Online Appendix for The Statistical Limit of Arbitrage

Rui Da* — Indiana University

Stefan Nagel† — University of Chicago; NBER and CEPR

Dacheng Xiu‡ — University of Chicago; NBER

This version: October 13, 2024

Abstract

The appendix contains additional empirical and theoretical results, as well as all the mathematical proofs.

*Address: 1275 E 10th St, Bloomington, IN 47405. E-mail: ruida@iu.edu.

†Address: 5807 S Woodlawn Avenue, Chicago, IL 60637. E-mail: stefan.nagel@chicagobooth.edu.

‡Address: 5807 S Woodlawn Avenue, Chicago, IL 60637. E-mail: dacheng.xiu@chicagobooth.edu.

Appendix A Additional Empirical Results

In this appendix we expand our empirical analysis. We first estimate latent factor models of portfolio returns, outlined in Section 4.3, under various alternative estimation windows and numbers of factors. The estimation is conducted for both the original portfolios and the ones adjusted for publication effects. We report the results in Table A1, which include the cross-sectional in-sample explanatory power (R^2) of expected returns of the factor models, the infeasible Sharpe ratios ($S^\star$) corresponding to the portfolio alphas, and the distributions of portfolio alphas' t-statistics.

T	# of Factors	R^2	$S^\star$	$\lvert t\text{-stat}\rvert > 2$	$\lvert t\text{-stat}\rvert > 3$	R^2	$S^\star$	$\lvert t\text{-stat}\rvert > 2$	$\lvert t\text{-stat}\rvert > 3$
		Original Portfolios				Adjusted for Publication Effects			
3 yrs	5	0.41	27.58	12.17	2.63	0.48	13.47	9.93	1.67
	10	0.46	33.04	17.90	5.19	0.58	16.31	15.53	3.91
	15	0.50	39.69	24.44	8.98	0.63	19.77	21.78	7.29
5 yrs	5	0.41	20.60	11.54	2.37	0.47	10.05	9.11	1.37
	10	0.47	23.11	14.78	3.68	0.59	11.25	12.06	2.45
	15	0.51	25.53	17.73	5.07	0.66	12.46	14.94	3.55
7 yrs	5	0.38	17.62	12.49	2.76	0.43	8.56	9.57	1.45
	10	0.46	19.01	14.15	3.45	0.59	9.14	11.12	1.99
	15	0.50	20.44	16.02	4.39	0.65	9.93	13.27	2.81

Table A1: Latent Factor Models of Portfolio Returns

Note: The left panel displays the results of estimating latent factor models with the original portfolio returns as test assets. Across estimation windows and numbers of factors, the panel reports the cross-sectional R^2, the infeasible Sharpe ratios ($S^\star$), and the percentages of portfolio alphas with large t-statistics. Similarly, the right panel details the result when the test assets are portfolio returns adjusted for publication effects.

The cross-sectional R^2s by these models are similar across estimation windows, and moderately increase with more factors included and publication effects adjusted for. Indeed, in both cases, a larger proportion of cross-sectional variation of expected returns will be attributed to risk premia. Considering that these R^2s are in-sample ones, we choose to focus on the 10-factor model in Section 4. Regarding the strength and rareness of alphas, both the percentage of portfolio alphas with large t-statistics and the infeasible Sharpe ratio ($S^\star$) increase with the number of factors. It suggests that, when we control for more factors, the removal of factor risks is perhaps more prominent than the attribution of expected returns to risk premia (thereby less to alphas), which ultimately allows alphas to stand out and to be profited from more easily. On the other hand, with longer estimation window, we see both the infeasible Sharpe ratio and the percentage of large t-statistics reduce. This result could

originate from that many alphas are only moderately persistent and is behind our choice of 5-year window in Section 4. Lastly, since fewer alphas are left after the publication effects are adjusted for, we observe increased cross-section R^2, and both the infeasible Sharpe ratios and the proportion of strong alpha signals shrink.

With the portfolio alphas obtained from the factor models, we show in Table A2 the investment performance of the four different trading strategies: OPT method, CSR, BH, and Lasso, measured by the average Sharpe ratios between January 1970 and December 2020.[1] It complements the analysis of Section 4.3, which focuses on 5-year estimation window and 10-factor specification. As a robustness check for the analysis on individual equity returns by Section 4.2, we also report in Table A2 the average Sharpe ratios generated by applying the four trading strategies to individual equity alphas.

T	# of Factors		5	10	15	5	10	15
		Individual Equities	Original Portfolios			Adjusted for Publication Effects		
3 yrs	OPT	0.71	1.23	1.33	1.25	0.54	0.59	0.54
	CSR	0.65	1.22	1.30	1.27	0.52	0.57	0.56
	BH	0.49	1.06	1.22	1.25	0.25	0.43	0.57
	Lasso	0.56	0.87	1.08	1.09	0.45	0.56	0.47
5 yrs	OPT	0.71	1.50	1.71	1.71	0.51	0.66	0.67
	CSR	0.70	1.47	1.67	1.70	0.55	0.61	0.66
	BH	0.65	1.52	1.49	1.67	0.25	0.35	0.48
	Lasso	0.62	0.95	1.18	1.23	0.48	0.45	0.47
7 yrs	OPT	0.81	1.71	1.95	2.01	0.72	0.73	0.79
	CSR	0.90	1.64	1.93	1.98	0.77	0.74	0.76
	BH	0.38	1.68	1.75	1.87	0.42	0.60	0.48
	Lasso	0.83	1.10	1.32	1.54	0.66	0.60	0.74

Table A2: Sharpe Ratios of Arbitrage Portfiolios

Note: The table reports the average Sharpe ratios generated by arbitrage portfolios constructed using OPT method, CSR, BH, and Lasso, with estimation windows set as 3, 5, and 7 years. The left panel displays the results when we invest in individual equity alphas. The center panel provides the Sharpe ratios generated when the investment universe is composed of portfolio alphas. The right panel displays the results when the publication effects are adjusted for when using portfolio alphas. The center and right panels include outputs under various numbers of factors in the latent factor model estimation.

Consistent with the evidence on the strength and rareness of alphas from Table A1, the investment gains from trading portfolio alphas significantly decline after the adjustment for publication effects. OPT method and CSR yield leading performance under most model specifications and for both individual equity alphas and portfolio alphas. Lasso generates

[1]The Sharpe ratio is the average between January 1972 and December 2020 under the 7-year estimation window, since our return data start from at January 1965.

second-tier outputs, due to the challenge of locating the optimal tuning parameter based on highly noisy return data. Notably, for portfolio alphas, we generally capture smaller Sharpe ratios when we control for fewer factors, which echoes the suggestive evidence from Table A1 on the impact of factor risks.

Appendix B Additional Theoretical Results

Proposition B1. *Suppose the same assumptions and information set $\mathcal{G}$ as in Theorem 2. Moreover, we assume that (a) $\|\beta\|_{\text{MAX}} \lesssim_{\text{P}} 1$ and $\lambda_{\min}(\beta^\intercal\beta) \gtrsim_{\text{P}} N$; (b) v_t is i.i.d. across t, $\text{E}(v_t) = 0$, and its covariance matrix Σ_v satisfies $1 \lesssim \lambda_{\min}(\Sigma_v) \leq \lambda_{\max}(\Sigma_v) \lesssim 1$. Then it holds that, for any strategy w that is $\mathcal{G}$-measurable,*

$$S(w) \leq (S(\mathcal{G})^2 + \gamma^\intercal\Sigma_v^{-1}\gamma)^{1/2} + o_{\text{P}}(1 + S(\mathcal{G})).$$

Proposition B2. *Suppose that r_t follows (1), Assumption 1 holds, $u_{i,t} \sim \mathcal{N}(0, \sigma^2)$, and α following (5) as in Example 1. We also assume $\mu \lesssim N^{-d}$, $N^d \lesssim T \lesssim N^{1-d}$, and $N^{-d'} \lesssim \rho \lesssim N^{-d}$ for some fixed $d' > d > 0$. We denote the Sharpe ratio of the arbitrage portfolio given by (23), as $\widehat{S}^{\text{CSR}}$. Then it satisfies $\widehat{S}^{\text{CSR}} - S^{\text{CSR}} = o_{\text{P}}(1)$, where*

$$S^{\text{CSR}} = \frac{N^{1/2}\rho\mu^2\sigma^{-2}}{(T^{-1} + \rho\mu^2\sigma^{-2})^{1/2}}.$$

Suppose further that S^{OPT} does not vanish. It follows that $S^{\text{CSR}} = S^{\text{OPT}}(1 + o(1))$, if and only if there exists $c_N \to 0$, such that $\sqrt{T}\mu/\sigma \leq c_N$ or $\sqrt{\rho T}\mu/\sigma \geq c_N^{-1}$ holds for all large N.

Proposition B3. *Suppose the same assumptions as in Proposition B2 hold. For any given τ, the Sharpe ratio of the arbitrage portfolio with weights given by (24) denoted by $\widehat{S}_\tau^{\text{BH}}$, satisfies $\widehat{S}_\tau^{\text{BH}} = S_\tau^{\text{BH}} + o_{\text{P}}(1 + S^{\text{OPT}})$, where*[2]

$$S_\tau^{\text{BH}} = \frac{N^{1/2}\text{E}(\alpha_i\check{s}_i\mathbb{1}_{\{|\check{s}_i|\geq s^*\}})}{\sigma\sqrt{\text{E}(\check{s}_i^2\mathbb{1}_{\{|\check{s}_i|\geq s^*\}})}} = \frac{N^{1/2}\rho\mu\int_{-\infty}^{\infty} x\mathbb{1}_{\{|x|\geq s^*\}}\phi_{1/T}(x - \mu\sigma^{-1})dx}{\sigma\sqrt{\int_{-\infty}^{\infty} x^2\mathbb{1}_{\{|x|\geq s^*\}}((1-\rho)\phi_{1/T}(x) + \rho\phi_{1/T}(x - \mu\sigma^{-1}))dx}},$$

$\check{s}_i = (\alpha_i + \bar{u}_i)/\sigma_i$, and s^* is the smallest positive solution of the equation

$$2(1-\tau)\Phi(-T^{1/2}s) = \tau\rho\Phi(T^{1/2}(\mu/\sigma - s)).^{[3]} \tag{B.1}$$

Suppose further that S^{OPT} does not vanish. Then there exists, for any fixed $\epsilon > 0$, some $\tau > 0$

[2] If $\widehat{w}^{\text{BH}} = 0$, i.e., no asset is selected, we set $\widehat{S}^{\text{BH}} = 0$ by convention.

[3] $\Phi(\cdot)$ is the standard normal cumulative distribution function.

such that, as $N, T \to \infty$, $S_\tau^{\mathrm{BH}} > (1-\epsilon)S^{\mathrm{OPT}}$, *if and only if* $\sqrt{T}\mu/\sigma \geq \sqrt{-2\log\rho}(1+o(1))$.

Proposition B4. *Suppose the same assumptions as in Proposition B2 hold. The Sharpe ratio of the arbitrage portfolio with weights given by* $\widehat{w}_q(\lambda)$, *denoted as* $\widehat{S}_{q,\lambda}$ *for* $q = 1, 2$, *satisfies* $\widehat{S}_{1,\lambda} - S_\lambda^{\mathrm{Lasso}} = o_{\mathrm{P}}(1 + S^{\mathrm{OPT}})$ *and* $\widehat{S}_{2,\lambda} - S^{\mathrm{CSR}} = o_{\mathrm{P}}(1 + S^{\mathrm{OPT}})$, *where*

$$S_\lambda^{\mathrm{Lasso}} = \frac{N^{1/2}\mathrm{E}(\alpha_i\psi_1(\check{s}_i,\lambda))}{\sigma\sqrt{\mathrm{E}(\psi_1(\check{s}_i,\lambda)^2)}} = \frac{N^{1/2}\rho\mu\int_{-\infty}^{\infty}\mathrm{sgn}(x)(|x|-\lambda)_+\phi_{1/T}(\mu\sigma^{-1}-x)dx}{\sigma\sqrt{\int_{-\infty}^{\infty}((|x|-\lambda)_+)^2((1-\rho)\phi_{1/T}(x)+\rho\phi_{1/T}(\mu\sigma^{-1}-x))dx}}, \tag{B.2}$$

and S^{CSR} *is defined in Proposition B2.*

Furthermore, if S^{OPT} *does not vanish, then* $S_\lambda^{\mathrm{Lasso}} = S^{\mathrm{OPT}}(1+o(1))$, *under some deterministic non-negative sequence of* λ, *if and only if, for some* $c_N \to 0$, *either* $\sqrt{T}\mu/\sigma \leq c_N$ *or* $\sqrt{T}\mu/\sigma - \sqrt{-2\log\rho} \geq c_N^{-1}$ *holds for all large* N.

Moreover, when $\sqrt{T}\mu/\sigma \leq c_N$, $S_\lambda^{\mathrm{Lasso}}$ *approaches* S^{OPT} *if and only if* λ *satisfies* $T^{1/2}\lambda \to 0$. *When* $\sqrt{T}\mu/\sigma - \sqrt{-2\log\rho} \geq c_N^{-1}$, $S_\lambda^{\mathrm{Lasso}}$ *approaches* S^{OPT} *if and only if* λ *satisfies* $\sqrt{T}(\mu/\sigma - \lambda) \to \infty$ *and* $\frac{\phi(\sqrt{T}\lambda)}{\rho(1+T\lambda^2)T(\mu/\sigma-\lambda)^2} \to 0$.

Appendix C Mathematical Proofs

C.1 Proof of Theorem 1

Since β is $\mathcal{G}$-measurable, we can write, for all w satisfying $w^\intercal\beta = 0$,

$$w^\intercal r_{t+1} = w^\intercal(\alpha + u_{t+1}).$$

Because $\mathrm{E}(u_{t+1}|\mathcal{F}_t) = 0$ according to the specification of (1) and $\mathcal{G}$ is a subset of $\mathcal{F}_t$, we have $\mathrm{E}(u_{t+1}|\mathcal{G}) = 0$. Moreover, since α is $\mathcal{F}$-measurable, it holds that $\mathrm{E}(u_{t+1}\alpha^\intercal|\mathcal{G}) = \mathrm{E}(\mathrm{E}(u_{t+1}\alpha^\intercal|\mathcal{F})|\mathcal{G}) = 0$, i.e., u_{t+1} and α are $\mathcal{G}$-conditionally uncorrelated. Therefore, $\mathrm{Var}(\alpha + u_{t+1}|\mathcal{G}) = \mathrm{Var}(\alpha|\mathcal{G}) + \mathrm{Var}(u_{t+1}|\mathcal{G})$. On the other hand, since $\mathrm{E}(u_{t+1}|\mathcal{G}) = 0$, we have $\mathrm{Var}(u_{t+1}|\mathcal{G}) = \mathrm{E}(u_{t+1}u_{t+1}^\intercal|\mathcal{G}) = \mathrm{E}(\mathrm{E}(u_{t+1}u_{t+1}^\intercal|\mathcal{F})|\mathcal{G}) = \mathrm{E}(\Sigma_u|\mathcal{G})$. Finally, we obtain $\mathrm{Var}(\alpha + u_{t+1}|\mathcal{G}) = \widetilde{\Sigma}_u$. Then, it holds that, for all $w \in \mathcal{G}$ satisfying $w^\intercal\beta = 0$,

$$U(w) = \mathrm{E}(w^\intercal r_{t+1}|\mathcal{G}) - \frac{\kappa}{2}\mathrm{Var}(r_{t+1}|\mathcal{G}) = w^\intercal\widetilde{\alpha} - \frac{\kappa}{2}w^\intercal\widetilde{\Sigma}_u w. \tag{C.3}$$

We first find w that maximize $w^\intercal\widetilde{\alpha} - \frac{\kappa}{2}w^\intercal\widetilde{\Sigma}_u w$ only under the constraint $w^\intercal\beta = 0$. Next, we will verify that the solution is unique and is indeed $\mathcal{G}$-measurable, which would prove the theorem. The first step is equivalent with maximizing

$$w^\intercal \widetilde{\alpha} - \frac{\kappa}{2} w^\intercal \widetilde{\Sigma}_u w + \lambda w^\intercal \beta,$$

where λ is the Lagrange multiplier such that the solution w satisfies $w^\intercal \beta = 0$. As the objective function is clearly strictly concave, we only need to look at the first-order condition, which reads

$$w = \frac{1}{\kappa} \widetilde{\Sigma}_u^{-1} (\widetilde{\alpha} + \lambda \beta).$$

We can pin down λ by requiring $\beta^\intercal \widetilde{\Sigma}_u^{-1} (\widetilde{\alpha} + \lambda \beta) = 0$, which gives $\lambda = (\beta^\intercal \widetilde{\Sigma}_u^{-1} \beta)^{-1} \beta^\intercal \widetilde{\Sigma}_u^{-1} \widetilde{\alpha}$ and thereby

$$w = \frac{1}{\kappa} \widetilde{\Sigma}_u^{-1} (\mathbb{I}_N + \beta (\beta^\intercal \widetilde{\Sigma}_u^{-1} \beta)^{-1} \beta^\intercal \widetilde{\Sigma}_u^{-1}) \widetilde{\alpha} = \frac{1}{\kappa} \widetilde{\Sigma}_u^{-1/2} \mathbb{M}_{\widetilde{\Sigma}_u^{-1/2} \beta} \widetilde{\Sigma}_u^{-1/2} \widetilde{\alpha}.$$

C.2 Proof of Theorem 2

To simplify the notation, we omit the dependence of β, Σ on N, and w on N and T. All limits are taken as $N \to \infty$. The derivation applies to either fixed T or $T \to \infty$ together with N.

We first note that, given (1), conditioning on $\mathcal{G}$ is equivalent to conditioning on the information set generated by

$$\{(\alpha_i + u_{i,s}, \beta_i, v_s, \sigma_i) : t - T + 1 \leq s \leq t, i \leq N\}.$$

According to Assumption 1, conditionally on Σ_u, $\{(\alpha_i, \alpha_i + u_{i,s}) : t - T + 1 \leq s \leq t\}$ is independent of $\{(\alpha_j + u_{j,s}, \beta_{j'}, v_s) : t - T + 1 \leq s \leq t, j, j' \leq N, j \neq i\}$. Therefore, the $\mathcal{G}$-conditional distribution of α_i is the same as the distribution of α_i conditional on $\{\alpha_i + u_{i,s} : t - T + 1 \leq s \leq t\}$ and Σ_u. Because σ_j is independent with (α_i, u_i) for $j \neq i$, the $\mathcal{G}$-conditional distribution of α_i is the same as the the $\mathcal{G}_i$-conditional distribution of α_i, where $\mathcal{G}_i$ is the information set generated by $\{(\alpha_i + u_{i,s}, \sigma_i) : t - T + 1 \leq s \leq t\}$. Since $\mathcal{G}_i$ is independent across i by Assumption 1, we conclude that, conditionally on $\mathcal{G}$, α_i remains independent across i.

Now define $\mathcal{E} = \mathrm{E}(w^\intercal r_{t+1} | \mathcal{F}_t) - \mathrm{E}(w^\intercal r_{t+1} | \mathcal{G})$. By the definition of $S(\widehat{w})$, we have

$$S(\widehat{w}) = \mathrm{E}(w^\intercal r_{t+1} | \mathcal{G}) / \mathrm{Var}(w^\intercal r_{t+1} | \mathcal{F}_t)^{1/2} + \mathcal{E} / \mathrm{Var}(w^\intercal r_{t+1} | \mathcal{F}_t)^{1/2}. \tag{C.4}$$

Since w is $\mathcal{G}$-measurable, it follows that $\mathcal{E} = w^\intercal (\alpha - \mathrm{E}(\alpha | \mathcal{G}))$ and that $\mathrm{E}(\mathcal{E}^2 | \mathcal{G}) = w^\intercal \mathrm{Var}(\alpha | \mathcal{G}) w$. Then, using Chebyshev's inequality, we have, for all positive fixed ϵ,

$$\mathrm{P}(|\mathcal{E}| / \|w\| \geq \epsilon) \leq \mathrm{E}(\mathcal{E}^2 / \|w\|^2) / \epsilon^2 = \mathrm{E}(w^\intercal \mathrm{Var}(\alpha | \mathcal{G}) w / \|w\|^2) / \epsilon^2. \tag{C.5}$$

Because conditionally on $\mathcal{G}$, α_i is independent across i, we have $\mathrm{Var}(\alpha|\mathcal{G})_{i,j} = \delta_{i,j}\mathrm{Var}(\alpha_i|\mathcal{G})$. It thereby follows that

$$\mathrm{E}(w^\intercal \mathrm{Var}(\alpha|\mathcal{G})w/\|w\|^2) \leq \mathrm{E}(\max_{i\leq N} \mathrm{Var}(\alpha_i|\mathcal{G})) \leq \mathrm{E}(\max_{i\leq N} \alpha_i^2) = o(1), \tag{C.6}$$

where the last step comes from condition (a) of Assumption 1. Combining (C.5) and (C.6), and using $\mathrm{Var}(w^\intercal r_{t+1}|\mathcal{F}_t) = w^\intercal \Sigma w \geq \lambda_{\min}(\Sigma_u)\|w\|^2 \gtrsim_{\mathrm{P}} \|w\|^2$, we obtain

$$|\mathcal{E}|/\mathrm{Var}(w^\intercal r_{t+1}|\mathcal{F}_t)^{1/2} \lesssim_{\mathrm{P}} |\mathcal{E}|/\|w\| = o_{\mathrm{P}}(1). \tag{C.7}$$

(C.7) and (C.4) lead to

$$S(w) = w^\intercal \mathrm{E}(r_{t+1}|\mathcal{G})(w^\intercal \Sigma w)^{-1/2} + o_{\mathrm{P}}(1). \tag{C.8}$$

Furthermore, if $w^\intercal \beta = 0$, then it follows that $w^\intercal r_t = w^\intercal(\alpha + u_t)$ and $w^\intercal \Sigma w = w^\intercal \Sigma_u w$. Equation (C.8) then becomes

$$S(w) = w^\intercal \widetilde{\alpha}(w^\intercal \Sigma_u w)^{-1/2} + o_{\mathrm{P}}(1). \tag{C.9}$$

Applying Cauchy-Schwarz inequality, we obtain

$$|w^\intercal \widetilde{\alpha}|^2 (w^\intercal \Sigma_u w)^{-1} \leq \widetilde{\alpha}^\intercal \Sigma_u^{-1} \widetilde{\alpha} = S(\mathcal{G})^2,$$

which, combined with (C.9), proves the inequality in (11).

We move on to the remaining results: the equality in (11), (12), and the optimal utility. They all hinge on the property of $\widetilde{w}$. As a first step, we introduce short-hand notation

$$\check{w} := \frac{1}{\kappa}\Sigma_u^{-1/2}\mathbb{P}_{\Sigma_u^{-1/2}\beta}\Sigma_u^{-1/2}\widetilde{\alpha}, \quad \check{w}' := \frac{1}{\kappa}\mathbb{P}_\beta \Sigma_u^{-1}\widetilde{\alpha}.$$

Given the specification of $\mathcal{G}$, we have $\widetilde{\Sigma}_u = \Sigma_u$ and thereby

$$\widetilde{w} = \frac{1}{\kappa}\Sigma_u^{-1}\widetilde{\alpha} - \check{w}, \quad \widetilde{w} - \frac{1}{\kappa}\mathbb{M}_\beta \Sigma_u^{-1}\widetilde{\alpha} = \check{w}' - \check{w}. \tag{C.10}$$

To establish the equality in (11), (12), and the optimal utility, we now prove that both $\|\check{w}\|$ and $\|\check{w}'\|$ are $o_{\mathrm{P}}(1)$, which will then quickly leads to those results.

