Biometrika (2018), **xx**, x, *pp*. 1–19
Printed in Great Britain

Supplementary material for 'Robust estimation of high-dimensional covariance and precision matrices'

BY MARCO AVELLA-MEDINA
Sloan School of Management, Massachusetts Institute of Technology, 30 Memorial Drive, Cambridge, Massachusetts 02142, U.S.A.
mavella@mit.edu

HEATHER S. BATTEY
Department of Mathematics, Imperial College London, 545 Huxley Building, South Kensington Campus, London SW7 2AZ, U.K.
h.battey@imperial.ac.uk

JIANQING FAN
Department of Operations Research and Financial Engineering, Princeton University, 205 Sherred Hall, Princeton, New Jersey 08540, U.S.A.
jqfan@princeton.edu

AND QUEFENG LI
Department of Biostatistics, University of North Carolina at Chapel Hill, 3105D McGavran-Greenberg Hall, Chapel Hill, North Carolina 27599, U.S.A.
quefeng@email.unc.edu

SUMMARY

This supplementary file contains proofs of the theoretical results in the main paper, complementary discussions of different technical aspects of our theory, and an additional plot for our real-data example.

S1. PROOFS

Proof of Proposition 1. We adapt the arguments of Lemma 3 in Bubeck et al. (2013) to our setting. Without loss of generality, we assume $E(X) = 0$ and construct the distribution pr as

$$\mathrm{pr}(X_{1u}X_{1v} = 0) = 1 - \alpha^{1+\gamma}, \quad \mathrm{pr}(X_{1u}X_{1v} = 1/\alpha) = \alpha^{1+\gamma}$$

for some $\alpha \in (0, 1)$. It is easy to check that $\sigma^*_{uv} = E(X_u X_v) = \alpha^\gamma$ and $E(|X_u X_v - \sigma^*_{uv}|^{1+\gamma}) \leqslant 2$. Let $\eta = 2^{-1}\varepsilon^{-1/(1+\gamma)}n^{-\gamma/(1+\gamma)}$ and take $\alpha = (2n\eta)^{-1}$. If $\varepsilon < 1/2$, we have $\sigma^*_{uv} = \alpha^\gamma =$

$(2n\eta)^{-\gamma} < \eta$, which implies $1/\alpha = 2n\eta > n(\eta + \sigma^*_{uv})$. This leads to the bound

$$\begin{aligned}
\operatorname{pr}\big(|\hat{\sigma}_{uv} - \sigma^*_{uv}| > \eta\big) &\geqslant \operatorname{pr}\big\{\hat{\sigma}_{uv} - \sigma^*_{uv} > \eta\big\} \\
&\geqslant \operatorname{pr}\big(\exists\, i \in [n] : X_{iu}X_{iv} > n(\eta + \sigma^*_{uv})\big) \\
&\geqslant \operatorname{pr}\big(\exists\, i \in [n] : X_{iu}X_{iv} = 1/\alpha\big) \\
&= 1 - (1 - \alpha^{1+\gamma})^n \\
&= 1 - \exp\Big[n \log\Big\{1 - \frac{1}{(2n\eta)^{1+\gamma}}\Big\}\Big] \\
&\geqslant 1 - \exp\Big\{-\frac{1}{n^\gamma (2\eta)^{1+\gamma}}\Big\} \\
&\geqslant \frac{1}{n^\gamma (2\eta)^{1+\gamma}} \\
&= \varepsilon. \qquad \square
\end{aligned}$$

Proof of Theorem 1. Once the critical condition (1) is satisfied by the pilot estimator, the proof follows similar arguments as in the proof of Theorem 1 in Rothman et al. (2009) and the proof of Theorem 2 in Cai & Liu (2011). For ease of reference, the complete argument is given here.

Consider the event $\mathcal{E}_n = \mathsf{E}_1 \cap \mathsf{E}_2 = \{|\tilde{\sigma}_{uv} - \sigma^*_{uv}| \leqslant \lambda_{uv} \text{ for all } u, v\} \cap \{\tilde{\sigma}_{uu}\tilde{\sigma}_{vv} \leqslant 2\sigma^*_{uu}\sigma^*_{vv} \text{ for all } u, v\}$. We first show that on the event $\mathcal{E}_n$, $\|\hat{\Sigma}^{\mathcal{T}} - \Sigma^*\|_2 \leqslant C_0 s_0(p)\{(\log p)/n\}^{(1-q)/2}$, where C_0 is a constant depending only on q and λ. We then show that $\operatorname{pr}(\mathcal{E}_n) \geqslant 1 - \varepsilon_{n,p}$ for $\varepsilon_{n,p}$ a positive deterministic sequence converging to zero when $(\log p)/n \to 0$.

Under conditions (ii) and (iii) on τ_λ,

$$\begin{aligned}
&\sum_{v=1}^p \big|\tau_{\lambda_{uv}}(\tilde{\sigma}_{uv}) - \sigma^*_{uv}\big| = \sum_{v=1}^p \big|\tau_{\lambda_{uv}}(\tilde{\sigma}_{uv}) - \sigma^*_{uv}\big| \mathbb{1}(|\tilde{\sigma}_{uv}| \geqslant \lambda_{uv}) + \sum_{v=1}^p \big|\sigma^*_{uv}\big| \mathbb{1}(|\tilde{\sigma}_{uv}| < \lambda_{uv}) \\
&\leqslant 2\sum_{v=1}^p \lambda_{uv}\mathbb{1}(|\sigma^*_{uv}| \geqslant \lambda_{uv}) + \sum_{v=1}^p \big|\tau_{\lambda_{uv}}(\tilde{\sigma}_{uv}) - \sigma^*_{uv}\big| \mathbb{1}(|\tilde{\sigma}_{uv}| \geqslant \lambda_{uv}, |\sigma^*_{uv}| < \lambda_{uv}) \\
&\quad + \sum_{v=1}^p |\sigma^*_{uv}| \mathbb{1}(|\tilde{\sigma}_{uv}| < \lambda_{uv}) \\
&= T_1 + T_2 + T_3.
\end{aligned}$$

On the event E_1, $T_2 \leqslant 2\sum_{j=1}^p |\sigma^*_{uv}|\mathbb{1}(|\sigma^*_{uv}| < \lambda_{uv})$ and $T_3 \leqslant \sum_{v=1}^p |\sigma^*_{uv}|\mathbb{1}(|\sigma^*_{uv}| < 2\lambda_{uv})$, where the first inequality uses condition (i) on τ_λ and the second uses the reversed triangle inequality. Therefore, combining the above inequalities and using the fact that $\lambda_{uv}\mathbb{1}(|\sigma^*_{uv}| \geqslant \lambda_{uv}) \leqslant \lambda_{uv}^{1-q}|\sigma^*_{uv}|^q$, on the event E_2 we have

$$\begin{aligned}
\sum_{v=1}^p \big|\tau_{\lambda_{uv}}(\tilde{\sigma}_{uv}) - \sigma^*_{uv}\big| &\leqslant (4 + 2^{1-q}) \sum_{v=1}^p \lambda_{uv}^{1-q}|\sigma^*_{uv}|^q \\
&\leqslant C_0 s_0(p) \Big(\frac{\log p}{n}\Big)^{(1-q)/2}.
\end{aligned}$$

Since $\|A\|_2 \leqslant \|A\|_1$ for any symmetric matrix A, it remains to bound from below the probability of the event $\mathcal{E}_n$. Notice that

$$\tilde{\sigma}_{uu}\tilde{\sigma}_{vv} = \sigma^*_{uu}\sigma^*_{vv} + (\tilde{\sigma}_{uu} - \sigma^*_{uu})\tilde{\sigma}_{vv} + (\tilde{\sigma}_{vv} - \sigma^*_{vv})\tilde{\sigma}_{uu} - (\tilde{\sigma}_{uu} - \sigma^*_{uu})(\tilde{\sigma}_{vv} - \sigma^*_{vv}). \qquad \text{(S1)}$$

It therefore follows from Condition 1 that for large n,

$$\mathrm{pr}\Big(\tilde\sigma_{uu}\tilde\sigma_{vv} \geqslant \gamma^2/2 \text{ for all } u,v=1,\ldots,p\Big) \geqslant 1-\frac{\varepsilon_{n,p}}{4}, \tag{S2}$$

Then we have

$$\begin{aligned}\mathrm{pr}\left\{\max_{u,v}\frac{|\tilde\sigma_{uv}-\sigma^*_{uv}|}{(\tilde\sigma_{uu}\tilde\sigma_{vv})^{1/2}} \geqslant \lambda\left(\frac{\log p}{n}\right)^{1/2}\right\} &\leqslant \mathrm{pr}\left[\max_{u,v}|\tilde\sigma_{uv}-\sigma^*_{uv}| \geqslant \lambda\left\{\frac{\min_{uv}(\tilde\sigma_{uu}\tilde\sigma_{vv})\log p}{n}\right\}^{1/2}\right]\\ &\leqslant \mathrm{pr}\left\{\max_{u,v}|\tilde\sigma_{uv}-\sigma^*_{uv}| \geqslant \lambda\left(\frac{2^{-1}\gamma^2\log p}{n}\right)^{1/2}\right\}+\frac{\varepsilon_{n,p}}{4}\\ &\leqslant \frac{\varepsilon_{n,p}}{2}, \end{aligned}\tag{S3}$$

where the second inequality follows from (S2) and the last one from Condition 1. Since Condition 1 and (S1) imply that for n large enough, $\mathrm{pr}(\mathsf{E}_2) \geqslant 1-\varepsilon_{n,p}/2$, it follows from (S3) that

$$\mathrm{pr}(\mathcal{E}_n) \geqslant \mathrm{pr}(\mathcal{E}_n|\mathsf{E}_2)\,\mathrm{pr}(\mathsf{E}_2) \geqslant 1-\varepsilon_{n,p}$$

for large enough n. □

Proof of Theorem 2. The proof makes use of the following lemmas from Cai et al. (2016).

LEMMA S1. *Let A be a symmetric matrix. Then $\|A\|_2 \leqslant \|A\|_w \leqslant \|A\|_1$ for all $1 \leqslant w \leqslant \infty$.*

LEMMA S2. *Let $\bar\Omega$ be any estimator of Ω^* and set $t_n = \|\bar\Omega-\Omega^*\|_{\max}$. Then on the event $\{\|\bar\omega_v\|_1 \leqslant \|\omega^*_v\|_1$ for $1\leqslant v\leqslant p\}$, we have $\|\bar\Omega-\Omega^*\|_1 \leqslant 12c_{n,p}t_n^{1-q}$ where $c_{n,p}$ is the sequence in Condition* 4.

