{W}}\mathbf{e}^T\mathbf{1}_n}{n^2h}\right)^{1/2} = O_p(\kappa^2) + O_p(\omega).$$

Term $\mathbf{T}_3 - \mathbf{T}_4$.

$$\begin{aligned}\mathbf{T}_3 - \mathbf{T}_4 &= \{h^{-1}\widehat{\mathbf{G}}_h^{-1}\widehat{\boldsymbol{\Gamma}}^T\widehat{\mathbf{W}} - (\boldsymbol{\Gamma}^T\boldsymbol{\Delta}^{-1}\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^T\boldsymbol{\Delta}^{-1}\}\boldsymbol{\varepsilon}_N \\ &= h^{-1}\widehat{\mathbf{G}}_h^{-1}(\widehat{\boldsymbol{\Gamma}}^T\widehat{\mathbf{W}} - \boldsymbol{\Gamma}^T\mathbf{W})\boldsymbol{\varepsilon}_N \\ &\quad + h^{-1}\widehat{\mathbf{G}}_h^{-1}\left(\mathbf{G}_h - \widehat{\mathbf{G}}_h\right)\mathbf{G}_h^{-1}\boldsymbol{\Gamma}^T\mathbf{W}\boldsymbol{\varepsilon}_N \\ &\quad + \{h^{-1}\mathbf{G}_h^{-1}\boldsymbol{\Gamma}^T\mathbf{W} - (\boldsymbol{\Gamma}^T\boldsymbol{\Delta}^{-1}\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^T\boldsymbol{\Delta}^{-1}\}\boldsymbol{\varepsilon}_N \\ &\quad = \mathbf{T}_{341} + \mathbf{T}_{342} + \boldsymbol{\nu} \in \mathbb{R}^d,\end{aligned}$$

where, as defined at the outset of Section 5.1, $\boldsymbol{\nu} = \mathbf{R}_{\mathbf{W}}(\boldsymbol{\varepsilon}_N) - \mathbf{R}(\boldsymbol{\varepsilon}_N)$.

We now consider $\mathbf{T}_{341}$ and $\mathbf{T}_{342}$ separately. To study $\mathbf{T}_{341}$ recall that, from Proposition F.1(ii), $\widehat{\mathbf{G}}_h = O_p(1)$, so the factor $\widehat{\mathbf{G}}_h^{-1}$ is neglected. Expanding the other factor and using the same term designation, we have

$$\begin{aligned}\widehat{\mathbf{G}}_h\mathbf{T}_{341} &= h^{-1}(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})^T\mathbf{W}\boldsymbol{\varepsilon}_N + h^{-1}\boldsymbol{\Gamma}^T(\widehat{\mathbf{W}} - \mathbf{W})\boldsymbol{\varepsilon}_N \\ &\quad + h^{-1}(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma})^T(\widehat{\mathbf{W}} - \mathbf{W})\boldsymbol{\varepsilon}_N \\ &= \mathbf{T}_{3411} + \mathbf{T}_{3412} + \mathbf{T}_{3413}.\end{aligned}$$

Using (F.3),

$$\mathbf{T}_{3411} = h^{-1}\left(\mathbf{O}_{\kappa\omega}\boldsymbol{\beta}^T\boldsymbol{\Gamma}^T + n^{-1}\boldsymbol{\beta}\mathbb{F}^T\mathbf{e} + n^{-1}\mathbf{O}_{\kappa\omega}\mathbb{F}^T\mathbf{e}\right)\mathbf{W}\boldsymbol{\varepsilon}_N. \tag{F.5}$$

Now,

$$\begin{aligned}\mathrm{var}(h^{-1}\boldsymbol{\beta}^T\boldsymbol{\Gamma}^T\mathbf{W}\boldsymbol{\varepsilon}_N) &= h^{-2}\boldsymbol{\beta}^T\boldsymbol{\Gamma}^T\mathbf{W}\boldsymbol{\Delta}\mathbf{W}\boldsymbol{\Gamma}\boldsymbol{\beta} = h^{-2}\boldsymbol{\beta}^T\boldsymbol{\Gamma}^T\mathbf{W}^{1/2}\boldsymbol{\rho}\mathbf{W}^{1/2}\boldsymbol{\Gamma}\boldsymbol{\beta} \\ &\leq \|\boldsymbol{\rho}\|h^{-1}\boldsymbol{\beta}^T\mathbf{G}_h\boldsymbol{\beta} = O(1)\end{aligned}$$

since $\|\boldsymbol{\rho}\| = O(h)$ by hypothesis. Therefore since $\mathrm{E}\left(h^{-1}\boldsymbol{\beta}^T\boldsymbol{\Gamma}^T\mathbf{W}\boldsymbol{\varepsilon}_N\right) = 0$, the first term of (F.5) is of order $\mathbf{O}_{\kappa\omega}$. The third term of (F.5) is a smaller order than the second so it is left to consider $n^{-1}\boldsymbol{\beta}\mathbb{F}^T\mathbf{e}\mathbf{W}\boldsymbol{\varepsilon}_N$. Again the expectation is 0 and for the variance we have using Lemma F.2(ii),

$$\begin{aligned}\mathrm{var}\left((hn)^{-1}\boldsymbol{\beta}\mathbb{F}^T\mathbf{e}\mathbf{W}\boldsymbol{\varepsilon}_N\right) &= (hn)^{-2}\mathrm{E}\left(\boldsymbol{\beta}\mathbb{F}^T\mathbf{e}\mathbf{W}\boldsymbol{\Delta}\mathbf{W}\mathbf{e}^T\mathbb{F}\boldsymbol{\beta}^T\right) \\ &= (h^2n)^{-1}\boldsymbol{\beta}\boldsymbol{\Phi}_n\boldsymbol{\beta}^T\,\mathrm{tr}(\mathbf{W}\boldsymbol{\Delta}\mathbf{W}\boldsymbol{\Delta}) \\ &= (h^2n)^{-1}\,\mathrm{tr}(\boldsymbol{\rho}^2) = O_p(\psi^2),\end{aligned}$$

where the penultimate equality uses (F.2). From this we get that $\mathbf{T}_{3411} = O_p(\psi) + O_p(\kappa) + O_p(\omega)$.

Next, $\mathrm{E}(\mathbf{T}_{3412}) = 0$ and

$$\begin{aligned} \mathrm{var}(\mathbf{T}_{3412}) &= h^{-2}\mathbf{\Gamma}^T\mathbf{W}^{1/2}\mathrm{E}(\mathbf{S}\boldsymbol{\rho}\mathbf{S})\mathbf{W}^{1/2}\mathbf{\Gamma} \leq (\|\boldsymbol{\rho}\|/h)\|\mathrm{E}(\mathbf{S}^2)\|(\mathbf{\Gamma}^T\mathbf{W}\mathbf{\Gamma}/h) \\ &= O(\omega^2), \end{aligned}$$

and thus by the hypothesis $\|\boldsymbol{\rho}\| = O(h)$, condition S.2 and universal condition (4.1) we have $\mathbf{T}_{3412} = O_p(\omega)$.