To this end, we start by analyzing $\mathrm{E}(\widetilde{\alpha}|\beta, \Sigma_u)$ and $\mathrm{Cov}(\widetilde{\alpha}|\Sigma_u, \beta)$. We note that

$$\mathrm{E}(\widetilde{\alpha}|\beta, \Sigma_u) = \mathrm{E}(\alpha|\beta, \Sigma_u) = \mathrm{E}(\alpha|\Sigma_u) = 0. \tag{C.11}$$

The first equality comes from that $\widetilde{\alpha} \coloneqq \mathrm{E}(\alpha|\mathcal{G})$ and that β and Σ_u are $\mathcal{G}$-measurable. The second equality comes from Assumption 1 (b). The last equality comes from Assumption 1 (a). Moreover, from the analysis above (C.4), $\widetilde{\alpha}_i$ is a function of $\{(\alpha_i + u_{i,s}, \sigma_i) : t-T+1 \leq s \leq t\}$. According to Assumption 1 (b), $\{(\alpha_i + u_{i,s}, \sigma_i) : t-T+1 \leq s \leq t\}$ and β are, conditionally on Σ_u, independent. Therefore,

$$\mathrm{Cov}(\widetilde{\alpha}_i, \widetilde{\alpha}_j|\Sigma_u, \beta) = \mathrm{Cov}(\widetilde{\alpha}_i, \widetilde{\alpha}_j|\Sigma_u) = \delta_{i,j}\mathrm{Var}(\widetilde{\alpha}_i|\Sigma_u). \tag{C.12}$$

The second equality comes from $\mathrm{Cov}(\widetilde{\alpha}_i, \widetilde{\alpha}_j|\Sigma_u) = 0$ for $i \neq j$, which is because (α_i, u_i) is i.i.d. across i per Assumption 1 (a). Therefore, it holds that

$$\mathrm{Cov}(\widetilde{\alpha}|\Sigma_u, \beta) \leq \mathrm{E}(\|\alpha\|^2_{\mathrm{MAX}}|\Sigma_u)\mathbb{I}_N \tag{C.13}$$

It hence holds that

$$\begin{aligned} \mathrm{E}(\|\check{w}\|^2|\Sigma_u, \beta) &= \mathrm{Tr}(\Sigma_u^{-1/2}\mathbb{P}_{\Sigma_u^{-1/2}\beta}\Sigma_u^{-1/2}\mathrm{Cov}(\widetilde{\alpha}|\Sigma_u, \beta)\Sigma_u^{-1/2}\mathbb{P}_{\Sigma_u^{-1/2}\beta}\Sigma_u^{-1/2}) \\ &\leq \mathrm{E}(\|\alpha\|^2_{\mathrm{MAX}}|\Sigma_u)\mathrm{Tr}(\Sigma_u^{-1/2}\mathbb{P}_{\Sigma_u^{-1/2}\beta}\Sigma_u^{-1}\mathbb{P}_{\Sigma_u^{-1/2}\beta}\Sigma_u^{-1/2}). \end{aligned} \tag{C.14}$$

The equality comes from (C.11). The inequality comes from (C.12), (C.13), and the well-known result on Loewner order (see, e.g., Theorem 7.7.2 at Horn and Johnson (2012)). On the other hand, it holds that

$$\begin{aligned} &\mathrm{Tr}(\Sigma_u^{-1/2}\mathbb{P}_{\Sigma_u^{-1/2}\beta}\Sigma_u^{-1}\mathbb{P}_{\Sigma_u^{-1/2}\beta}\Sigma_u^{-1/2}) \lesssim_{\mathrm{P}} \mathrm{Tr}(\Sigma_u^{-1/2}\mathbb{P}_{\Sigma_u^{-1/2}\beta}\mathbb{P}_{\Sigma_u^{-1/2}\beta}\Sigma_u^{-1/2}) \\ = \ &\mathrm{Tr}(\mathbb{P}_{\Sigma_u^{-1/2}\beta}\Sigma_u^{-1}\mathbb{P}_{\Sigma_u^{-1/2}\beta}) \lesssim_{\mathrm{P}} \mathrm{Tr}(\mathbb{P}_{\Sigma_u^{-1/2}\beta}) = \mathrm{rank}(\beta) = K. \end{aligned} \tag{C.15}$$

Both inequalities come from $\Sigma_u \gtrsim_{\mathrm{P}} \mathbb{I}_N$ by Assumption 1 (a). Substituting (C.15) into (C.14), we obtain

$$\mathrm{E}(\|\check{w}\|^2|\Sigma_u, \beta) \lesssim_{\mathrm{P}} \mathrm{E}(\|\alpha\|^2_{\mathrm{MAX}}|\Sigma_u) = o_{\mathrm{P}}(1), \tag{C.16}$$

where the last bound comes from Assumption 1 (a). Symmetrically, we have

$$\begin{aligned} \mathrm{E}(\|\check{w}'\|^2|\Sigma_u, \beta) &= \mathrm{Tr}(\mathbb{P}_\beta\Sigma_u^{-1}\mathrm{Cov}(\widetilde{\alpha}|\Sigma_u, \beta)\Sigma_u^{-1}\mathbb{P}_\beta) \leq \mathrm{E}(\|\alpha\|^2_{\mathrm{MAX}}|\Sigma_u)\mathrm{Tr}(\mathbb{P}_\beta\Sigma_u^{-2}\mathbb{P}_\beta) \\ &\lesssim_{\mathrm{P}} \mathrm{E}(\|\alpha\|^2_{\mathrm{MAX}}|\Sigma_u)\mathrm{Tr}(\mathbb{P}_\beta) = \mathrm{E}(\|\alpha\|^2_{\mathrm{MAX}}|\Sigma_u)K = o_{\mathrm{P}}(1). \end{aligned} \tag{C.17}$$

Applying Chebyshev's inequality to (C.16) and (C.17), we obtain

$$\|\check{w}\| + \|\check{w}'\| = o_{\mathrm{P}}(1). \tag{C.18}$$

Combining (C.18) with the second part of (C.10), we establish (12) using the triangle inequality.

To prove the equality in (11), we note that, using the first part of (C.10),

$$\widetilde{w}^\intercal\widetilde{\alpha} = \frac{1}{\kappa}\widetilde{\alpha}^\intercal\Sigma_u^{-1}\widetilde{\alpha} - \check{w}^\intercal\widetilde{\alpha}, \quad \widetilde{w}^\intercal\Sigma_u\widetilde{w} = \frac{1}{\kappa^2}\widetilde{\alpha}^\intercal\Sigma_u^{-1}\widetilde{\alpha} + \check{w}^\intercal\Sigma_u\check{w} - \frac{2}{\kappa}\check{w}^\intercal\widetilde{\alpha}. \tag{C.19}$$

On the other hand, applying that $\Sigma_u \lesssim_{\mathrm{P}} \mathbb{I}_N$ by Assumption 1 (a) and (C.18), we obtain

$$\check{w}^\intercal\Sigma_u\check{w} \lesssim_{\mathrm{P}} \|\check{w}\| = o_{\mathrm{P}}(1), \quad |\check{w}^\intercal\widetilde{\alpha}| \leq \sqrt{\check{w}^\intercal\Sigma_u\check{w}}S(\mathcal{G}) \lesssim_{\mathrm{P}} \|\check{w}\|S(\mathcal{G}) = o_{\mathrm{P}}(S(\mathcal{G})). \tag{C.20}$$

Substituting (C.19) into (C.20), we establish

$$\widetilde{w}^\intercal\widetilde{\alpha} = \frac{1}{\kappa}S(\mathcal{G})^2 + o_{\mathrm{P}}(S(\mathcal{G})), \quad \widetilde{w}^\intercal\Sigma_u\widetilde{w} = \frac{1}{\kappa^2}S(\mathcal{G})^2 + o_{\mathrm{P}}(1 + S(\mathcal{G})). \tag{C.21}$$

Suppose $S(\mathcal{G}) \gtrsim_{\mathrm{P}} 1$. Then, if $w = \widetilde{w}$, the equality in (11) indeed follows from (C.21). Suppose $S(\mathcal{G}) = o_{\mathrm{P}}(1)$. Then $S(\widetilde{w}) \leq o_{\mathrm{P}}(1)$ and $-S(\widetilde{w}) = S(-\widetilde{w}) \leq o_{\mathrm{P}}(1)$ according to (the inequality part of) (11) as $\widetilde{w}$ is both $\mathcal{G}$-measurable and factor-neutral. In other words, we have $S(\widetilde{w}) = o_{\mathrm{P}}(1)$ and that the equality in (11) also holds. Applying the classic subsequence argument, we establish that the equality in (11) holds if $w = \widetilde{w}$.

Finally, substituting (C.21) into (C.3), we obtain (recall $\widetilde{\Sigma}_u = \Sigma_u$ here)

$$U(\widetilde{w}) = \widetilde{w}^\intercal\widetilde{\alpha} - \frac{\kappa}{2}\widetilde{w}^\intercal\Sigma_u\widetilde{w} = \frac{1}{2\kappa}S(\mathcal{G})^2 + o_{\mathrm{P}}(1 + S(\mathcal{G})).$$

Taking square root, we establish the optimal utility result.

C.3 Proof of Theorem 3, Corollary 2, and Corollary 3

Proof of Theorem 3. We have established in the beginning of the proof of Theorem 2 that the $\mathcal{G}$-conditional distribution of α_i is the same as the $\mathcal{G}_i$-conditional distribution of α_i, where $\mathcal{G}_i$ is the information set generated by $\{(\alpha_i + u_{i,s}) : t - T + 1 \leq s \leq t\}$ and σ_i. Note that $u_{i,s}$ is centered normal, we have that the conditional probability density of $\{r_{i,s}^* := \alpha_i + u_{i,s}, t - T + 1 \leq s \leq t\}$ given s_i and σ_i, denoted by $p(r_i^*|s_i, \sigma_i)$, is

$$p(r_i^*|s_i, \sigma_i) = \prod_{t-T+1\leq s\leq t} \sigma_i^{-1}\phi(\sigma_i^{-1}r_{i,s}^* - s_i) = \phi(T^{1/2}(\check{s}_i - s_i))f(r_i^*). \tag{C.22}$$

Here $f(r_i^*)$ is a function of r_i^* and σ_i that does not depend on s_i. Hence, applying Bayes' theorem, we have

$$
\begin{aligned}
\widetilde{\alpha}_i = \mathrm{E}(\alpha_i|\mathcal{G}_i) = \sigma_i \mathrm{E}(s_i|\mathcal{G}_i) &= \sigma_i \int x p(s_i = x|r_i^*, \sigma_i) dx \\
&= \sigma_i \int x \frac{p(r_i^*|s_i = x, \sigma_i) p(s_i = x|\sigma_i)}{\int p(r_i^*|s_i = x', \sigma_i) p(s_i = x'|\sigma_i) dx'} dx \\
&= \sigma_i \int x \frac{p(r_i^*|s_i = x, \sigma_i) p_s(x)}{\int p(r_i^*|s_i = x', \sigma_i) p_s(x') dx'} dx \\
&= \sigma_i \int x \frac{\phi(T^{1/2}(\check{s}_i - x)) p_s(x)}{\int \phi(T^{1/2}(\check{s}_i - x')) p_s(x') dx'} dx = \sigma_i \mathrm{E}(s_i|\check{s}_i),
\end{aligned}
$$

The second line comes from the Bayes' theorem. In the third line, $p_s(\cdot)$ is the marginal density of s_i that is invariant across i, and we use the fact that s_i and σ_i are independent, given by condition (a) of Assumption 2. The first equality in the last line comes from (C.22). The second equality in the last line comes from the Bayes' theorem and that $\phi(T^{1/2}(\check{s}_i - x))$ is, up to a constant, the density of $\check{s}_i$ conditional on $s_i = x$. We hence establish the first statement in the theorem.

Next, we note that $\psi(a) = \mathrm{E}(s_i|\check{s}_i = a)$ and, conditional on $s_i = a$, $\check{s}_i \approx \mathcal{N}(a, T^{-1})$. The marginal density of $\check{s}_i$ is then indeed $\mathrm{E}(\phi_{1/T}(a - s_i))$. Therefore, (18) directly comes from the Tweedie's formula, see, e.g., Robbins (1956) . The proof ends. ∎

Proof of Corollary 2. Apparently, $\check{s}_i = s_i + \sigma_i^{-1}\bar{u}_i$ is i.i.d. across i. By direction calculation and the definition of S^{OPT} in the statement of the corollary, we have

$$
\mathrm{E}(\widetilde{\alpha}^\intercal \Sigma_u^{-1} \widetilde{\alpha}) = \sum_i \mathrm{E}(\mathrm{E}(s_i|\mathcal{G})^2) = N \int \psi(a)^2 p(a) da = (S^{\mathrm{OPT}})^2. \tag{C.23}
$$

Now we study $S(\mathcal{G}) = \widetilde{\alpha}^\intercal \Sigma_u^{-1} \widetilde{\alpha} = \sum_i \mathrm{E}(s_i|\mathcal{G})^2$. Using the fact that $a^2 - b^2 = (a-b)^2 + 2b(a-b)$, we have

$$
\begin{aligned}
& \mathrm{E}(|\mathrm{E}(s_i \mathbb{1}_{\{|s_i| \le c_N\}}|\mathcal{G})^2 - \mathrm{E}(s_i|\mathcal{G})^2|) \\
\le\ & \mathrm{E}(\mathrm{E}(s_i \mathbb{1}_{\{|s_i| > c_N\}}|\mathcal{G})^2) + 2\mathrm{E}(|\mathrm{E}(s_i|\mathcal{G})\mathrm{E}(s_i \mathbb{1}_{\{|s_i| > c_N\}}|\mathcal{G})|) \\
\le\ & \mathrm{E}(s_i^2 \mathbb{1}_{\{|s_i| > c_N\}}) + 2\sqrt{\mathrm{E}(\mathrm{E}(s_i|\mathcal{G})^2)\mathrm{E}(\mathrm{E}(s_i \mathbb{1}_{\{|s_i| > c_N\}}|\mathcal{G})^2)} \\
\le\ & \mathrm{E}(s_i^2 \mathbb{1}_{\{|s_i| > c_N\}}) + 2\sqrt{\mathrm{E}(\mathrm{E}(s_i|\mathcal{G})^2)\mathrm{E}(s_i^2 \mathbb{1}_{\{|s_i| > c_N\}})} \\
\le\ & c_N N^{-1} + \sqrt{\mathrm{E}(\mathrm{E}(s_i|\mathcal{G})^2) c_N N^{-1}},
\end{aligned} \tag{C.24}
$$

where the last step holds by the assumption $\mathrm{E}(s_i^2 \mathbb{1}_{\{|s_i| \ge c_N\}}) = o(N^{-1})$. Then we have

$$
\mathrm{E}\left(|\sum_i \mathrm{E}(s_i \mathbb{1}_{\{|s_i| \le c_N\}}|\mathcal{G})^2 - \sum_i \mathrm{E}(s_i|\mathcal{G})^2| \right)
$$

$$
\begin{aligned}
&\leq \sum_i \mathrm{E}(|\mathrm{E}(s_i \mathbb{1}_{\{|s_i| \leq c_N\}}|\mathcal{G})^2 - \mathrm{E}(s_i|\mathcal{G})^2|) \\
&\leq c_N + \sqrt{\sum_i \mathrm{E}(\mathrm{E}(s_i|\mathcal{G})^2) c_N} = o(1 + S^{\mathrm{OPT}}),
\end{aligned} \tag{C.25}
$$

where the second inequality is a direct result of (C.24), and the last estimate is given by (C.23). From (C.25) and (C.23), it follows, respectively, using Markov's inequality and triangle inequality that

$$
\sum_i \mathrm{E}(s_i \mathbb{1}_{\{|s_i| \leq c_N\}}|\mathcal{G})^2 = \sum_i \mathrm{E}(s_i|\mathcal{G})^2 + o_{\mathrm{P}}(1 + S^{\mathrm{OPT}}), \tag{C.26}
$$

$$
\mathrm{E}\left(\sum_i \mathrm{E}(s_i \mathbb{1}_{\{|s_i| \leq c_N\}}|\mathcal{G})^2\right) = (S^{\mathrm{OPT}})^2 + o(1 + S^{\mathrm{OPT}}). \tag{C.27}
$$

Further, we have

$$
\begin{aligned}
\mathrm{Var}\left(\sum_i \mathrm{E}(s_i \mathbb{1}_{\{|s_i| \leq c_N\}}|\mathcal{G})^2\right) &= \sum_i \mathrm{Var}(\mathrm{E}(s_i \mathbb{1}_{\{|s_i| \leq c_N\}}|\mathcal{G})^2) \\
&\leq c_N^2 \sum_i \mathrm{E}(\mathrm{E}(s_i \mathbb{1}_{\{|s_i| \leq c_N\}}|\mathcal{G})^2) = o(1 + (S^{\mathrm{OPT}})^2).
\end{aligned} \tag{C.28}
$$

For the first line, we use that $\mathrm{E}(s_i \mathbb{1}_{\{|s_i| \leq c_N\}}|\mathcal{G})$ is independent across i. The second line is obvious as $|s_i| \mathbb{1}_{\{|s_i| \leq c_N\}} \leq c_N$. The last line comes from (C.27). Combining (C.27) and (C.28), we obtain

$$
\sum_i \mathrm{E}(s_i \mathbb{1}_{\{|s_i| \leq c_N\}}|\mathcal{G})^2 = (S^{\mathrm{OPT}})^2 + o(1 + S^{\mathrm{OPT}}) + o_{\mathrm{P}}(1 + (S^{\mathrm{OPT}})^2)^{1/2}.
$$

Along with (C.26), we obtain

$$
\sum_i \mathrm{E}(s_i|\mathcal{G})^2 = (S^{\mathrm{OPT}})^2 + o_{\mathrm{P}}(1 + S^{\mathrm{OPT}}).
$$

In light of the definition of $S(\mathcal{G})$, and the fact that

$$
((S^{\mathrm{OPT}})^2 + o_{\mathrm{P}}(1 + S^{\mathrm{OPT}}))^{1/2} = S^{\mathrm{OPT}} + o_{\mathrm{P}}(1),
$$

we conclude the proof. ∎

Proof of Corollary 3. Because of the tail condition $\mathrm{E}(s_i^2 \mathbb{1}_{\{|s_i| \geq c_N\}}) \leq c_N N^{-1}$ for some sequence $c_N \to 0$ and that σ_i is constant across i, we have

$$\mathrm{E}\left|\alpha^\intercal\alpha - \sum_i \alpha_i^2 \mathbb{1}_{\{|\alpha_i|<c_N\}}\right| = \mathrm{E}\left|\sum_i \alpha_i^2 \mathbb{1}_{\{|\alpha_i|\geq c_N\}}\right| = o(1),$$

which, by Markov's inequality and triangle inequality, respectively, leads to

$$\alpha^\intercal\alpha = \sum_i \alpha_i^2 \mathbb{1}_{\{|\alpha_i|<c_N\}} + o_{\mathrm{P}}(1), \quad \mathrm{E}\left(\sum_i \alpha_i^2 \mathbb{1}_{\{|\alpha_i|<c_N\}}\right) = \mu^2\rho N. \tag{C.29}$$

On the other hand, it holds that

$$\mathrm{Var}\left(\sum_i \alpha_i^2 \mathbb{1}_{\{|\alpha_i|<c_N\}}\right) \leq \sum_i \mathrm{E}(\alpha_i^4 \mathbb{1}_{\{|\alpha_i|<c_N\}}) \leq c_N^2 \sum_i \mathrm{E}(\alpha_i^2) = c_N^2\mu^2\rho N. \tag{C.30}$$

Combining (C.29) and (C.30), we obtain

$$\alpha^\intercal\alpha = \mu^2\rho N + o_{\mathrm{P}}(1 + \mu\sqrt{\rho N}).$$

As a result, it holds that

$$S^\star = \sigma^{-1}\sqrt{\alpha^\intercal\alpha} = \sigma^{-1}\mu(\rho N)^{1/2} + o_{\mathrm{P}}(1). \tag{C.31}$$

Further, in light of the explicit distribution of α in Example 1, we have

$$\psi(a) = \frac{\mu\rho\phi(a - T^{1/2}\mu/\sigma) - \mu\rho\phi(a + T^{1/2}\mu/\sigma)}{(2-2\rho)\phi(a) + \rho\phi(a - T^{1/2}\mu/\sigma) + \rho\phi(a + T^{1/2}\mu/\sigma)}, \tag{C.32}$$

$$\left(S^{\mathrm{OPT}}\right)^2 = \frac{\mu\rho N}{2\sigma^2}\int \psi(a)(\phi(a - T^{1/2}\mu/\sigma) - \phi(a + T^{1/2}\mu/\sigma))da. \tag{C.33}$$

Suppose that $T^{1/2}\mu\sigma^{-1} - \sqrt{-2\log\rho} \leq C < \infty$. Then we have

$$\begin{aligned}\sup_{a\geq C}\frac{\rho\phi(a)}{\phi(a - T^{1/2}\mu/\sigma)} &= \exp\left(\log\rho + T^{1/2}\mu\sigma^{-1}\left(\frac{1}{2}T^{1/2}\mu\sigma^{-1} - C\right)\right) \\ &\leq \exp\left(\log\rho + \frac{1}{2}\left(\sqrt{-2\log\rho} + C\right)\left(\sqrt{-2\log\rho} - C\right)\right) \leq 1.\end{aligned} \tag{C.34}$$

On the other hand, in light of (C.32) and (C.33), we have

$$\begin{aligned}\left(S^{\mathrm{OPT}}\right)^2 &= \frac{\mu\rho N}{\sigma^2}\int \psi(a)\phi(a - T^{1/2}\mu/\sigma)da \\ &\leq \frac{\mu\rho N}{\sigma^2}\int \frac{\mu\rho\phi(a - T^{1/2}\mu/\sigma)}{(2-2\rho)\phi(a) + \rho\phi(a - T^{1/2}\mu/\sigma)}\phi(a - T^{1/2}\mu/\sigma)da \\ &= \frac{\mu^2\rho N}{\sigma^2}\int \frac{\rho\phi(a)}{(2-2\rho)\phi(a - T^{1/2}\mu/\sigma) + \rho\phi(a)}\phi(a)da.\end{aligned}$$

We hence obtain from (C.34) that, for N sufficiently large,

$$\left(S^{\mathrm{OPT}}\right)^2 \leq \frac{\mu^2\rho N}{\sigma^2}\left(\int_{a\geq C}\frac{1}{3-2\rho}\phi(a)da + \int_{a\leq C}\phi(a)da\right) \leq \frac{\mu^2\rho N}{\sigma^2}\left(1-\frac{1}{2}\Phi(-C)\right).$$

This proves the "if" part, given (C.31) and that $\mu^2\rho N/\sigma^2$ does not vanish. Now suppose $T^{1/2}\mu\sigma^{-1} - \sqrt{-2\log\rho} \to \infty$. Then, for all fixed $x > 0$, we have, for sufficiently large N,

$$\begin{aligned}
\sup_{a:|a|\leq x}\frac{\phi(a+T^{1/2}\mu/\sigma)}{\rho\phi(a)} &= \exp\left(-\log\rho - T^{1/2}\mu\sigma^{-1}\left(\frac{1}{2}T^{1/2}\mu\sigma^{-1}-x\right)\right) \\
&\leq \exp\left(-\log\rho - \frac{1}{2}\left(\sqrt{-2\log\rho}+c_N^{-1}\right)\left(\sqrt{-2\log\rho}+c_N^{-1}\right)\right) \\
&\leq \exp\left(-c_N^{-2}/2\right) \to 0, \qquad \text{(C.35)} \\
\sup_{a:|a|\leq x}\frac{\phi(a+2T^{1/2}\mu/\sigma)}{\phi(a)} &= \exp\left(-2T^{1/2}\mu\sigma^{-1}(T^{1/2}\mu\sigma^{-1}-x)\right) \to 0. \qquad \text{(C.36)}
\end{aligned}$$

Given (C.32), it holds that

$$\psi\left(a+T^{1/2}\mu/\sigma\right) = \mu\frac{1-\frac{\phi(a+2T^{1/2}\mu/\sigma)}{\phi(a)}}{1+\frac{(2-2\rho)\phi(a+T^{1/2}\mu/\sigma)}{\rho\phi(a)}+\frac{\phi(a+2T^{1/2}\mu/\sigma)}{\phi(a)}}.$$

Substituting (C.36) into the numerator, and (C.35) and (C.36) into the denominator, we obtain that, for all fixed $x > 0$,

$$\sup_{a:|a|\leq x}\left|\mu^{-1}\psi\left(a+T^{1/2}\mu/\sigma\right)-1\right| \to 0. \qquad \text{(C.37)}$$

Since the integrand of (C.33) is always positive and even in a, it holds that, for all fixed $x > 0$,

$$\begin{aligned}
\left(S^{\mathrm{OPT}}\right)^2 &\geq \frac{\mu\rho N}{\sigma^2}\int_{|a-T^{1/2}\mu/\sigma|\leq x}\psi(a)(\phi(a-T^{1/2}\mu/\sigma)-\phi(a+T^{1/2}\mu/\sigma))da \\
&\geq \frac{\mu\rho N}{\sigma^2}\int_{|a-T^{1/2}\mu/\sigma|\leq x}\psi(a)\phi(a-T^{1/2}\mu/\sigma)(1-c_N)da \\
&\geq \frac{\mu\rho N}{\sigma^2}\int_{|a-T^{1/2}\mu/\sigma|\leq x}\mu\phi(a-T^{1/2}\mu/\sigma)(1-c_N)da \\
&\geq \frac{\mu^2\rho N}{\sigma^2}(1-c_N-2\Phi(-x)).
\end{aligned}$$

Here the second inequality comes from (C.36), the third inequality is a result of (C.37), and the last inequality is obvious. Because this result holds for all fixed $x > 0$, the "only if" part

is proved. ∎

C.4 Proof of Theorem 4

Given the length of the proof, a briefly explanation is warranted to clarify the key ideas and structure.