The same argument as provided in equation (5) of the main paper shows that $\tilde\Sigma^+$ achieves the same rate of convergence as $\tilde\Sigma$. By Lemma S1, it suffices to prove that

$$\inf_{\Omega^*\in\mathcal{U}_q}\mathrm{pr}\left\{\|\tilde\Omega-\Omega^*\|_1 \leqslant C_0M_{n,p}^{1-q}c_{n,p}\left(\frac{\log p}{n}\right)^{(1-q)/2}\right\} \geqslant 1-\varepsilon_{n,p}.$$

To this end, consider $\|\check\Omega^{(1)}-\Omega^*\|_{\max}$. Since $1/M_1 \leqslant \lambda_{\min}(\Omega^*) \leqslant \lambda_{\max}(\Omega^*) \leqslant M_1$, we have $\max_v(\sigma^*_{vv})\max_v(\omega^*_{vv}) \leqslant M_1^2$. Therefore, by Condition 3 the following bound holds with probability at least $1-\varepsilon_{n,p}$:

$$\begin{aligned}\|\check\Omega^{(1)}-\Omega^*\|_{\max} &= \|(\Omega^*\tilde\Sigma^+-I_p)\check\Omega^{(1)}+\Omega^*(I_p-\tilde\Sigma^+\check\Omega^{(1)})\|_{\max}\\ &\leqslant \|(\Omega^*\tilde\Sigma^+-I_p)\|_{\max}\|\check\Omega^{(1)}\|_1+\delta_{n,p}\max_v\tilde\sigma^+_{vv}\max_v\check\omega^{(1)}_{vv}\|\Omega^*\|_1\\ &\leqslant C\|\check\Omega^{(1)}\|_1\left(\frac{\log p}{n}\right)^{1/2}+2\delta\|\Omega^*\|_1\max_v\sigma^*_{vv}\max_v\check\omega^{(1)}_{vv}\left(\frac{\log p}{n}\right)^{1/2}\\ &\leqslant C\|\check\Omega^{(1)}\|_1\left(\frac{\log p}{n}\right)^{1/2}+2\delta\|\Omega^*\|_1\max_v\sigma^*_{vv}\max_v\omega^*_{vv}\left(\max_v\frac{\check\omega^{(1)}_{vv}}{\omega^*_{vv}}\right)\left(\frac{\log p}{n}\right)^{1/2}\\ &\leqslant C\|\check\Omega^{(1)}\|_1\left(\frac{\log p}{n}\right)^{1/2}+2\delta M_1^2\|\Omega^*\|_1\left(\max_v\frac{\check\omega^{(1)}_{vv}}{\omega^*_{vv}}\right)\left(\frac{\log p}{n}\right)^{1/2}\end{aligned}\tag{S4}$$

for some positive constant C. By construction and Condition 4, $\|\check\Omega^{(1)}\|_1 \leqslant \|\Omega^*\|_1 \leqslant c_{n,p}\max_v(|\omega_{vv}*|^{1-q})$ and $\max_v\omega^*_{vv}/\min_v\omega^*_{vv} \leqslant \lambda_{\max}(\Omega^*)/\lambda_{\min}(\Omega^*) \leqslant M_1^2$. Provided $\check\omega_{vv} = O(1)$, with probability at least $1-\varepsilon_{n,p}$, $\|\tilde\Omega-\Omega^*\|_1 \leqslant KM_{n,p}\{(\log p)/n\}^{1/2}$ for some constant K, where $M_{n,p}$ is as in Condition 4.

By Condition 3, Ω^* belongs to the feasible set (9) on an event of probability at least $1-\varepsilon_{n,p}$. Thus, by an analogous argument to that appearing in (S4), we obtain that on the same event, $\|\tilde{\Omega}-\Omega^*\|_{\max} \leqslant CM_{n,p}\{(\log p)/n\}^{1/2}$. It follows from Lemma S2 that on an event of probability at least $1-\varepsilon_{n,p}$, $\|\tilde{\Omega}-\Omega^*\|_1 \leqslant C_0 M_{n,p}^{1-q} c_{n,p}\{(\log p)/n\}^{(1-q)/2}$, where $c_{n,p}$ is as in Condition 4.□

Proof of Proposition 2. The proof relies on the following lemmas due to Liu et al. (2012, Theorem 4.2) and Serfling & Mazumder (2009, Theorem 1).

LEMMA S3. *Let $X_1,\ldots,X_n$ be independent and identically distributed copies of the elliptically distributed random vector X with covariance matrix* $\mathrm{var}(X)=\Sigma^*=D^*R^*D^*$*, and let $R^*=(r^*_{uv})$ and $\tilde{R}=(\tilde{r}_{uv})$ be as defined in* § 4·1 *of the main paper. For any $\delta\in(0,1)$,*

$$\mathrm{pr}\left\{\max_{u,v}|\tilde{r}_{uv}-r^*_{uv}| \geqslant 3\pi\left(\frac{2\log\delta^{-1}}{5n}\right)^{1/2}\right\} \leqslant \delta.$$

LEMMA S4. *Let $\tilde{\sigma}_{uu}$ be the median absolute deviation estimator of σ^*_{uu} defined in* § 4·1 *of the main paper. Then, for $\delta=C\varepsilon$ for every $\varepsilon>0$ and $C=1/F^{-1}(3/4)$,*

$$\mathrm{pr}\Big(|\tilde{\sigma}_{uu}-\sigma^*_{uu}|>\delta\Big) \leqslant 6\exp(-2n\Delta^2_{\delta,n}),$$

where $\Delta_{\delta,n}=\min(a^*_n,b^*_n,c^*_n,d^*_n)$ with

$$\begin{aligned}
a^*_n &= \big[F_u(\nu_u+\delta/2)-\{\lfloor(n+1)/2\rfloor-1\}/n\big]_+,\\
b^*_n &= \lfloor(n+1)/2\rfloor/n - F_u(\nu_u-\delta/2),\\
c^*_n &= \big\{F_u(\nu_u+\sigma^*_{uu}+\delta/2)-F_u(\nu_u-\sigma^*_{uu}-\delta/2)-\lfloor(n+1)/2\rfloor/n\big\}_+,\\
d^*_n &= \lfloor(n+1)/2\rfloor/n - \big[F_u(\nu_u+\sigma^*_{uu}-\delta/2)-F_u(\nu_u-\sigma^*_{uu}+\delta/2)\big].
\end{aligned}$$

Here F_u is the distribution function of X_u and $\nu_u=F_u^{-1}(1/2)$.

We now prove Proposition 2. Denote the components of $\tilde{\Sigma}_R$ by $\tilde{\Sigma}_R=(\tilde{\sigma}^R_{uv})$. Then we have

$$\begin{aligned}
\max_{u,v}|\tilde{\sigma}^R_{uv}-\sigma^*_{uv}| \leqslant{}& \max_{u,v}\big(|\tilde{\sigma}^{1/2}_{uu}-\sigma^{*1/2}_{uu}||\tilde{r}_{uv}||\tilde{\sigma}^{1/2}_{vv}-\sigma^{*1/2}_{vv}|\big)+2\max_{u,v}\big(|\tilde{\sigma}^{1/2}_{uu}-\sigma^{*1/2}_{uu}||\tilde{r}_{uv}|\sigma^{*1/2}_{vv}\big)\\
&+\max_{u,v}\big(\sigma^{*1/2}_{uu}|\tilde{r}_{uv}-r^*_{uv}|\sigma^{*1/2}_{vv}\big).
\end{aligned}$$

Thus, with $\varsigma=\max_u\sigma^{*1/2}_{uu}$ and noting that $\tilde{r}_{uv}\in[-1,1]$, we have

$$\max_{u,v}|\tilde{\sigma}^R_{uv}-\sigma^*_{uv}| \leqslant 2\varsigma\max_u|\tilde{\sigma}^{1/2}_{uu}-\sigma^{*1/2}_{uu}|+\varsigma^2\max_{u,v}|\tilde{r}_{uv}-r^*_{uv}|+\max_u|\tilde{\sigma}^{1/2}_{uu}-\sigma^{*1/2}_{uu}|^2.$$

We hence obtain, by the Bonferroni inequality,

$$\begin{aligned}
\mathrm{pr}\big(\max_{u,v}|\tilde{\sigma}^R_{uv}-\sigma^*_{uv}|>t\big) \leqslant{}& \mathrm{pr}\Big(\max_u|\tilde{\sigma}^{1/2}_{uu}-\sigma^{*1/2}_{uu}|>\frac{t}{6\varsigma}\Big)+\mathrm{pr}\Big(\max_{u,v}|\tilde{r}_{uv}-r^*_{uv}|>\frac{t}{3\varsigma^2}\Big)\\
&+\mathrm{pr}\Big(\max_u|\tilde{\sigma}^{1/2}_{uu}-\sigma^{*1/2}_{uu}|^2>\frac{t}{3}\Big)=P_1+P_2+P_3,
\end{aligned}$$

Taking $t=C\{(\log p)/n\}^{1/2}$ with $C=\pi\sqrt{2}/\varsigma^2$ ensures $P_2\leqslant p^{-1}$ by Lemma S3. For P_1, through a first-order Taylor series expansion with Lagrange remainder, we obtain

$$\max_u|\tilde{\sigma}^{1/2}_{uu}-\sigma^{*1/2}_{uu}| \leqslant 2\max_u\bar{Q}_u^{-1/2}|\tilde{\sigma}_{uu}-\sigma^*_{uu}|,$$

where $\bar{Q}_u$ lies on a line segment between $\tilde{\sigma}_{uu}^{1/2}$ and $\sigma_{uu}^{*1/2}$. So

$$\begin{aligned} P_1 &\leqslant \mathrm{pr}\Big(2\max_u \bar{Q}_u^{-1/2}|\tilde{\sigma}_{uu}-\sigma^*_{uu}| > \frac{t}{6\varsigma},\ \max_u \bar{Q}_u^{-1/2}\leqslant U\Big) + \mathrm{pr}\Big(\max_u \bar{Q}_u^{-1/2} > U\Big) \\ &\leqslant \mathrm{pr}\Big(\max_u |\tilde{\sigma}_{uu}-\sigma^*_{uu}| > \frac{t}{12U\varsigma}\Big) + \mathrm{pr}\Big(\max_u \bar{Q}_u^{-1/2} > U\Big) \\ &= P_{11}+P_{12}. \end{aligned}$$