Using (F.3) again, the final addend of $\mathbf{T}_{341}$ can be written as

$$\mathbf{T}_{3413} = h^{-1}\left(\mathbf{O}_{\kappa\omega}\boldsymbol{\beta}^T\mathbf{\Gamma}^T + n^{-1}\boldsymbol{\beta}\mathbb{F}^T\mathbf{e} + n^{-1}\mathbf{O}_{\kappa\omega}\mathbb{F}^T\mathbf{e}\right)(\widehat{\mathbf{W}} - \mathbf{W})\boldsymbol{\varepsilon}_N.$$

The order of the first term follows from $\mathbf{T}_{3412}$ and the third term is a smaller order than the second, so it is sufficient to find the magnitude of $\mathbf{A} = (hn)^{-1}\boldsymbol{\beta}\mathbb{F}^T\mathbf{e}(\widehat{\mathbf{W}} - \mathbf{W})\boldsymbol{\varepsilon}_N$: $\mathrm{E}(\mathbf{A}) = 0$ and

$$\begin{aligned} \mathrm{var}(\mathbf{A}) &= (hn)^{-2}\boldsymbol{\beta}\mathrm{E}(\mathbb{F}^T\mathbf{e}(\widehat{\mathbf{W}} - \mathbf{W})\mathbf{\Delta}(\widehat{\mathbf{W}} - \mathbf{W})\mathbf{e}^T\mathbb{F})\boldsymbol{\beta}^T \\ &= (hn)^{-2}\boldsymbol{\beta}\mathrm{E}(\mathbf{F}^T\mathbf{e}\mathbf{W}^{1/2}\mathbf{S}\boldsymbol{\rho}\mathbf{S}\mathbf{W}^{1/2}\mathbf{e}^T\mathbb{F})\boldsymbol{\beta}^T \\ &\leq (hn)^{-2}\|\boldsymbol{\rho}\| \times \|\mathrm{E}(\mathbf{S}^2)\| \times \boldsymbol{\beta}\mathrm{E}(\mathbb{F}^T\mathbf{e}\mathbf{W}\mathbf{e}^T\mathbb{F})\boldsymbol{\beta}^T \\ &= (p/hn)(\|\boldsymbol{\rho}\|/h)(\boldsymbol{\beta}\mathbf{\Phi}_n\boldsymbol{\beta}^T)(\mathrm{tr}(\boldsymbol{\rho})/p)O(\omega^2) \\ &= O(\kappa^2)O(\omega^2), \end{aligned}$$

where the last two equalities make use of condition S.2, the hypothesis $\|\boldsymbol{\rho}\| = O(h)$ and universal condition W.1. Comparing the orders of $\mathbf{T}_{3411}$, $\mathbf{T}_{3412}$ and $\mathbf{T}_{3413}$, we have $\mathbf{T}_{341} = O_p(\psi) + O_p(\kappa) + O_p(\omega)$.

Turning to term $\mathbf{T}_{342}$ and using parts (ii) and (iv) of Proposition F.1, we have $\widehat{\mathbf{G}}_h^{-1} = O_p(1)$ and $\mathbf{G}_h - \widehat{\mathbf{G}}_h = O_p(\omega) + O_p(\kappa)$. Then by universal condition (4.1), it follows that $\widehat{\mathbf{G}}_h^{-1}(\widehat{\mathbf{G}}_h - \mathbf{G}_h)\mathbf{G}_h^{-1} = O_p(\omega) + O_p(\kappa)$. Next, $\mathrm{var}(h^{-1}\mathbf{\Gamma}^T\mathbf{W}\boldsymbol{\varepsilon}_N) = h^{-2}\mathbf{\Gamma}^T\mathbf{W}\mathbf{\Delta}\mathbf{W}\mathbf{\Gamma} \leq h^{-1}\|\boldsymbol{\rho}\|\mathbf{G}_h = O(1)$. It follows that $\mathbf{T}_{342} = O_p(\omega) + O_p(\kappa)$.

Putting the various parts together we have in summary, $\mathbf{T}_3 - \mathbf{T}_4 = \boldsymbol{\nu} + O_p(\psi) + O_p(\kappa) + O_p(\omega)$, which when compared with the orders of $\mathbf{T}_1$ and $\mathbf{T}_2$ gives the desired conclusions. □

APPENDIX G: PROOFS OF LEMMAS 5.1 AND 5.2

The order of $\|\mathbf{S}\|$ given in Lemma 5.1 is a consequence of Theorem 2 of Rothman *et al.* (2008), where we replace Lemma 1 in Rothman *et al.* (2008) with the following Proposition G.1, which accounts for the use of the residual sample covariance matrix $\widehat{\mathbf{\Delta}}$ based on sub-Gaussian errors, in place of a marginal sample covariance matrix based on i.i.d. Gaussian errors.

PROPOSITION G.1. *Suppose that $\mathbb{Z} = [\mathbf{z}_{ij}]$ is an $n \times p$ matrix with i.i.d rows having mean zero, covariance matrix $\mathbf{\Sigma} = [\sigma_{jk}]$ and uniformly sub-Gaussian elements. Let $\mathbf{Z}_j$ denote the jth column of $\mathbb{Z}$, let $\mathbf{P}$ denote the projection onto an subspace of $\mathbb{R}^n$ with fixed dimension $r < n$, and let $\mathbf{Q} = \mathbf{I}_n - \mathbf{P}$. Then for all $t > 0$ we have*

$$\Pr\left(\left|\frac{\mathbf{Z}_j^T\mathbf{Q}\mathbf{Z}_k}{n} - \sigma_{jk}\right| > t\right) \le C_1 e^{-C_2 n \min(t^2,t)}$$

where C_1, C_2 are positive constants.

PROOF. Adapting an approach by Bickel and Levina (2008a),

$$\begin{aligned}
&\Pr\left(\left|\frac{\mathbf{Z}_k^T\mathbf{Q}\mathbf{Z}_j}{n} - \sigma_{jk}\right| > t\right) \\
&\quad = \Pr\left[\frac{1}{4}\left|\left(\frac{(\mathbf{Z}_k+\mathbf{Z}_j)^T\mathbf{Q}(\mathbf{Z}_k+\mathbf{Z}_j)}{n} - \ell_{jk}\right) - \left(\frac{(\mathbf{Z}_k-\mathbf{Z}_j)^T\mathbf{Q}(\mathbf{Z}_k-\mathbf{Z}_j)}{n} - m_{jk}\right)\right| > t\right] \\
&\quad \le \Pr\left(\left|\frac{(\mathbf{Z}_k+\mathbf{Z}_j)^T(\mathbf{Z}_k+\mathbf{Z}_j)}{n} - \ell_{jk}\right| > t\right) + \Pr\left(\left|\frac{(\mathbf{Z}_k+\mathbf{Z}_j)^T\mathbf{P}(\mathbf{Z}_k+\mathbf{Z}_j)}{n}\right| > t\right) \\
&\quad + \Pr\left(\left|\frac{(\mathbf{Z}_k-\mathbf{Z}_j)^T(\mathbf{Z}_k-\mathbf{Z}_j)}{n} - m_{jk}\right| > t\right) + \Pr\left(\left|\frac{(\mathbf{Z}_k-\mathbf{Z}_j)^T\mathbf{P}(\mathbf{Z}_k-\mathbf{Z}_j)}{n}\right| > t\right) \\
&\quad = \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}
\end{aligned}$$

where $\ell_{jk} = \sigma_{jj} + \sigma_{kk} + 2\sigma_{jk}$, $m_{jk} = \sigma_{jj} + \sigma_{kk} - 2\sigma_{jk}$, and $\sigma_{jk} = 0.25(\ell_{jk} - m_{jk})$.

We begin by bounding I. The n random variables $\mathbf{z}_{1k} + \mathbf{z}_{1j}, \ldots, \mathbf{z}_{nk} + \mathbf{z}_{nj}$ are independent and sub-Gaussian. Also $\mathrm{E}(\mathbf{z}_{ik} + \mathbf{z}_{ij})^2 = \ell_{jk}$.