The whole proof is organized into 5 steps. Steps 1 - 4 demonstrate that the distance between the conditional expectation vector $\psi := \Sigma_u^{-1/2}\mathrm{E}(\alpha|\mathcal{G}) = (\psi(\check{s}_1), \ldots, \psi(\check{s}_N))^\intercal$, and the estimate $\breve{\psi} := (\breve{\psi}(\widehat{s}_1), \ldots, \breve{\psi}(\widehat{s}_N))^\intercal$, measured by L2 norm, is small compared to S^{OPT}. Based on the result, step 5 shows the gap between monotonicity-preserving estimate $\widehat{\psi}$ and ψ in L2 norm is also small compared to S^{OPT}. This leads to that the Sharpe ratio generated by $\widehat{w}^{\mathrm{OPT}} = \mathbb{M}_\beta \widehat{\Sigma}_u^{-1/2}\widehat{\psi}$ converges to S^{OPT}, proved in the last step.

We note that, because of the rare and weak nature of alphas, $\mathrm{E}(\alpha_i|\mathcal{G})$ converges to zero in probability for each individual i, despite their large collective contribution to Sharpe ratio. Therefore, we need instead the L2 norm of errors involved in $\widehat{\psi}$ to converge to zero.

Step 1. Throughout the proof, we use the following notation, introduced in the main text of the paper,

$$\check{s}_i = T^{-1}\sum_{s\in\mathcal{T}}(s_i + \varepsilon_{i,s}), \quad \widehat{s}_i = \widehat{\alpha}_i/\widehat{\sigma}_i, \quad p(a) = \mathrm{E}(\phi_{1/T}(a - s_i)), \quad \psi(a) = \mathrm{E}(s_i|\check{s}_i = a). \tag{C.38}$$

As in that statement, $p(a)$ is the density of $\check{s}_i$, and $\psi(a)$ is the expectation of s_i, conditional on $\check{s}_i = a$.

Intuitively, for assets with large $\check{s}_i$, $\check{s}_i$ is a relatively precisely estimate the true s_i. In contrast, for assets with small $\check{s}_i$, more likely $\check{s}_i$ is driven by noise. As a result, we introduce $B = \{i \le N : |\check{s}_i| \le \widetilde{k}_N T^{-1/2}\}$ to separate the two cases, where $\widetilde{k}_N = k_N^{-2}$. Moreover, we set $\breve{\psi}$ and ψ as the N-dimensional vectors with entries $\breve{\psi}_i := \breve{\psi}(\widehat{s}_i)$ and $\psi_i := \psi(\check{s}_i)$. It holds that

$$\|\breve{\psi} - \psi\|^2 \le \sum_{i\in B}(\breve{\psi}_i - \psi_i)^2 + \sum_{i\in B^c}(\breve{\psi}_i - \psi_i)^2. \tag{C.39}$$

The majority of the proof (steps 2 - 4) is to establish that $\widehat{\psi}$ constructed by us estimates conditional expectation vector ψ sufficiently precisely in the following sense:

$$\|\breve{\psi} - \psi\|^2 = o_{\mathrm{P}}(1 + (S^{\mathrm{OPT}})^2). \tag{C.40}$$

The last two steps prove optimality of our portfolio strategy based on the above result. We end this step by noting that Corollary 2 states

$$\|\psi\| = S(\mathcal{G}) = S^{\mathrm{OPT}} + o_{\mathrm{P}}(1). \tag{C.41}$$

Step 2. This step control the magnitude of $\sum_{i\in B}(\breve{\psi}_i - \psi_i)^2$ of (C.39). It does so by showing

$$\sum_{i\in B^c}(\psi_i - \check{s}_i)^2 = o_{\mathrm{P}}(1 + (S^{\mathrm{OPT}})^2) \quad \text{and} \quad \sum_{i\in B^c}(\breve{\psi}_i - \check{s}_i)^2 = o_{\mathrm{P}}(1 + (S^{\mathrm{OPT}})^2). \tag{C.42}$$

Since $\psi_i \coloneqq \psi(\check{s}_i)$, to bound $\sum_{i\in B^c}(\psi_i - \check{s}_i)^2$, we show that $|\psi(a) - a|$ is small. On the other hand, Tweedie's formula reads

$$\psi(a) - a = T^{-1}\frac{p'(a)}{p(a)}. \tag{C.43}$$

Moreover, we have, for all positive sequence b_N and all a,

$$\begin{aligned} |p'(a)| &\leq T\int |x-a|\phi_{1/T}(a-x)p_s(x)dx \\ &\leq b_N T\int_{|x-a|\leq b_N}\phi_{1/T}(a-x)p_s(x)dx + \sup_{x:|x-a|>b_N} T|x-a|\phi_{1/T}(a-x) \\ &\leq b_N T p(a) + \sup_{y:|y|>b_N} T^{3/2}|y|\exp(-Ty^2/2). \end{aligned} \tag{C.44}$$

The second inequality comes from the $p_s(x)$, as a density, integrates to one. Then, choosing b_N that satisfies $b_N \gtrsim T^{-1/2}(\log N)^d$ with $d > 1/2$ and $b_N = o(T^{-1/2}\widetilde{k}_N)$, which is always possible, we obtain, for all a,

$$|p'(a)| \leq c_N T^{1/2}\widetilde{k}_N p(a) + c_N T N^{-2}. \tag{C.45}$$

It hence holds that

$$\max_i \frac{|p'(\check{s}_i)|}{p(\check{s}_i)} \lesssim_{\mathrm{P}} \sup_a \frac{|p'(a)|}{p(a)}\mathbb{1}_{\{p(a)\geq T^{1/2}N^{-3/2}\}} \leq c_N T^{1/2}\widetilde{k}_N. \tag{C.46}$$

The first inequality comes from (D.13) of Lemma D3. The second directly follows from (C.45). Combining (C.46) and (C.43), we obtain

$$\mathrm{P}((\check{s}_i - \psi(\check{s}_i))^2 \leq c_N T^{-1}\widetilde{k}_N^2, \forall i \leq N) \geq 1 - c_N. \tag{C.47}$$

As a result,

$$\sum_{i\in B^c}(\check{s}_i - \psi(\check{s}_i))^2 \lesssim_{\mathrm{P}} c_N T^{-1}\widetilde{k}_N^2|B^c| \leq c_N\sum_{i\in B^c}\check{s}_i^2 \lesssim_{\mathrm{P}} c_N\sum_{i\in B^c}\psi(\check{s}_i)^2. \tag{C.48}$$

Here the first inequality is simply (C.47), the second holds since $\check{s}_i^2 \geq T^{-1}\widetilde{k}_N^2$ for all $i \in B^c$ by definition, and the last inequality is a direct implication of the first two. Given (C.48), we obtain the first part of (C.42) by noting $\sum_{i\in B^c}\psi(\check{s}_i)^2 \lesssim_{\mathrm{P}} (S^{\mathrm{OPT}})^2 + 1$ due to (C.41).

Now we establish the second part of (C.42). By construction we have

$$\breve{\psi}(a) - a = \frac{1 + k_N^2\,\widehat{p}'(a)}{T\,\widehat{p}(a)}, \quad \text{with} \quad \widehat{p}(a) = \frac{1}{Nk_N}\sum_i \phi_{1/T}\left(\frac{\widehat{s}_i - a}{k_N}\right). \tag{C.49}$$

Similar to (C.44), we have, for all positive sequence b_N and all a,

$$\begin{aligned}
|\widehat{p}'(a)| &\leq \frac{T}{Nk_N^2}\sum_i \frac{|\widehat{s}_i - a|}{k_N}\phi_{1/T}\left(\frac{\widehat{s}_i - a}{k_N}\right) \\
&\leq \frac{T}{Nk_N^2}\sum_{i:|\widehat{s}_i - a|/k_N \leq b_N} \frac{|\widehat{s}_i - a|}{k_N}\phi_{1/T}\left(\frac{\widehat{s}_i - a}{k_N}\right) + \frac{T}{k_N^2}\sup_{i:|\widehat{s}_i - a|/k_N > b_N} \frac{|\widehat{s}_i - a|}{k_N}\phi_{1/T}\left(\frac{\widehat{s}_i - a}{k_N}\right) \\
&\leq \frac{Tb_N}{k_N}\widehat{p}(a) + \frac{T}{k_N^2}\sup_{y:|y|>b_N} |y|\exp(-Ty^2/2).
\end{aligned}$$

Choosing b_N that satisfies $b_N \gtrsim T^{-1/2}(\log N)^d$ with $d > 1/2$ and $b_N = o(T^{-1/2}\widetilde{k}_N k_N)$, which is always possible, we obtain, for all a,

$$|\widehat{p}'(a)| \leq c_N T^{1/2}\widetilde{k}_N\widehat{p}(a) + c_N T^{1/2}N^{-2}. \tag{C.50}$$

Therefore, it holds that

$$\max_i \frac{|\widehat{p}'(\widehat{s}_i)|}{\widehat{p}(\widehat{s}_i)} \leq c_N T^{1/2}\widetilde{k}_N, \tag{C.51}$$

which comes from (C.50) and that $\widehat{p}(\widehat{s}_i) \geq \frac{1}{Nk_N}\phi_{1/T}(0) \gtrsim \frac{\sqrt{T}}{Nk_N}$ for all i. As a result, we obtain the second part of (C.42):

$$\sum_{i\in B^c}(\breve{\psi}_i - \check{s}_i)^2 \leq c_N T^{-1}|B^c|\widetilde{k}_N^2 + |B^c|\max_{i\leq N}|\widehat{s}_i - \check{s}_i|^2 \leq c_N T^{-1}|B^c|\widetilde{k}_N^2 \lesssim_{\mathrm{P}} c_N(S^{\mathrm{OPT}})^2 + c_N.$$

Here the first inequality is simply substituting (C.51) into (C.49), the second inequality comes from $\max_{i\leq N}|\widehat{s}_i - \check{s}_i| \leq c_N T^{-1/2}\widetilde{k}_N$ by Lemma D2, the last inequality holds by (C.41) and (the last two inequalities of) (C.48).

Step 3. To analyze $\sum_{i\in B}(\widehat{\psi}_i - \psi_i)^2$ of (C.39), we introduce an auxiliary function:

$$\bar{\psi}(a) = \frac{\int x\phi_{v^2/T}(a - x)p_s(x)dx}{\int \phi_{v^2/T}(a - x)p_s(x)dx}, \quad \text{with} \quad v := \sqrt{1 + k_N^2}. \tag{C.52}$$

$\bar{\psi}(a)$ is essentially the expectation of s_i, conditional on $\check{s}'_i = a$, where $\check{s}'_i \approx \mathcal{N}(s_i, v^2)$, i.e., $\check{s}'_i$ has slightly more noisy than $\check{s}_i$. The goal is to establish

$$\sum_{i\in B}(\psi_i - \bar{\psi}(\check{s}_i))^2 = o_{\mathrm{P}}(1 + (S^{\mathrm{OPT}})^2) \quad \text{and} \quad \sum_{i\in B}(\breve{\psi}_i - \bar{\psi}(\check{s}_i))^2 = o_{\mathrm{P}}(1 + (S^{\mathrm{OPT}})^2). \tag{C.53}$$

Then the triangle inequality would give us the desired bound on $\sum_{i\in B}(\breve{\psi}_i - \psi_i)^2$. The current step proves the first part, whereas the next step will be devoted to show the second part.

We use $\bar{p}(a)$ and $\bar{\pi}(a)$ to denote the denominator and numerator of $\bar{\psi}(a)$ as in (C.52), and use $\pi(a)$ to denote $\int x\phi_{1/T}(a-x)p_s(x)dx$ so that $\psi(a) = \pi(a)/p(a)$. The goal is to show that $\bar{p}(a)$ and $\bar{\pi}(a)$ are, respectively, close to $p(a)$ and $\pi(a)$. We first note that $\phi_{v^2}(y)$ and $\phi(y)$ are close in that, for all y,

$$\begin{aligned} |\phi_{v^2}(y) - \phi(y)| &\leq \sup_{y:|y|\leq k_N^{-1}} |\phi_{v^2}(y) - \phi(y)| + \sup_{y:|y|>k_N^{-1}} |\phi_{v^2}(y) + \phi(y)| \\ &\leq c_N k_N^{-1}\phi(y) \sup_{y:|y|\leq k_N^{-1}} |y/v - y| + c_N N^{-2} \leq c_N\phi(y) + c_N N^{-2}. \end{aligned} \tag{C.54}$$

Here we use (D.17) of Lemma D4 (choose $j = 0$) and that $|v^{-1} - 1| \asymp k_N^2$. We note $\phi_{1/T}(y) = \sqrt{T}\phi\left(\sqrt{T}y\right)$ (and symmetrically for $\phi_{v^2/T}(y)$). Hence, using (C.54), we directly obtain that, for all a,

$$\begin{aligned} |\bar{p}(a) - p(a)| &\leq \int |\phi_{v^2/T}(a-x) - \phi_{1/T}(a-x)|p_s(x)dx \\ &\leq c_N \int \phi_{1/T}(a-x)p_s(x)dx + c_N\sqrt{T}N^{-2} = c_N p(a) + c_N\sqrt{T}N^{-2}. \end{aligned} \tag{C.55}$$

Now we bound the difference $|\pi(a) - \bar{\pi}(a)|$. Because $p_s(x)$ is a even function, we note that, for all $a \geq 0$,

$$\pi(a) = \int_0^\infty x\bar{\phi}(|a|, x)p_s(x)dx, \quad \text{and} \quad \bar{\pi}(a) = \int_0^\infty x\bar{\phi}(a/v, x/v)p_s(x)dx, \tag{C.56}$$

where

$$\bar{\phi}(a, x) \coloneqq \phi_{1/T}(a-x) - \phi_{1/T}(a+x) = \phi_{1/T}(a-x)(1 - e^{-2Txa}).$$

Since $|(1 - e^{-y}) - (1 - e^{-y/v^2})| \leq c_N(1 - e^y)$ for all $y \geq 0$, it follows from (C.54) and direct calculations that, for all $a \geq 0$ and $x \geq 0$,

$$|\bar{\phi}(a/v, x/v) - \bar{\phi}(a, x)| \leq c_N\bar{\phi}(a, x) + c_N\sqrt{T}N^{-2}. \tag{C.57}$$

Substituting (C.57) into (C.56), we obtain, for all $a \geq 0$,

$$|\bar{\pi}(a) - \pi(a)| \leq c_N|\pi(a)| + c_N\sqrt{T}N^{-2}\int_0^\infty x p_s(s)dx \leq |\pi(a)| + c_N\sqrt{T}N^{-2}. \tag{C.58}$$

Here the last inequality holds by $\mathrm{E}(|s|) \leq \sqrt{\mathrm{E}(s^2)} \leq c_N$ due to condition (a) of Assumption 2. Because $\pi(a)$ and $\bar{\pi}(a)$ are both odd functions in a due to that $p_s(x)$ is a even function of x, (C.58) apparently holds for all a.

To establish from (C.55) and (C.58) that $\bar{\psi}(a)$ and $\psi(a)$ are close, we set $A \coloneqq \{a : |a| \leq \widetilde{k}_N T^{-1/2}, p(a) \geq \sqrt{T}N^{-3/2}\}$. Then we obtain that, for all $a \in A$,

$$\begin{aligned} |\bar{\psi}(a) - \psi(a)| &= \left|\frac{\bar{\pi}(a)}{\bar{p}(a)} - \frac{\pi(a)}{\bar{p}(a)}\right| + \left|\frac{\pi(a)}{\bar{p}(a)} - \frac{\pi(a)}{p(a)}\right| \\ &\leq (1 + c_N)\frac{|\bar{\pi}(a) - \pi(a)|}{p(a)} + c_N\frac{|\pi(a)|}{p(a)} \leq c_N\frac{N^{-2}}{p(a)} + c_N\psi(a). \end{aligned} \tag{C.59}$$

Here the first equality is obvious, the first inequality comes from the lower bound of $p(a)$ (by the definition of A) and (C.55), the second inequality is a result of (C.58). From (C.59), it follows that, for all a satisfying $a \in A$,

$$|\bar{\psi}(a) - \psi(a)|^2 \leq c_N\frac{N^{-2}}{p(a)} + c_N\psi(a)^2. \tag{C.60}$$

where we use Cauchy-Schwarz inequality and the lower bound of $p(a)$. Therefore, we arrive at

$$N\int_A |\bar{\psi}(a) - \psi(a)|^2 p(a)da \leq c_N + c_N N\int_{-\infty}^\infty \psi(a)^2 p(a)da \leq c_N + c_N(S^{\mathrm{OPT}})^2, \tag{C.61}$$

which comes from (C.60) and that $\int_A da \leq 2\widetilde{k}_N$. Therefore, using Chebyshev's inequality and comparing the definitions of sets A and B, we obtain

$$\begin{aligned} \sum_{i\in B}(\psi_i - \bar{\psi}(\check{s}_i))^2 \mathbb{1}_{\{p(\check{s}_i)\geq\sqrt{T}N^{-3/2}\}} &\lesssim_{\mathrm{P}} \mathrm{E}\sum_{i\in B}(\psi_i - \bar{\psi}(\check{s}_i))^2 \mathbb{1}_{\{p(\check{s}_i)\geq\sqrt{T}N^{-3/2}\}} \\ &= N\int_A |\bar{\psi}(a) - \psi(a)|^2 p(a)da \leq c_N + c_N(S^{\mathrm{OPT}})^2, \end{aligned}$$

where the last inequality holds by (C.61). Given (D.13) of Lemma D3, we obtain the first part of (C.53).

Step 4. This step proves the second part of (C.53), i.e., we bound $\sum_{i\in B}(\breve{\psi}_i - \bar{\psi}(\check{s}_i))^2$. We introduce $\widetilde{p}(a)$ and $\widetilde{\psi}(a)$ that mimick $\widehat{p}(a)$ and $\widehat{\psi}(a)$ by replacing the data input $\check{s}_i$ with $\widehat{s}_i$:

$$\widetilde{p}(a) = \frac{1}{Nk_N}\sum_i \phi_{1/T}\left(\frac{\check{s}_i - a}{k_N}\right), \quad \text{and} \quad \widetilde{\psi}(a) = a + \frac{v^2}{T}\frac{\widetilde{p}'(a)}{\widetilde{p}(a)}. \tag{C.62}$$

Then we can decompose the quantity of interest:

$$\sum_{i\in B}(\check{\psi}_i - \bar{\psi}(\check{s}_i))^2 \le \sum_{i\in B}(\widetilde{\psi}(\check{s}_i) - \bar{\psi}(\check{s}_i))^2 + \sum_{i\in B}(\check{\psi}_i(\widehat{s}_i) - \widetilde{\psi}(\check{s}_i))^2. \tag{C.63}$$

We first show that $\sum_{i\in B}(\widetilde{\psi}(\check{s}_i) - \bar{\psi}(\check{s}_i))^2$ is small. Since we have $\widetilde{p}(\check{s}_i) \ge \frac{1}{Nk_N}\phi_{1/T}(0) \gtrsim \frac{\sqrt{T}}{Nk_N}$ for all i, symmetric to the derivation of (C.51), we have

$$\max_i \frac{\widetilde{p}'(\check{s}_i)}{\widetilde{p}(\check{s}_i)} \le c_N T^{1/2}\widetilde{k}_N. \tag{C.64}$$

On the other hand, symmetric to the derivation of (C.46), we obtain

$$\max_i \frac{\bar{p}'(\check{s}_i)}{\bar{p}(\check{s}_i)} \lesssim_{\mathrm{P}} \max_i \frac{\bar{p}'(a)}{\bar{p}(a)}\mathbb{1}_{\{p(a)\ge N^{-3/2}\}} \lesssim c_N T^{1/2}\widetilde{k}_N. \tag{C.65}$$

where for the second inequality we note $\bar{p}(a) \gtrsim p(a)$ for all a due to $1 \le v \lesssim 1$. Substituting (C.64) and (C.65) into the definitions of $\widetilde{\psi}(a)$ and $\bar{\psi}(a)$ ((C.62) and (C.52)), we obtain

$$\max_i |\widetilde{\psi}(\check{s}_i) - \bar{\psi}(\check{s}_i)| \lesssim_{\mathrm{P}} c_N\widetilde{k}_N T^{-1/2}. \tag{C.66}$$

According to Lemma 3 of Brown and Greenshtein (2009), with the additional condition that $\max_{i\le N}\sqrt{T}|s_i| = o(N^{d'})$ for every $d' > 0$, we have (in our notation) that, for every $d > 0$,

$$\mathrm{E}\left(\sum_i T(\widetilde{\psi}(\check{s}_i) - \bar{\psi}(\check{s}_i))^2\right) \lesssim N^d.$$

A scrutiny of their proof of the lemma reveals that this additional condition is only indispensable (a) to derive three equalities: (48), (59), and (62) (the way it is used is similar across the three), and (b) to guarantee that $\max_{i\le N}\sqrt{T}\bar{\psi}(\check{s}_i) = o(N^d)$ for every $d > 0$. In the absence of this additional condition, a weaker result holds: for every $d > d' > 0$,

$$\mathrm{E}\left(\sum_i \min\{T(\widetilde{\psi}(\check{s}_i) - \bar{\psi}(\check{s}_i))^2, N^{d'}\}\mathbb{1}_{\{\sqrt{T}|\check{s}_i|\le N^{d'}, p(\check{s}_i)\ge\sqrt{T}N^{d'-1}\}}\right) \lesssim N^d. \tag{C.67}$$

(C.67) turns out sufficient for establishing a desired bound on $\sum_{i\in B}(\widetilde{\psi}(\check{s}_i) - \bar{\psi}(\check{s}_i))^2$, which we demonstrate now. Then we have, for every $d > d' > 0$,

$$\sum_{i\in B} T(\widetilde{\psi}(\check{s}_i) - \bar{\psi}(\check{s}_i))^2 \lesssim_{\mathrm{P}} \sum_{i\in B} \min\{T(\widetilde{\psi}(\check{s}_i) - \bar{\psi}(\check{s}_i))^2, N^{d'}\}$$
$$\lesssim_{\mathrm{P}} N^d + \sum_{i\in B} \min\{T(\widetilde{\psi}(\check{s}_i) - \bar{\psi}(\check{s}_i))^2, N^{d'}\} \mathbb{1}_{\{p(\widetilde{z}_i)\geq \sqrt{T}N^{d'-1}\}} \lesssim N^d. \tag{C.68}$$

Here the first inequality comes from (C.66), the second inequality comes from that $\mathrm{E}\left(\sum_{i\in B} \mathbb{1}_{\{p(\check{s}_i)<\sqrt{T}N^{d'-1}\}}\right) \lesssim \widetilde{k}_N N^{d'}$ (by the definition of set B), and the last is simply (C.67).