We have $P_{11} \leqslant 6p\exp(-2n\Delta^2_{\varepsilon,n})$ by the Bonferroni inequality and Lemma S4, where $\varepsilon = C_0^*\{(\log p)/n\}^{1/2}$ with $C_0^* = \pi\sqrt{2}/(4U\varsigma)$ and $\Delta_{\delta,n}$ is defined in Lemma S4. Let m_u be the population median of the uth variable and let f_u be the corresponding density function. Then b_n^* of Lemma S4 satisfies $b_n^* = f_u(\tilde{m}_u)\delta + O(n^{-1})$, where $\tilde{m}_u$ lies on a line segment between m_u and $m_u+\delta$. Since f_u is continuous in a neighbourhood of the median and $f_u(m) > 0$ by the elliptical assumption, for $(\log p)/n$ sufficiently small we have $f_u(\tilde{m}) = C_1 > 0$. Then $b_n^* \geqslant C_1C_0^*\{(\log p)/n\}^{1/2}$. Similar calculations for a_n^*, c_n^* and d_n^* show that $\Delta_{\varepsilon,n} \geqslant C_1C_0^*\{(\log p)/n\}^{1/2}$, and so $P_{11} \leqslant 6p^{1-L}$ for $L = C_1C_0^*$, which is greater than 1 for C_0^* sufficiently large. This entails $U < C_1\pi\sqrt{2}/(4\varsigma)$. For the control over P_{12}, we have

$$\begin{aligned} P_{12} &\leqslant \mathrm{pr}\Big(\max_u \bar{Q}_u^{-1/2} > U,\ \max_u |\tilde{\sigma}_{uu}-\sigma^*_{uu}| \leqslant \xi\Big) + \mathrm{pr}\Big(\max_u |\tilde{\sigma}_{uu}-\sigma^*_{uu}| > \xi\Big) \\ &= P_{121}+P_{122}. \end{aligned}$$

By Bonferroni's inequality and Lemma S4, $P_{122} \leqslant 6p\exp(-2n\Delta^2_{\xi,n})$. Therefore, by similar arguments to those given for P_{11}, P_{122} can be bounded by a term of order p^{-1} by taking $\xi \geqslant C_2\{(\log p)/n\}^{1/2}$ for C_2 a sufficiently large constant depending only on U and ς. For P_{121},

$$\max_u \bar{Q}_u^{-1/2} \leqslant \max_u\,[\{\min(\tilde{\sigma}_{uu}^{1/2}, \sigma_{uu}^{*1/2})\}^{-1/2}] = \max_u\,\{\max(\tilde{\sigma}_{uu}^{-1/4}, \sigma_{uu}^{*-1/4})\}.$$

But on the event $\{\max_u(|\tilde{\sigma}_{uu}-\sigma^*_{uu}|) \leqslant \xi\}$, $\sigma^*_{uu}-\xi \leqslant \tilde{\sigma}_{uu} \leqslant \sigma^*_{uu}+\xi$ for all $u \in [p]$. Therefore

$$P_{121} \leqslant \mathrm{pr}\Big\{\max_u(\sigma^*_{uu}-\xi)^{-1/4} > U\Big\} = \mathrm{pr}\Big\{(\min_u \sigma^*_{uu}-\xi)^{-1/4} > U\Big\} = 0$$

because $\min_u \sigma^*_{uu} > \xi + U^{-4}$ by assumption. □

Proof of Proposition 3. We use the following lemma in the proof.

LEMMA S5 (THEOREM 5 IN FAN ET AL., 2017). *Let $Z_1,\ldots,Z_n$ be real-valued independent identically distributed random variables with $EZ_1 = \mu^*$ and* $\mathrm{var}(Z_1) = \sigma^{*2}$. *Let $\delta \in (0,1)$ be such that $n^{-1}\log\delta^{-1} \leqslant 1/8$ and $H = \{(nv^2)/\log\delta^{-1}\}^{1/2}$, where $v \geqslant \sigma^*$. Then Huber's M-estimator $\tilde{\mu}^H$ satisfies*

$$\mathrm{pr}\bigg\{|\tilde{\mu}^H-\mu| \geqslant 4v\Big(\frac{\log\delta^{-1}}{n}\Big)^{1/2}\bigg\} \leqslant 2\delta.$$

If $(2+L)(\log p)/n < 1/8$, Lemma S5 and the union bound guarantee that

$$\mathrm{pr}\Big[\max_u |\tilde{\mu}_u^H-\mu_u^*| \geqslant K\{(\log p)/n\}^{1/2}\Big] \leqslant 2p^{-(1+L)} \tag{S5}$$

and

$$\mathrm{pr}\Big[\max_{u,v} |\tilde{\mu}_{uv}^H-\mu_{uv}^*| \geqslant K\{(\log p)/n\}^{1/2}\Big] \leqslant 2p^{-L}. \tag{S6}$$

From (S5) we have, with probability at least $1-2p^{-(1+L)}$,

$$\max_{u,v}|\tilde\mu_u^H\tilde\mu_v^H-\mu_u^*\mu_v^*|\leqslant 2K\max_u|\mu_u^*|\{(\log p)/n\}^{1/2}+K^2(\log p)/n. \tag{S7}$$

Combining (S6) and (S7), for large n we have

$$\begin{aligned}&\mathrm{pr}\Big[\max_{u,v}|\tilde\sigma_{uv}^H-\sigma_{uv}^*|\geqslant C\{(\log p)/n\}^{1/2}\Big]\\ &\leqslant \mathrm{pr}\Big[\max_{u,v}|\tilde\mu_{uv}^H-\mu_{uv}^*|+\max_{u,v}|\tilde\mu_u^H\tilde\mu_v^H-\mu_u^*\mu_v^*|\geqslant C\{(\log p)/n\}^{1/2}\Big]\\ &\leqslant 2p^{-L}(1+p^{-1}).\end{aligned}$$

□

Proof of Proposition 4. We will start by stating and proving three auxiliary results. Lemma S6 extends a large deviation result of Petrov (1995) to random variables satisfying Cramér's condition only in a shrinking neighbourhood of zero. Lemma S7 shows that the weights associated with Huber's estimator are uniformly close to $1/n$ for large n and for all entries of the pilot covariance estimator. Lemma S8 shows that a truncated empirical covariance matrix estimator satisfies Condition 3. More specifically, it considers the truncated empirical mean pilot estimator $\check\Sigma_H=(\check\sigma_{uv}^H)$ where $\check\sigma_{uv}^H=n^{-1}\sum_i \mathbb{1}(|X_{iu}X_{iv}|\leqslant H)X_{iu}X_{iv}$ and $H=K(n/\log p)^{1/2}$ for some positive constant K. Finally, we show that the difference between $\check\Sigma_H$ and $\tilde\Sigma_H$ is sufficiently small. Thus, the proposition is proved.

LEMMA S6. *Let X_1 be a zero-mean random variable with finite variance σ^2 and let $E\{\exp(tX_1)\}<\infty$ in the interval $|t|<H$, where $H=O(n^{-\alpha})$ and $\alpha\in(0,1/2)$. If $x\geqslant 0$ and $x=o(n^{1/2}/\log n)$, then*

$$\frac{1-F_n(x)}{1-\Phi(x)}=\exp\bigg\{\Big(\frac{x}{n^{1/2}}\Big)^3\lambda\Big(\frac{x}{n^{1/2}}\Big)\bigg\}\bigg\{1+O\Big(\frac{x\log n}{n^{1/2}}\Big)\bigg\}, \tag{S8}$$

$$\frac{F_n(-x)}{\Phi(-x)}=\exp\bigg\{-\Big(\frac{x}{n^{1/2}}\Big)^3\lambda\Big(\frac{x}{n^{1/2}}\Big)\bigg\}\bigg\{1+O\Big(\frac{x\log n}{n^{1/2}}\Big)\bigg\}, \tag{S9}$$

where $F_n(x)=\mathrm{pr}(\sum_{i=1}^n X_i<n^{1/2}\sigma x)$. Here $\lambda(t)=\sum_{k=0}^\infty c_k t^k$ is a power series which will be defined in (S13).

Proof. We adapt the arguments of Theorem 5.23 in Petrov (1995) to account for a vanishing constant H. Let h be an arbitrary number in the interval $(-H,H)$ so that $E\{\exp(hX_1)\}<\infty$. Denoting the distribution function of X_1 by $V(x)$, we introduce the sequence of random variables $\bar X_n,\ldots,\bar X_n$, with common distribution function defined as

$$\mathrm{d}\bar V(x)=\frac{1}{R(h)}\exp(hx)\,\mathrm{d}V(x),$$

where $R(h)=E\{\exp(hX_1)\}=\int_{-\infty}^{\infty}\exp(hx)\,\mathrm{d}V(x)$. Write

$$\bar m=E(\bar X_1),\quad \bar\sigma^2=\mathrm{var}(\bar X_1),\quad \bar S_n=\sum_{j=1}^n\bar X_j,\quad \bar F_n(x)=\mathrm{pr}(\bar S_n-n\bar m<x\bar\sigma n^{1/2}).$$

For sufficiently small h, we have

$$\log\{R(h)\}=\sum_{\nu=1}^\infty\frac{\gamma_\nu}{\nu!}h^\nu,$$

where γ_ν is the cumulant of order ν of the random variable X_1, so $\gamma_1 = 0$ and $\gamma_2 = \sigma^2$. By Petrov (1995, pp. 179–80), choosing h as the unique real root of

$$\sigma t = \bar{m}(h) \tag{S10}$$

for a sufficiently small t, the following expansion of this root h in a power series in t holds:

$$h = \frac{t}{\sigma} - \frac{\gamma_3}{2\sigma^4}\, t^2 + o(t^2). \tag{S11}$$

Then we obtain

$$\frac{t^2}{2} + \log\{R(h)\} - h\bar{m} = t^3\lambda(t), \tag{S12}$$

where

$$\lambda(t) = \frac{\gamma_3}{6\sigma^3} + \frac{\gamma_4\sigma^2 - 3\gamma_3^2}{24\sigma^6}\, t + o(t). \tag{S13}$$

Furthermore, by equation (5.78) in Petrov (1995),

$$1 - F_n(\bar{m}\sigma^{-1}n^{1/2}) = \exp\bigl[n\log\{R(h)\} - nh\bar{m}\bigr] \int_0^\infty \exp(-hy\bar{\sigma}n^{1/2})\,\mathrm{d}\bar{F}_n(y). \tag{S14}$$