We can write

$$\frac{(\mathbf{Z}_k+\mathbf{Z}_j)^T(\mathbf{Z}_k+\mathbf{Z}_j)}{n} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{z}_{ik} + \mathbf{z}_{ij})^2.$$

It is clear that the random variable $(\mathbf{z}_{ik} + \mathbf{z}_{ij})^2 - \ell_{jk}$ is a centered sub-exponential random variable and, $\mathrm{I} \le C_1 e^{-C_2 n \min(t^2,t)}$ (Vershynin 2011, corollary 17). To bound II, note that we can write $\mathbf{P} = \sum_{m=1}^{r} \mathbf{c}_m \mathbf{c}_m^T$, where

$\mathbf{c}_m$ is an n-dimensional vector with $\mathbf{c}_m^T\mathbf{c}_m = 1$ and $\mathbf{c}_1, \ldots, \mathbf{c}_r$ are orthogonal.

$$\begin{aligned}
\Pr\left(\left|\frac{(\mathbf{Z}_k+\mathbf{Z}_j)^T\mathbf{P}(\mathbf{Z}_k+\mathbf{Z}_j)}{n}\right| > t\right) &\le \Pr\left(\frac{1}{n}\sum_{m=1}^{r}\{\mathbf{c}_m^T(\mathbf{Z}_k+\mathbf{Z}_j)\}^2 > t\right)\\
&\le r\max_{1\le m\le r}\Pr\left(\frac{1}{n}\{\mathbf{c}_m^T(\mathbf{Z}_k+\mathbf{Z}_j)\}^2 > \frac{t}{r}\right)\\
&\le r\max_{1\le m\le r}\Pr\left(\left|\mathbf{c}_m^T(\mathbf{Z}_k+\mathbf{Z}_j)\right| > \sqrt{\frac{nt}{r}}\right)\\
&\le C_3 e^{-C_4 nt},
\end{aligned}$$

where the last inequality follows from Proposition 10 in Vershynin (2011). Terms III and IV are bounded similarly, and the proposition follows by setting $\mathbf{P} = \mathbf{P}_{\mathbb{F}}$. □

Although the result in Proposition G.1 is for the residual covariance matrix, a similar result is used for the residual correlation matrix as it was in Rothman *et al.* (2008).

The order of $\|\mathrm{E}(\mathbf{S}^2)\|$ given in Lemma 5.1 can be seen as follows. First, using results in either Rothman *et al.* (2008) or Ravikumar *et al.* (2008) and Proposition G.1 it can be shown that for positive constants c_1, c_2 and c_3, $\Pr(\|\mathbf{S}\| > M\omega_{\text{spice}}) \le c_1 p^{2-c_2M^{c_3}}$ for all M. With this we have

$$\|\mathrm{E}(\mathbf{S}^2)/\omega_{\text{spice}}^2\| \le \mathrm{E}\|\mathbf{S}/\omega_{\text{spice}}\|^2 = \int_0^\infty \Pr(\|\mathbf{S}/\omega_{\text{spice}}\|^2 > t)dt.$$

Select t_0 sufficiently large so that $2 - c_2t^{c_3} < 0$ for all $t \ge t_0$. Then

$$\begin{aligned}
\|\mathrm{E}(\mathbf{S}^2)/\omega_{\text{spice}}^2\| &\le \int_0^{t_0}\Pr(\|\mathbf{S}/\omega_{\text{spice}}\|^2 > t)dt + \int_{t_0}^\infty \Pr(\|\mathbf{S}/\omega_{\text{spice}}\|^2 > t)dt\\
&\le \int_0^{t_0}\Pr(\|\mathbf{S}/\omega_{\text{spice}}\|^2 > t)dt + \int_{t_0}^\infty c_1 p^{(2-c_2t^{(c_3/2)})}dt = O(1).
\end{aligned}$$

Turning to the rate for $\|\mathbf{S}\|$ in Lemma 5.2, let $S_{jj} = \mathbf{Z}_j^T\mathbf{Q}\mathbf{Z}_j/n\sigma_{jj}$ denote the j-th diagonal element of $\mathbf{S}$. Using the notation of Lemma 5.1, it can be shown that using a similar argument that $\Pr(|S_{jj}-1| > t) \le C_1e^{-C_2t^2}$, $j = 1, \ldots, p$, and thus $\Pr(\max_j |S_{jj}-1| > t) \le C_1pe^{-C_2t^2}$ from which the rate follows. □

APPENDIX H: PROOF OF PROPOSITION 6.2

H.1. Preliminary results. Let $W_p(\boldsymbol{\Delta}, n)$ denote the Wishart distribution with covariance matrix $\boldsymbol{\Delta}$ and n degrees of freedom.

LEMMA H.1. *(von Rosen, 1988) Let* $\mathbf{W_I} \sim W_p(\mathbf{I}_p, n)$*. If* $n > p + 3$*, then* $\mathrm{E}(\mathbf{W_I}^{-1}) = a\mathbf{I}_p$*,* $\mathrm{E}(\mathbf{W_I}^{-1}\mathbf{W_I}^{-1}) = b\mathbf{I}_p$ *and* $\mathrm{var}\{\mathrm{vec}(\mathbf{W_I}^{-1})\} = c_1(\mathbf{I}_{p^2} + \mathbf{C}_{p^2}) + 2c_2\mathrm{vec}(\mathbf{I}_p)\mathrm{vec}^T(\mathbf{I}_p)$*, where* $\mathbf{C}_{p^2}$ *is the* $p^2 \times p^2$ *commutation matrix,* $a = (n-p-1)^{-1}$*,* $b = (n-1)c_1$*,* $c_1^{-1} = (n-p)(n-p-1)(n-p-3)$*, and* $c_2^{-1} = (n-p)(n-p-1)^2(n-p-3)$*.*

The following corollary gives the implications of this lemma that are relevant for studying $\widehat{\mathbf{R}}(\mathbf{X}_N) - \mathbf{R}(\mathbf{X}_N)$.

COROLLARY H.1. *If* $p/n \to [0, 1)$ *as* $p, n \to \infty$ *then* $a \asymp n^{-1}$*,* $b \asymp n^{-2}$*,* $c_1 \asymp n^{-3}$ *and* $c_2 \asymp n^{-4}$*.*

LEMMA H.2. *Let* $\mathbf{W_I} \sim W_p(\mathbf{I}, n)$*, let the elements of* $\mathbf{N} \in \mathbb{R}^{r\times p}$ *be i.i.d. N(0,1) random variables, and let* $\mathbf{L}_1 \in \mathbb{R}^{p\times q}$ *and* $\mathbf{L}_2 \in \mathbb{R}^{r\times s}$ *be fixed matrices so that* $\mathbf{W_I}$ *is independent of* $\mathbf{L}_2^T\mathbf{N}$*. Then, with* a*,* b*,* c_1 *and* c_2 *as defined in Lemma H.1,*

$$\mathrm{var}\{\mathrm{tr}(\mathbf{W_I}^{-1})\} = 2c_1p + 2c_2p^2 \tag{H.1}$$

$$\mathrm{var}\{\mathrm{vec}(\mathbf{L}_2^T\mathbf{N}\mathbf{W_I}^{-1}\mathbf{L}_1)\} = b\mathbf{L}_1^T\mathbf{L}_1 \otimes \mathbf{L}_2^T\mathbf{L}_2 \tag{H.2}$$

$$\mathrm{E}(\mathbf{L}_2^T\mathbf{N}\mathbf{W_I}^{-1}\mathbf{N}^T\mathbf{L}_2) = ap\mathbf{L}_2^T\mathbf{L}_2 \tag{H.3}$$

$$\begin{aligned}\mathrm{var}\{\mathrm{vec}(\mathbf{L}_2^T\mathbf{N}\mathbf{W_I}^{-1}\mathbf{N}^T\mathbf{L}_2)\} &= (2c_1p + 2c_2p^2)\mathrm{vec}(\mathbf{L}_2^T\mathbf{L}_2)\mathrm{vec}^T(\mathbf{L}_2^T\mathbf{L}_2) \\ &\quad + bp(\mathbf{I}_{s^2} + \mathbf{C}_{s^2})(\mathbf{L}_2^T\mathbf{L}_2 \otimes \mathbf{L}_2^T\mathbf{L}_2).\end{aligned} \tag{H.4}$$