Next, we show that $\sum_{i\in B}(\widetilde{\psi}(\check{s}_i) - \bar{\psi}(\check{s}_i))^2$ is small. Lemma D2 states that

$$\max_{i\in B} |\widehat{s}_i - \check{s}_i| \lesssim_{\mathrm{P}} \chi_N := N^{-1/2}(\epsilon_N + \mathrm{E}(s_j^2)^{1/2}), \quad \text{with} \quad \epsilon_N := k_N^5. \tag{C.69}$$

Since $\breve{\psi}(\widehat{s}_i)$ and $\widetilde{\psi}(\check{s}_i)$ depends on $\{\widehat{s}_j\}$ and $\{\check{s}_j\}$ in the exactly same way, we can obtain the desired result by exploiting that such dependence is sufficiently "continuous". Concretely, we write, uniformly over $i \in B$,

$$\begin{aligned} & |\widehat{p}(\widehat{s}_i) - \widetilde{p}(\check{s}_i)| \\ \leq\ & \frac{1}{Nk_N}\sum_j \left|\phi_{1/T}\left(\frac{\widehat{s}_j - \widehat{s}_i}{k_N}\right) - \phi_{1/T}\left(\frac{\check{s}_j - \check{s}_i}{k_N}\right)\right| \lesssim_{\mathrm{P}} \sqrt{T}\chi_N k_N^{-2}\widetilde{p}(\check{s}_i) + \sqrt{T}N^{-2}k_N^{-1}, \end{aligned} \tag{C.70}$$

$$\begin{aligned} & |\widehat{p}'(\widehat{s}_i) - \widetilde{p}'(\check{s}_i)| \\ \leq\ & \frac{T}{Nk_N^3}\sum_j \left|(\widehat{s}_j - \widehat{s}_i)\phi_{1/T}\left(\frac{\widehat{s}_j - \widehat{s}_i}{k_N}\right) - (\check{s}_j - \check{s}_i)\phi_{1/T}\left(\frac{\check{s}_j - \check{s}_i}{k_N}\right)\right| \\ \lesssim_{\mathrm{P}}\ & T\chi_N k_N^{-4}\widetilde{p}(\widetilde{z}_i) + TN^{-2}k_N^{-1}. \end{aligned} \tag{C.71}$$

Here the first inequalities for both results hold by definition (note $\phi_{1/T}(a) = \sqrt{T}\phi(\sqrt{T}a)$ and $\phi'(a) = -a\phi(a)$). The second inequalities for both results comes from substituting (C.69) into (D.17) of Lemma D4. Since $\widetilde{p}(\check{s}_i) \geq \frac{1}{Nk_N}\phi_{1/T}(0) \gtrsim \frac{\sqrt{T}}{Nk_N}$ by definition, we obtain from (C.70) and (C.71) that

$$\max_{i\in B} \frac{|\widehat{p}(\widehat{s}_i) - \widetilde{p}(\check{s}_i)|}{\widetilde{p}(\check{s}_i)} \lesssim_{\mathrm{P}} \sqrt{T}\chi_N k_N^{-2} + N^{-1} \lesssim \sqrt{T}\chi_N k_N^{-2}, \tag{C.72}$$

$$\max_{i\in B} \frac{|\widehat{p}'(\widehat{s}_i) - \widetilde{p}'(\check{s}_i)|}{\widetilde{p}(\check{s}_i)} \lesssim_{\mathrm{P}} T\chi_N k_N^{-4} + \sqrt{T}N^{-1} \lesssim T\chi_N k_N^{-4}. \tag{C.73}$$

Then we have

$$\begin{aligned} & \max_{i\in B}\left|\frac{\widehat{p}'(\widehat{s}_i)}{\widehat{p}(\widehat{s}_i)} - \frac{\widetilde{p}'(\check{s}_i)}{\widetilde{p}(\check{s}_i)}\right| \\ \leq\ & \max_{i\in B}\frac{\widetilde{p}(\check{s}_i)}{\widehat{p}(\widehat{s}_i)}\frac{|\widehat{p}'(\widehat{s}_i) - \widetilde{p}'(\check{s}_i)|}{\widetilde{p}(\check{s}_i)} + \max_{i\in B}\frac{\widetilde{p}(\check{s}_i)}{\widehat{p}(\widehat{s}_i)}\frac{\widetilde{p}'(\check{s}_i)}{\widetilde{p}(\check{s}_i)}\frac{|\widehat{p}(\widehat{s}_i) - \widetilde{p}(\check{s}_i)|}{\widetilde{p}(\check{s}_i)} \lesssim_{\mathrm{P}} T\chi_N k_N^{-4}. \end{aligned} \tag{C.74}$$

The first inequality is direct algebra. Substituting (C.64), (C.72), and (C.73) into the right-hand-side of the first inequality, we obtain the second inequality. Combining (C.69) and (C.74) with the definitions of $\widehat{\psi}$ and $\widetilde{\psi}$ ((C.49) and (C.62)), we obtain

$$\sum_{i\in B}(\breve{\psi}(\widehat{s}_i)-\widetilde{\psi}(\check{s}_i))^2 \leq N\max_{i\in B}|\widehat{s}_i-\check{s}_i|^2+\frac{N}{T^2}\max_{i\in B}\left|\frac{\widetilde{p}'(\widehat{s}_i)}{\widehat{p}(\widehat{s}_i)}-\frac{\widetilde{p}'(\check{s}_i)}{\widetilde{p}(\check{s}_i)}\right|^2 \lesssim_{\mathrm{P}} k_N^{-8}(\epsilon_N^2+\mathrm{E}(s_j^2)). \quad \text{(C.75)}$$

The goal is to show $\sum_{i\in B}(\breve{\psi}(\widehat{s}_i)-\widetilde{\psi}(\check{s}_i))^2 = o_{\mathrm{P}}(1+(S^{\mathrm{OPT}})^2)$, which is apparently true from (C.75) if $\mathrm{E}(s_j^2)\leq\epsilon_N^2$. For the case $\mathrm{E}(s_j^2)>\epsilon_N^2$, we observe

$$\mathrm{E}(s_j^2)=\mathrm{E}(s_j^2\mathbb{1}_{\{\epsilon_N/2<|s_i|\leq 1\}})+\mathrm{E}(s_j^2\mathbb{1}_{\{|s_i|\leq\epsilon_N/2\}})+\mathrm{E}(s_j^2\mathbb{1}_{\{|s_i|>1\}})\leq \mathrm{P}(|s_i|>\epsilon_N/2)+\epsilon_N^2/4+c_N N^{-1},$$

where the last step comes from condition (a) of Assumption 2. We hence obtain $\mathrm{P}(|s_i|>\epsilon_N/2)\gtrsim\epsilon_N^2$, which futher indicates $\sum_i \mathbb{1}_{\{|s_i|\geq\epsilon_N/2\}}\gtrsim_{\mathrm{P}} N\epsilon_N^2$ (the sum follows binomial distribution with its standard deviation dominated by its mean). As a result, we write

$$N\epsilon_N^4 \lesssim_{\mathrm{P}} \sum_i \epsilon_N^2\mathbb{1}_{\{|s_i|\geq\epsilon_N/2\}} \lesssim_{\mathrm{P}} \sum_i \check{s}_i^2\mathbb{1}_{\{|\check{s}_i|\geq\epsilon_N/4\}} \lesssim \sum_{i\in B^c}\check{s}_i^2 \lesssim_{\mathrm{P}} 1+(S^{\mathrm{OPT}})^2. \quad \text{(C.76)}$$

Here the second inequality comes from $\check{s}_i-s_i=\bar{\varepsilon}_i$ and $\max_i|\bar{\varepsilon}_i|\lesssim\sqrt{(\log N)/T}$ by the uniform bound on i.i.d normal variables. The thrid inequality holds by the definition of B, and the last inequality can be established from holds by (C.41) and (the last two inequalities of) (C.48). Since $\mathrm{E}(s_j^2)\leq 1+\mathrm{E}(s_j^2\mathbb{1}_{\{|s_i|>1\}})\lesssim 1$ by condition (a) of Assumption 2, it follows from (C.76) that $k_N^{-8}\mathrm{E}(s_j^2)=o_{\mathrm{P}}(1+(S^{\mathrm{OPT}})^2)$. Given (C.75), we prove $\sum_{i\in B}(\breve{\psi}(\widehat{s}_i)-\widetilde{\psi}(\check{s}_i))^2 = o_{\mathrm{P}}(1+(S^{\mathrm{OPT}})^2)$. Substituting this result and (C.68) into (C.63), we obtain $\sum_{i\in B}(\breve{\psi}(\widehat{s}_i)-\widetilde{\psi}(\check{s}_i))^2 = o_{\mathrm{P}}(1+(S^{\mathrm{OPT}})^2)$, i.e., the second part of (C.53). Substituting (C.42) and (C.53) into (C.39), we finally establish (C.40).

Step 5. The goal of this step is to establish

$$\|\widehat{\psi}-\psi\| = o_{\mathrm{P}}(1+S^{\mathrm{OPT}}). \quad \text{(C.77)}$$

We start by defining

$$l_j = \arg\min_{i\leq N}|\check{s}_i-\widehat{s}_j|. \quad \text{(C.78)}$$

Suppose $\check{s}_{l_i}>\check{s}_{l_j}$ and $\widehat{s}_i\leq\widehat{s}_j$ for some (i,j). According to (C.78), we have $|\check{s}_{l_i}-\widehat{s}_i|\leq|\check{s}_{l_j}-\widehat{s}_i|$ and $|\check{s}_{l_j}-\widehat{s}_j|\leq|\check{s}_{l_i}-\widehat{s}_j|$. The former would lead to $\check{s}_{l_i}+\check{s}_{l_j}<2\widehat{s}_i$, and the latter would lead to $\check{s}_{l_i}+\check{s}_{l_j}>2\widehat{s}_j$. They together contradict $\widehat{s}_i\leq\widehat{s}_j$. Hence we have $\check{s}_{l_i}\leq\check{s}_{l_j}$ if $\widehat{s}_i\leq\widehat{s}_j$ for all (i,j). Since $\psi(a)$ is increasing in a (see, e.g., Efron (2011)), letting $\psi_i^*=\psi(\check{s}_{l_i})$, we have

$\psi_i^* \leq \psi_j^*$ if $\widehat{s}_i \leq \widehat{s}_j$ for all (i,j). As a result, we have

$$\|\widehat{\psi} - \breve{\psi}\| \leq \|\psi^* - \breve{\psi}\| \leq \|\psi^* - \psi\| + o_{\mathrm{P}}(1 + S^{\mathrm{OPT}}). \tag{C.79}$$

The first inequality comes from the definition of $\widehat{\psi}$ and that $\psi_i^* \leq \psi_j^*$ if $\widehat{s}_i \leq \widehat{s}_j$ for all (i,j) established above. The second inequality comes from (C.40) proved in steps 2 - 4.

Given (C.40) and (C.79), to obtain (C.77) we only need to bound $\|\psi^* - \psi\|$. For this purpose, we note that (C.78) leads to $|\check{s}_{l_i} - \widehat{s}_i| \leq |\check{s}_i - \widehat{s}_i|$ for all i. Applying Lemma D2, we obtain

$$\max_{1\leq i\leq N} |\check{s}_{l_i} - \check{s}_i| \lesssim_{\mathrm{P}} c_N \widetilde{k}_N T^{-1/2}, \quad \max_{i\in B} |\check{s}_{l_i} - \check{s}_i| \lesssim_{\mathrm{P}} \chi_N, \tag{C.80}$$

where $\chi_N = N^{-1/2}(\epsilon_N + \mathrm{E}(s_j^2)^{1/2})$ with $\epsilon_N \coloneqq k_N^5$. Similar to (C.48), we have

$$\sum_{i\in B^c}((\check{s}_{l_i} - \psi(\check{s}_{l_i}))^2 + (\check{s}_{l_i} - \check{s}_i)^2) \lesssim_{\mathrm{P}} c_N T^{-1}\widetilde{k}_N^2 |B^c| \leq c_N \sum_{i\in B^c} \psi(\check{s}_i)^2 \lesssim_{\mathrm{P}} o_{\mathrm{P}}(1 + (S^{\mathrm{OPT}})^2). \tag{C.81}$$

Here the first inequality comes from (C.47) and the first part of (C.80). The second inequality is already shown in (C.48). The last inequality is established right below (C.48). Combining (C.81) and the first part of (C.42), and applying the triangle inequality, we obtain

$$\sum_{i\in B^c}(\psi(\check{s}_{l_i}) - \psi(\check{s}_i))^2 = o_{\mathrm{P}}(1 + (S^{\mathrm{OPT}})^2). \tag{C.82}$$

Next, it holds that, uniformly over $i \in B$,

$$\begin{aligned}
|p(\check{s}_{l_i}) - p(\check{s}_i)| &\leq \int |\phi_{1/T}(\check{s}_{l_i} - x) - \phi_{1/T}(\check{s}_i - x)| p_s(x) ds \\
&\lesssim (T \log N)^{1/2} |\check{s}_{l_i} - \check{s}_i| p(\check{s}_i) + c_N \sqrt{T} N^{-2} \\
&\lesssim_{\mathrm{P}} (T \log N)^{1/2} \chi_N p(\check{s}_i) + c_N \sqrt{T} N^{-2}. \qquad \text{(C.83)} \\
|p(\check{s}_{l_i})' - p(\check{s}_i)'| &\leq T \int |(\check{s}_{l_i} - x)\phi_{1/T}(\check{s}_{l_i} - x) - (\check{s}_i - x)\phi_{1/T}(\check{s}_i - x)| p_s(x) ds \\
&\lesssim (T \log N) |\check{s}_{l_i} - \check{s}_i| p(\check{s}_i) + c_N T N^{-2} \\
&\lesssim_{\mathrm{P}} (T \log N) \chi_N p(\check{s}_i) + c_N T N^{-2}. \qquad \text{(C.84)}
\end{aligned}$$

For both (C.83) and (C.84), the first inequalities holds by definition of $p(a)$, the second inequalities come from (D.17) of Lemma D4 (choose $j = 0$ and $j = 1$ respectively), and the last inequalities come from the second part of (C.80). Then we have, uniformly over i for which $\check{s}_i \in A$ (A is defined above (C.59)),

$$
\begin{aligned}
|\psi(\check{s}_{l_i}) - \psi(\check{s}_i)| &\leq |\check{s}_{l_i} - \check{s}_i| + \frac{1}{T}\frac{|p(\check{s}_{l_i})' - p(\check{s}_i)'|}{p(\check{s}_{l_i})} + \frac{|p(\check{s}_i)'|}{T}\left|\frac{1}{p(\check{s}_{l_i})} - \frac{1}{p(\check{s}_i)}\right| \\
&\lesssim_{\mathrm{P}} \chi_N + \chi_N + \frac{\chi_N}{\sqrt{T}}\frac{p(\check{s}_i)'}{p(\check{s}_i)} \lesssim \chi_N\left(\widetilde{k}_N + \sqrt{T}|\psi(\check{s}_i)|\right). \qquad \text{(C.85)}
\end{aligned}
$$

The first inequality is obvious. The second inequality comes from the second part of (C.80), (C.83), (C.84), that $p(\check{s}_i) \geq \sqrt{T}N^{-3/2}$ by the definition of A, and that $1/\sqrt{NT} \lesssim \chi_N$ as $T \gtrsim N^d$ for fixed $d > 1/2$ by assumption. The last inequality comes from (C.43) and $|\check{s}_i| \leq \widetilde{k}_N$ as $\check{s}_i \in A$. Therefore, we have

$$
\begin{aligned}
&\sum_{i\in B}(\psi(\check{s}_{l_i}) - \psi(\check{s}_i))^2 \mathbb{1}_{\{p(\check{s}_i)\geq\sqrt{T}N^{-3/2}\}} \\
= &\sum_{i:\check{s}_i\in A}(\psi(\check{s}_{l_i}) - \psi(\check{s}_i))^2 \lesssim_{\mathrm{P}} N\chi_N^2\widetilde{k}_N^2 + T\chi_N^2\sum_{i\in B}\psi(\widetilde{z}_i)^2 \\
\lesssim_{\mathrm{P}} &(\epsilon_N^2 + \mathrm{E}(s_j^2))\widetilde{k}_N^2 + o_{\mathrm{P}}(1 + (S^{\mathrm{OPT}})^2) = o_{\mathrm{P}}(1 + (S^{\mathrm{OPT}})^2). \qquad \text{(C.86)}
\end{aligned}
$$

The equality is obvious and the first inequality comes from (C.85). The second inequality comes from (C.41) and that $T\chi_N^2 = o(1)$ due to $\mathrm{E}(s_j^2) = o(1)$ and $T \lesssim N^{d'}$ for fixed $d' < 1$. The last inequality is established in the analysis after (C.75). Combining (C.86), Lemma D3, and (C.82), we establish $\|\psi^* - \psi\| = o_{\mathrm{P}}(1 + S^{\mathrm{OPT}})$. Substituting this into (C.79) and using (C.40) lead to (C.77).

Step 6. This step combines (C.77) with (C.41) to prove the theorem, i.e., that the Sharpe ratio of the strategy $\widehat{w}^{\mathrm{OPT}}$ we construct achieves $(S^{\mathrm{OPT}})^2$ asymptotically. Since Sharpe ratio is invariant to scaling of weights, we can treat $\widehat{w}^{\mathrm{OPT}}$ as $\widehat{w}^{\mathrm{OPT}} = \mathbb{M}_\beta\breve{w}$, where $\breve{w} := \widehat{\Sigma}_u^{-1/2}\widehat{\psi}$.

Using condition (a) of Assumption 1 and (D.1) of Lemma D1, we have $\max_{i\leq N}|\widehat{\sigma}_i/\sigma_i - 1| \lesssim_{\mathrm{P}} c_N$. As a result, we obtain $\|\Sigma^{1/2}\breve{w} - \widehat{\psi}\| \lesssim_{\mathrm{P}} c_N\|\widehat{\psi}\|$. Then, it follows from (C.77) and (C.41) that

$$
\|\Sigma_u^{1/2}\breve{w} - \psi\| \leq \|\Sigma^{1/2}\breve{w} - \widehat{\psi}\| + \|\widehat{\psi} - \psi\| \lesssim_{\mathrm{P}} c_N\|\psi\| + \|\widehat{\psi} - \psi\| = o_{\mathrm{P}}(1 + S^{\mathrm{OPT}}). \qquad \text{(C.87)}
$$

Hence we have

$$
\begin{aligned}
|(\breve{w}^\intercal\Sigma_u^{1/2} - \psi^\intercal)\psi| &\leq \|\Sigma_u^{1/2}\breve{w} - \psi\|\|\psi\| = o_{\mathrm{P}}(1 + (S^{\mathrm{OPT}})^2), \qquad \text{(C.88)} \\
|\breve{w}^\intercal\Sigma_u\breve{w} - \psi^\intercal\psi| &\leq \|\Sigma_u^{1/2}\breve{w} - \psi\|^2 + 2\|\Sigma_u^{1/2}\breve{w} - \psi\|\|\psi\| = o_{\mathrm{P}}(1 + (S^{\mathrm{OPT}})^2). \qquad \text{(C.89)}
\end{aligned}
$$

Here for both (C.88) and (C.89), the first inequalities come from Cauchy-Schwarz, whereas the last equalities come from (C.87) and (C.41). Further, substituting (C.41) into (C.88) and

(C.89), we obtain

$$\breve{w}^\intercal \Sigma_u^{1/2}\psi = (S^{\mathrm{OPT}})^2 + o_{\mathrm{P}}(1+(S^{\mathrm{OPT}})^2), \quad \breve{w}^\intercal \Sigma_u \breve{w} = (S^{\mathrm{OPT}})^2 + o_{\mathrm{P}}(1+(S^{\mathrm{OPT}})^2). \tag{C.90}$$

Next, we note that

$$\begin{aligned} \|\mathbb{P}_\beta \breve{w}\| &\leq \|\mathbb{P}_\beta \Sigma_u^{-1/2}\psi\| + \|\mathbb{P}_\beta(\breve{w} - \Sigma_u^{-1/2}\psi)\| \\ &= \|\mathbb{P}_\beta \Sigma_u^{-1}\widetilde{\alpha}\| + (\breve{w} - \Sigma_u^{-1/2}\psi)^\intercal \mathbb{P}_\beta(\breve{w} - \Sigma_u^{-1/2}\psi) \\ &\leq \|\mathbb{P}_\beta \Sigma_u^{-1}\widetilde{\alpha}\| + \|\breve{w} - \Sigma_u^{-1/2}\psi\| = o_{\mathrm{P}}(1+S^{\mathrm{OPT}}). \end{aligned} \tag{C.91}$$

To obtain the last equality, we note that the first term can be bounded using (C.18) as $\breve{w}' := \frac{1}{\kappa}\mathbb{P}_\beta \Sigma_u^{-1}\widetilde{\alpha}$ and that the second term can be bounded using (C.87) and $1 \lesssim_{\mathrm{P}} \lambda_{\min}(\Sigma_u)$. As a result, we obtain

$$\breve{w}^\intercal \mathbb{M}_\beta \Sigma_u \mathbb{M}_\beta \breve{w} = \breve{w}^\intercal \Sigma_u \breve{w} + \breve{w}^\intercal \mathbb{P}_\beta \Sigma_u \mathbb{P}_\beta \breve{w} - 2\breve{w}^\intercal \Sigma_u \mathbb{P}_\beta \breve{w} = \breve{w}^\intercal \Sigma_u \breve{w} + o_{\mathrm{P}}(1+(S^{\mathrm{OPT}})^2). \tag{C.92}$$

For the last equality, we use (C.89), (C.91), and $\lambda_{\max}(\Sigma_u) \lesssim_{\mathrm{P}} 1$ by condition (a) of Assumption 1. Similarly, we write

$$\breve{w}^\intercal \mathbb{M}_\beta \widetilde{\alpha} = \breve{w}^\intercal \widetilde{\alpha} - \breve{w}^\intercal \mathbb{P}_\beta \widetilde{\alpha} = \breve{w}^\intercal \widetilde{\alpha} + o_{\mathrm{P}}(1+(S^{\mathrm{OPT}})^2), \tag{C.93}$$

where for the last equality, we use (C.91) and $\|\widetilde{\alpha}\|^2 \lesssim_{\mathrm{P}} \widetilde{\alpha}\Sigma_u^{-1}\widetilde{\alpha} = (S^{\mathrm{OPT}})^2$.

We now conclude that, when S^{OPT} does not vanish,

$$\begin{aligned} \widehat{S}^{\mathrm{OPT}} &= \frac{(\widehat{w}^{\mathrm{OPT}})^\intercal \alpha}{\sqrt{(\widehat{w}^{\mathrm{OPT}})^\intercal \Sigma \widehat{w}^{\mathrm{OPT}}}} = \frac{\breve{w}^\intercal \mathbb{M}_\beta \alpha}{\sqrt{\breve{w}^\intercal \mathbb{M}_\beta \Sigma \mathbb{M}_\beta \breve{w}}} = \frac{\breve{w}^\intercal \mathbb{M}_\beta \widetilde{\alpha}}{\sqrt{\breve{w}^\intercal \mathbb{M}_\beta \Sigma_u \mathbb{M}_\beta \breve{w}}} + o_{\mathrm{P}}(1) \\ &= \frac{\breve{w}^\intercal \widetilde{\alpha}}{\sqrt{\breve{w}^\intercal \Sigma_u \breve{w}}} + o_{\mathrm{P}}(1+S^{\mathrm{OPT}}) = S^{\mathrm{OPT}} + o_{\mathrm{P}}(1+S^{\mathrm{OPT}}). \end{aligned} \tag{C.94}$$

The first two equalities hold by definition. The third one comes (C.9) and that $\mathbb{M}_\beta \breve{w}$ is $\mathcal{G}$-measurable and factor-neutral. The fourth one comes from (C.92), (C.93), and the second part of (C.90). The last equality comes from (C.90) ($\widetilde{\alpha} = \Sigma_u^{1/2}\psi$). Because $\widehat{w}^{\mathrm{OPT}}$ is $\mathcal{G}$-measurable and $\beta^\intercal \widehat{w}^{\mathrm{OPT}} = 0$, Theorem 2 and Corollary 2 apply. We hence have $\widehat{S}^{\mathrm{OPT}} \leq S(\mathcal{G}) + o_{\mathrm{P}}(1) = S^{\mathrm{OPT}} + o_{\mathrm{P}}(1)$. Because $-\widehat{S}^{\mathrm{OPT}}$ is the Sharpe ratio generated by $-\widehat{w}^{\mathrm{OPT}}$, we also have $-\widehat{S}^{\mathrm{OPT}} \leq S^{\mathrm{OPT}} + o_{\mathrm{P}}(1)$. As a result, when S^{OPT} does vanish, we have $\widehat{S}^{\mathrm{OPT}} = o_{\mathrm{P}}(1)$. Therefore, given (C.94) and using the subsequence argument (see, e.g., Andrews and Cheng (2012)), we have

$$\widehat{S}^{\mathrm{OPT}} = S^{\mathrm{OPT}} + o_{\mathrm{P}}(1+S^{\mathrm{OPT}}).$$

In other words, we have, for all $\mathrm{P} \in \mathbb{P}$,

$$\lim_{N,T\to\infty} \mathrm{P}(|\widehat{S}^{\mathrm{OPT}} - S^{\mathrm{OPT}}| \geq \epsilon S^{\mathrm{OPT}} + \epsilon) = 0. \tag{C.95}$$

Suppose the theorem does not hold, then there is a sequence of data-generating processes P_k with $\mathrm{P}_k \in \mathbb{P}$ for each $k \in \{1, 2, ...\}$ such that

$$\limsup_{N,T\to\infty} \lim_{k\to\infty} \mathrm{P}_k(|\widehat{S}^{\mathrm{OPT}} - S^{\mathrm{OPT}}| \geq \epsilon S^{\mathrm{OPT}} + \epsilon) > 0.$$

This contradicts (C.95), and the theorem is proved.