We now show that

$$\int_0^\infty \exp(-hy\bar{\sigma}n^{1/2})\,\mathrm{d}\bar{F}_n(y) = (2\pi)^{-1/2}\psi(\bar{m}\sigma^{-1}n^{1/2})\Bigl\{1 + O\bigl(h\log n\bigr)\Bigr\}, \tag{S15}$$

where $\psi(x)$ is the Mills ratio

$$\psi(x) = \frac{1-\Phi(x)}{\Phi'(x)} = \exp(x^2/2)\int_x^\infty \exp(-y^2/2)\,\mathrm{d}y.$$

Writing $Q_n(y) = \bar{F}_n(y) - \Phi(y)$,

$$\begin{aligned}\int_0^\infty \exp(-hy\bar{\sigma}n^{1/2})\,\mathrm{d}\bar{F}_n(y) &= (2\pi)^{-1/2}\int_0^\infty \exp(-h\bar{\sigma}n^{1/2}y - y^2/2)\,\mathrm{d}y \\ &\quad + \int_0^\infty \exp(-h\bar{\sigma}n^{1/2}y)\,\mathrm{d}Q_n(y) \\ &= (2\pi)^{-1/2}I_1 + I_2.\end{aligned}$$

For some universal positive constant A and $n \geqslant 3$, Theorem 5.12 in Petrov (1995) gives the bound

$$|Q_n(y)| \leqslant \frac{An^{-1/2}\log n}{1+y^2}.$$

It follows that

$$\begin{aligned}|I_2| &= \left|Q_n(0) + \int_0^\infty Q_n(y)\,\mathrm{d}\exp(-h\bar{\sigma}n^{1/2}y)\right| \\ &\leqslant An^{-1/2}\log n + An^{-1/2}\log(n)\int_0^\infty \frac{1}{1+y^2}\,\mathrm{d}\exp(-h\bar{\sigma}n^{1/2}y) \\ &= O\bigl(n^{-1/2}\log n\bigr).\end{aligned}$$

Substituting $u = h\bar{\sigma}n^{1/2}y$ into I_1, we obtain

$$I_1 = (h\bar{\sigma}n^{1/2})^{-1}\int_0^\infty \exp\{-u - u^2/(2h^2\bar{\sigma}^2 n)\}\,\mathrm{d}u$$

and

$$h\bar{\sigma}n^{1/2}I_1 \leqslant \int_0^\infty \exp(-u)\,\mathrm{d}u = 1, \quad h\bar{\sigma}n^{1/2}I_1 > \int_0^\infty \exp\{-u - u^2/(c^2\sigma^2)\}\,\mathrm{d}u,$$

where the last inequality follows from $hn^{1/2} > c$ for large n and some positive constant c. Therefore, $hn^{1/2}I_1 \asymp 1$ and

$$\begin{aligned}
&\int_0^\infty \exp(-hy\bar{\sigma}n^{1/2})\,\mathrm{d}\bar{F}_n(y) \\
&= (2\pi)^{-1/2}I_1 + O\big(n^{-1/2}\log n\big) = (2\pi)^{-1/2}I_1\Big\{1 + \frac{1}{I_1}O\big(n^{-1/2}\log n\big)\Big\} \\
&= (2\pi)^{-1/2}I_1\big\{1 + O\big(h\log n\big)\big\}.
\end{aligned}$$

By substituting into this last expression $I_1 = \psi(\bar{m}\sigma^{-1}n^{1/2})\{1 + O(h)\}$, which holds by equation (5.85) in Petrov (1995), we establish (S15).

It follows from equations (S14) and (S15) that

$$\begin{aligned}
&1 - F_n(\bar{m}\sigma^{-1}n^{1/2}) \\
&= (2\pi)^{-1/2}\exp\big[n\log\{R(h)\} - hn\bar{m}\big]\psi(\bar{m}\sigma^{-1}n^{1/2})\Big\{1 + O\big(h\log n\big)\Big\} \\
&= \exp\left[\frac{n\bar{m}^2}{2\sigma^2} + n\log\{R(h)\} - hn\bar{m}\right]\psi(\bar{m}\sigma^{-1}n^{1/2})\{1 - \Phi(\bar{m}\sigma^{-1}n^{1/2})\}\Big\{1 + O\big(h\log n\big)\Big\}.
\end{aligned}$$

Choosing $t = xn^{-1/2}$ in equation (S10), we have $x = \bar{m}\sigma^{-1}n^{1/2}$. Taking into account (S11) and $h/(t/\sigma) \to 1$, we finally obtain that for large n,

$$1 - F_n(x) = \{1 - \Phi(x)\}\exp\bigg\{\Big(\frac{x}{n^{1/2}}\Big)^3\lambda\Big(\frac{x}{n^{1/2}}\Big)\bigg\}\bigg\{1 + O\Big(\frac{x\log n}{n^{1/2}}\Big)\bigg\}.$$

This proves (S8). The same arguments also imply (S9). □

LEMMA S7. *Assume* $(6 + L)(\log p)/n < 1/8$ *and the conditions of Proposition* 3. *Let* $\tilde{\mathrm{w}}_i^{uv} = \min(1, H/|X_{iu}X_{iv} - \tilde{\sigma}_{uv}^H|)$ *for* $u, v = 1, \ldots, p$ *and* $i = 1, \ldots, n$, *and write* $a_n \asymp_{\mathrm{p}} b_n$ *if* $a_n = O_{\mathrm{p}}(b_n)$ *and* $b_n = O_{\mathrm{p}}(a_n)$ *as* $n \to \infty$. *Then, for some fixed* $t > 0$, *if* $H = K(n/\log p)^{1/2}$,

$$\max_{u,v}\mathrm{pr}\left[\Big|n - \sum_{i=1}^n \tilde{\mathrm{w}}_i^{uv}\Big| \leqslant \left\{\frac{(2+L)\log p}{n}\right\}^{1/2}\right] \geqslant 1 - O\{p^{-(2+L)}\}.$$

Proof. It suffices to show that

$$\mathrm{pr}\bigg(n - \sum_{i=1}^n \tilde{\mathrm{w}}_i^{uv} \leqslant t\bigg) \geqslant \bigg[1 - \exp\bigg\{-\frac{nt^2}{2} + o(1)\bigg\} - O\{p^{-(4+L)}\}\bigg], \qquad \text{(S16)}$$

since if we choose $t^2 = (4+L)(\log p)/n$, Bonferroni's inequality would then give

$$\begin{aligned}
\mathrm{pr}\biggl(n - \sum_{i=1}^{n} \tilde{\mathrm{w}}_i^{uv} \leqslant t \ \ \forall u, v = 1, \ldots, p\biggr) &\geqslant 1 - \sum_{u \geqslant v} \mathrm{pr}\biggl(n - \sum_{i=1}^{n} \tilde{\mathrm{w}}_i^{uv} \geqslant t\biggr) \\
&\geqslant 1 - \frac{p(p-1)}{2} \max_{u,v} \mathrm{pr}\biggl(n - \sum_{i=1}^{n} \tilde{\mathrm{w}}_i^{uv} \geqslant t\biggr) \\
&\geqslant 1 - p^2 \exp\Bigl\{-\frac{tn}{2} + o(1)\Bigr\} - O\{p^{-(2+L)}\} \\
&\geqslant 1 - O\{p^{-(2+L)}\}.
\end{aligned}$$

To prove (S16), first observe that

$$\begin{aligned}
&\mathrm{pr}\biggl(n - \sum_{i=1}^{n} \tilde{\mathrm{w}}_i^{uv} \leqslant t\biggr) \\
&= \mathrm{pr}\biggl(\frac{1}{n} \sum_{i=1}^{n} \tilde{\mathrm{w}}_i^{uv} \geqslant 1 - t/n\biggr) \\
&= \mathrm{pr}\biggl\{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(|X_{iu}X_{iv} - \tilde{\sigma}_{uv}^H| \leqslant H) + \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{1}(|X_{iu}X_{iv} - \tilde{\sigma}_{uv}^H| > H)H}{|X_{iu}X_{iv} - \tilde{\sigma}_{uv}^H|} \geqslant 1 - t/n\biggr\} \\
&= \mathrm{pr}\biggl\{1 - \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(|X_{iu}X_{iv} - \tilde{\sigma}_{uv}^H| > H)\biggl(1 - \frac{H}{|X_{iu}X_{iv} - \tilde{\sigma}_{uv}^H|}\biggr) \geqslant 1 - t/n\biggr\} \\
&\geqslant \mathrm{pr}\biggl\{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(|X_{iu}X_{iv} - \tilde{\sigma}_{uv}^H| > H) \leqslant t/n\biggr\}.
\end{aligned}$$

Further, the arguments of Proposition 3 under the assumption $(6+L)(\log p)/n < 1/8$ imply that $\max_{u,v} |\tilde{\sigma}_{uv}^H - \sigma_{uv}^*| \leqslant C\{(\log p)/n\}^{1/2}$ with probability at least $1 - O\{p^{(-4+L)}\}$. Combining this max norm bound with Hoeffding's inequality, we obtain

$$\begin{aligned}
&\mathrm{pr}\biggl\{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(|X_{iu}X_{iv} - \tilde{\sigma}_{uv}^H| > H) \leqslant t/n\biggr\} &\text{(S17)} \\
&\geqslant \mathrm{pr}\biggl\{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(|X_{iu}X_{iv} - \sigma_{uv}^*| > H - \max_{u,v} |\tilde{\sigma}_{uv}^H - \sigma_{uv}^*|) \leqslant t/n\biggr\} \\
&\geqslant \mathrm{pr}\biggl\{\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(|X_{iu}X_{iv} - \sigma_{uv}^*| > H - \eta) \leqslant t/n\biggr\}\Bigl\{1 - O(p^{-(4+L)})\Bigr\} \\
&\geqslant \biggl[1 - \exp\Bigl\{-\frac{n(t/n - \delta_H)^2}{2}\Bigr\}\biggr]\Bigl[1 - O\{p^{-(4+L)}\}\Bigr], &\text{(S18)}
\end{aligned}$$

where $\eta \geqslant C\{(\log p)/n\}^{1/2}$ and $\delta_H = E\{\mathbb{1}(|X_{iu}X_{iv} - \sigma_{uv}^*| < H - \eta)\}$. From Markov's inequality, $\delta_H = O[\{(\log p)/n\}^{1/2}]$. Then (S17) and simple manipulations establish (S16). □

LEMMA S8. *Let* $\check{\Sigma}_H := (\check{\sigma}_{uv}^H) = n^{-1} \sum_{i=1}^{n} \mathbb{1}(|X_{iu}X_{iv}| \leqslant H) X_{iu}X_{iv}$. *Under the conditions of Proposition* 4, *for large* n *and* $H = K(n/\log p)^{1/2}$ *with* $K \geqslant 0$ *and* $C \geqslant 0$,

$$\mathrm{pr}\Bigl[\max_{u,v} \bigl|(\check{\Sigma}_H \Omega^* - I_p)_{uv}\bigr| \leqslant C\{(\log p)/n\}^{1/2}\Bigr] \geqslant 1 - \varepsilon_{n,p},$$

where $\varepsilon_{n,p} \leqslant C_0(\log p)^{-1/2}p^{-L}$ *for positive constants* C_0 *and* L.