PROOF. The first conclusion (H.1) is found as follows:

$$\begin{aligned}\mathrm{var}\left(\mathrm{tr}(\mathbf{W_I}^{-1})\right) &= \mathrm{E}\left(\mathrm{tr}^2(\mathbf{W_I}^{-1})\right) - \mathrm{E}^2\left(\mathrm{tr}(\mathbf{W_I}^{-1})\right) \\ \mathrm{tr}(\mathbf{W_I}^{-1}) &= \mathrm{vec}^T(\mathbf{W_I}^{-1})\mathrm{vec}(\mathbf{I}_p) \\ \mathrm{tr}^2(\mathbf{W_I}^{-1}) &= \mathrm{vec}^T(\mathbf{I}_p)\mathrm{vec}(\mathbf{W_I}^{-1})\mathrm{vec}^T(\mathbf{W_I}^{-1})\mathrm{vec}(\mathbf{I}_p).\end{aligned}$$

Therefore, using Lemma H.1 and letting $\mathbf{A} = \mathrm{E}\{\mathrm{vec}(\mathbf{W_I}^{-1})\}$,

$$\begin{aligned}\mathrm{E}\,\mathrm{tr}^2(\mathbf{W_I}^{-1}) &= \mathrm{vec}^T(\mathbf{I}_p)\mathrm{E}\left(\mathrm{vec}(\mathbf{W_I}^{-1})\mathrm{vec}^T(\mathbf{W_I}^{-1})\right)\mathrm{vec}(\mathbf{I}_p) \\ &= \mathrm{vec}^T(\mathbf{I}_p)\mathrm{var}\left(\mathrm{vec}(\mathbf{W_I}^{-1})\right)\mathrm{vec}(\mathbf{I}_p) \\ &\quad + \mathrm{vec}^T(\mathbf{I}_p)\mathbf{A}\mathbf{A}^T\mathrm{vec}(\mathbf{I}_p) \\ &= \mathrm{vec}^T(\mathbf{I}_p)\left[c_1(\mathbf{I}_{p^2} + \mathbf{C}_{p^2}) + 2c_2\mathrm{vec}(\mathbf{I}_p)\mathrm{vec}^T(\mathbf{I}_p)\right]\mathrm{vec}(\mathbf{I}_p) \\ &\quad + \mathrm{vec}^T(\mathbf{I}_p)\mathbf{A}\mathbf{A}^T\mathrm{vec}(\mathbf{I}_p) \\ &= 2c_1p + 2c_2p^2 + \mathrm{vec}^T(\mathbf{I}_p)\mathbf{A}\mathbf{A}^T\mathrm{vec}(\mathbf{I}_p) \\ \mathrm{E}^2\left(\mathrm{tr}(\mathbf{W_I}^{-1})\right) &= \mathrm{vec}^T(\mathbf{I}_p)\mathbf{A}\mathbf{A}^T\mathrm{vec}(\mathbf{I}_p).\end{aligned}$$

For (H.2) and (H.3) we have, using Lemma H.1 and F.2,

$$\begin{aligned} \mathrm{var}\{\mathrm{vec}(\mathbf{L}_2^T\mathbf{N}\mathbf{W}_{\mathbf{I}}^{-1}\mathbf{L}_1)\} &= \mathrm{E}\{(\mathbf{L}_1^T\mathbf{W}_{\mathbf{I}}^{-1}\otimes\mathbf{L}_2^T)(\mathbf{W}_{\mathbf{I}}^{-1}\mathbf{L}_1\otimes\mathbf{L}_2)\} \\ &= \mathbf{L}_1^T\mathrm{E}(\mathbf{W}_{\mathbf{I}}^{-1}\mathbf{W}_{\mathbf{I}}^{-1})\mathbf{L}_1\otimes\mathbf{L}_2^T\mathbf{L}_2 = b\mathbf{L}_1^T\mathbf{L}_1\otimes\mathbf{L}_2^T\mathbf{L}_2. \\ \mathrm{E}(\mathbf{L}_2^T\mathbf{N}\mathbf{W}_{\mathbf{I}}^{-1}\mathbf{N}\mathbf{L}_2) &= a\mathbf{L}_2^T\mathrm{E}(\mathbf{N}\mathbf{N}^T)\mathbf{L}_2 = ap\mathbf{L}_2^T\mathbf{L}_2. \end{aligned}$$

Let $\mathbf{A} = \mathrm{var}\{\mathrm{vec}(\mathbf{L}_2^T\mathbf{N}\mathbf{W}_{\mathbf{I}}^{-1}\mathbf{N}^T\mathbf{L}_2)\}$. Result (H.4) makes use of Lemmas H.1 and F.2:

$$\begin{aligned} \mathbf{A} &= \mathrm{var}\{\mathrm{vec}[\mathrm{E}(\mathbf{L}_2^T\mathbf{N}\mathbf{W}_{\mathbf{I}}^{-1}\mathbf{N}^T\mathbf{L}_2)|\mathbf{W}_{\mathbf{I}}]\} + \mathrm{E}[\mathrm{var}\{\mathrm{vec}(\mathbf{L}_2^T\mathbf{N}\mathbf{W}_{\mathbf{I}}^{-1}\mathbf{N}^T\mathbf{L}_2)|\mathbf{W}_{\mathbf{I}}\}] \\ &= \mathrm{var}\{\mathrm{vec}(\mathbf{L}_2^T\mathbf{L}_2)\,\mathrm{tr}(\mathbf{W}_{\mathbf{I}}^{-1})\} + \mathrm{E}[(\mathbf{I}_{s^2}+\mathbf{C}_{s^2})(\mathbf{L}_2^T\mathbf{L}_2\otimes\mathbf{L}_2^T\mathbf{L}_2)\,\mathrm{tr}(\mathbf{W}_{\mathbf{I}}^{-1}\mathbf{W}_{\mathbf{I}}^{-1})] \\ &= \mathrm{var}\{\,\mathrm{tr}(\mathbf{W}_{\mathbf{I}}^{-1})\}\mathrm{vec}(\mathbf{L}_2^T\mathbf{L}_2)\mathrm{vec}^T(\mathbf{L}_2^T\mathbf{L}_2) + bp(\mathbf{I}_{s^2}+\mathbf{C}_{s^2})(\mathbf{L}_2^T\mathbf{L}_2\otimes\mathbf{L}_2^T\mathbf{L}_2) \\ &= (2c_1p+2c_2p^2)\mathrm{vec}(\mathbf{L}_2^T\mathbf{L}_2)\mathrm{vec}^T(\mathbf{L}_2^T\mathbf{L}_2) + bp(\mathbf{I}_{s^2}+\mathbf{C}_{s^2})(\mathbf{L}_2^T\mathbf{L}_2\otimes\mathbf{L}_2^T\mathbf{L}_2). \end{aligned}$$

□

The following lemma specializes Lemma F.4 for normal errors. Recall that $\widehat{\boldsymbol{\Delta}}$ was defined in (5.3).

LEMMA H.3. *Assume the context and conditions of Proposition 6.2. Let* $\boldsymbol{\alpha}\in\mathbb{R}^{n\times s}$ *and* $\boldsymbol{\gamma}\in\mathbb{R}^{d\times l}$ *be nonstochastic matrices with* s, l *and* d *fixed,* $\boldsymbol{\gamma}=O(1)$ *and* $\boldsymbol{\alpha}^T\boldsymbol{\alpha}=O(n)$. *Assume that* $\boldsymbol{\alpha}^T\mathbf{e}\perp\!\!\!\perp\widehat{\boldsymbol{\Delta}}$ *and* $n>p+r+4$. *Then*

(i). $\mathrm{E}(\boldsymbol{\alpha}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma}\boldsymbol{\gamma})=0$ *and* $(h(p)n)^{-1}\boldsymbol{\alpha}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma}\boldsymbol{\gamma}=O_p((nh)^{-1/2})$.