C.5 Proof of Proposition 1

By definition we have

$$(\widehat{S}^{\star})^2 = \alpha^{\intercal}\mathbb{M}_\beta\widehat{\Sigma}_u^{-1}\mathbb{M}_\beta\alpha + 2\alpha^{\intercal}\mathbb{M}_\beta\widehat{\Sigma}_u^{-1}\mathbb{M}_\beta\bar{u} + \bar{u}^{\intercal}\mathbb{M}_\beta\widehat{\Sigma}_u^{-1}\mathbb{M}_\beta\bar{u} - T^{-1}N.$$

We start with the analysis of $\alpha^{\intercal}\mathbb{M}_\beta\widehat{\Sigma}_u^{-1}\mathbb{M}_\beta\alpha$. From (D.1) of Lemma D1, it follows

$$\|\widehat{\Sigma}_u - \Sigma_u\|_{\mathrm{MAX}} \lesssim_{\mathrm{P}} \sqrt{T^{-1}\log N}. \tag{C.96}$$

As a result, noting $\mathrm{P}(0 \leq \mathbb{M}_\beta \leq \mathbb{I}_N) \to 1$ and $\mathrm{P}(\Sigma_u \approx \mathbb{I}_N) \to 1$ by the assumption $\|\beta\|_{\mathrm{MAX}} \lesssim_{\mathrm{P}} 1$ and $\lambda_{\min}(\beta^{\intercal}\beta) \gtrsim_{\mathrm{P}} N$, and recalling $(S^{\star})^2 = \alpha^{\intercal}\Sigma_u^{-1}\alpha$, we have

$$|\alpha^{\intercal}\mathbb{M}_\beta\widehat{\Sigma}_u^{-1}\mathbb{M}_\beta\alpha - \alpha^{\intercal}\mathbb{M}_\beta\Sigma_u^{-1}\mathbb{M}_\beta\alpha| \lesssim_{\mathrm{P}} \sqrt{T^{-1}\log N}(S^{\star})^2. \tag{C.97}$$

On the other hand, it holds that

$$\begin{aligned}
|\alpha^{\intercal}\mathbb{M}_\beta\Sigma_u^{-1}\mathbb{M}_\beta\alpha - \alpha^{\intercal}\Sigma_u^{-1}\alpha| &\leq \alpha^{\intercal}\mathbb{P}_\beta\Sigma_u^{-1}\mathbb{P}_\beta\alpha + 2\sqrt{(\alpha^{\intercal}\Sigma_u^{-1}\alpha)(\alpha^{\intercal}\mathbb{P}_\beta\Sigma_u^{-1}\mathbb{P}_\beta\alpha)} \\
&\lesssim_{\mathrm{P}} \alpha^{\intercal}\mathbb{P}_\beta\alpha + \sqrt{(\alpha^{\intercal}\Sigma_u^{-1}\alpha)(\alpha^{\intercal}\mathbb{P}_\beta\alpha)} \\
&\lesssim_{\mathrm{P}} N^{-1}\|\alpha^{\intercal}\beta\|^2 + \sqrt{N^{-1}(\alpha^{\intercal}\Sigma_u^{-1}\alpha)\|\alpha^{\intercal}\beta\|^2} \\
&\lesssim_{\mathrm{P}} \mathrm{E}(s_i^2) + S^{\star}\mathrm{E}(s_i^2)^{1/2}.
\end{aligned} \tag{C.98}$$

Here the first inequality comes from Cauchy-Schwarz inequality. The second comes from $\mathrm{P}(\Sigma_u \approx \mathbb{I}_N) \to 1$ and $\mathbb{P}_\beta^2 = \mathbb{P}_\beta$. We obtain the third line by using $\lambda_{\min}(\beta^{\intercal}\beta) \gtrsim N$. The last line holds because of Chebyshev's inequality and that $\mathrm{E}(\|\alpha^{\intercal}\beta\|^2|\beta, \Sigma_u) \lesssim N\|\beta\|_{\mathrm{MAX}}^2\mathrm{E}(\alpha_i^2|\Sigma_u) \leq N\|\beta\|_{\mathrm{MAX}}^2\lambda_{\max}(\Sigma_u)\mathrm{E}(s_i^2)$ by condition (a) of Assumption 1 and condition (b) of Assumption 2.

On the other hand, because of the assumption $\mathrm{E}(s_i^2 \mathbb{1}_{\{|s_i| \geq c_N\}}) \leq c_N N^{-1}$, we have

$$\mathrm{E}\left|\alpha^\intercal \Sigma_u^{-1} \alpha - \sum_i s_i^2 \mathbb{1}_{\{|s_i| < c_N\}}\right| = \mathrm{E}\left|\sum_i s_i^2 \mathbb{1}_{\{|s_i| \geq c_N\}}\right| = o(1),$$

which, by Markov's inequality, leads to

$$\alpha^\intercal \Sigma_u^{-1} \alpha = \sum_i s_i^2 \mathbb{1}_{\{|s_i| < c_N\}} + o_{\mathrm{P}}(1).$$

Moreover, it holds that

$$\mathrm{Var}\left|\sum_i s_i^2 \mathbb{1}_{\{|s_i| < c_N\}}\right| \leq \sum_i \mathrm{E}(s_i^4 \mathbb{1}_{\{|s_i| < c_N\}}) \leq c_N^2 \sum_i \mathrm{E}(s_i^2 \mathbb{1}_{\{|s_i| < c_N\}}).$$

Using Chebyshev's inequality, we obtain

$$(S^\star)^2 = \alpha^\intercal \Sigma_u^{-1} \alpha \geq \sum_i s_i^2 \mathbb{1}_{\{|s_i| < c_N\}} \gtrsim_{\mathrm{P}} \sum_i \mathrm{E}(s_i^2 \mathbb{1}_{\{|s_i| < c_N\}}) \geq N\mathrm{E}(s_i^2) + o(1).$$

Combining this result with (C.98) and (C.97), we have

$$\alpha^\intercal \mathbb{M}_\beta \widehat{\Sigma}_u^{-1} \mathbb{M}_\beta \alpha = (S^\star)^2 + o_{\mathrm{P}}\left(\sqrt{T^{-1} \log N}((S^\star)^2 + 1)\right). \tag{C.99}$$

Next, we study $\alpha^\intercal \mathbb{M}_\beta \widehat{\Sigma}_u^{-1} \mathbb{M}_\beta \bar{u}$. It holds that

$$\alpha^\intercal \mathbb{M}_\beta \widehat{\Sigma}_u^{-1} \mathbb{M}_\beta \bar{u} \lesssim \alpha^\intercal \mathbb{M}_\beta \widehat{\Sigma}_u^{-1} \mathbb{M}_\beta \widehat{\Sigma}_u^{-1} \mathbb{M}_\beta \alpha \sqrt{\bar{u}^\intercal \bar{u}} \lesssim \alpha^\intercal \alpha \sqrt{\bar{u}^\intercal \bar{u}} = O_{\mathrm{P}}(((S^\star)^2 + 1) T^{-1/2}). \tag{C.100}$$

The first inequality comes from Cauchy-Schwarz. The second inequality holds because $\mathrm{P}(\mathbb{M}_\beta \approx \mathbb{I}_N) \to 1$, $\mathbb{M}_\beta^2 = \mathbb{M}_\beta$, and $\mathrm{P}(\widehat{\Sigma}_u \approx \mathbb{I}_N) \to 1$ due to $\mathrm{P}(\Sigma_u \approx \mathbb{I}_N) \to 1$ and (C.96). The third inequality holds by $\mathrm{P}(\widehat{\Sigma}_u \approx \mathbb{I}_N) \to 1$ as well.

Now we analyze $\bar{u}^\intercal \mathbb{M}_\beta \widehat{\Sigma}_u^{-1} \mathbb{M}_\beta \bar{u} - T^{-1} N$. We write

$$\begin{aligned} N = \mathrm{tr}(\widehat{\Sigma}_u^{-1} \widehat{\Sigma}_u) &= \sum_{i \leq N} (\widehat{\Sigma}_u^{-1})_{i,i} \left(T^{-1} \sum_{s \in T} (\mathbb{M}_\beta u_s)_i^2 - (\mathbb{M}_\beta \bar{u})_i^2 \right) \\ &= T^{-1} \sum_{s \in T} u_s^\intercal \mathbb{M}_\beta \widehat{\Sigma}_u^{-1} \mathbb{M}_\beta u_s - \bar{u}^\intercal \mathbb{M}_\beta \widehat{\Sigma}_u^{-1} \mathbb{M}_\beta \bar{u} \\ &= T^{-1} \sum_{s \in T} u_s^\intercal \mathbb{M}_\beta \widehat{\Sigma}_u^{-1} \mathbb{M}_\beta u_s + O_{\mathrm{P}}(N/T). \end{aligned} \tag{C.101}$$

The last line comes from $\bar{u}^\intercal \mathbb{M}_\beta \widehat{\Sigma}_u^{-1} \mathbb{M}_\beta \bar{u} \lesssim_{\mathrm{P}} \bar{u}^\intercal \bar{u}$ because of $\mathbb{M}_\beta^2 = \mathbb{M}_\beta$ and $\mathrm{P}(\widehat{\Sigma}_u \approx \mathbb{I}_N) \to 1$.

Furthermore, I have

$$\begin{aligned}\bar{u}^\intercal \mathbb{M}_\beta \widehat{\Sigma}_u^{-1} \mathbb{M}_\beta \bar{u} - \bar{u}^\intercal \widehat{\Sigma}_u^{-1} \bar{u} &\leq 2|\bar{u}^\intercal \widehat{\Sigma}_u^{-1} \mathbb{P}_\beta \bar{u}| + \bar{u}^\intercal \mathbb{P}_\beta \widehat{\Sigma}_u^{-1} \mathbb{P}_\beta \bar{u} \\ &\lesssim_{\mathrm{P}} \sqrt{\bar{u}^\intercal \bar{u}} \sqrt{\bar{u}^\intercal \mathbb{P}_\beta \bar{u}} + \bar{u}^\intercal \mathbb{P}_\beta \bar{u} \lesssim_{\mathrm{P}} N^{1/2}/T. \end{aligned} \tag{C.102}$$

Here we obtain the second inequality using $\mathrm{P}(\widehat{\Sigma}_u \approx \mathbb{I}_N) \to 1$ and the last inequality using $\mathrm{P}(\mathbb{P}_\beta \lesssim \mathbb{I}_N) \to 1$. Similarly, it holds that

$$\begin{aligned} & T^{-2} \sum_{t \in T} (u_t^\intercal \mathbb{M}_\beta \widehat{\Sigma}_u^{-1} \mathbb{M}_\beta u_t - u_t^\intercal \widehat{\Sigma}_u^{-1} u_t) \\ = \; & T^{-2} \sum_{t \in T} \left(2\sqrt{u_t^\intercal u_t} \sqrt{u_t^\intercal \mathbb{P}_\beta \widehat{\Sigma}_u^{-1} \mathbb{P}_\beta u_t} + u_t^\intercal \mathbb{P}_\beta \widehat{\Sigma}_u^{-1} \mathbb{P}_\beta u_t \right) \\ \lesssim_{\mathrm{P}} \; & T^{-2} \sum_{t \in T} \left(\sqrt{u_t^\intercal u_t} \sqrt{u_t^\intercal \mathbb{P}_\beta u_t} + u_t^\intercal \mathbb{P}_\beta u_t \right) \lesssim_{\mathrm{P}} \frac{N^{1/2}}{T}. \end{aligned} \tag{C.103}$$

From (C.101), (C.102), and (C.103), it directly follows

$$\bar{u}^\intercal \mathbb{M}_\beta \widehat{\Sigma}_u^{-1} \mathbb{M}_\beta \bar{u} - T^{-1} N = \bar{u}^\intercal \widehat{\Sigma}_u^{-1} \bar{u} - T^{-2} \sum_{s \in T} u_s^\intercal \widehat{\Sigma}_u^{-1} u_s + O_{\mathrm{P}}(N^{1/2}/T + N/T^2). \tag{C.104}$$

On the other hand, we have

$$\widehat{\Sigma}_u^{-1} = -\Sigma_u^{-2}(\widehat{\Sigma}_u - 2\Sigma_u) + \Sigma_u^{-2} \widehat{\Sigma}_u^{-1} (\widehat{\Sigma}_u - \Sigma_u)^2.$$

It then follows from (C.96) and $\mathrm{P}(\Sigma_u \approx \mathbb{I}_N) \to 1$ that

$$\begin{aligned} \bar{u}^\intercal \widehat{\Sigma}_u^{-1} \bar{u} &= -\bar{u}^\intercal \Sigma_u^{-2} (\widehat{\Sigma}_u - 2\Sigma_u) \bar{u} + O_{\mathrm{P}}(T^{-1} (\log N) \bar{u}^\intercal \bar{u}) \\ &= -\bar{u}^\intercal \Sigma_u^{-2} (\widehat{\Sigma}_u - 2\Sigma_u) \bar{u} + O_{\mathrm{P}}(T^{-2} N \log N). \end{aligned} \tag{C.105}$$

Similarly, we have

$$T^{-2} \sum_{t \in T} u_t^\intercal \widehat{\Sigma}_u^{-1} u_t = -T^{-2} \sum_{t \in T} u_t^\intercal \Sigma_u^{-2} (\widehat{\Sigma}_u - 2\Sigma_u) u_t + O_{\mathrm{P}}(T^{-2} N \log N). \tag{C.106}$$

Substituting (C.105) and (C.106) into (C.104), we have

$$\begin{aligned} \bar{u}^\intercal \mathbb{M}_\beta \widehat{\Sigma}_u^{-1} \mathbb{M}_\beta \bar{u} - T^{-1} N &= -\bar{u}^\intercal \Sigma_u^{-2} (\widehat{\Sigma}_u - 2\Sigma_u) \bar{u} + T^{-2} \sum_{t \in T} u_t^\intercal \Sigma_u^{-2} (\widehat{\Sigma}_u - 2\Sigma_u) u_t \\ &\quad + O_{\mathrm{P}}(T^{-1} N^{1/2} + T^{-2} N \log N). \end{aligned} \tag{C.107}$$

Now we analyze $\widehat{\Sigma}_u$. We write

$$\begin{aligned}(\widehat{\Sigma}_u)_{i,i} &= (T^{-1}uu^\intercal)_{i,i} + (\mathbb{M}_\beta \bar{u}\bar{u}^\intercal \mathbb{M}_\beta)_{i,i} \\ &\quad + (\mathbb{P}_\beta T^{-1}uu^\intercal)_{i,i} + (T^{-1}uu^\intercal \mathbb{P}_\beta)_{i,i} + (\mathbb{P}_\beta T^{-1}uu^\intercal \mathbb{P}_\beta)_{i,i}. \end{aligned} \tag{C.108}$$

From the uniform bound on i.i.d. random variables and $\|\mathbb{P}_\beta\|_{\text{MAX}} \lesssim N^{-1}$ by the assumption $\|\beta\|_{\text{MAX}} \lesssim_{\text{P}} 1$ and $\lambda_{\min}(\beta^\intercal\beta) \gtrsim_{\text{P}} N$, it follows

$$\|\mathbb{M}_\beta \bar{u}\bar{u}^\intercal \mathbb{M}_\beta\|_{\text{MAX}} \lesssim_{\text{P}} \|\bar{u}\bar{u}^\intercal\|_{\text{MAX}} = \|\bar{u}\|_{\text{MAX}}^2 \lesssim T^{-1}\log N.$$

Using $\text{P}(\Sigma_u \approx \mathbb{I}_N) \to 1$, this gives

$$\sum_{i\le N} |\bar{u}_i^2 (\Sigma_u^{-2})_{i,i} (\mathbb{M}_\beta \bar{u}\bar{u}^\intercal \mathbb{M}_\beta)_{i,i}| \lesssim_{\text{P}} T^{-1}(\log N) \sum_{i\le N} \bar{u}_i^2 \lesssim_{\text{P}} T^{-2} N \log N, \tag{C.109}$$

and

$$T^{-2}\sum_{t\in T}\sum_{i\le N} |u_{i,t}^2 (\Sigma_u^{-2})_{i,i} (\mathbb{M}_\beta \bar{u}\bar{u}^\intercal \mathbb{M}_\beta)_{i,i}| \lesssim_{\text{P}} T^{-3}(\log N) \sum_{t\in T}\sum_{i\le N} u_{i,t}^2 \lesssim_{\text{P}} T^{-2} N \log N. \tag{C.110}$$

Further, we obtain

$$\begin{aligned} &\sum_{i\le N, j\le K} \text{E}(|(T^{-1}uu^\intercal\beta)_{i,j}||\beta, \Sigma_u) \\ \le &\sum_{i\le N, j\le K} \sqrt{\text{E}((T^{-1}uu^\intercal)_{i,i}|\Sigma_u)\text{E}((T^{-1}\beta^\intercal uu^\intercal\beta)_{j,j}|\beta,\Sigma_u)} \\ = &\sum_{i\le N, j\le K} \sqrt{(\Sigma_u)_{i,i}(\beta^\intercal\Sigma_u\beta)_{j,j}} \le N^{3/2}K\|\beta\|_{\text{MAX}}\lambda_{\max}(\Sigma_u) \lesssim_{\text{P}} N^{3/2}. \end{aligned} \tag{C.111}$$

The first inequality comes from Cauchy-Schwarz. The last inequality directly follows from condition (a) of Assumption 1. Similarly,

$$\begin{aligned} &\sum_{j\le K, k\le K} \text{E}(|(T^{-1}\beta uu^\intercal\beta)_{j,k}||\beta, \Sigma_u) \\ \le &\sum_{j\le K, k\le K} \sqrt{\text{E}((T^{-1}\beta^\intercal uu^\intercal\beta)_{j,j}|\beta, \Sigma_u)\text{E}((T^{-1}\beta^\intercal uu^\intercal\beta)_{j,j}|\beta,\Sigma_u)} \\ = &\sum_{j\le K, k\le K} \sqrt{(\beta^\intercal\Sigma_u\beta)_{j,j}(\beta^\intercal\Sigma_u\beta)_{k,k}} \le KN\|\beta\|_{\text{MAX}}\lambda_{\max}(\Sigma_u) \lesssim_{\text{P}} N. \end{aligned} \tag{C.112}$$

From (C.111) and (C.112), it directly follows

$$\sum_{i\leq N, j\leq K} |(T^{-1}uu^\intercal\beta)_{i,j}| \lesssim_{\mathrm{P}} N^{3/2}, \qquad \sum_{j\leq K, k\leq K} |(T^{-1}\beta uu^\intercal\beta)_{j,k}| \lesssim_{\mathrm{P}} N. \tag{C.113}$$

Using (C.113), and noting $\max_{i\leq N}|\bar{u}_i| \lesssim_{\mathrm{P}} \sqrt{T^{-1}\log N}$ from the uniform bound on i.i.d. random variables and $\|(\beta^\intercal\beta)^{-1}\beta\|_{\mathrm{MAX}} \lesssim N^{-1}$ by assumption, we obtain

$$\begin{aligned} & \sum_{i\leq N} |\bar{u}_i^2 (T^{-1}uu^\intercal \mathbb{P}_\beta)_{i,i}| \\ \leq & \sum_{i\leq N, j\leq K} |\bar{u}_i^2 (T^{-1}uu^\intercal\beta)_{i,j}| \|(\beta^\intercal\beta)^{-1}\beta\|_{\mathrm{MAX}} \\ \lesssim_{\mathrm{P}} & N^{-1}T^{-1}\log N \sum_{i\leq N, j\leq K} |(T^{-1}uu^\intercal\beta)_{i,j}| \lesssim_{\mathrm{P}} N^{1/2}T^{-1}\log N, \end{aligned} \tag{C.114}$$

and

$$\begin{aligned} & \sum_{i\leq N} |\bar{u}_i^2 (\mathbb{P}_\beta(T^{-1}uu^\intercal)\mathbb{P}_\beta)_{i,i}| \\ \leq & \sum_{i\leq N, j\leq K, k\leq K} \bar{u}_i^2 |(T^{-1}\beta uu^\intercal\beta)_{j,k}| \|(\beta^\intercal\beta)^{-1}\beta\|_{\mathrm{MAX}}^2 \\ \lesssim_{\mathrm{P}} & N^{-1}T^{-1}\log N \sum_{j\leq K, k\leq K} |(T^{-1}\beta uu^\intercal\beta)_{j,k}| \lesssim_{\mathrm{P}} T^{-1}\log N. \end{aligned} \tag{C.115}$$

Symmetric reasoning leads to

$$\frac{1}{T^2}\sum_{s\in T}\sum_{i\leq N} |u_{i,s}^2 (T^{-1}uu^\intercal\mathbb{P}_\beta)_{i,i}| \lesssim_{\mathrm{P}} N^{1/2}T^{-1}, \tag{C.116}$$

$$\frac{1}{T^2}\sum_{s\in T}\sum_{i\leq N} |u_{i,s}^2 (\mathbb{P}_\beta(T^{-1}uu^\intercal)\mathbb{P}_\beta)_{i,i}| \lesssim_{\mathrm{P}} T^{-1}. \tag{C.117}$$

Substituting (C.109), (C.110), (C.114), (C.116), (C.115), and (C.117) into (C.107) and (C.108), we obtain

$$\bar{u}^\intercal\mathbb{M}_\beta\widehat{\Sigma}_u^{-1}\mathbb{M}_\beta\bar{u} - T^{-1}N = -T^{-2}\sum_{i:i\leq N} A_i + O_{\mathrm{P}}(T^{-1}N^{1/2}\log N + T^{-2}N\log N). \tag{C.118}$$

Here and only here we use short-hand notation

$$A_i = \sum_{t\in T}\sum_{t'\in T: t'\neq t} (\Sigma_u^{-2})_{i,i}(T^{-1}uu^\intercal - 2\Sigma_u)_{i,i}u_{i,t}u_{i,t'}.$$

Since A_i is i.i.d. across i, we only need to analyze it for a single i. It obviously holds that $\mathrm{E}(A_i|\Sigma_u) = 0$. We also note $\mathrm{E}(((T^{-1}uu^\intercal - 2\Sigma_u)_{i,i})^2 u_{i,t}u_{i,t'}u_{i,s}u_{i,s'}|\Sigma_u) = 0$ unless two elements

of $\{t, t', s, s'\}$ are the same, and $\mathrm{E}(((T^{-1}uu^\intercal - 2\Sigma_u)_{i,i})^2 u_{i,t}u_{i,t'}u_{i,s}u_{i,s'}|\Sigma_u) \lesssim T^{-2}\mathrm{E}(u_{i,t}^8|\Sigma_u)$ unless elements of $\{t, t', s, s'\}$ only take two different values. Then we obtain

$$\mathrm{E}(A_i^2|\Sigma_u) \lesssim T^2(\Sigma_u^{-4})_{i,i}\mathrm{E}(u_{i,t}^8|\Sigma_u).$$

It hence follows that

$$T^{-2}\sum_{i:i\leq N} A_i \lesssim_{\mathrm{P}} T^{-1}N^{1/2}\mathrm{E}(u_{i,t}^8(\Sigma_u^{-4})_{i,i}) \lesssim T^{-1}N^{1/2}, \tag{C.119}$$

where the last inequality comes from that $\varepsilon_{i,t}$ has finite eighth moment by condition (a) of Assumption 2. Subsituting (C.119) into (C.118), we obtain

$$\bar{u}^\intercal \mathbb{M}_\beta \widehat{\Sigma}_u^{-1}\mathbb{M}_\beta \bar{u} - T^{-1}N = O_{\mathrm{P}}(T^{-1}N^{1/2}\log N + T^{-2}N\log N). \tag{C.120}$$

Combining (C.99), (C.100), and (C.120), and noting $N^{1/2}T \leq c_N$ and $T \lesssim N$ by assumption, we obtain

$$\begin{aligned}(\widehat{S}^\star)^2 &= (S^\star)^2 + o_{\mathrm{P}}\left(T^{-1/2}\sqrt{\log N}((S^\star)^2+1) + T^{-1}N^{1/2}\log N\right)\\ &= (S^\star)^2 + o_{\mathrm{P}}(T^{-1}N^{1/2}\log N((S^\star)^2+1)).\end{aligned}$$

Therefore, we obtain, under $S^\star \geq C$,

$$(\widehat{S}^\star)^2 = (S^\star)^2(1+o_{\mathrm{P}}(T^{-1}N^{1/2}\log N)), \quad \Longrightarrow \quad \frac{\widehat{S}^\star - S^\star}{S^\star} = o_{\mathrm{P}}(T^{-1}N^{1/2}\log N).$$

And, under $S^* \leq c_N$, we have

$$(\widehat{S}^\star)^2 = (S^\star)^2 + o_{\mathrm{P}}(T^{-1}N^{1/2}\log N), \quad \Longrightarrow \quad \widehat{S}^\star - S^\star = o_{\mathrm{P}}\left(\sqrt{T^{-1}N^{1/2}\log N}\right).$$

We note by construction $(\widetilde{S}^\star)^2 = (\widehat{S}^\star)^2 + N/T$. Then, under $S^\star \geq C$, it holds that

$$(\widehat{S}^\star)^2 = (S^\star)^2 + \frac{N}{T} + (S^\star)^2 o_{\mathrm{P}}\left(\frac{\sqrt{N}\log N}{T}\right) \quad \Longrightarrow \quad \frac{\widehat{S}^\star - \sqrt{(S^\star)^2 + N/T}}{S^\star} = o_{\mathrm{P}}\left(\frac{\sqrt{N}\log N}{T}\right).$$

Similarly, under $S^* \leq c_N$, we have

$$(\widehat{S}^\star)^2 = (S^\star)^2 + \frac{N}{T} + o_{\mathrm{P}}\left(\frac{\sqrt{N}\log N}{T}\right) \quad \Longrightarrow \quad \widehat{S}^\star - \sqrt{(S^\star)^2 + N/T} = o_{\mathrm{P}}\left(\frac{\sqrt{N}\log N}{T}\right).$$

The proof concludes.