Proof. Let $\nu_{uv} = E\{\sum_{k=1}^{p} \mathbb{1}(|X_{iu}X_{ik}| \leqslant H)X_{iu}X_{ik}\omega^*_{kv}\}$; then $\|\Sigma_H\Omega^* - I\|_{\max} = O[\{(\log p)/n\}^{1/2}]$ implies

$$|\nu_{uv} - \mathbb{1}(u = v)| = O[\{(\log p)/n\}^{1/2}]. \tag{S19}$$

Note also that

$$\begin{aligned}
&\mathrm{pr}\bigg\{\bigg|\sum_{k=1}^{p}\tilde{\sigma}^H_{uk}\omega^*_{kv} - \mathbb{1}(u=v)\bigg| \geqslant \lambda_n(\sigma^*_{uu}\omega^*_{vv})^{1/2}\bigg\}\\
&\leqslant \mathrm{pr}\bigg\{\bigg|\sum_{i=1}^{n}\sum_{k=1}^{p}\mathbb{1}(|X_{iu}X_{ik}| \leqslant H)X_{iu}X_{ik}\omega^*_{kv} - n\nu_{uv}\bigg| \geqslant n\lambda_n(\sigma^*_{uu}\omega^*_{vv})^{1/2} - n|\mathbb{1}(u=v) - \nu_{uv}|\bigg\}\\
&\leqslant \mathrm{pr}\bigg[\frac{1}{n^{1/2}}\bigg|\sum_{i=1}^{n}\sum_{k=1}^{p}\mathbb{1}(|X_{iu}X_{ik}| \leqslant H)X_{iu}X_{ik}\omega^*_{kv} - n\nu_{uv}\bigg| \geqslant \delta(\sigma^*_{uu}\omega^*_{vv}\log p)^{1/2} - O\{(\log p)^{1/2}\}\bigg].
\end{aligned} \tag{S20}$$

Since $|\mathbb{1}(|X_{iu}X_{ik}| \leqslant H)X_{iu}X_{ik}| \leqslant K(n/\log p)^{1/2}$ and $\max_{u,v}|\omega^*_{uv}| \leqslant M$ for some $M > 0$, we have

$$\begin{aligned}
\bigg|\sum_{k=1}^{p}\mathbb{1}(|X_{iu}X_{ik}| \leqslant H)X_{iu}X_{ik}\omega^*_{kv}\bigg| &\leqslant K(n/\log p)^{1/2}\max_v\sum_{k=1}^{p}|\omega^*_{kv}|\\
&\leqslant \tilde{K}(n/\log p)^{1/2}\max_v\sum_{k=1}^{p}|\omega^*_{kv}|^{1-q}\\
&\leqslant \tilde{K}(n/\log p)^{1/2}c_{n,p}\\
&= O\{n^{(2-q)/2}/(\log p)^{(4-q)/2}\}.
\end{aligned}$$

Therefore $E[\exp\{t\sum_{k=1}^{p}\mathbb{1}(|X_{iu}X_{ik}| \leqslant H)X_{iu}X_{ik}\omega^*_{kv}\}] < \infty$ as long as we choose $|t| < O\{(\log p)^{(4-q)/2}/n^{(2-q)/2}\} = O(n^{-(1-q)/3})$. Thus, taking $|t| < Cn^{-(1-q)/3}$ for some sufficiently large positive constant C, applying Lemma S6 to (S20) yields

$$\begin{aligned}
&\mathrm{pr}\bigg\{\bigg|\sum_{k=1}^{p}\tilde{\sigma}^H_{uk}\omega^*_{kv} - \mathbb{1}(u=v)\bigg| \geqslant \lambda_n(\sigma^*_{uu}\omega^*_{vv})^{1/2}\bigg\}\\
&\leqslant \{1 + o(1)\}\,\mathrm{pr}\Big\{|N(0,1)| \geqslant \delta(\log p)^{1/2}\Big\} \leqslant C(\log p)^{-1/2}p^{-\delta^2/2}.
\end{aligned}$$

Taking $\delta \leqslant 2$, the last inequality and an application of Bonferroni's inequality complete the proof. □

We now prove Proposition 4. The strategy is to rewrite the Huber estimator as a weighted mean that behaves essentially like a truncated mean with a diverging level of truncation as n goes to infinity. Writing $\tilde{\mathrm{w}}^{uv}_i = \min(1, H/|X_{iu}X_{iv} - \tilde{\sigma}^H_{uv}|)$ for $u, v = 1, \ldots, p$ and $i = 1, \ldots, n$, it is easy to see that

$$\tilde{\sigma}^H_{uv} = \frac{\sum_{i=1}^{n}\tilde{\mathrm{w}}^{uv}_i X_{iu}X_{iv}}{\sum_{i=1}^{n}\tilde{\mathrm{w}}^{uv}_i}.$$

Hence, letting $\tilde{\mathrm{w}}_i^u = (\tilde{\mathrm{w}}_i^{u1}, \ldots, \tilde{\mathrm{w}}_i^{up})^{\mathrm{T}}$ and $n_{uv} = \sum_{i=1}^n \tilde{\mathrm{w}}_i^{uv}$, it follows from Lemma S7 and Proposition 3 that

$$\begin{aligned}
&\mathrm{pr}\Big\{\Big|\sum_{k=1}^p \tilde{\sigma}_{uk}^H \omega_{kv}^* - \mathbb{1}(u=v)\Big| \geqslant \lambda_n (\sigma_{uu}^* \omega_{vv}^*)^{1/2}\Big\} \\
&= \mathrm{pr}\Big\{\Big|\sum_{i=1}^n \sum_{k=1}^p \frac{\tilde{\mathrm{w}}_i^{uk}}{\sum_{j=1}^n \tilde{\mathrm{w}}_j^{uk}} X_{iu} X_{ik} \omega_{kv}^* - \mathbb{1}(u=v)\Big| \geqslant \lambda_n (\sigma_{uu}^* \omega_{vv}^*)^{1/2}\Big\} \\
&\leqslant \mathrm{pr}\Big\{\Big|\sum_{i=1}^n \sum_{k=1}^p \tilde{\mathrm{w}}_i^{uk} X_{iu} X_{ik} \omega_{kv}^* - \mathbb{1}(u=v)\Big| \geqslant n\lambda_n (\sigma_{uu}^* \omega_{vv}^*)^{1/2}\Big\} + O\{p^{-(2+L)}\}. \qquad \text{(S21)}
\end{aligned}$$

In order to further bound (S21),

$$\begin{aligned}
&\Big|\sum_{i=1}^n \sum_{k=1}^p \tilde{\mathrm{w}}_i^{uk} X_{iu} X_{ik} \omega_{kv}^*\Big| \\
&= \Big|\sum_{i=1}^n \sum_{k=1}^p \Big\{\mathbb{1}(|X_{iu}X_{ik} - \tilde{\sigma}_{uk}^H| \leqslant H) X_{iu} X_{ik} \omega_{kv}^* \qquad \text{(S22)} \\
&\qquad + \mathbb{1}(|X_{iu}X_{ik} - \tilde{\sigma}_{uk}^H| > H) \frac{H}{|X_{iu}X_{ik} - \tilde{\sigma}_{uk}^H|} \operatorname{sgn}(X_{iu}X_{ik} - \tilde{\sigma}_{uk}^H) \omega_{kv}^* \Big\}\Big| \\
&\leqslant \Big|\sum_{i=1}^n \sum_{k=1}^p \mathbb{1}(|X_{iu}X_{ik} - \tilde{\sigma}_{uk}^H| \leqslant H) X_{iu} X_{ik} \omega_{kv}^*\Big| \\
&\quad + \Big|\sum_{i=1}^n \sum_{k=1}^p \mathbb{1}(|X_{iu}X_{iv} - \tilde{\sigma}_{uv}^H| > H) \frac{H}{(X_{iu}X_{ik} - \tilde{\sigma}_{uk}^H)^2} (X_{iu}X_{ik} - \tilde{\sigma}_{uk}^H) \omega_{kv}^*\Big| \\
&= T_1 + T_2. \qquad \text{(S23)}
\end{aligned}$$

For the second term in (S22), note first that

$$\begin{aligned}
\|\Sigma_H \Omega^* - I\|_{\max} = \|(\Sigma_H - \Sigma^*)\Omega^*\|_{\max} &= \max_{u,v} \Big|\sum_{k=1}^p E\big\{\mathbb{1}(|X_{1u}X_{1k}| > H) X_{1u} X_{1k}\big\} \omega_{kp}^*\Big| \\
&= O[\{(\log p)/n\}^{1/2}].
\end{aligned}$$