(ii). $(h(p)n^2)^{-1}\mathrm{E}(\boldsymbol{\alpha}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\mathbf{e}^T\boldsymbol{\alpha})=O(\kappa^2)$ *and*

$$(h(p)n^2)^{-1}\left(\boldsymbol{\alpha}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\mathbf{e}^T\boldsymbol{\alpha}-\mathrm{E}(\boldsymbol{\alpha}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\mathbf{e}^T\boldsymbol{\alpha})\right)=O_p((h\sqrt{n})^{-1}).$$

(iii). $h^{-1}\mathrm{E}(\boldsymbol{\gamma}^T\boldsymbol{\Gamma}^T\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma}\boldsymbol{\gamma})=a\widetilde{n}\boldsymbol{\gamma}^T\mathbf{G}_h\boldsymbol{\gamma}=O(1)$ *and*

$$h^{-1}\left(\boldsymbol{\gamma}^T\boldsymbol{\Gamma}^T\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma}\boldsymbol{\gamma}-\mathrm{E}(\boldsymbol{\gamma}^T\boldsymbol{\Gamma}^T\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma}\boldsymbol{\gamma})\right)=O_p(n^{-1/2}).$$

PROOF. We will use repeatedly that $\widehat{\boldsymbol{\Delta}}^{-1}\sim\widetilde{n}\boldsymbol{\Delta}^{-1/2}\mathbf{W}_{\mathbf{I}}^{-1}\boldsymbol{\Delta}^{-1/2}$ with $\mathbf{W}_{\mathbf{I}}\sim W_p(\mathbf{I},\widetilde{n})$ and $\widetilde{n}=n-r-1$.

(i). Since $\widehat{\boldsymbol{\Delta}}\perp\!\!\!\perp\boldsymbol{\alpha}^T\mathbf{e}$ we have $\mathrm{E}(\boldsymbol{\alpha}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma}\boldsymbol{\gamma})=0$, so we need to consider only its variance. Applying (H.2) of Lemma H.2 with $\mathbf{L}_2=\boldsymbol{\alpha}$ and $\mathbf{L}_1=\boldsymbol{\Delta}^{-1/2}\boldsymbol{\Gamma}\boldsymbol{\gamma}$ we get

$$\begin{aligned} \mathrm{var}\{(nh)^{-1}\mathrm{vec}(\boldsymbol{\alpha}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma}\boldsymbol{\gamma})\} &= (\widetilde{n}/nh)^2\mathrm{var}\{\mathrm{vec}(\boldsymbol{\alpha}^T\mathbf{e}\boldsymbol{\Delta}^{-1/2}\mathbf{W}_{\mathbf{I}}^{-1}\boldsymbol{\Delta}^{-1/2}\boldsymbol{\Gamma}\boldsymbol{\gamma})\} \\ &= (\widetilde{n}/nh)^2(b\boldsymbol{\gamma}^T\boldsymbol{\Gamma}^T\boldsymbol{\Delta}^{-1}\boldsymbol{\Gamma}\boldsymbol{\gamma}\otimes\boldsymbol{\alpha}^T\boldsymbol{\alpha})) \\ &= O_p(1/nh), \end{aligned}$$

where we use $b\asymp n^{-2}$ from Corollary H.1.

(ii). From (H.3) of Lemma H.2 with $\mathbf{L}_2 = \boldsymbol{\alpha}$ we have

$$\mathrm{E}((hn^2)^{-1}\boldsymbol{\alpha}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\mathbf{e}^T\boldsymbol{\alpha}) = (\widetilde{n}ap/hn^2)\boldsymbol{\alpha}^T\boldsymbol{\alpha} = O_p(\kappa^2),$$

where we use $a \asymp n^{-1}$ from Corollary H.1. Next, using (H.4) of Lemma H.2 and letting $\mathbf{A} = \boldsymbol{\alpha}^T\boldsymbol{\alpha}$, we have

$$\begin{aligned}\mathrm{var}((hn^2)^{-1}\mathrm{vec}(\boldsymbol{\alpha}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\mathbf{e}^T\boldsymbol{\alpha})) &= (\widetilde{n}/hn^2)^2(2c_1p + 2c_2p^2)\mathrm{vec}(\mathbf{A})\mathrm{vec}^T(\mathbf{A}) \\ &\quad +(\widetilde{n}/hn^2)^2bp(\mathbf{I}_{s^2} + \mathbf{C}_{s^2})(\mathbf{A}\otimes\mathbf{A}) \\ &= O(1/nh^2),\end{aligned}$$

where we used $b \asymp n^{-2}$, $c_1 \asymp n^{-3}$, $c_2 \asymp n^{-4}$ from Corollary H.1 and $n > p + r + 4$.

(iii). Let $\mathbf{T} = \boldsymbol{\gamma}^T\boldsymbol{\Gamma}^T\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma}\boldsymbol{\gamma}$. From Lemma H.1, $\mathrm{E}(h^{-1}\mathbf{T}) = a\widetilde{n}\boldsymbol{\gamma}^T\mathbf{G}_h\boldsymbol{\gamma} = O(1)$, where we use $a \asymp n^{-1}$ from Corollary H.1. Let $\mathbf{A} = \boldsymbol{\Delta}^{-1/2}\boldsymbol{\Gamma}\boldsymbol{\gamma} \otimes \boldsymbol{\Delta}^{-1/2}\boldsymbol{\Gamma}\boldsymbol{\gamma}$. Then

$$\begin{aligned}\mathrm{var}\{h^{-1}\mathrm{vec}(\mathbf{T})\} &= (\widetilde{n}/h)^2\mathbf{A}^T\left(c_1(\mathbf{I}_{p^2} + \mathbf{C}_{p^2}) + 2c_2\mathrm{vec}(\mathbf{I}_p)\mathrm{vec}^T(\mathbf{I}_p)\right)\mathbf{A} \\ &= O_p(1/n),\end{aligned}$$

where we used $c_1 \asymp n^{-3}$ and $c_2 \asymp n^{-4}$ from Corollary H.1. □

H.2. Proposition 6.2 parts (i) and (ii). Recall that the model for this proposition is $\mathbb{X} = \mathbf{1}_n\boldsymbol{\mu}^T + \mathbb{F}\boldsymbol{\beta}^T\boldsymbol{\Gamma}^T + \mathbf{e}$ where $\mathbf{e}$ is normally distributed with mean 0 and variance $\mathrm{var}(\mathrm{vec}(\mathbf{e}^T)) = \mathbf{I}_n \otimes \boldsymbol{\Delta}$. Consequently, we have $\mathbb{F}^T\mathbb{Z} = \mathbb{F}^T\mathbb{F}\boldsymbol{\beta}^T\boldsymbol{\Gamma}^T + \mathbb{F}^T\mathbf{e}$. Additionally, we will encounter terms that contain the factor $a\widetilde{n}\boldsymbol{\Delta}^{-1}$, which arises because $\mathrm{E}(\widehat{\boldsymbol{\Delta}}^{-1}) = a\widetilde{n}\boldsymbol{\Delta}^{-1}$, where a is as defined in Lemma H.1.

Consider now the convergence of $\widehat{\mathbf{K}}$ with $\widehat{\mathbf{W}} = \widehat{\boldsymbol{\Delta}}^{-1} = \widetilde{n}(\mathbb{Z}^T\mathbf{Q}_{\mathbb{F}}\mathbb{Z})^{-1}$, where $\mathbb{Z}^T\mathbf{Q}_{\mathbb{F}}\mathbb{Z} \sim W_p(\boldsymbol{\Delta}, \widetilde{n})$, where $\widetilde{n} = n - r - 1$ as defined previously. Our constructions require that the moments given in Lemma H.1 exist. Accordingly we require throughout that $\widetilde{n} > p+3$. Recall that $\kappa = (p/nh)^{1/2}$.