C.6 Proof of Proposition B1

We note that all the assumptions in Theorem 2, other than the factor-neutrality, are assumed here as well. Therefore, equation (C.8) in the proof of Theorem 2 stays valid, since factor-neutrality is used only after the derivation of (C.8). Applying Cauchy-Schwarz inequality to (C.8), we obtain

$$|S(w)| \leq \sqrt{\mathrm{E}(r_{t+1}|\mathcal{G})^\intercal \Sigma^{-1} \mathrm{E}(r_{t+1}|\mathcal{G})} + o_{\mathrm{P}}(1). \tag{C.121}$$

On the other hand, it implies by Woodbury matrix identity and from the fact that $\Sigma = \beta\Sigma_v\beta^\intercal + \Sigma_u$,

$$\Sigma^{-1} = \Sigma_u^{-1} - \Sigma_u^{-1}\beta(\Sigma_v^{-1} + \beta^\intercal\Sigma_u^{-1}\beta)^{-1}\beta^\intercal\Sigma_u^{-1}. \tag{C.122}$$

By direct calculations, we have

$$\beta^\intercal\Sigma^{-1}\beta = ((\beta^\intercal\Sigma_u^{-1}\beta)^{-1} + \Sigma_v)^{-1}.$$

Let $H_1 = (\beta^\intercal\Sigma_u^{-1}\beta)^{-1}$ and $H_2 = \Sigma_v$, and using the fact that $(H_1+H_2)^{-1} - H_2^{-1} = -(H_1+H_2)^{-1}H_1H_2^{-1}$, we have

$$\beta^\intercal\Sigma^{-1}\beta - \Sigma_v^{-1} = -((\beta^\intercal\Sigma_u^{-1}\beta)^{-1} + \Sigma_v)^{-1}(\beta^\intercal\Sigma_u^{-1}\beta)^{-1}\Sigma_v^{-1}.$$

Therefore, using the fact that $\lambda_{\min}(\beta^\intercal\beta) \gtrsim_{\mathrm{P}} N$ and that $\lambda_{\max}(\Sigma_u) \lesssim_{\mathrm{P}} 1$ by assumption, we have

$$\lambda_{\max}((\beta^\intercal\Sigma_u^{-1}\beta)^{-1}) = \lambda_{\min}^{-1}(\beta^\intercal\Sigma_u^{-1}\beta) \leq \lambda_{\min}^{-1}(\beta^\intercal\beta)\lambda_{\max}(\Sigma_u) \lesssim_{\mathrm{P}} N^{-1}. \tag{C.123}$$

Also, note that $\lambda_{\max}(\Sigma_v^{-1}) = \lambda_{\min}^{-1}(\Sigma_v) \lesssim 1$, and that

$$\lambda_{\max}(((\beta^\intercal\Sigma_u^{-1}\beta)^{-1} + \Sigma_v)^{-1}) = \lambda_{\min}^{-1}((\beta^\intercal\Sigma_u^{-1}\beta)^{-1} + \Sigma_v) \leq \lambda_{\min}^{-1}(\Sigma_v) \lesssim 1,$$

we have

$$\|\beta^\intercal\Sigma^{-1}\beta - \Sigma_v^{-1}\| \lesssim_{\mathrm{P}} N^{-1},$$

which in turn leads to

$$\gamma^\intercal\beta^\intercal\Sigma^{-1}\beta\gamma = \gamma^\intercal\Sigma_v^{-1}\gamma + o_{\mathrm{P}}(1). \tag{C.124}$$

Next, we show

$$\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma^{-1}\beta\gamma = o_{\mathrm{P}}(1). \tag{C.125}$$

Notice that $\mathrm{E}(\mathrm{E}(\alpha|\mathcal{G})|\Sigma,\beta) = \mathrm{E}(\alpha|\Sigma,\beta) = \mathrm{E}(\alpha|\Sigma) = 0$ (by conditions (a) and (b) of Assumption 1), and that, conditionally on (Σ,β), $\mathrm{E}(\alpha_i|\mathcal{G})$ is independent across i as demonstrated in the beginning of the proof of Theorem 2. Therefore,

$$\mathrm{E}((\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma^{-1}\beta\gamma)^2|\Sigma,\beta) \leq \sum_{i\leq N}\mathrm{E}(\mathrm{E}(\alpha_i|\mathcal{G})^2|\Sigma,\beta)\max_{j\leq N}(\gamma^\intercal\beta^\intercal\Sigma^{-1})_j^2. \tag{C.126}$$

On the other hand, from (C.122), we obtain

$$\gamma^\intercal\beta^\intercal\Sigma^{-1} = \gamma^\intercal\Sigma_v^{-1}(\Sigma_v^{-1}+\beta^\intercal\Sigma_u^{-1}\beta)^{-1}\beta^\intercal\Sigma_u^{-1}.$$

Because of $\lambda_{\min}(\Sigma_v) \gtrsim 1$, $\|\beta\|_{\mathrm{MAX}} \lesssim_{\mathrm{P}} 1$, $\|\Sigma_u\|_{\mathrm{MAX}} \leq \|\Sigma_u\| \lesssim_{\mathrm{P}} 1$, $\lambda_{\min}(\Sigma_u) \gtrsim_{\mathrm{P}} 1$, Σ_u is diagonal, and (C.123), we have

$$\|\gamma^\intercal\beta^\intercal\Sigma^{-1}\|_{\mathrm{MAX}} \lesssim \|(\Sigma_v^{-1}+\beta^\intercal\Sigma_u^{-1}\beta)^{-1}\|\|\beta^\intercal\Sigma_u^{-1}\|_{\mathrm{MAX}} \lesssim_{\mathrm{P}} \lambda_{\max}((\beta^\intercal\Sigma_u^{-1}\beta)^{-1}) \lesssim_{\mathrm{P}} N^{-1}.$$

Hence, we have, for all positive fixed ϵ,

$$\mathrm{P}(|\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma^{-1}\beta\gamma| \geq \epsilon|\Sigma,\beta) \leq \mathrm{E}((\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma^{-1}\beta\gamma)^2|\Sigma,\beta)/\epsilon^2 = o_{\mathrm{P}}(1), \tag{C.127}$$

where the last equality comes from (C.126) and that $\mathrm{E}\left(\sum_{i\leq N}\mathrm{E}(\mathrm{E}(\alpha_i|\mathcal{G})^2|\Sigma,\beta)\right) \leq \sum_{i\leq N}\mathrm{E}(\alpha_i^2) = o(N)$ by condition (a) of Assumption 1. Since $\mathrm{P}(|\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma^{-1}\beta\gamma| \geq \epsilon|\Sigma,\beta) \leq 1$ are uniformly bounded for all N (by definition), we obtain by taking expectations on both sides of (C.127) that, for all positive fixed ϵ,

$$\mathrm{P}(|\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma^{-1}\beta\gamma| \geq \epsilon) = o(1),$$

which is equivalent to (C.125).

Finally, we derive

$$\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma^{-1}\mathrm{E}(\alpha|\mathcal{G}) = \mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma_u^{-1}\mathrm{E}(\alpha|\mathcal{G}) + o_{\mathrm{P}}(1). \tag{C.128}$$

Following the same derivation for (C.126), we obtain

$$\mathrm{E}(|\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma_u^{-1}\beta|_{\mathrm{F}}^2|\Sigma,\beta) \leq \sum_{i\leq N}\mathrm{E}(\mathrm{E}(\alpha_i|\mathcal{G})^2|\Sigma,\beta)\max_j(\Sigma_u^{-1}\beta\beta^\intercal\Sigma_u^{-1})_{j,j}.$$

Because $\|\beta\|_{\mathrm{MAX}} \lesssim_{\mathrm{P}} 1$ and $\lambda_{\min}(\Sigma_u) \gtrsim_{\mathrm{P}} 1$, we have

$$\max_j(\Sigma_u^{-1}\beta\beta^\intercal\Sigma_u^{-1})_{j,j} \lesssim \|\Sigma_u^{-1}\beta\|_{\mathrm{MAX}}^2 \lesssim_{\mathrm{P}} 1.$$

Then given the above result that $\mathrm{E}\left(\sum_{i\leq N}\mathrm{E}(\mathrm{E}(\alpha_i|\mathcal{G})^2|\Sigma,\beta)\right) = o(N)$, we obtain that $\mathrm{E}(|\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma_u^{-1}\beta|_{\mathrm{F}}^2|\Sigma,\beta) = o_{\mathrm{P}}(N)$. Therefore, similar to the derivation of (C.125), we ob-

tain

$$|\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma_u^{-1}\beta|_{\mathrm{F}}^2 = o_{\mathrm{P}}(N).$$

On the other hand, using (C.123), we obtain

$$\|(\Sigma_v^{-1}+\beta^\intercal\Sigma_u^{-1}\beta)^{-1}\| = \lambda_{\min}^{-1}(\Sigma_v^{-1}+\beta^\intercal\Sigma_u^{-1}\beta) \leq \lambda_{\max}((\beta^\intercal\Sigma_u^{-1}\beta)^{-1}) \lesssim_{\mathrm{P}} N^{-1}. \tag{C.129}$$

Then, using (C.129), we have

$$\begin{aligned}&\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma_u^{-1}\beta(\Sigma_v^{-1}+\beta^\intercal\Sigma_u^{-1}\beta)^{-1}\beta^\intercal\Sigma_u^{-1}\mathrm{E}(\alpha|\mathcal{G})\\ \leq&|\mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma_u^{-1}\beta|_{\mathrm{F}}^2\|(\Sigma_v^{-1}+\beta^\intercal\Sigma_u^{-1}\beta)^{-1}\| = o_{\mathrm{P}}(1),\end{aligned}$$

and hence, in light of (C.122), we obtain (C.128).

Given that $\mathrm{E}(r_{t+1}|\mathcal{G}) = \mathrm{E}(\alpha|\mathcal{G}) + \beta\gamma$, it follows from (C.124), (C.125), and (C.128) that

$$\mathrm{E}(r_{t+1}|\mathcal{G})^\intercal\Sigma^{-1}\mathrm{E}(r_{t+1}|\mathcal{G}) = \mathrm{E}(\alpha|\mathcal{G})^\intercal\Sigma_u^{-1}\mathrm{E}(\alpha|\mathcal{G}) + \gamma^\intercal\Sigma_v^{-1}\gamma + o_{\mathrm{P}}(1).$$

In light of (C.121), we conclude the proof.

C.7 Proof of Propositions B2, B3, and B4

We present the notation that is used throughout Section C.7 and Appendix E to facilitate the exposition of our proofs. We write

$$\zeta \coloneqq \sqrt{T}\mu/\sigma, \quad \check{z}_i \coloneqq \sqrt{T}\check{s}_i, \quad \widehat{z}_i \coloneqq \sqrt{T}\widehat{s}_i.$$

ζ represents the signal strengh of our alphas. $\widehat{z}_i$ is simply the t-statistic and $\check{z}_i$ is hypothetically what the t-statistic would be in the absence of risk factors. Next, we introduce soft- and hard-thresholding functions:

$$\widetilde{\psi}_1(a,\lambda) \coloneqq \mathrm{sgn}(a)(|a|-\lambda), \quad \widetilde{\psi}_2(a,\lambda) \coloneqq a\mathbb{1}_{\{|a|\geq\lambda\}}.$$

Then, for $q \in \{1,2\}$ and $i \in \{1,\ldots,N\}$,

$$\widetilde{\psi}_{q,i}(\lambda) \coloneqq \widetilde{\psi}_q(\check{z}_1,\lambda), \quad \widehat{\psi}_{q,i}(\lambda) \coloneqq \widetilde{\psi}_q(\widehat{z}_i,\lambda).$$

Further, $\widetilde{\psi}_q(\lambda)$ and $\widehat{\psi}_q(\lambda)$ to represent, repectively, the vectors $(\widetilde{\psi}_q(\check{z}_1,\lambda),\ldots,\widetilde{\psi}_q(\check{z}_N,\lambda))^\intercal$ and $(\widehat{\psi}_{q,1}(\lambda),\ldots,\widehat{\psi}_{q,N}(\lambda))^\intercal$. Based on $\widehat{\psi}_q(\lambda)$, we introduce $\widehat{w}_q'(\lambda) \coloneqq \widehat{\Sigma}_u^{-1/2}\widehat{\psi}_q(\lambda)$. The purpose of doing so is that $\widehat{w}_q'(\lambda)$, with appropriate choice of λ, incorporates the weights of Lasso, CSR,

and BH methods all as special cases. Lastly, we define

$$\widehat{S}'_q(\lambda) := \frac{\widehat{w}'_q(\lambda)\mathbb{M}_\beta \alpha}{\sigma \|\mathbb{M}_\beta \widehat{w}'_q(\lambda)\|}, \quad S_q(\lambda) := \frac{N^{1/2}\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)\psi_i)}{\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2)^{1/2}}.$$

$\widehat{S}'_q(\lambda)$ is the Sharpe ratio generated by the strategy $\widehat{w}'_q(\lambda)$. $S_q(\lambda)$, again with appropriate choice of λ, incorporates as special cases S^{CSR}, S^{BH}, and S^{LASSO} (the probability limits of Sharpe ratios) all as special cases.

C.7.1 Proof of Proposition B2

By definition it holds that $\widehat{\Sigma}_u^{-1}\widehat{\alpha} = \breve{w}_1(\lambda)$ under $\lambda = 0$. Hence we have $\widehat{S}^{\mathrm{CSR}} = \widehat{S}_1(\lambda)$, choosing $\lambda = 0$. In other words, CSR is a special case of Lasso. On the other hand, by definition we have, choosing $\lambda = 0$,

$$S_\lambda^{\mathrm{Lasso}} = N^{1/2}\mathrm{E}(\check{s}_i s_i)\mathrm{E}(\check{s}_i^2)^{-1/2}. \tag{C.130}$$

Moreover, we can write

$$\sqrt{T} \times \mathrm{E}(\check{s}_i s_i) = \frac{\rho\mu}{2\sigma}\int a(\phi(a-\zeta) - \phi(a+\zeta))da = \frac{\rho\mu\zeta}{\sigma}, \tag{C.131}$$

$$T \times \mathrm{E}(\check{s}_i^2) = \int a^2(\phi(a) + \rho\phi(a-\zeta))da = 1 + \rho(1+\zeta^2) = (1+o(1))(1+\rho\zeta^2). \tag{C.132}$$

For both lines, the first equalities come from the distribution of α_i imposed by Example 1, $s_i = \alpha_i\sigma^{-1}$ by definition, and $\sqrt{T}(\check{s}_i - s_i)$ is standard normal. The second equalities hold by direct calculations. The last equality of the second line comes from $\rho \to 0$. Therefore, it holds that, choosing $\lambda = 0$,

$$\widehat{S}^{\mathrm{CSR}} = \widehat{S}_1(\lambda) = S_\lambda^{\mathrm{Lasso}} + o_{\mathrm{P}}(1 + S^{\mathrm{OPT}}) = S^{\mathrm{CSR}} + o_{\mathrm{P}}(1 + S^{\mathrm{OPT}}). \tag{C.133}$$

Here the second equality comes from $\widehat{S}_1(\lambda) = S_\lambda^{\mathrm{Lasso}} + o_{\mathrm{P}}(1 + S^{\mathrm{OPT}})$ as in Proposition B4. The third equality comes from (C.130), (C.131), (C.132), $\zeta = \sqrt{T}\mu/\sigma$ by definition, and that $|S^{\mathrm{CSR}}| \leq S^{\mathrm{OPT}} + o_{\mathrm{P}}(1)$ by Theorem 2 and Corollary 2.

Now we verify the sufficient and necessary condition. We utilize the last equality of (C.133), i.e., that CSR is a special case of Lasso, and the optimality conditions for Lasso provided in Proposition B4. We start with the "if" part. Invoking the classic subsequence argument, we only need to consider the two cases $\sqrt{T}\mu/\sigma \to 0$ and $\sqrt{\rho T}\mu/\sigma \to \infty$ separately. The former case is obvious since $\lambda = 0$ certainly satisfies $T^{1/2}\lambda \to 0$ condition required by Proposition B4. For the latter case, noting that $\sqrt{\rho T}\mu/\sigma \to \infty$ always leads to $\sqrt{T}\mu/\sigma - \sqrt{-2\log\rho} \to \infty$ due to $-\log\rho \asymp \log N$ by assumption, we only need to verify that, under

$\lambda = 0$, we have both $\sqrt{T}(\mu/\sigma - \lambda) \to \infty$ and $\frac{\phi(\sqrt{T}\lambda)}{\rho(1+T\lambda^2)T(\mu/\sigma-\lambda)^2} \to 0$ as in Proposition B4. Both are obviously true.

Next we show the "only if" part. As $\sqrt{T}\mu/\sigma \to 0$ is violated, we only need to show that, if $\sqrt{\rho T}\mu/\sigma \to \infty$ is violated, we would not have $\frac{\phi(\sqrt{T}\lambda)}{\rho(1+T\lambda^2)T(\mu/\sigma-\lambda)^2} \to 0$ under $\lambda = 0$. This is also obvious. The proof ends.

C.7.2 Proof of Proposition B3

Proof of Proposition B3, Part 1. In this part we prove $\widehat{S}_\tau^{\mathrm{BH}} = S_\tau^{\mathrm{BH}} + o_{\mathrm{P}}(1 + S^{\mathrm{OPT}})$. By assumption there exists fixed $d > 0$ such that $\mu \le N^{-d}$. Then there exists fixed (d_1, d_2, d_3) such that $0 < d_1 < d_2 < d_3 < \min\{d, 1\}$. We also set $c_x = \sqrt{2(1-x)\log N}$.

We let $\check{z} = -\Phi^{-1}(p_{(\widehat{k})}/2)$ and $\check{z}' = -\Phi^{-1}(p_{(\widehat{k}+1)}/2)$, where we recall Φ is the standard normal cdf. In other words, $\check{z}$ and $\check{z}'$ are the t-statistics whose p-values, calculated based on standard normal distribution, are $p_{(\widehat{k})}$ and $p_{(\widehat{k}+1)}$. We further define, for $j \in \{0, +, -\}$,

$$m_j(a) = \sum_{i \in H_j} \mathbb{1}_{\{|\widehat{z}_i| \ge a\}}, \quad \check{m}_j(a) = \sum_{i \in H_j} \mathbb{1}_{\{|\check{z}_i| \ge a\}},$$

where $H_0 = \{i \le N : \alpha_i = 0\}$, $H_+ = \{i \le N : \alpha_i = \zeta\}$, $H_- = \{i \le N : \alpha_i = -\zeta\}$, and $\widehat{z}_i = T^{1/2}\widehat{s}_i$ is the t-statistic of stock i. From the definitions of $\check{z}$ and $\check{z}'$, we obtain

$$\frac{2N\Phi(-\check{z})}{\sum_{j\in\{0,+,-\}} m_j(\check{z})} \le \tau, \quad \frac{2N\Phi(-\check{z}')}{\sum_{j\in\{0,+,-\}} m_j(\check{z}')} > \tau, \quad \sum_{j\in\{0,+,-\}} m_j(\check{z}') = \sum_{j\in\{0,+,-\}} m_j(\check{z}) + 1. \tag{C.134}$$

Given (E.74) and the monotonicity of $m_j(z)$ in z, there exists a positive sequence $b_N = o\left(1/\sqrt{\log N}\right)$ such that, with high probability, uniformly over x, and for $j \in \{0, +, -\}$,

$$\check{m}_j(x + b_N) \le m_j(x) \le \check{m}_j(x - b_N). \tag{C.135}$$

Further, noting $\check{z}_i$ is i.i.d. across i, we obtain from equation (13) of Liu and Shao (2014) that, for all deterministic sequences (a_N, a'_N) satisfying $N\Phi(-a_N) \to \infty$ and $\rho N\Phi(\zeta - a'_N) \to \infty$,

$$\sup_{0\le z\le a_N} \left|\frac{\check{m}_0(z)}{2N\Phi(-z)} - 1\right| = o_{\mathrm{P}}(1), \quad \sup_{0\le z\le a'_N} \left|\frac{2\check{m}_\pm(z)}{\rho N\Phi(\zeta - z) + \rho N\Phi(-\zeta - z)} - 1\right| = o_{\mathrm{P}}(1). \tag{C.136}$$

Using (E.15) and $-\Phi^{-1}(N^{-1}) \asymp \sqrt{\log N}$, we note that, for $b_N = o\left(1/\sqrt{\log N}\right)$, $\Phi(-z \pm b_N) = \Phi(-z)(1 + o(1))$ uniformly over $|z| \le -\Phi^{-1}(N^{-1})$. Hence we obtain from (C.135) and (C.136) that, for all deterministic sequences (a_N, a'_N) satisfying $N\Phi(-a_N) \to \infty$ and $\rho N\Phi(\zeta - a'_N) \to \infty$,

$$\sup_{0\le z\le a_N}\left|\frac{m_0(z)}{2N\Phi(-z)}-1\right|=o_{\mathrm{P}}(1),\quad \sup_{0\le z\le a'_N}\left|\frac{2m_{\pm}(z)}{\rho N\Phi(\zeta-z)+\rho N\Phi(-\zeta-z)}-1\right|=o_{\mathrm{P}}(1). \tag{C.137}$$

In addition, it follows from (E.15) and the second part of (E.1) that, uniformly over $z\ge 0$,

$$\frac{\rho\Phi(\zeta-z)}{\Phi(-z)}\asymp h(z),\quad \text{with}\quad h(z):=\frac{z}{1+(z-\zeta)_+}e^{(z-a^*)\zeta}. \tag{C.138}$$

Here a^* is introduced in the statement of Lemma E1, so is ζ^* which appears below. We now establish that, for all fixed (x,x') satisfying $0<x'<x\le 2$, if $\sup_{0\le z\le c_x}h(z)\gtrsim 1$, then

$$\inf_{z\ge c_{x'}}h(z)\to\infty. \tag{C.139}$$

Since $h(z)\lesssim e^{\log z+(z-a^*)\zeta}$ by definition, it holds that, if $\sup_{0\le z\le c_x}h(z)\gtrsim 1$, there exists $0\le a_N\le c_x$ such that $h(a_N)\gtrsim 1$ and thereby

$$e^{\log a_N+(a_N-a^*)\zeta}\gtrsim 1.$$

Because $a^*\zeta-\log a_N\gtrsim\log N$ by the first part of (E.1), $\zeta^*\gtrsim\sqrt{\log N}$ due to our assumption on ρ, and $a_N=o(\log N)$, we have $\zeta\gtrsim a_N^{-1}\log N\gtrsim\sqrt{\log N}$. For all z satisfying $z-a_N\ge c_{x'}-c_x\gtrsim\sqrt{\log N}$, it holds that $h(z)\gtrsim h(a_N)z^{-1}e^{(z-a_N)\zeta}\to\infty$ as $\zeta\gtrsim\sqrt{\log N}$. We hence establish (C.139).