Also observe that by Proposition 3 and Hoeffding's inequality,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \mathbb{1}(|X_{iu}X_{ik} - \tilde{\sigma}_{uk}^H| > H) &\leqslant \frac{1}{n}\sum_{i=1}^n \mathbb{1}(|X_{iu}X_{ik} - \sigma_{uk}^*| > H - \eta) \\
&\leqslant (H-\eta)^{-1} E|X_{1u}X_{1k} - \sigma_{uk}^*| + \Big\{\frac{\log(p^{2+L})}{2n}\Big\}^{1/2} \\
&= O\Big\{\Big(\frac{\log p}{n}\Big)^{1/2}\Big\}
\end{aligned}$$

with probability at least $1-2p^{-(2+L)}$, where $\eta \geqslant C\{(\log p)/n\}^{1/2}$ and $C>0$. Thus, with probability at least $1-2p^{-(2+L)}$,

$$\begin{aligned}
T_2 &\leqslant H^{-1}\Big|\sum_{i=1}^{n}\sum_{k=1}^{p}\mathbb{1}(|X_{iu}X_{iv}-\tilde\sigma_{uv}^H|>H)(X_{iu}X_{ik}-\tilde\sigma_{uk}^H)\omega_{kv}^*\Big| \\
&\quad+\Big|\sum_{i=1}^{n}\sum_{k=1}^{p}\mathbb{1}(|X_{iu}X_{iv}-\tilde\sigma_{uv}^H|>H)\frac{(X_{iu}X_{ik}-\tilde\sigma_{uk}^H)^2-H^2}{H(X_{iu}X_{ik}-\tilde\sigma_{uk}^H)^2}(X_{iu}X_{ik}-\tilde\sigma_{uk}^H)\omega_{kv}^*\Big| \\
&\leqslant O\{(n\log p)^{1/2}\}+nH^{-1}\sum_{k=1}^{p}\Big[\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(|X_{iu}X_{iv}-\tilde\sigma_{uv}^H|>H)\big|X_{iu}X_{ik}-\tilde\sigma_{uk}^H\big|\Big]|\omega_{kv}^*| \\
&\leqslant O\{(n\log p)^{1/2}\}+nH^{-2}M^q\sum_{k=1}^{p}\Big[\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(|X_{iu}X_{iv}-\tilde\sigma_{uv}^H|>H)\big(X_{iu}X_{ik}-\tilde\sigma_{uk}^H\big)^2\Big]|\omega_{kv}^*|^{1-q} \\
&\leqslant O\{(n\log p)^{1/2}\}+C_{q,\kappa}nH^{-2}c_{n,p} \\
&\leqslant O\{(n\log p)^{1/2}\}+o\{(n\log p)^{1/2}\}. \qquad \text{(S24)}
\end{aligned}$$

Similar arguments to those used in (S24) can also be used to show that the first term in (S22) satisfies

$$T_1 \leqslant \Big|\sum_{i=1}^{n}\sum_{k=1}^{p}\mathbb{1}(|X_{iu}X_{ik}-\sigma_{uk}^*|\leqslant H)X_{iu}X_{ik}\omega_{kv}^*\Big|+O\{(n\log p)^{1/2}\} \qquad \text{(S25)}$$

with probability at least $1-4p^{-(2+L)}$ because for some constant $\eta \geqslant C\{(\log p)/n\}^{1/2}$,

$$\begin{aligned}
&\Big|\sum_{i=1}^{n}\sum_{k=1}^{p}\Big\{\mathbb{1}(|X_{iu}X_{ik}-\tilde\sigma_{uk}^H|\leqslant H)-\mathbb{1}(|X_{iu}X_{ik}-\sigma_{uk}^*|\leqslant H)\Big\}X_{iu}X_{ik}\omega_{kv}^*\Big| \\
&\leqslant nM^q\sum_{k=1}^{p}\Big\{\frac{1}{n}\sum_{i=1}^{n}\Big|\mathbb{1}(|X_{iu}X_{ik}-\tilde\sigma_{uk}^H|\leqslant H)-\mathbb{1}(|X_{iu}X_{ik}-\sigma_{uk}^*|\leqslant H)\Big||X_{iu}X_{ik}|\Big\}|\omega_{kv}^*|^{1-q} \\
&= nM^q\sum_{k=1}^{p}\Big\{\frac{1}{n}\sum_{i=1}^{n}\Big|\mathbb{1}(|X_{iu}X_{ik}-\sigma_{uk}^*|>H)\Big||X_{iu}X_{ik}|-\mathbb{1}(|X_{iu}X_{ik}-\tilde\sigma_{uk}^H|>H)\Big\}|\omega_{kv}^*|^{1-q} \\
&\leqslant 2nM^q\sum_{k=1}^{p}\Big\{\frac{1}{n}\sum_{i=1}^{n}\Big|\mathbb{1}(|X_{iu}X_{ik}-\sigma_{uk}^*|>H-\eta)\Big||X_{iu}X_{ik}|\Big\}|\omega_{kv}^*|^{1-q}.
\end{aligned}$$

In addition, consider the term $\nu_{uv}=E\{\sum_{k=1}^{p}\mathbb{1}(|X_{iu}X_{ik}-\sigma_{uk}^*|\leqslant H)X_{iu}X_{ik}\omega_{kv}^*\}$; then $\|\Sigma_H\Omega^*-I\|_{\max}=O[\{(\log p)/n\}^{1/2}]$ implies

$$|\nu_{uv}-\mathbb{1}(u=v)|=O[\{(\log p)/n\}^{1/2}]. \qquad \text{(S26)}$$

Therefore, upon combining (S21)–(S26) we see that

$$\begin{aligned}
&\mathrm{pr}\bigg\{\bigg|\sum_{k=1}^{p}\tilde{\sigma}_{uk}^{H}\omega_{kv}^{*}-\mathbb{1}(u=v)\bigg|\geqslant\lambda_n(\sigma_{uu}^{*}\omega_{vv}^{*})^{1/2}\bigg\}\\
&\leqslant \mathrm{pr}\bigg[\bigg|\sum_{i=1}^{n}\sum_{k=1}^{p}\tilde{\mathrm{w}}_i^{uk}X_{iu}X_{ik}\omega_{kv}^{*}-n\nu_{uv}\bigg|+n\Big|\nu_{uv}-\mathbb{1}(u=v)\{1-o(1)\}\Big|\geqslant n\lambda_n(\sigma_{uu}^{*}\omega_{vv}^{*})^{1/2}\{1-o(1)\}\bigg]\\
&\qquad\qquad + O\{p^{-(2+L)}\}\\
&\leqslant \mathrm{pr}\bigg[\bigg|\sum_{i=1}^{n}\sum_{k=1}^{p}\mathbb{1}(|X_{iu}X_{ik}-\sigma_{uk}^{*}|\leqslant H)X_{iu}X_{ik}\omega_{kv}^{*}-\nu_{uv}\bigg|\geqslant n\lambda_n(\sigma_{uu}^{*}\omega_{vv}^{*})^{1/2}-O\{(n\log p)^{1/2}\}\bigg]\\
&\qquad\qquad + O\{p^{-(2+L)}\}.
\end{aligned}$$

The proof follows from the last inequality and the arguments of Lemma S8. □

Remark 1. Unlike the proof of Lemma 1 in Cai et al. (2016), we cannot compute explicitly the mean and variance constants required to standardize $\sum_{k=1}^{p}\mathbb{1}(|X_{iu}X_{ik}-\sigma_{uk}^{*}|\leqslant H)X_{iu}X_{ik}\omega_{kv}^{*}$.

Proof of Proposition 5. The following auxiliary lemma is a simplified version of Proposition 1 in Lerasle & Oliveira (2011).

LEMMA S9. *Let $Z_1,\ldots,Z_n$ be independent identically distributed random variables with $E(Z_1)=\mu^*$ and $\mathrm{var}(Z_1)=(\sigma^*)^2$. Let $\delta\in(0,1)$ and $M\leqslant n/2$, and let $B_1,\ldots,B_M$ be a regular partition of $[n]$. Then if $M\geqslant\log(\delta^{-1})$, we have that for some constant $K\leqslant 2(6\mathrm{e})^{1/2}$,*

$$\mathrm{pr}\bigg\{\tilde{\mu}_M-\mu^*\geqslant K\bigg(\frac{\sigma^{*2}\log\delta^{-1}}{n}\bigg)^{1/2}\bigg\}\leqslant\delta.$$

By Lemma S9 we have that for all $u,v\in[p]$,

$$\mathrm{pr}\Big[|\tilde{\mu}_u^M-\mu_u^*|\geqslant K\{\sigma_{uu}^*(\log p)/n\}^{1/2}\Big]\leqslant 2p^{-(2+L)}$$

and

$$\mathrm{pr}\Big[|\tilde{\mu}_{uv}^M-\mu_{uv}^*|\geqslant K\{\theta_{uv}^*(\log p)/n\}^{1/2}\Big]\leqslant 2p^{-(2+L)}, \tag{S27}$$

where $\theta_{uv}^*=\mathrm{var}(X_uX_v)$. Hence, with probability at least $1-2p^{-(1+L)}$,

$$\max_{u,v}\big|\tilde{\mu}_u^M\tilde{\mu}_v^M-\mu_u^*\mu_v^*\big|\leqslant 2K\max_{u,v}|\mu_u^*|\{\sigma_{vv}^*(\log p)/n\}^{1/2}+K^2\max_u\sigma_{uu}^*(\log p)/n. \tag{S28}$$

Combining (S27) and (S28), it follows from the union bound that

$$\begin{aligned}
&\mathrm{pr}\Big[\max_{u,v}|\tilde{\sigma}_{uv}^M-\sigma_{uv}^*|\geqslant C\{(\log p)/n\}^{1/2}\Big]\\
&\leqslant \mathrm{pr}\Big[\max_{u,v}|\tilde{\mu}_{uv}^M-\mu_{uv}^*|+\max_{u,v}|\tilde{\mu}_u^M\tilde{\mu}_v^M-\mu_u^*\mu_v^*|\geqslant C\{(\log p)/n\}^{1/2}\Big]\\
&\leqslant 2p^{-L}(1+p^{-1}).
\end{aligned}$$