To justify Proposition 6.2 it is enough to consider

$$h^{-1}\boldsymbol{\Phi}_n^{1/2}\widehat{\mathbf{K}}\boldsymbol{\Phi}_n^{1/2} = (n^2h)^{-1}\mathbb{F}^T\mathbb{Z}\widehat{\boldsymbol{\Delta}}^{-1}\mathbb{Z}^T\mathbb{F},$$

where $\widehat{\boldsymbol{\Delta}} \perp\!\!\!\perp \mathbb{F}^T\mathbb{Z}$. Substituting for $\mathbb{F}^T\mathbb{Z}$ it is sufficient to study the following three terms: $\mathbf{T}_1 = h^{-1}\boldsymbol{\Phi}_n\boldsymbol{\beta}^T\boldsymbol{\Gamma}^T\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma}\boldsymbol{\beta}\boldsymbol{\Phi}_n$, $\mathbf{T}_2 = (nh)^{-1}\mathbb{F}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma}\boldsymbol{\beta}\boldsymbol{\Phi}_n$, and $\mathbf{T}_3 = (n^2h)^{-1}\mathbb{F}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\mathbf{e}^T\mathbb{F}$. Using Lemma H.3(iii) with $\boldsymbol{\gamma} = \boldsymbol{\beta}\boldsymbol{\Phi}_n$, $\mathrm{E}(\mathbf{T}_1) = a\widetilde{n}\boldsymbol{\Phi}_n\boldsymbol{\beta}^T\mathbf{G}_h\boldsymbol{\beta}\boldsymbol{\Phi}_n = O(1)$, and $\mathbf{T}_1 - \mathrm{E}(\mathbf{T}_1) = O_p(n^{-1/2})$. From Lemma H.3(i) with $\boldsymbol{\alpha} = \mathbb{F}$ and $\boldsymbol{\gamma} = \boldsymbol{\beta}\boldsymbol{\Phi}_n$, $\mathrm{E}(\mathbf{T}_2) = 0$ and $\mathbf{T}_2 - \mathrm{E}(\mathbf{T}_2) = O_p((nh)^{-1/2})$. To address $\mathbf{T}_3$ we use Lemma H.3(ii) with $\boldsymbol{\alpha} = \mathbb{F}$ to get $\mathrm{E}(\mathbf{T}_3) = (\widetilde{n}ap/hn^2)\mathbb{F}^T\mathbb{F} = O(\kappa^2)$ and $\mathbf{T}_3 - \mathrm{E}(\mathbf{T}_3) = O_p((h\sqrt{n})^{-1})$. □

H.3. Proposition 6.2 part (iii).

H.3.1. *Preliminary results.* The following proposition summarizes the orders of various terms that we will encounter in our study of $\widehat{\mathbf{R}}_{\widehat{\boldsymbol{\Delta}}}(\mathbf{X}_N) - \mathbf{R}(\mathbf{X}_N)$. In this section, $\mathbf{G}_h = \boldsymbol{\Gamma}^T\boldsymbol{\Delta}^{-1}\boldsymbol{\Gamma}/h$ and $\widehat{\mathbf{G}}_h = \widehat{\boldsymbol{\Gamma}}^T\widehat{\boldsymbol{\Delta}}^{-1}\widehat{\boldsymbol{\Gamma}}/h$

PROPOSITION H.1. *Assume the conditions of Proposition 6.2 and again let a, b, c_1 and c_2 be as defined in Lemma H.1.*

(i) $\widehat{\mathbf{V}}_d = \boldsymbol{\Phi}_n^{1/2}\boldsymbol{\beta}^T + O_p(1/\sqrt{n})$, *where the columns* $\widehat{\mathbf{V}}_d$ *are the first d eigenvectors of* $\widehat{\mathbf{K}}$.
(ii) $\widehat{\mathbf{G}}_h = O_p(1)$.
(iii) $h^{-1}(p)\boldsymbol{\Gamma}^T(\widehat{\boldsymbol{\Delta}}^{-1} - a\widetilde{n}\boldsymbol{\Delta}^{-1})\boldsymbol{\Gamma} = O_p(1/\sqrt{n})$.
(iv) $\widehat{\mathbf{G}}_h - a\widetilde{n}\mathbf{G}_h = O_p(\kappa^2) + O_p(n^{-1/2})$.
(v) $\widehat{\mathbf{G}}_h - h^{-1}(p)\widehat{\boldsymbol{\Gamma}}^T\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma} = O_p(\kappa^2) + O_p(n^{-1/2})$.

PROOF. The proofs for parts (i) and (ii) follow the same logic as the proofs for parts (i) and (ii) of Proposition F.1, only the order changes. Part (iii) follows immediately from Lemma H.3(iii) with $\boldsymbol{\gamma} = \mathbf{I}_d$.

(iv). The proof of part (iv) follows the same general steps as the proof of Proposition F.1(iv). In particular, the expansion at (F.4) is valid here, replacing $\widehat{\mathbf{W}}$ by $\widehat{\boldsymbol{\Delta}}^{-1}$ and defining the $d \times r$ matrix $\mathbf{O}_n = O_p(n^{-1/2})$, again leaving three terms to study: Define the transient notation $\mathbf{A}_1 = \boldsymbol{\Gamma}\boldsymbol{\beta}(\boldsymbol{\beta}\boldsymbol{\Phi}_n + \mathbf{O}_{\kappa\omega})^T$, $\mathbf{A}_2 = \mathbf{e}^T\mathbb{F}(\boldsymbol{\beta} + \mathbf{O}_{\kappa\omega})^T$, $\mathbf{T}_1 = \mathbf{A}_1^T\widehat{\mathbf{W}}\mathbf{A}_1$, $\mathbf{T}_2 = n^{-1}\mathbf{A}_1^T\widehat{\mathbf{W}}\mathbf{A}_2$ and $\mathbf{T}_3 = n^{-2}\mathbf{A}_2^T\widehat{\mathbf{W}}\mathbf{A}_2$. Then

$$\begin{aligned}\widehat{\boldsymbol{\Gamma}}^T\widehat{\boldsymbol{\Delta}}^{-1}\widehat{\boldsymbol{\Gamma}} &= \mathbf{A}_1^T\widehat{\boldsymbol{\Delta}}^{-1}\mathbf{A}_1 + n^{-1}\mathbf{A}_2^T\widehat{\boldsymbol{\Delta}}^{-1}\mathbf{A}_1 + n^{-1}\mathbf{A}_1^T\widehat{\boldsymbol{\Delta}}^{-1}\mathbf{A}_2 + n^{-2}\mathbf{A}_2^T\widehat{\boldsymbol{\Delta}}^{-1}\mathbf{A}_2 \\ &= \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_2^T + \mathbf{T}_3.\end{aligned}$$