We first suppose $z^*\ge c_{d_2}$. From (C.138) and the definition of z^* (B.1) we have $h(z^*)\asymp 1$. Then, since $d_2<d_3$, it follows from (C.139) (by contradiction) that

$$\sup_{0\le z\le c_{d_3}}\frac{\rho\Phi(\zeta-z)}{\Phi(-z)}\to 0. \tag{C.140}$$

Using (C.140), and applying (C.137) and the monotonicity of $m_j(z)$ and $\Phi(-z)$ in z, it holds that, for all (random) sequence z satisfying $0\le z\le c_{d_3}$,

$$m_0(z)=2N\Phi(-z)(1+o_{\mathrm{P}}(1)),\quad m_{\pm}(z)=o_{\mathrm{P}}(N\Phi(-z)). \tag{C.141}$$

Here we also use $\rho\Phi(-\zeta-z)\le\Phi(-z)$. Comparing (C.141) with (C.134), we have that $\mathrm{P}(\check{z}\le c_{d_3})\to 0$. We hence have, with high probability,

$$\mu^2\rho N\Phi(\zeta-\check{z})\le\mu^2\rho N\Phi(\zeta-c_{d_3})\le\mu^2 N\Phi(-c_{d_3})=o_{\mathrm{P}}(\mu^2N^d)=o_{\mathrm{P}}(1). \tag{C.142}$$

Here the second inequality comes from (C.140), the first equality comes from $d_3<d$, and the last equality comes from $\mu\le N^{-d}$. (The situation where $\check{z}$ does not exists is equivalent with

$\breve{z}$ being so large that $\sup_i |\widehat{z}_i| < \breve{z}$ and is hence a special case.) Given (C.142), Lemma E7 applies and we obtain $\widehat{S}_\tau^{\mathrm{BH}} = \widetilde{S}_2(\breve{z}) = o_{\mathrm{P}}(1)$.

On the other hand, we have

$$\begin{aligned} S_\tau^{\mathrm{BH}} &= \frac{N^{1/2}\mathrm{E}(\widetilde{\psi}_{2,i}(z^*)s_i)}{\mathrm{E}(\widetilde{\psi}_{2,i}(z^*)^2)^{1/2}} \le \frac{\mu N^{1/2}\mathrm{E}(|\widetilde{\psi}_{2,i}(z^*)|\mathbb{1}_{\{\alpha_i\neq 0,|\breve{z}_i|\ge z^*\}})}{\sigma \mathrm{E}(\widetilde{\psi}_{2,i}(z^*)^2)^{1/2}} \\ &\le \frac{\mu}{\sigma}\sqrt{N\mathrm{E}(\mathbb{1}_{\{\alpha_i\neq 0,|\breve{z}_i|\ge z^*\}})} \le \frac{\mu}{\sigma}\sqrt{2\rho N\Phi(\zeta - z^*)} \lesssim \mu\sqrt{N\Phi(-z^*)} = o(1). \end{aligned}$$

The second inequality comes from Cauchy-Schwarz inequality, the last inequality comes from (B.1), and the last equality comes $z^* \ge c_{d_2}$ and $N\Phi(-c_{d_2}) = o_{\mathrm{P}}(N^d)$ as $d_2 < d$. Hence we prove $\widehat{S}_\tau^{\mathrm{BH}} - S_\tau^{\mathrm{BH}} = o_{\mathrm{P}}(1)$ under $z^* \ge c_{d_2}$.

Now suppose $z^* \le c_{d_2}$. Then it follows from (C.138), (B.1) and (C.139) that, as $d_1 < d_2$,

$$\inf_{z\ge c_{d_1}} \frac{\rho\Phi(\zeta - z)}{\Phi(-z)} \to \infty. \tag{C.143}$$

Since $d_1 > 0$, we have $N\Phi(-c_{d_1}) \to \infty$. Given (C.143), we further have $\rho N\Phi(\zeta - c_{d_1}) \to \infty$. Applying (C.137), and using (C.143), it holds that

$$m_0(c_{d_1}) = o_{\mathrm{P}}(\rho\Phi(\zeta - c_{d_1})), \quad m_\pm(c_{d_1}) = \rho\Phi(\zeta - c_{d_1})(1 + o_{\mathrm{P}}(1)). \tag{C.144}$$

Let $\widetilde{h}(z) := 2N\Phi(-z)/\sum_{j\in\{0,+,-\}} m_j(z)$. From (C.144), we know that $h(c_{d_1}) = o_{\mathrm{P}}(1)$. Since $\widetilde{h}(0) = 1$, we know that with high probability $\breve{z}$ exists and satisfy $0 \le \breve{z} \le c_{d_1}$. Since $\Phi(-z)$ is decreasing in z, we have $N\Phi(-\breve{z}) \to \infty$ and $\rho N\Phi(\zeta - \breve{z}) \to \infty$ in probability. Therefore, we obtain from (C.137) that

$$m_0(\breve{z}) = 2N\Phi(-\breve{z})(1 + o_{\mathrm{P}}(1)), \quad m_\pm(\breve{z}) = \frac{\rho}{2}N(\Phi(\zeta - \breve{z}) + \Phi(-\zeta - \breve{z}))(1 + o_{\mathrm{P}}(1)). \tag{C.145}$$

Since $\breve{z}' \le \breve{z}$, (C.145) would still hold if all $\breve{z}$ are replaced by $\breve{z}'$. Hence, substituting (C.145) back into (C.134), and noting $\Phi(-\zeta - z) \le \Phi(-z)$, we have

$$\frac{2(1-\tau)\Phi(-\breve{z})}{\tau\rho\Phi(\zeta - \breve{z})} = 1 + o_{\mathrm{P}}(1). \tag{C.146}$$

Next, using (B.1) and (E.15), we note that $z^* \le c_{d_2}$ leads to that $\zeta \gtrsim \sqrt{\log N}$ (see a few lines after (C.139)). As a result, using (E.15), and comparing (B.1) and (C.146), we have

$$|\breve{z} - z^*| = o_{\mathrm{P}}\left(1/\sqrt{\log N}\right). \tag{C.147}$$

We finally obtain

$$\widehat{S}_\tau^{\mathrm{BH}} = \frac{\widehat{w}_q'(T^{-1/2}\check{z})^\intercal \mathbb{M}_\beta \alpha}{\sigma \|\mathbb{M}_\beta \widehat{w}_q'(T^{-1/2}\check{z})\|} = S_2(z^*) + o_{\mathrm{P}}(S^{\mathrm{OPT}}+1) = S_\tau^{\mathrm{BH}} + o_{\mathrm{P}}(S^{\mathrm{OPT}}+1).$$

where the first equality holds by definition, the second comes from (C.147), $\rho N\Phi(\zeta - z^*) \geq \rho N\Phi(\zeta - c_{d_1}) \to \infty$ (see after (C.143)), and Lemma E6 (choose $q=2$ and $(\lambda', \lambda) = (\check{z}, z^*)$)), and the last holds by $S_2(z^*) = S_\tau^{\mathrm{BH}}$ due to $\psi_i = \mathrm{E}(s_i|\mathcal{G})$ and that $\widetilde{\psi}_{q,i}(z^*)$ is $\mathcal{G}$-measurable. We hence prove $\widehat{S}_\tau^{\mathrm{BH}} - S_\tau^{\mathrm{BH}} = o_{\mathrm{P}}(S^{\mathrm{OPT}}+1)$ under $z^* \leq c_{d_2}$. ∎

Proof of Proposition B3, Part 2. In this part we prove that the sufficient and necessary condition is correct. By definition it holds that

$$\frac{S_\tau^{\mathrm{BH}}}{S^{\mathrm{OPT}}} = \frac{S_2(z^*)}{S^{\mathrm{OPT}}} = \frac{\mathrm{E}(\widetilde{\psi}_{2,i}(z^*)\psi_i)}{\mathrm{E}(\widetilde{\psi}_{2,i}(z^*)^2)^{1/2}\mathrm{E}(\psi_i^2)^{1/2}} = \mathrm{Corr}(\psi_i, \widetilde{\psi}_{2,i}(z^*)). \tag{C.148}$$

The first equality holds by $S_2(z^*) = S_\tau^{\mathrm{BH}}$ due to $\psi_i = \mathrm{E}(s_i|\mathcal{G})$ and that $\widetilde{\psi}_{q,i}(z^*)$ is $\mathcal{G}$-measurable. Since $\rho \to 0$ by assumption, we know that $z^* \to 0$ does not hold as it contradicts its definition (B.1). Moreover, we have that, using the first part of (E.1) and that $-\log\rho \asymp \log N$ by assumption, $\zeta - a^* \geq -c_N\sqrt{\log N}$ and $\zeta - \zeta^* \geq -c_N\sqrt{\log N}$ are equivalent. Hence, according to (E.23) and (E.24) of Lemma E3 (choose $q=2$), using (C.148) and the subsequence argument, and recalling $\zeta = \sqrt{T}\mu/\sigma$ and $\zeta^* = \sqrt{-2\log\rho}$, we readily obtain the "only if" part. Then we study the "if" part, according to (E.25), we only need to show that, for all fixed $\epsilon > 0$, there exists a fixed $\tau > 0$ such that, for large N,

$$\frac{\Phi(-z^*)}{\rho\Phi(\zeta - a^*)} \leq \epsilon. \tag{C.149}$$

The rest of the proof is to show (C.149). Rewritting the definition of z^*, and noting τ is fixed, we have

$$\frac{\rho\Phi(\zeta - z^*)}{2\Phi(-z^*)} = \frac{1-\tau}{\tau} \asymp 1. \tag{C.150}$$

Using (E.15) and the second part of (E.1) (recall $\chi(a)$ and a^* are introduced in the statement of Lemma E1), noting $\chi(a) \coloneqq \frac{\rho}{2}\phi(a-\zeta)/\phi(a)$, we obtain from (C.150) that

$$h(z^*) \coloneqq \frac{z^*}{1+(z^*-\zeta)_+} e^{(z^*-a^*)\zeta} \asymp 1. \tag{C.151}$$

Here we also introduce $h(z^*)$ as short-hand notation. Suppose $z^* \geq a^*+1$. Then $h(z^*) \geq e^{(z^*-a^*)\zeta} \to \infty$. Suppose $a^* \leq z^* \leq a^*+1$. Then $h(z^*) \geq \frac{1+a^*}{2+(a^*-\zeta)_+} \to \infty$. Suppose $z^* \leq a^*-1$.

Then $h(z^*) \le a^* e^{-\zeta} \to 0$. Hence we conclude $a^* - 1 \le z^* \le a^*$ for large N. Therefore, it follows from (C.151) that

$$e^{(a^*-z^*)\zeta} \asymp \frac{a^*}{1+(a^*-\zeta)_+}. \tag{C.152}$$

It then holds that, when $\zeta - \zeta^* \ge -c_N\sqrt{\log N}$,

$$\begin{aligned} a^* - z^* &\lesssim \frac{1}{\zeta}\log\frac{a^*}{1+(a^*-\zeta)_+} \\ &\lesssim \frac{1}{1+(a^*-\zeta)_+}\frac{1+(a^*-\zeta)_+}{a^*}\log\frac{a^*}{1+(a^*-\zeta)_+} \le o\left(\frac{1}{1+(a^*-\zeta)_+}\right). \end{aligned} \tag{C.153}$$

The first inequality follows from (C.152). The second inequality comes from $a^* \asymp \zeta$ due to the first part of (E.1) and $\zeta^* \asymp \sqrt{\log N}$ by the assumption on ρ. To obtain the last inequality, we utilize that $(a^* - \zeta)_+ = o(a^*)$, again due to the first part of (E.1), and that $\lim_{x\to\infty} x^{-1}\log x = 0$. We hence obtain

$$\frac{|\Phi(\zeta - z^*) - \Phi(\zeta - a^*)|}{\Phi(\zeta - a^*)} \lesssim (a^* - z^*)\max_{z^*\le a\le a^*}\phi(a-\zeta) = o\left(\frac{\phi(a^*-\zeta)}{1+(a^*-\zeta)_+}\right) = o(\Phi(\zeta - a^*)). \tag{C.154}$$

The first inequality is obvious given that $z^* \le a^*$ for large N. The first equality comes from (C.153) and that $\sup_{y:|y|\lesssim(1+|x|)^{-1}}\phi(x+y) \lesssim \phi(x)$ uniformly over x. The last equality is a result of (E.15). It follows from (C.154) that $\Phi(\zeta - a^*) = \Phi(\zeta - z^*)(1+o(1))$. Combining this result with (C.150), we further have $2\Phi(-z^*)/(\rho\Phi(\zeta - a^*)) = (1+o(1))\tau/(1-\tau)$. We hence prove (C.149) and the proof concludes. ∎

C.7.3 Proof of Proposition B4

By definition, we have $\breve{w}_{1,i}(\lambda) = \widehat{\sigma}_i^{-1}\psi_1(\widehat{s}_i,\lambda) = T^{-1/2}\widehat{\sigma}_i^{-1}\psi_1(\widehat{z}_i, T^{1/2}\lambda) = T^{-1/2}\widetilde{w}_1'(T^{1/2}\lambda)$ and $\psi_1(\check{s}_i,\lambda) = T^{-1/2}\widetilde{\psi}_1(T^{1/2}\lambda)$. Hence, it holds that

$$\widehat{S}_{1,\lambda} = \frac{\breve{w}_1(\lambda)^\intercal\mathbb{M}_\beta\alpha}{\sigma\|\mathbb{M}_\beta\breve{w}_1(\lambda)\|} = \widehat{S}_1'(T^{1/2}\lambda), \quad S_\lambda^{\text{Lasso}} = \frac{N^{1/2}\mathrm{E}(\psi_i\psi_1(\check{s}_i,\lambda))}{\sqrt{\mathrm{E}(\psi_1(\check{s}_i,\lambda)^2)}} = S_1(T^{1/2}\lambda), \tag{C.155}$$

where the first equality in the second part comes from $\psi_i = \mathrm{E}(s_i|\mathcal{G})$ and that $\psi_1(\check{s}_i,\lambda)$ is $\mathcal{G}$-measurable.

Suppose λ satisfies $\rho N\Phi(\zeta - T^{1/2}\lambda) \to \infty$. Then we have

$$\widehat{S}_{1,\lambda} = \widehat{S}_1'(T^{1/2}\lambda) = S_1(T^{1/2}\lambda) + o_{\mathrm{P}}(S^{\text{OPT}}+1) = S_\lambda^{\text{Lasso}} + o_{\mathrm{P}}(S^{\text{OPT}}+1). \tag{C.156}$$

Here the first and last equality comes from (C.155) and the second equality comes from

Lemma E6 as $\rho N\Phi(\zeta - T^{1/2}\lambda) \to \infty$ (replace both λ' and λ with $T^{1/2}\lambda$ and choose $q = 1$).

Now suppose $\rho N\Phi(\zeta - T^{1/2}\lambda) \lesssim 1$. It holds that

$$S_1(T^{1/2}\lambda) = \frac{N^{1/2}\mathrm{E}(\widetilde{\psi}_{1,i}(T^{1/2}\lambda)s_i)}{\mathrm{E}(\widetilde{\psi}_{1,i}(T^{1/2}\lambda)^2)^{1/2}} \leq \frac{\mu}{\sigma}\sqrt{N\mathrm{E}(\mathbb{1}_{\{|\check{s}_i|\geq 0, \alpha_i \neq 0\}})} \leq \frac{\mu}{\sigma}\sqrt{2\rho N\Phi(\zeta - T^{1/2}\lambda)} = o(1). \tag{C.157}$$

Here the first equality comes from $\psi_i = \mathrm{E}(s_i|\mathcal{G})$ and that $\widetilde{\psi}_{1,i}(\lambda)$ is $\mathcal{G}$-measurable. The first inequality holds by $|s_i| = \frac{\mu}{\sigma}\mathbb{1}_{\{\alpha_i \neq 0\}}$ and Cauchy-Schwarz inequality. The second inequality is obvious. The last equality comes from $\mu = o(1)$ by assumption. Since $\rho N\Phi(\zeta - T^{1/2}\lambda) \lesssim 1$ leads to $\mu^2\rho N\Phi(\zeta - T^{1/2}\lambda) \to 0$, Lemma E7 applies and hence $\widehat{S}_1'(T^{1/2}\lambda) = o_{\mathrm{P}}(1)$. Combining with (C.157), we obtain $\widehat{S}_1'(T^{1/2}\lambda) = S_1(T^{1/2}\lambda) + o_{\mathrm{P}}(1)$. Given (C.155), we obtain $\widehat{S}_1(\lambda) = S_\lambda^{\mathrm{Lasso}} + o_{\mathrm{P}}(1)$ under $\rho N\Phi(\zeta - T^{1/2}\lambda) \lesssim 1$. Hence we establish $\widehat{S}_1(\lambda) = S_\lambda^{\mathrm{Lasso}} + o_{\mathrm{P}}(1)$ with the classic subsequence argument.

Finally, it holds that

$$\frac{S_\lambda^{\mathrm{Lasso}}}{S^{\mathrm{OPT}}} = \frac{S_1(T^{1/2}\lambda)}{S^{\mathrm{OPT}}} = \frac{\mathrm{E}(\widetilde{\psi}_{1,i}(T^{1/2}\lambda)\psi_i)}{\mathrm{E}(\widetilde{\psi}_{1,i}(T^{1/2}\lambda)^2)^{1/2}\mathrm{E}(\psi_i^2)^{1/2}} = \mathrm{Corr}(\psi_i, \widetilde{\psi}_{1,i}(T^{1/2}\lambda)), \tag{C.158}$$

where the first equality comes from (C.155) and the second comes from the definition of S_1 and $S^{\mathrm{OPT}} = N^{1/2}\mathrm{E}(\psi_i^2)^{1/2}$. On the other hand, we have that, using the first part of (E.1) and that $-\log\rho \asymp \log N$ by assumption, $\zeta - a^* \to \infty$ and $\zeta - \zeta^* \to \infty$ are equivalent. Hence, according to (E.23) and (E.24) of Lemma E3 (replace λ with $T^{1/2}\lambda$ and choose $q = 2$), using (C.158), and recalling $\zeta = \sqrt{T}\mu/\sigma$ and $\zeta^* = \sqrt{-2\log\rho}$, we obtain the remaining statements of Proposition B4.

Appendix D Lemmas supporting Section C.4

Lemma D1. *We define $\bar{u}_i = T^{-1}\sum_{s\in\mathcal{T}} u_{i,s}$. Suppose Assumptions 1 and 2 hold and that $\|\beta\|_{\mathrm{MAX}} \lesssim_{\mathrm{P}} 1$ and $\lambda_{\min}(\beta^\intercal\beta) \gtrsim_{\mathrm{P}} N$. Also suppose $\log N \lesssim T \lesssim N^d$ with fixed $d < 1$. Then it holds that, as $N, T \to \infty$,*

$$\max_{1\leq i\leq N} |\widehat{\sigma}_i^2 - \sigma_i^2| = O_{\mathrm{P}}\left(\sqrt{(\log N)/T}\right), \tag{D.1}$$

$$\max_{1\leq i\leq N} |(\mathbb{P}_\beta \bar{u})_i| = O_{\mathrm{P}}\left(1/\sqrt{TN}\right), \tag{D.2}$$

$$\max_{1\leq i\leq N} |(\mathbb{P}_\beta \alpha)_i| = O_{\mathrm{P}}(N^{-1/2}\mathrm{E}(s_i^2)^{1/2}). \tag{D.3}$$

Proof. We start with (D.1). First of all, we write

$$\begin{aligned}\max_{1\le i\le N}|\widehat{\sigma}_i^2-\sigma_i^2| &\le \|\mathbb{M}_\beta(T^{-1}uu^\intercal-\bar{u}\bar{u}^\intercal)\mathbb{M}_\beta-\Sigma_u\|_{\text{MAX}}\\ &\le \|\mathbb{M}_\beta\Sigma_u\mathbb{M}_\beta-\Sigma_u\|_{\text{MAX}}+\|\mathbb{M}_\beta(T^{-1}uu^\intercal-\bar{u}\bar{u}^\intercal-\Sigma_u)\mathbb{M}_\beta\|_{\text{MAX}}. \quad \text{(D.4)}\end{aligned}$$

Now we establish the upper bounds of the two terms in the second line. We write

$$\begin{aligned}\|\mathbb{M}_\beta\Sigma_u\mathbb{M}_\beta-\Sigma_u\|_{\text{MAX}} &\le \|\mathbb{P}_\beta\Sigma_u\mathbb{P}_\beta\|_{\text{MAX}}+2\|\mathbb{P}_\beta\Sigma_u\|_{\text{MAX}}\\ &\le (N\|\mathbb{P}_\beta\|_{\text{MAX}}+2)\|\mathbb{P}_\beta\|_{\text{MAX}}\|\Sigma_u\|_{\text{MAX}}\lesssim_{\text{P}} N^{-1}. \quad \text{(D.5)}\end{aligned}$$

The last inequality comes from $\|\mathbb{P}_\beta\|_{\text{MAX}}\le C\|\beta\|_{\text{MAX}}^2\|(\beta^\intercal\beta)^{-1}\|_{\text{MAX}}\lesssim_{\text{P}} N^{-1}$, which is true by assumption. On the other hand, we have

$$\begin{aligned}&\|\mathbb{M}_\beta(T^{-1}uu^\intercal-\bar{u}\bar{u}^\intercal-\Sigma_u)\mathbb{M}_\beta\|_{\text{MAX}}\\ \le\ &\|T^{-1}uu^\intercal-\bar{u}\bar{u}^\intercal-\Sigma_u\|_{\text{MAX}}(1+2N\|\mathbb{P}_\beta\|_{\text{MAX}}+N^2\|\mathbb{P}_\beta\|_{\text{MAX}})\lesssim_{\text{P}}\sqrt{(\log N)/T}, \quad \text{(D.6)}\end{aligned}$$

where the last inequality comes from the uniform bound on i.i.d. normal variables, $\log N\lesssim T$ by assumption, and that $\lambda_{\max}(\Sigma_u)\lesssim_{\text{P}} 1$ by condition (a) of Assumption 1. Substituting (D.5) and (D.6) into (D.4), and noting $N^{-1}\le C\sqrt{(\log N)/T}$ by assumption, we obtain (D.1).

We obtain (D.2) by writing

$$\max_{1\le i\le N}|(\mathbb{P}_\beta\bar{u})_k|\le C\|\beta\|_{\text{MAX}}\|(\beta^\intercal\beta)^{-1}\|_{\text{MAX}}\max_{1\le k\le K}|(\beta^\intercal\bar{u})_k|\lesssim_{\text{P}}\max_{1\le k\le K}N^{-1}|(\beta^\intercal\bar{u})_k|\lesssim_{\text{P}} 1/\sqrt{TN}.$$

Here the last inequality comes from that K is fixed, $\text{E}(\bar{u}_i\bar{u}_j|\beta,\Sigma_u)\lesssim\delta_{i,j}\sigma_i^2T^{-1}$ by conditions (a) and (b) of Assumption 1, and $\lambda_{\max}(\Sigma_u)\lesssim_{\text{P}} 1$.

Finally, we write

$$\max_{1\le i\le N}|(\mathbb{P}_\beta\alpha)_i|\le C\|\beta\|_{\text{MAX}}\|(\beta^\intercal\beta)^{-1}\|_{\text{MAX}}\max_{1\le k\le K}|(\beta^\intercal\alpha)_k|\le CN^{-1}\max_{1\le k\le K}|(\beta^\intercal\alpha)_k|. \quad \text{(D.7)}$$

On the other hand, from condition (a) and (b) of Assumption 1 and condition (a) of Assumption 2, we have $\text{E}(\alpha_i\alpha_j|\beta,\Sigma_u)=\delta_{i,j}\sigma_i^2\text{E}(s_i^2)$. Therefore, as K is fixed and $\lambda_{\max}(\Sigma_u)\lesssim_{\text{P}} 1$, we have

$$\max_{1\le k\le K}|(\beta^\intercal\alpha)_k|\lesssim_{\text{P}} c_NN^{1/2}\|\beta\|_{\text{MAX}}\lesssim_{\text{P}} N^{1/2}\text{E}(s_i^2)^{1/2}. \quad \text{(D.8)}$$

Substituting (D.8) into (D.7), we obtain (D.3). ∎

Lemma D2. *Suppose Assumptions 1 and 2 hold and that* $\|\beta\|_{\text{MAX}}^2\lesssim_{\text{P}} 1$ *and* $\lambda_{\min}(\beta^\intercal\beta)\gtrsim_{\text{P}} N$. *Also assume* $N^d\lesssim T\lesssim N^{d'}$ *with fixed* $d>0$ *and* $d'<1$. *Then it holds that, as* $N,T\to\infty$,

$$\max_{1\le i\le N}|\widehat{s}_i-\check{s}_i|\lesssim_{\text{P}} c_NT^{-1/2}\widetilde{k}_N. \quad \text{(D.9)}$$

If we additionally have $T \gtrsim N^d$ with fixed $d > 1/2$, then it holds that, as $N, T \to \infty$,

$$\max_{i \in B} |\widehat{s}_i - \check{s}_i| \lesssim_{\mathrm{P}} \chi_N. \tag{D.10}$$

Here $\chi_N \coloneqq \sqrt{1/N}(k_N^5 + \mathrm{E}(s_j^2)^{1/2})$, and set B is $B \coloneqq \{i \in N : T^{1/2}|\check{s}_i| \leq \widetilde{k}_N\}$, with $\widetilde{k}_N \coloneqq k_N^{-2}$.