□

Proof of Proposition 6. It is easy to check that for any μ' and μ'' such that $\mu''-\mu'\geqslant 0$,

$$\psi_H(Z_i-\mu')-\psi_H(Z_i-\mu'')\geqslant(\mu''-\mu')\mathbb{1}(Z_i-\mu''<H,Z_i-\mu'>-H),$$

where $\psi_H(\cdot)$ is defined in (11). We consider the cases $\mu \geqslant \tilde{\mu}^H$ and $\tilde{\mu}^H \geqslant \mu$ in turn. If $\mu \geqslant \tilde{\mu}^H$, we obtain

$$\mu - \tilde{\mu}^H \leqslant \left| \frac{1}{n\tilde{p}^H} \sum_{i=1}^n \psi_H(Z_i - \mu) - \frac{1}{n\tilde{p}^H} \sum_{i=1}^n \psi_H(Z_i - \tilde{\mu}^H) \right| = \frac{1}{\tilde{p}^H} \left| \frac{1}{n} \sum_{i=1}^n \psi_H(Z_i - \mu) \right|, \tag{S29}$$

where $\tilde{p}^H = n^{-1} \sum_{i=1}^n \mathbb{1}(Z_i - \tilde{\mu}^H < H,\, Z_i - \mu > -H)$ and the last equality follows from the definition of $\tilde{\mu}^H$. If $\tilde{\mu}^H \geqslant \mu$, we obtain a similar expression to (S29) but with $\tilde{p}^H = n^{-1} \sum_{i=1}^n \mathbb{1}(Z_i - \mu < H,\, Z_i - \tilde{\mu}^H > -H)$. With both definitions of $\tilde{p}^H$, $\tilde{p}^H \geqslant 1 - n^{-1} \sum_{i=1}^n \mathbb{1}(|Z_i - \mu| \geqslant H)$. Hence, by Bernstein's inequality followed by Markov's inequality, with probability at least $1 - \delta/2$,

$$\begin{aligned} \tilde{p}^H &\geqslant 1 - \mathrm{pr}(|Z_i - \mu| \geqslant H) - \left\{ \frac{2\log(2\delta^{-1})}{n} \right\} - \frac{\log(2\delta^{-1})}{3n} \\ &= 1 - \mathrm{pr}\left\{ |Z_i - \mu|^{1+\varepsilon} \geqslant \frac{vn}{\log(2\delta^{-1})} \right\} - \left\{ \frac{2\log(2\delta^{-1})}{n} \right\} - \frac{\log(2\delta^{-1})}{3n} \\ &\geqslant 1 - \frac{4\log(2\delta^{-1})}{3n} - \left\{ \frac{2\log(2\delta^{-1})}{n} \right\}. \end{aligned} \tag{S30}$$

It remains to bound $n^{-1} \sum_{i=1}^n \psi_H(Z_i - \mu)$. Note that

$$\begin{aligned} E\{\psi_H(Z_i - \mu)\} &= E\{(Z_i - \mu)\mathbb{1}(|Z_i - \mu| \leqslant H)\} + HE\{\mathrm{sgn}(Z_i - \mu)\mathbb{1}(|Z_i - \mu| > H)\} \\ &\leqslant E\{(Z_i - \mu)\mathbb{1}(|Z_i - \mu| > H)\} + E\{|Z_i - \mu|\mathbb{1}(|Z_i - \mu| > H)\} \\ &\leqslant 2E(|Z_i - \mu|^{1+\varepsilon} H^{-\varepsilon}) \\ &\leqslant \frac{2v}{H^\varepsilon}. \end{aligned}$$

Therefore, by Bernstein's inequality, with probability at least $1 - \delta/2$,

$$\frac{1}{n} \sum_{i=1}^n \psi_H(Z_i - \mu) \leqslant \frac{2v}{H^\varepsilon} + \left\{ \frac{2vH^{1-\varepsilon} \log(2\delta^{-1})}{n} \right\} + \frac{H\log(2\delta^{-1})}{3n}. \tag{S31}$$

Combining (S29)–(S31) and Bonferroni's inequality, with probability at least $1 - \delta$,

$$\begin{aligned} \tilde{\mu}^H - \mu &\leqslant \left[1 - \frac{4\log(2\delta^{-1})}{3n} - \left\{ \frac{2\log(2\delta^{-1})}{n} \right\}^{1/2} \right]^{-1} \left[\frac{2v}{H^\varepsilon} + \left\{ \frac{2vH^{1-\varepsilon}\log(2\delta^{-1})}{n} \right\}^{1/2} + \frac{H\log(2\delta^{-1})}{3n} \right] \\ &= \left[1 - \frac{4\log(2\delta^{-1})}{3n} - \left\{ \frac{2\log(2\delta^{-1})}{n} \right\}^{1/2} \right]^{-1} \frac{7+\sqrt{2}}{3} v^{1/(1+\varepsilon)} \left\{ \frac{\log(2\delta^{-1})}{n} \right\}^{\varepsilon/(1+\varepsilon)}. \end{aligned} \tag{S32}$$

Finally, it follows from $n > 12\log(2\delta^{-1})$ that $4\log(2\delta^{-1})/(3n) - \{2\log(2\delta^{-1})/n\}^{1/2} \leqslant 1/3$ and therefore (S32) gives

$$\tilde{\mu}^H - \mu \leqslant \frac{7+\sqrt{2}}{2} v^{1/(1+\varepsilon)} \left\{ \frac{\log(2\delta^{-1})}{n} \right\}^{\varepsilon/(1+\varepsilon)}$$

with probability at least $1 - \delta$. □

Proof of Theorem 3. Using Corollary 1 and the arguments of Proposition 3, we obtain the elementwise max norm bound

$$\mathrm{pr}\Big[\max_{u,v}|\tilde{\sigma}_{uv}-\sigma^*_{uv}|\leqslant K\big\{(\log p)/n\big\}^{\varepsilon/(2+\varepsilon)}\Big]\geqslant 1-O(p^{-L}).$$

The proof is completed by inspection of the arguments given for the proof of Theorem 1. □

Proof of Theorem 4. Using Corollary 1 and the arguments of Proposition 4, we obtain the elementwise max norm bound

$$\mathrm{pr}\Big[\max_{u,v}\big|(\tilde{\Sigma}_H\Omega^*-I_p)_{uv}\big|\leqslant K\big\{(\log p)/n\big\}^{\varepsilon/(2+\varepsilon)}\Big]\geqslant 1-O(p^{-L}).$$

The proof is completed by inspection of the arguments given for the proof of Theorem 2. □

S2. Discussion

We first discuss Theorems 1 and 2 by emphasizing the consequences of the technical conditions we impose on the paramater spaces. We then justify a technical condition required in the proof of Proposition 4. Finally, we show that pilot estimators meeting Condition 1 but not Condition 3 lead to consistent estimation of the precision matrix with suboptimal rates of convergence.

S2·1. *Technical conditions of Theorems* 1 *and* 2

In Theorems 1 and 2, we assume that the diagonal elements of Σ^* and Ω^* are bounded away from zero and infinity. In this work, we prioritize the relaxation of sub-Gaussian assumptions as we believe they are unrealistic in high-dimensional settings.

Our assumption $\min_u(\sigma^*_{uu})=\gamma>0$, in Theorem 1, is very similar to the one appearing in equation (14) of Cai & Liu (2011). In that paper $\min_{u,v}\mathrm{var}(X_uX_v)\geqslant\tau>0$ is assumed, even when X is multivariate normal. Note that our assumptions on $\min_u\sigma^*_{uu}$ and $\max_u\sigma^*_{uu}$ are also implied by $0<c\leqslant\lambda_{\min}(\Sigma^*)\leqslant\lambda_{\max}(\Sigma^*)\leqslant C<\infty$. The latter is arguably a rather mild regularity condition, which has appeared in many papers (e.g., Bickel & Levina, 2008; Rothman et al., 2009; Fan et al., 2013).

S2·2. *Bounding the max-norm error of the partial population covariance matrix*

In Proposition 4 we verified that the Huber estimator satisfies Condition 3 under the assumption that $\|\Sigma_H\Omega^*-I\|_{\max}=O[\{(\log p)/n\}^{1/2}]$. The following lemma gives three sufficient conditions that guarantee this max-norm error bound.

Lemma S10. *Assuming $E(X)=0$ and $\log p=o(n)$, under either of the following scenarios we have that $\|\Sigma_H\Omega^*-I\|_{\max}=O[\{(\log p)/n\}^{1/2}]$ for $H=K(n/\log p)^{1/2}$, where K is a large enough constant.*

(i) *For any $\nu\in\mathbb{R}^p$, $X^{\mathrm{T}}\nu$ is sub-Gaussian with parameter $\kappa_0^2\|\nu\|_2^2$, i.e., $E\{\exp(tX^{\mathrm{T}}\nu)\}\leqslant E\{\exp(t^2\kappa_0^2\|\nu\|_2^2/2)\}$ for any $t\in\mathbb{R}$. Furthermore, $\log p=O(n^{\alpha})$ and $M_{n,p}=O(n^{\tau})$ for some $\alpha\in[0,1)$ and $\tau>0$.*
(ii) *For all $u=1,\ldots,p$, $E(X_u^4)<\infty$ and $M_{n,p}=O(1)$.*
(iii) *For all $u,v=1,\ldots,p$ and $\varepsilon\in(0,1)$, we have $E(|X_uX_v|^{2+\varepsilon})<\infty$ and $M_{n,p}=O\{(n/\log p)^{\varepsilon/2}\}$.*

Proof. First note that

$$
\begin{aligned}
\Big|(\Sigma_H\Omega^*)_{uv} - \mathbb{1}(u=v)\Big| = \Big|(\sigma^H_{u.})^{\mathrm{T}}\omega^*_{.v} - \mathbb{1}(u=v)\Big| &= \Big|\sum_{k=1}^p |\sigma^H_{uk}\omega^*_{kv} - \mathbb{1}(u=v)\Big| \\
&= \Big|\sum_{k=1}^p E\{X_uX_k\,\mathbb{1}(|X_uX_k| \leqslant H)\}\omega^*_{kv} - \mathbb{1}(u=v)\Big| \\
&= \Big|\sum_{k=1}^p E\{X_uX_k\,\mathbb{1}(|X_uX_k| > H)\}\omega^*_{kv}\Big| \\
&\leqslant \max_k E\{|X_uX_k|\,\mathbb{1}(|X_uX_k| > H)\}\sum_{k=1}^p |\omega^*_{kv}| \\
&\leqslant M_{n,p}\max_k E\{|X_uX_k|\,\mathbb{1}(|X_uX_k| > H)\}. \qquad \text{(S33)}
\end{aligned}
$$

We will complete the proof by separately showing that the right-hand side of the last inequality is $o(1)$ under the three different settings of the lemma.

(i) First we rewrite X_uX_k as a quadratic form $X_uX_k = X^{\mathrm{T}}BX$, where B is a $p\times p$ matrix with $B_{ku} = B_{uk} = 1$ and zeros in all other components. Exploiting the sub-Gaussian tails of X and Theorem 2.1 of Hsu et al. (2012), we see that for $t > 2$,

$$\mathrm{pr}\{|X_uX_k - E(X_uX_k)| > 2\kappa_0^2 t\} \leqslant \exp(-t). \qquad \text{(S34)}$$

Applying the Cauchy–Schwarz inequality followed by the triangle inequality and (S34), we obtain

$$
\begin{aligned}
&E\{|X_uX_k|\,\mathbb{1}(|X_uX_k| > H)\} \\
&\leqslant \{E(X_uX_k)^2\}^{1/2}\{\mathrm{pr}(|X_uX_k| > H)\}^{1/2} \\
&\leqslant \{E(X_uX_k)^2\}^{1/2}\{\mathrm{pr}(|X_uX_k - E(X_uX_k)| > H - |E(X_uX_k)|\}^{1/2} \\
&\leqslant \{E(X_uX_k)^2\}^{1/2}(\exp[-\{H - |E(X_uX_k)|\}/(2\kappa_0^2)])^{1/2} \\
&= O\big[\exp\{-c(n/\log p)^{1/2}\}\big] \\
&= O\big\{\exp(-n^{1-\alpha})\big\}
\end{aligned}
$$

for some constant $c > 0$. Combining this exponential tail decay bound with (S33), we see that the claimed result holds as long as $M_{n,p} = O(n^\tau)$ where $\tau > 0$.