Expanding $\mathbf{T}_1$ we have

$$\mathbf{T}_1 = \boldsymbol{\Gamma}^T\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma} + \mathbf{O}_n\boldsymbol{\beta}^T\boldsymbol{\Gamma}^T\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma} + \boldsymbol{\Gamma}^T\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma}\boldsymbol{\beta}\mathbf{O}_n^T + \mathbf{O}_n\boldsymbol{\beta}^T\boldsymbol{\Gamma}^T\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma}\boldsymbol{\beta}\mathbf{O}_n^T$$

and thus by part (iii) $h^{-1}(\mathbf{T}_1 - a\widetilde{n}\boldsymbol{\Gamma}^T\boldsymbol{\Delta}^{-1}\boldsymbol{\Gamma}) = O_p(1/\sqrt{n})$. Next, expanding

$$\begin{aligned}\mathbf{T}_2 = {} & n^{-1}\boldsymbol{\beta}\mathbb{F}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma} + n^{-1}\mathbf{O}_n\mathbb{F}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma} + n^{-1}\boldsymbol{\beta}\mathbb{F}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma}\boldsymbol{\beta}\mathbf{O}_n^T \\ & + n^{-1}\mathbf{O}_n\mathbb{F}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Gamma}\boldsymbol{\beta}\mathbf{O}_n^T\end{aligned}$$

and using Lemma H.3(i) with $\boldsymbol{\alpha} = \mathbb{F}\boldsymbol{\beta}^T$ and $\gamma = \mathbf{I}_d$ we have $h^{-1}\mathbf{T}_2 = O_p(1/\sqrt{hn})$. Finally, expanding $\mathbf{T}_3$ and using Lemma H.3(ii) with $\boldsymbol{\alpha} = \mathbb{F}\boldsymbol{\beta}^T$

and the transient notation $\mathbf{A} = \mathbb{F}^T \mathbf{e} \widehat{\boldsymbol{\Delta}}^{-1} \mathbf{e}^T \mathbb{F}$,

$$\begin{aligned} h^{-1}\mathbf{T}_3 &= h^{-1}n^{-2}\boldsymbol{\beta}\mathbf{A}\boldsymbol{\beta}^T + h^{-1}n^{-2}\boldsymbol{\beta}\mathbf{A}\mathbf{O}_n^T + h^{-1}n^{-2}\mathbf{O}_n\mathbf{A}\boldsymbol{\beta}^T \\ &\quad + h^{-1}n^{-2}\mathbf{O}_n\mathbf{A}\mathbf{O}_n^T \\ &= O_p(\kappa^2) + O_p((h\sqrt{n})^{-1}). \end{aligned}$$

(v). As before we expand $\widehat{\boldsymbol{\Gamma}}^T \widehat{\boldsymbol{\Delta}}^{-1} \boldsymbol{\Gamma}$ and we get

$$\begin{aligned} \widehat{\boldsymbol{\Gamma}}^T \widehat{\boldsymbol{\Delta}}^{-1} \boldsymbol{\Gamma} &= \boldsymbol{\Gamma}^T \widehat{\boldsymbol{\Delta}}^{-1} \boldsymbol{\Gamma} + \mathbf{O}_n \boldsymbol{\beta}^T \boldsymbol{\Gamma}^T \widehat{\boldsymbol{\Delta}}^{-1} \boldsymbol{\Gamma} + n^{-1} \boldsymbol{\beta} \mathbb{F}^T \mathbf{e} \widehat{\boldsymbol{\Delta}}^{-1} \boldsymbol{\Gamma} \\ &\quad + n^{-1} \mathbf{O}_n \mathbb{F}^T \mathbf{e} \widehat{\boldsymbol{\Delta}}^{-1} \boldsymbol{\Gamma} \\ h^{-1} \widehat{\boldsymbol{\Gamma}}^T \widehat{\boldsymbol{\Delta}}^{-1} \boldsymbol{\Gamma} &= h^{-1} \boldsymbol{\Gamma}^T \widehat{\boldsymbol{\Delta}}^{-1} \boldsymbol{\Gamma} + O_p(1/\sqrt{n}), \end{aligned}$$

where the final equation follows from (i) and (iii) of Lemma H.3. Next, it follows immediately from the proof of part (iv) that $\widehat{\mathbf{G}}_h = h^{-1} \boldsymbol{\Gamma}^T \widehat{\boldsymbol{\Delta}}^{-1} \boldsymbol{\Gamma} + O_p(\kappa^2) + O_p(n^{-1/2})$, and thus that $\widehat{\mathbf{G}}_h - h^{-1} \widehat{\boldsymbol{\Gamma}}^T \widehat{\boldsymbol{\Delta}}^{-1} \boldsymbol{\Gamma} = O_p(\kappa^2) + O_p(n^{-1/2})$. □

H.3.2. *Proposition 6.2 part (iii).* We need to find the limiting behavior of

$$\begin{aligned} \widehat{\mathbf{R}}_{\widehat{\boldsymbol{\Delta}}}(\mathbf{X}_N) - \mathbf{R}(\mathbf{X}_N) &= (h^{-1} \widehat{\mathbf{G}}_h^{-1} \widehat{\boldsymbol{\Gamma}}^T \widehat{\boldsymbol{\Delta}}^{-1} \boldsymbol{\Gamma} - \mathbf{I}_d) \boldsymbol{\beta} \mathbf{f}_N \\ &\quad + h^{-1} \widehat{\mathbf{G}}_h^{-1} \widehat{\boldsymbol{\Gamma}}^T \widehat{\boldsymbol{\Delta}}^{-1} (\boldsymbol{\mu} - \bar{\mathbf{X}}) + h^{-1} \widehat{\mathbf{G}}_h^{-1} \widehat{\boldsymbol{\Gamma}}^T \widehat{\boldsymbol{\Delta}}^{-1} \boldsymbol{\varepsilon}_N \\ &\quad - h^{-1} \mathbf{G}_h^{-1} \boldsymbol{\Gamma}^T \boldsymbol{\Delta}^{-1} \boldsymbol{\varepsilon}_N \\ &= \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3 - \mathbf{T}_4. \end{aligned}$$

Term $\mathbf{T}_1$. It follows from Proposition H.1 (ii) and (v) that $\mathbf{T}_1 = O_p(\kappa^2) + O_p(n^{-1/2})$.

Term $\mathbf{T}_2$. We have $\mathbf{T}_2 = \widehat{\mathbf{G}}_h^{-1} \{\widehat{\boldsymbol{\Gamma}}^T \widehat{\boldsymbol{\Delta}}^{-1} (\boldsymbol{\mu} - \bar{\mathbf{X}})/h\}$. From Proposition H.1(ii), $\widehat{\mathbf{G}}_h = O_p(1)$. To find the order of the second factor we have

$$h^{-1} \left(\widehat{\boldsymbol{\Gamma}}^T \widehat{\boldsymbol{\Delta}}^{-1} (\boldsymbol{\mu} - \bar{\mathbf{X}}) \right) = \widehat{\mathbf{V}}_d^T \boldsymbol{\Phi}_n^{-1/2} (\mathbb{F}^T \mathbb{Z}/nh) \widehat{\boldsymbol{\Delta}}^{-1} (\boldsymbol{\mu} - \bar{\mathbf{X}}),$$

and, letting $\mathbf{A} = \mathbf{e} \widehat{\boldsymbol{\Delta}}^{-1} \mathbf{e}^T$,

$$\begin{aligned} (nh)^{-1} \mathbb{F}^T \mathbb{Z} \widehat{\boldsymbol{\Delta}}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}) &= (n^2 h)^{-1} \mathbb{F}^T \mathbb{X} \widehat{\boldsymbol{\Delta}}^{-1} \mathbf{e}^T \mathbf{1}_n \\ &= (n^2 h)^{-1} \mathbb{F}^T \mathbb{F} \boldsymbol{\beta}^T \boldsymbol{\Gamma}^T \widehat{\boldsymbol{\Delta}}^{-1} \mathbf{e}^T \mathbf{1}_n + (n^2 h)^{-1} \mathbb{F}^T \mathbf{e} \widehat{\boldsymbol{\Delta}}^{-1} \mathbf{e}^T \mathbf{1}_n \\ &= (nh)^{-1} \boldsymbol{\Phi}_n \boldsymbol{\beta}^T \boldsymbol{\Gamma}^T \widehat{\boldsymbol{\Delta}}^{-1} \mathbf{e}^T \mathbf{1}_n + (n^2 h)^{-1} \mathbb{F}^T \mathbf{e} \widehat{\boldsymbol{\Delta}}^{-1} \mathbf{e}^T \mathbf{1}_n \\ &= \mathbf{T}_{21} + \mathbf{T}_{22}. \end{aligned}$$

From Lemma H.3(i) with $\boldsymbol{\alpha} = \mathbf{1}_n$ and $\boldsymbol{\gamma} = \boldsymbol{\beta}\boldsymbol{\Phi}_n$ follows that $\mathbf{T}_{21} = O(1/\sqrt{nh})$. For $\mathbf{T}_{22}$ we get using Lemma H.3(ii) with $\boldsymbol{\alpha} = \mathbb{F}$ and with $\boldsymbol{\alpha} = \mathbf{1}_n$ that $\mathbf{T}_{22} = O_p(\kappa^2) + O_p(n^{-1/2})$.

Term $\mathbf{T}_3 - \mathbf{T}_4$.

$$\begin{aligned}
\mathbf{T}_3 - \mathbf{T}_4 &= \{h^{-1}\widehat{\mathbf{G}}_h^{-1}\widehat{\boldsymbol{\Gamma}}^T\widehat{\boldsymbol{\Delta}}^{-1} - h^{-1}\mathbf{G}_h^{-1}\boldsymbol{\Gamma}^T\boldsymbol{\Delta}^{-1}\}\boldsymbol{\epsilon}_N \\
&= h^{-1}\widehat{\mathbf{G}}_h^{-1}\left[\widehat{\boldsymbol{\Gamma}}^T\widehat{\boldsymbol{\Delta}}^{-1} - a\widetilde{n}\boldsymbol{\Gamma}^T\boldsymbol{\Delta}^{-1}\right]\boldsymbol{\varepsilon}_N \\
&\quad -h^{-1}\widehat{\mathbf{G}}_h^{-1}(\widehat{\mathbf{G}}_h - a\widetilde{n}\mathbf{G}_h)\mathbf{G}_h^{-1}\boldsymbol{\Gamma}^T\boldsymbol{\Delta}^{-1}\boldsymbol{\varepsilon}_N \\
&= \mathbf{T}_{341} + \mathbf{T}_{342}.
\end{aligned}$$

Now, using (F.3) we have

$$\begin{aligned}
\widehat{\mathbf{G}}_h\mathbf{T}_{341} &= h^{-1}\boldsymbol{\Gamma}^T\left(\widehat{\boldsymbol{\Delta}}^{-1} - a\widetilde{n}\boldsymbol{\Delta}^{-1}\right)\boldsymbol{\varepsilon}_N + h^{-1}\mathbf{O}_n\boldsymbol{\beta}^T\boldsymbol{\Gamma}^T\left(\widehat{\boldsymbol{\Delta}}^{-1} - a\widetilde{n}\boldsymbol{\Delta}^{-1}\right)\boldsymbol{\varepsilon}_N \\
&\quad +(a\widetilde{n}/h)\mathbf{O}_n\boldsymbol{\beta}^T\boldsymbol{\Gamma}^T\boldsymbol{\Delta}^{-1}\boldsymbol{\varepsilon}_N + (nh)^{-1}\boldsymbol{\beta}\mathbb{F}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\varepsilon}_N \\
&\quad +(nh)^{-1}\mathbf{O}_n\mathbb{F}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\varepsilon}_N \\
&= \mathbf{T}_{3411} + \mathbf{T}_{3412} + \mathbf{T}_{3413} + \mathbf{T}_{3414} + \mathbf{T}_{3415}
\end{aligned}$$

$\mathrm{E}(\mathbf{T}_{3411}) = 0$ and, letting $\mathbf{A} = \boldsymbol{\Delta}^{-1/2}(\widetilde{n}\mathbf{W}_{\mathbf{I}}^{-1} - a\widetilde{n}\mathbf{I}_p)\boldsymbol{\Delta}^{-1/2}$ and using Lemma H.1,

$$\begin{aligned}
\mathrm{var}\,(\mathbf{T}_{3411}) &= h^{-2}\boldsymbol{\Gamma}^T\mathrm{E}\left[\left(\widehat{\boldsymbol{\Delta}}^{-1} - a\widetilde{n}\boldsymbol{\Delta}^{-1}\right)\boldsymbol{\Delta}\left(\widehat{\boldsymbol{\Delta}}^{-1} - a\widetilde{n}\boldsymbol{\Delta}^{-1}\right)\right]\boldsymbol{\Gamma} \\
&= h^{-2}\boldsymbol{\Gamma}^T\mathrm{E}(\mathbf{A}\boldsymbol{\Delta}\mathbf{A})\boldsymbol{\Gamma} \\
&= h^{-2}\boldsymbol{\Gamma}^T\left[b\widetilde{n}^2\boldsymbol{\Delta}^{-1} - 2(a\widetilde{n})^2\boldsymbol{\Delta}^{-1} + (a\widetilde{n})^2\boldsymbol{\Delta}^{-1}\right]\boldsymbol{\Gamma} \\
&= \frac{b\widetilde{n}^2 - (a\widetilde{n})^2}{h}\mathbf{G}_h = O(\kappa^2).
\end{aligned}$$

With this we have that $\mathbf{T}_{3411}$ and $\mathbf{T}_{3412}$ are both $O_p(\kappa)$. Since $\mathbf{G}_h = O(1)$ we have that $\mathbf{T}_{3413} = O_p(1/\sqrt{n})$. The expectations of $\mathbf{T}_{3414}$ and $\mathbf{T}_{3415}$ are both 0, and their common factor has variance, using Lemmas F.2 and H.1,

$$\begin{aligned}
\mathrm{var}\left((hn)^{-1}\mathbb{F}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\varepsilon}_N\right) &= (nh)^{-2}\mathrm{E}\left(\mathbb{F}^T\mathbf{e}\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Delta}\widehat{\boldsymbol{\Delta}}^{-1}\mathbf{e}\mathbb{F}\right) \\
&= (nh)^{-2}\mathbb{F}^T\mathbb{F}\mathrm{E}\left(\mathrm{tr}(\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Delta}\widehat{\boldsymbol{\Delta}}^{-1}\boldsymbol{\Delta}\right) \\
&= (\widetilde{n}^2/nh^2)\boldsymbol{\Phi}_n\mathrm{E}\left(\mathrm{tr}(\mathbf{W}_{\mathbf{I}}^{-2})\right) \\
&= (bp\widetilde{n}^2/nh^2)\boldsymbol{\Phi}_n = O_p(p/nh^2) = O_p(\kappa^2/h).
\end{aligned}$$

and therefore $\mathbf{T}_{3414}$ and $\mathbf{T}_{3415}$ are both $O_p(\kappa/\sqrt{h})$ and, consequently, $\mathbf{T}_{341} = O_p(\kappa)$.

Term $\mathbf{T}_{342}$. From parts (ii) and (iv) of Proposition H.1 and the stability condition of Section 4.3, we get that $\mathbf{T}_{342} = O_p(\kappa^2) + O_p(n^{-1/2})$.

Combining the above results, we have that $\widehat{\mathbf{R}}_{\widehat{\boldsymbol{\Delta}}}(\mathbf{X}_N) - \mathbf{R}(\mathbf{X}_N) = O_p(\kappa^2) + O_p(n^{-1/2}) + O_p(\kappa) = O_p(\kappa)$. □

REFERENCES

Magnus, J. R., and Neudecker, H. (1979). The commutation matrix: Some properties and applications. *The Annals of Statistics* **7**, 381–394.

Tyler, D. E. (1981). Asymptotic inference for eigenvectors. *The Annals of Statistics* **9**, 725–736.

Vershynin, R. (2011). Introduction to the non-asymptotic analysis of random matrices. In Y. Eldar and G. Kutyniok, editors, *Compressed sensing: theory and applications*. Cambridge University Press. Submitted.

School of Statistics
University of Minnesota
Minneapolis, Minnesota 55455, USA
E-mail: dennis@stat.umn.edu; rothman@stat.umn.edu

Instituto de Matemática Aplicada del Litoral
and Facultad de Ingeniería Química, CONICET
and UNL, Güemes 3450, (3000) Santa Fe, Argentina.
E-mail: liliana.forzani@gmail.com