(ii) Applying the Cauchy–Schwarz, triangle and Chebychev's inequalities gives

$$
\begin{aligned}
&E\{|X_uX_k|\,\mathbb{1}(|X_uX_k| > H)\} \\
&\leqslant \{E(X_uX_k)^2\}^{1/2}\{\mathrm{pr}(|X_uX_k| > H)\}^{1/2} \\
&\leqslant \{E(X_uX_k)^2\}^{1/2}\{\mathrm{pr}(|X_uX_k - E(X_uX_k)| > H - |E(X_uX_k)|\}^{1/2} \\
&\leqslant \frac{\{E(X_uX_k)^2\}^{1/2}\{\mathrm{var}(X_uX_k)\}^{1/2}}{\big|H - |E(X_uX_k)|\big|} \\
&= O\Big\{\Big(\frac{\log p}{n}\Big)^{1/2}\Big\}.
\end{aligned}
$$

Using this result and $M_{n,p} = O(1)$ in (S33) establishes the second claim.

(iii) Applying the Cauchy–Schwarz inequality yields

$$\begin{aligned} &E\{|X_uX_k|\,\mathbb{1}(|X_uX_k|>H)\}\\ &=E\{|X_uX_k|\,\mathbb{1}(|X_uX_k|>H)\mathbb{1}(|X_uX_k|>H)\}\\ &\leqslant [E\{(X_uX_k)^2\,\mathbb{1}(|X_uX_k|>H)\}]^{1/2}\{\mathrm{pr}(|X_uX_k|>H)\}^{1/2}. \end{aligned} \tag{S35}$$

Furthermore,

$$\begin{aligned} E\{(X_uX_k)^2\,\mathbb{1}(|X_uX_k|>H)\} &= E\{|X_uX_k|^{2+\varepsilon}|X_uX_k|^{-\varepsilon}\,\mathbb{1}(|X_uX_k|>H)\}\\ &\leqslant E(|X_uX_k|^{2+\varepsilon})H^{-\varepsilon}\\ &= O\bigg\{\Big(\frac{\log p}{n}\Big)^{\varepsilon/2}\bigg\}, \end{aligned} \tag{S36}$$

and by the triangle inequality followed by Markov's inequality we obtain

$$\begin{aligned} \mathrm{pr}(|X_uX_k|>H) &\leqslant \mathrm{pr}\{|X_uX_k-E(X_uX_k)|>H-|E(X_uX_k)|\}\\ &=\mathrm{pr}\big[|X_uX_k-E(X_uX_k)|^{2+\varepsilon}>\big|H-|E(X_uX_k)|\big|^{2+\varepsilon}\big]\\ &\leqslant \frac{E\{|X_uX_k-E(X_uX_k)|^{2+\varepsilon}\}}{\big|H-|E(X_uX_k)|\big|^{2+\varepsilon}}\\ &= O\bigg\{\Big(\frac{\log p}{n}\Big)^{1+\varepsilon/2}\bigg\}. \end{aligned} \tag{S37}$$

From (S35), (S36) and (S37) we see that

$$E\{|X_uX_k|\,\mathbb{1}(|X_uX_k|>H)\} = O\bigg\{\Big(\frac{\log p}{n}\Big)^{(1+\varepsilon)/2}\bigg\},$$

which, combined with $M_{n,p}=O\{(n/\log p)^{\varepsilon/2}\}$ and (S33), verifies the third claim. □

Scenarios (i) and (ii) of the above lemma are benchmark settings under which other pilot estimators satisfy Condition 3. More specifically, on the one hand, (i) assumes sub-Gaussian tails and is therefore a scenario where the sample covariance leads to optimal rates of convergence (Cai et al., 2016); on the other hand, (ii) entails that $M_{n,p}$ is bounded, and, as mentioned in § 3 of the main paper, in this case the three robust pilot estimators presented in the paper can be shown to satisfy Condition 3 while the sample covariance fails as shown in Proposition 1. Scenario (iii) is the most interesting as it allows for both heavy tails and a diverging $M_{n,p}$.

S2·3. *Convergence rate of pilot estimators that satisfy Condition* 1 *but not Condition* 3

The following proposition shows that pilot estimators satisfying Condition 1 but not Condition 3 can still lead to consistent precision matrix estimators but with a slower rate.

PROPOSITION S1. *Suppose that Conditions* 1 *and* 4 *are satisfied. Under the scaling condition* $\log p = o(n)$ *we have, for a constant* $C_0>0$*,*

$$\inf_{\Omega^*\in\mathcal{G}_q}\mathrm{pr}\bigg\{\|\tilde\Omega-\Omega^*\|_2\leqslant C_0M_{n,p}^{2-2q}c_{n,p}\Big(\frac{\log p}{n}\Big)^{(1-q)/2}\bigg\}\geqslant 1-\varepsilon_{n,p},$$

where $\varepsilon_{n,p}$ *is a deterministic sequence that decreases to zero as* $n,p\to\infty$ *and* $\tilde\Omega$ *is the robust adaptively constrained* ℓ_1*-minimization estimator described in* (7)–(10) *of the main paper with* $\lambda_{n,p}=\lambda M_{n,p}\{(\log p)/n\}^{1/2}$ *and* $\lambda>0$*.*

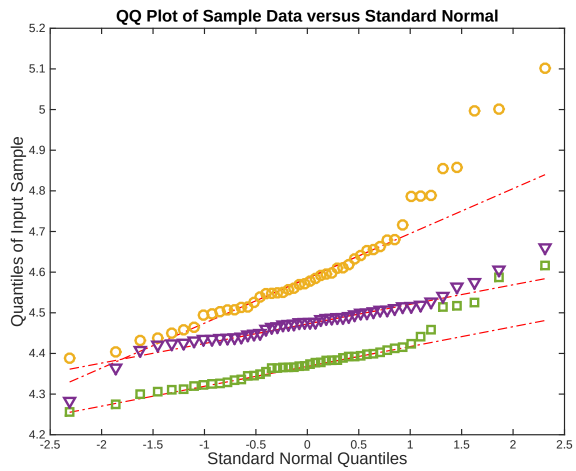

Fig. S1. QQ-plots for the three genes ME1, IRF3 and ACSL1.

Proof. We only sketch the arguments since they are essentially the same as those for Theorem 2. Following (S4) and the subsequent arguments but using the weaker bound $\|\tilde{\Sigma}\Omega^* - I\|_{\max} = O(M_{n,p}\{(\log p)/n\}^{1/2})$, we can show that with probability $1 - \varepsilon_{n,p}$,

$$\|\tilde{\Omega} - \Omega^*\|_{\max} \leqslant CM_{n,p}^2 \left(\frac{\log p}{n}\right)^{1/2}$$

for some constant $C > 0$. Then it follows from Lemma S2 that on an event of probability at least $1 - \varepsilon_{n,p}$, $\|\tilde{\Omega} - \Omega^*\|_2 \leqslant \|\tilde{\Omega} - \Omega^*\|_1 \leqslant C_0 M_{n,p}^{2-2q} c_{n,p} (\log p/n)^{(1-q)/2}$. □

S3. ADDITIONAL PLOT FOR THE REAL-DATA EXAMPLE

Figure S1 shows a QQ-plot of the three genes (ME1, IRF3 and ACSL1) whose expression contains outliers.

REFERENCES

BICKEL, P. J. & LEVINA, E. (2008). Covariance regularization by thresholding. *Ann. Statist.* **36**, 2577–604.

BUBECK, S., CESA-BIANCHI, N. & LUGOSI, G. (2013). Bandits with heavy tail. *IEEE Trans. Info. Theory* **59**, 7711–7.

CAI, T. & LIU, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *J. Am. Statist. Assoc.* **106**, 672–84.

CAI, T., LIU, W. & LUO, X. (2011). A constrained ℓ_1 minimization approach to sparse precision matrix estimation. *J. Am. Statist. Assoc.* **106**, 594–607.

CAI, T., LIU, W. & ZHOU, H. (2016). Estimating sparse precision matrix: Optimal rates of convergence and adaptive estimation. *Ann. Statist.* **44**, 455–88.

FAN, J., LIAO, Y. & MINCHEVA, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *J. R. Statist. Soc.* B **75**, 603–80.

FAN, J., LI, Q. & WANG, Y. (2017). Estimation of high-dimensional mean regression in the absence of symmetry and light-tail assumptions. *J. R. Statist. Soc.* B **79**, 747–65.

HSU, D., KAKADE, S. M. & ZHANG, T. (2012). A tail inequality for quadratic forms of subgaussian random vectors. *Electron. Commun. Prob.* **17**, 1–6.

LERASLE, M. & OLIVEIRA, R. (2011). Robust empirical mean estimators. *arXiv:* 1112.3914.

LIU, H., HAN, F., YUAN, M., LAFFERTY, J. & WASSERMAN, L. (2012). High-dimensional semiparametric Gaussian copula graphical models. *Ann. Statist.* **40**, 2293–326.

PETROV, V. (1995). *Limit Theorems of Probability Theory.* Oxford: Clarendon Press.

ROTHMAN, A., BICKEL, P., LEVINA, E. & ZHU, J. (2008). Sparse permutation invariant covariance estimation. *Electron. J. Statist.* **2**, 494–515.

ROTHMAN, A., LEVINA, E. & ZHU, J. (2009). Generalized thresholding of large covariance matrices. *J. Am. Statist. Assoc.* **104** 177–86.

SERFLING, R. & MAZUMDER, S. (2009). Exponential probability inequality and convergence results for the median absolute deviation and its modifications. *Statist. Prob. Lett.* **79**, 1767–73.

SHAO, Q. M. (1999). A Cramér type large deviation result for Student's t-statistic. *J. Theor. Prob.* **12**, 385–98.