Proof. By definition we have

$$\widehat{s}_i - \check{s}_i = -\frac{(\mathbb{P}_\beta \alpha)_i}{\widehat{\sigma}_i} - \frac{(\mathbb{P}_\beta \bar{u})_i}{\widehat{\sigma}_i} + \left(\frac{\sigma_i}{\widehat{\sigma}_i} - 1\right)\check{s}_i. \tag{D.11}$$

Since $T \gtrsim N^d$ with $d > 0$ by assumption, (D.1) of Lemma D1 leads to $\max_{1 \leq i \leq N} |\widehat{\sigma}_i^2 - \sigma_i^2| = o_{\mathrm{P}}(1)$. Then, noting $\min_i \sigma_i \gtrsim_{\mathrm{P}} 1$ by condition (d) of Assumption 1, we obtain $\min_i \widehat{\sigma}_i \gtrsim_{\mathrm{P}} 1$. Applying (D.1) again, we have $\max_i \left|\frac{\sigma_i}{\widehat{\sigma}_i} - 1\right| \lesssim_{\mathrm{P}} \sqrt{(\log N)/T}$. Using these two results, and substituting (D.2) and (D.3) of Lemma D1 into (D.11), we obtain

$$\max_{1 \leq i \leq N} |\widehat{s}_i - \check{s}_i| \lesssim_{\mathrm{P}} \chi_N + \sqrt{(\log N)/T} \max_{1 \leq i \leq N} |\check{s}_i|, \qquad \max_{i \in B} |\widehat{s}_i - \check{s}_i| \lesssim_{\mathrm{P}} \chi_N + \sqrt{(\log N)/T} \max_{i \in B} |\check{s}_i|. \tag{D.12}$$

Since $\mathrm{P}(|s_i| \geq 1) \leq \mathrm{E}(s_i^2 \mathbb{1}_{\{|s_i| \geq 1\}}) \leq c_N N^{-1}$ by condition (a) of Assumption 2, we have $\mathrm{P}(\max_i |s_i| \geq 1) \leq c_N$. Combining this result with $\max_i |\bar{\varepsilon}_i| \lesssim_{\mathrm{P}} \sqrt{(\log N)/T}$ by the uniform bound on i.i.d. normal variables, we obtain $\max_{1 \leq i \leq N} |\check{s}_i| \lesssim_{\mathrm{P}} 1$ (again noting $T \gtrsim N^d$ with $d > 0$ by assumption). Then we have $\sqrt{(\log N)/T} \max_{1 \leq i \leq N} |\check{s}_i| \lesssim_{\mathrm{P}} T^{-1/2}\sqrt{\log N}$. Also, we have $\chi_N \leq c_N T^{-1/2}$ since $T = o(N)$ by assumption and $\mathrm{E}(s_j^2) \lesssim 1 + \mathrm{E}(s_i^2 \mathbb{1}_{\{|s_i| \geq 1\}}) \lesssim 1$ by condition (a) of Assumption 2. Substituting the two bounds into the first part of (D.12), we achieve (D.9).

On the other hand, the definition of set B leads to $\max_{i \in B} |\check{s}_i| \leq T^{-1/2}\widetilde{k}_N$. Then, noting $T \gtrsim N^d$ with $d > 1/2$ by assumption, we have $\sqrt{(\log N)/T} \max_{i \in B} |\check{s}_i| \lesssim \sqrt{1/N} k_N^5 \leq \chi_N$. Given the second part of (D.12), we obtain (D.10). ∎

Lemma D3. *Suppose Assumptions 1 and 2 hold and that $\|\beta\|_{\mathrm{MAX}}^2 \lesssim_{\mathrm{P}} 1$ and $\lambda_{\min}(\beta^\intercal \beta) \gtrsim_{\mathrm{P}} N$. $N^d \lesssim T \lesssim N^{d'}$ with fixed $d > 1/2$ and $d' < 1$. Then it holds that, as $N, T \to \infty$,*

$$\mathrm{P}(p(\check{s}_i) \geq T^{1/2} N^{-3/2}, \forall i \leq N) \geq 1 - c_N. \tag{D.13}$$

Proof. Note that when $|x| < 1$, we can find $C > 1$ such that $a \geq C$ implies $|a - x| \geq C - 1$. Therefore, for $|x| < 1$, we have

$$\int_{|a| \geq C} \phi_{1/T}(x - a) da \leq \int_{|a| \geq C-1} \sqrt{T} \exp(-Ta^2/2) da \lesssim \exp(-T) \leq c_N N^{-1}. \tag{D.14}$$

The last step comes from $T \gtrsim N^d$ for some $d > 1/2$ by assumption. Then we can bound

$$\begin{aligned}\int_{|a|\geq C} p(a)da &\leq \int_{|x|\geq 1} p_s(x)dx + \int_{|x|<1}\int_{|a|\geq C} \phi_{1/T}(x-a)da p_s(x)dx \\ &\leq \int_{|x|\geq 1} p_s(x)dx + \sup_{x:|x|<1}\int_{|a|\geq C} \phi_{1/T}(x-a)da \leq c_N N^{-1}. \qquad \text{(D.15)}\end{aligned}$$

Here the last inequality comes from (D.14) and $\int_{|x|\geq 1} p_s(x)dx \leq \mathrm{E}(s_i^2 \mathbb{1}_{\{|s_i|\geq 1\}}) \leq \mathrm{E}(s_i^2 \mathbb{1}_{\{|s_i|\geq c_N\}}) \leq c_N N^{-1}$ by condition (a) of Assumption 2. It follows from (D.15) that

$$\begin{aligned}\mathrm{P}(p(\check{s}_i) < T^{1/2}N^{-3/2}) &= \int \mathbb{1}_{\{p(a)<T^{1/2}N^{-3/2}\}} p(a)da \\ &\leq c_N N^{-1} + \int_{|a|<C} \mathbb{1}_{\{p(a)<T^{1/2}N^{-3/2}\}} p(a)da \leq c_N N^{-1}. \qquad \text{(D.16)}\end{aligned}$$

The last inequality also uses $T = o(N)$ by assumption. (D.16) proves the lemma by Bonferroni inequalities. ∎

Lemma D4. *It holds that, for $j \in \{0,1\}$ and for all $(a,\bar{a})$ satisfying $|\bar{a}-a| \leq 1/\sqrt{6\log N}$, as $N \to \infty$,*

$$|a^j\phi(a) - \bar{a}^j\phi(\bar{a})| \lesssim (\log N)^{(j+1)/2}|\bar{a}-a|\phi(a) + c_N N^{-2}, \qquad \text{(D.17)}$$

$$\phi(\bar{a}) \lesssim (\log N)^{(j+1)/2}(1+|a|)^{-(j+1)}(\phi(a) + c_N N^{-2}). \qquad \text{(D.18)}$$

Proof. We first write that, for all a and for $j \in \{0,1\}$,

$$\begin{aligned}|a^j\phi(a) - \bar{a}^j\phi(\bar{a})| &\leq |\bar{a}^j - a^j|\phi(a) + (|\bar{a}^j - a^j| + |a|^j)|\phi(\bar{a}) - \phi(a)| \\ &\leq |\bar{a}-a|\phi(a) + (|\bar{a}-a| + |a|^j)\phi(a)|e^{-(a^2-\bar{a}^2)/2} - 1|.\end{aligned}$$

On the other hand, for all sequence b_N satisfying $b_N \geq 1$, and for all $(a,\bar{a})$ satisfying $|a| \leq b_N$ and $|\bar{a}-a| \leq b_N^{-1}$, we have $|e^{-(a^2-\bar{a}^2)/2} - 1| \lesssim |\bar{a}-a|b_N$. As a result, for all such b_N and $(a,\bar{a})$, it holds that, for $j \in \{0,1\}$,

$$|a^j\phi(a) - \bar{a}^j\phi(\bar{a})| \lesssim b_N^{j+1}|\bar{a}-a|\phi(a). \qquad \text{(D.19)}$$

Moreover, it holds that, for $j \in \{0,1,2\}$ and for all b_N' that satisfies $b_N' \geq \sqrt{5\log N}$,

$$\sup_{a:|a|\geq b_N'} |a^j\phi(a)| \leq c_N N^{-2}. \qquad \text{(D.20)}$$

Then, choosing $b_N = \sqrt{6\log N}$, which ensures $b_N \geq \sqrt{5\log N}$ and $b_N - b_N^{-1} \geq \sqrt{5\log N}$ for

large N, we obtain (D.17) by combining (D.19) and (D.20). Further, (D.18) would directly come from (D.17) by choosing $j = 0$ when $|a| \leq \sqrt{6 \log N}$, which guarantees $(\log N)(1 + |a|)^{-2} \gtrsim 1$. For $|a| \geq \sqrt{6 \log N}$, we obtain (D.18) from (D.20) by choosing $j = 2$, because in this case we have $|\bar{a}| \geq \sqrt{5 \log N}$ for large N, and $1 + |a| \sim 1 + |\bar{a}|$. The proof ends. ∎

Appendix E Lemmas supporting Section C.7

Lemma E1. *Suppose that r_t follows (1), Assumption 1 holds, $u_{i,t} \approx \mathcal{N}(0, \sigma^2)$, and α following (5) as in Example 1. Also assume $N^{-d'} \leq \rho \leq N^{-d}$ with fixed $d' > d > 0$. Recall we define $p(a)$ and $\psi(a)$ in Corollary 2 and $\Phi(x)$ is the standard normal cumulative distribution function. Then it holds that (i) when $\zeta - \zeta^* \lesssim -\sqrt{\log N}$,*

$$\int \psi(a)^2 p(a) da \asymp (1 \wedge \zeta^2) \mu^2 \rho \iota \Phi(a^* - 2\zeta);$$

(ii) when $\zeta - \zeta^* \geq -c_N \sqrt{\log N}$,

$$\int \psi(a)^2 p(a) da \asymp \mu^2 \rho \Phi(\zeta - a^*).$$

Here $\iota := e^{3\zeta^2/2 - \zeta a^*}$, ζ^* and $\chi(a)$ are defined by $\zeta^* := \sqrt{-2 \log(\rho/2)}$ and $\chi(a) := \frac{\rho}{2} \phi(a - \zeta)/\phi(a)$, and a^* is the solution of $\chi(a) = 1$.

Proof. It follows from definition that

$$a^* = \frac{\zeta}{2} + \frac{\zeta^{*2}}{2\zeta} \quad \text{and} \quad \chi(a) = \exp((a - a^*)\zeta). \tag{E.1}$$

By definition we have

$$T^{-1/2} p(T^{-1/2} a) = (1 - \rho)\phi(a) + \frac{\rho}{2}\phi(|a| - \zeta) + \frac{\rho}{2}\phi(|a| + \zeta).$$

Because $\phi(|a| + \zeta) \leq \phi(a)$ and $\rho \to 0$ we have, uniformly over $a \geq 0$,

$$T^{-1/2} p(T^{-1/2} a) = (1 + o(1)) \frac{\rho}{2} \phi(a - \zeta)(1 + \chi(a)^{-1}). \tag{E.2}$$

We also note $\phi(a + \zeta) = \exp(-2a\zeta)\phi(a - \zeta)$. Therefore, uniformly over $a \geq 0$,

$$\psi(T^{-1/2} a) = (1 + o(1)) \mu \frac{1 - \exp(-2a\zeta)}{1 + \chi(a)^{-1}}. \tag{E.3}$$

As a result of (E.2) and (E.3), it holds that

$$\int \psi(a)^2 p(a) da = (1+o(1))\mu^2 \rho B^*, \quad \text{with} \quad B^* \coloneqq \int_0^\infty (1-\exp(-2a\zeta))^2 \frac{\chi(a)}{1+\chi(a)} \phi(a-\zeta) da. \tag{E.4}$$

Then the goal becomes characterization of B^* across different cases. For this purpose, we introduce two set sequences $A_1 = (0, a^*)$ and $A_2 = (a^*, \infty)$. Based on these sets, we define, for $j \in \{1, 2\}$,

$$B_j = \int_{A_j} (1-\exp(-2a\zeta))^2 \frac{\chi(a)}{1+\chi(a)} \phi(a-\zeta) da, \quad B_j' \coloneqq \int_{A_j} (1-\exp(-2a\zeta))^2 \phi(a-2\zeta) da. \tag{E.5}$$

It apparently holds that $B^* = B_1 + B_2$. Furthermore, from the first part of (E.1), we note $\zeta a^* \to \infty$ and thereby $\sup_{a \geq a^*} \exp(-2a\zeta) \to 0$. From the second part of (E.1), $\chi(a) \leq 1$ for all $a \leq a^*$ and $\chi(a) \geq 1$ for all $a \geq a^*$. Hence, it follows from (E.5) that

$$B_1 \eqsim \int_0^a (1-\exp(-2a\zeta))^2 \chi(a) \phi(a-\zeta) da, \quad B_2 \eqsim \int_{a^*}^\infty \phi(a-\zeta) da. \tag{E.6}$$

We note $\chi(a)\phi(a-\zeta) = \iota\phi(a-2\zeta)$. Applying this relation to the two parts of (E.6), we obtain, respectively,

$$B_1 \eqsim \iota B_1', \quad B_2 \lesssim \iota B_2'. \tag{E.7}$$

where, for the second result, we use that $\inf_{a \geq a^*}(1-\exp(-2a\zeta))^2 \chi(a) \gtrsim 1$ due to $\inf_{a \geq a^*} \chi(a) \geq 1$ and $\inf_{a \geq a^*} a\zeta \to \infty$ by the first part of (E.1).

Now we gauge the magnitude of B_1 and B_2 for cases (i) and (ii). We further separate case (i) into two subcases: case (ia) where $\zeta \lesssim 1$ (under which $\zeta - \zeta^* \lesssim -\sqrt{\log N}$ always holds, as $\zeta^* \gtrsim \sqrt{\log N}$ by the assumption $\rho \leq N^{-d}$ with fixed $d > 0$), and case (ib) where $\zeta \to \infty$ and $\zeta - \zeta^* \lesssim -\sqrt{\log N}$. We start with case (ia) where $\zeta \lesssim 1$. We note

$$\phi(a-2\zeta)\exp(-2a\zeta) = \phi(a)\exp(-2\zeta^2), \quad \phi(a-2\zeta)\exp(-4a\zeta) = \phi(a+2\zeta). \tag{E.8}$$

Using (E.8), we obtain

$$B_1' + B_2' = \int (1-\exp(-2a\zeta))^2 \phi(a-2\zeta) da = 1 - \exp(-2\zeta^2). \tag{E.9}$$

Moreover, it apparently holds

$$B_2' \leq \int_{A_2} \phi(a-2\zeta) da = \Phi(2\zeta - a^*). \tag{E.10}$$

On the other hand, when $\zeta \lesssim 1$, we have $a^* - 2\zeta \geq \zeta^{*2}/(3\zeta)$ for N sufficiently large. Then, when $\zeta \lesssim 1$

$$\zeta^{-2}\Phi(2\zeta - a^*) \lesssim \frac{1}{\zeta^{*4}}\frac{\zeta^{*2}}{\zeta}\exp(-\zeta^{*4}/(18\zeta^2)) \to 0. \tag{E.11}$$

where the last convergence comes from $\zeta^* \to \infty$ (by the assumption $\rho \le N^{-d}$ with fixed $d > 0$), $\zeta^{*2}/\zeta \to \infty$ (as $\zeta \lesssim 1$), and $\lim_{x\to\infty} x\exp(-x^2) = 0$. Since $1 - \exp(-2\zeta^2) \asymp \zeta^2$ under $\zeta \lesssim 1$, combining (*E.9*), (E.10), and (E.11), we obtain that, when $\zeta \lesssim 1$,

$$B_1' \asymp \zeta^2, \quad B_2' = o(\zeta^2). \tag{E.12}$$

Substituting (E.12) into (E.7), we obtain (recall $B^* = B_1 + B_2$) that, when $\zeta \lesssim 1$,

$$B^* \asymp \zeta^2 e^{-\zeta a^*}. \tag{E.13}$$

(E.4) and (E.13) together lead to $\int \psi(a)^2 p(a) da \asymp \mu^2\zeta^2\rho e^{-\zeta a^*}$. Given $\iota \asymp e^{-\zeta a^*}$ and $\Phi(a^* - 2\zeta) \asymp 1$ (by the first part of (E.1)) when $\zeta \lesssim 1$, we prove the lemma for case (ia).

Next, we study cases (ib) and (ii). We note in both cases we have $\zeta \to \infty$. Using (E.8) and $\Phi(x) = 1 - \Phi(-x)$, we obtain by direct calculation

$$B_1' = \int_0^{a^*} \phi(a-2\zeta)da = \Phi(a^*-2\zeta) - 2\Phi(-2\zeta) + \Phi(-a^*-2\zeta) - \exp(-2\zeta^2)(2\Phi(a^*)-1). \tag{E.14}$$

We note that, uniformly over x,

$$\Phi(-x) \asymp \frac{1}{1+x_+}\phi(x_+), \tag{E.15}$$

where $x_+ = \max\{x, 0\}$. Using (E.15), we obtain $\Phi(a^* - 2\zeta) \ge \Phi(-3\zeta/2) \asymp \frac{1}{1+\zeta}\phi(3\zeta/2) \asymp \frac{1}{1+\zeta}\exp(-9\zeta^2/8)$. Then we have, $\Phi(-2\zeta) = o(\Phi(a^* - 2\zeta))$ and $\exp(-2\zeta^2) = o(\Phi(a^* - 2\zeta))$ when $\zeta \to \infty$. Since $\Phi(-a^* - 2\zeta) \le \Phi(-2\zeta)$ and $0 \le \Phi(a^*) \le 1$, we obtain from (E.14) that, when $\zeta \to \infty$,

$$B_1' = (1 + o(1))\Phi(a^* - 2\zeta). \tag{E.16}$$

Substituting (E.16) into the first part of (E.7), and noting the second part of (E.6), we obtain that, when $\zeta \to \infty$,

$$B_1 \asymp \iota\Phi(a^* - 2\zeta), \quad B_2 \asymp \Phi(\zeta - a^*). \tag{E.17}$$

We first investigate case (ib), in which it holds that $a^* - \zeta \gtrsim \sqrt{\log N}$ according to the first part of (E.1) and $\zeta^* \asymp \sqrt{\log N}$ by the assumption on ρ. Then from (E.17) and (E.15), it follows that, in case (ib),

$$B_1 \gtrsim \frac{1}{1 + (2\zeta - a^*)_+}\phi(a^* - \zeta), \quad B_2 \asymp \frac{1}{1 + (a^* - \zeta)_+}\phi(a^* - \zeta). \tag{E.18}$$

Here the first part comes from $\phi(x_+) \geq \phi(x)$ for all x, and $\phi(2\zeta - a^*)e^{3\zeta^2/2-\zeta a^*} = \phi(a^* - \zeta)$. The second part comes from $a^* - \zeta \gtrsim \sqrt{\log N}$. Since $1 + (2\zeta - a^*)_+ \lesssim (a^* - \zeta)_+$ according to the first part of (E.1) and $\zeta^* \asymp \sqrt{\log N}$ by the assumption on ρ, we obtain from (E.18) that, in case (ib),

$$B_2 \lesssim B_1. \tag{E.19}$$

Now we study case (ii), in which it holds that $a^* - \zeta \leq c_N\sqrt{\log N}$ by the first part of (E.1) and $\zeta^* \asymp \sqrt{\log N}$. Hence, we have $2\zeta - a^* \gtrsim 1$. Therefore, from (E.17) and (E.15), it follows that, in case (ii),

$$B_1 \asymp \frac{1}{1 + (2\zeta - a^*)_+}\phi(a^* - \zeta), \quad B_2 \gtrsim \frac{1}{1 + (a^* - \zeta)_+}\phi(a^* - \zeta). \tag{E.20}$$

Here the first part comes from $2\zeta - a^* \gtrsim 1$ and $\phi(2\zeta - a^*)e^{3\zeta^2/2-\zeta a^*} = \phi(a^* - \zeta)$. The second part comes from $\phi(x_+) \geq \phi(x)$ for all x. Since $1 + (a^* - \zeta)_+ = o((2\zeta - a^*)_+)$, we obtain from (E.20) that, in case (ii),

$$B_1 = o(B_2). \tag{E.21}$$

Combining (E.19) and (E.21) with (E.17), we prove the lemma for cases (ib) and (ii). The proof concludes. ∎

Lemma E2. *Suppose a real-valued function sequence $h_N : R \to R$, a sequnce a_N, and a positive sequence Δ_N satisfy $\Delta_N^{-1/2}(\mathbb{1}_{A_N} - h_N) \to 0$ in L^2 with $A_N \coloneqq (a_N, a_N + \Delta_N)$. Then there exists a set sequence $\bar{A}_N$ such that*

$$\int_{A_N - \bar{A}_N} da = o(\Delta_N) \quad \textit{and} \quad \sup_{a \in A'_N} |h_N(a) - 1| = o(1).$$

Proof. We define $h_N^*(a) = h_N(a\Delta_N^{-1})$ and $A_N^* = (a_N\Delta_N^{-1}, a_N\Delta_N^{-1} + 1)$. Then clearly $(\mathbb{1}_{A_N^*} - h_N^*) \to 0$ in L^2. This further leads to $h_N^* \to \mathbb{1}_{A_N^*}$ in measure (see, e.g., Theorem 2.15.a of Folland (2009)), i.e., for every $\epsilon > 0$, $\int_{A_N^*} \mathbb{1}_{\{|1-h_N^*(a)|\geq\epsilon\}}da \to 0$. This is equivalent to that, for every $\epsilon > 0$,

$$\Delta_N^{-1}\int_{A_N} \mathbb{1}_{\{|1-h_N(a)|\geq\epsilon\}}da \to 0. \tag{E.22}$$

We hence prove the lemma. ∎

Lemma E3. *Suppose the same assumptions as in Lemma E1 and let a^* be as defined therein. Then there exists a deterministic sequence $\lambda \geq 0$ such that $\mathrm{Corr}(\psi_i, \widetilde{\psi}_{q,i}(\lambda)) \to 1$ if and only if, for some $c_N \to 0$,*

$$\begin{cases} (\mu,\rho) \in \left\{(\mu,\rho) : \zeta \le c_N \, or \, \zeta - a^* \ge c_N^{-1}\right\} & for \; q = 1, \\ (\mu,\rho) \in \left\{(\mu,\rho) : \zeta \le c_N \, or \, \zeta - a^* \ge -c_N\sqrt{\log N}\right\} & for \; q = 2. \end{cases} \tag{E.23}$$

Moreover, if sequence $\lambda \ge 0$ *is determnistic,* $\mathrm{Corr}(\psi_i, \widetilde{\psi}_{q,i}(\lambda)) \to 1$ *if and only if*

$$\begin{cases} \lambda \to 0, & when \; \zeta \to 0 and \; q \in \{1,2\}, \\ \zeta - \lambda \to \infty and \; \frac{\phi(\lambda)}{\rho(1+\lambda^2)(\zeta-\lambda)^2} \to 0, & when \; \zeta - a^* \to \infty and \; q = 1. \end{cases} \tag{E.24}$$

Further, when $\zeta - a^* \ge -c_N\sqrt{\log N}$*, it holds that*

$$\mathrm{Corr}(\psi_i, \widetilde{\psi}_{q,i}(\lambda)) \ge 1 + o(1) + O\left(\frac{\Phi(-\lambda)}{\rho\Phi(\zeta - a^*)}\right). \tag{E.25}$$

Proof. Step 1. As a result of (E.2) and (E.3), it holds that, by the symmetry of $\psi(a)$, $\widetilde{\psi}_q(a,\lambda)$, and $p(a)$,

$$\int \psi(a)^2 p(a) da = (1+o(1))\mu^2\rho \int_0^\infty (1-\exp(-2a\zeta))^2 \frac{\chi(a)}{1+\chi(a)} \phi(a-\zeta) da, \tag{E.26}$$

$$\int \widetilde{\psi}_q(a,\lambda)^2 p(a) da = (1+o(1))\rho \int_0^\infty \widetilde{\psi}_q(a,\lambda)^2 \frac{1+\chi(a)}{\chi(a)} \phi(a-\zeta) da, \tag{E.27}$$

$$\int \psi(a)\widetilde{\psi}_q(a,\lambda) p(a) da = (1+o(1))\mu\rho \int_0^\infty \widetilde{\psi}_q(a,\lambda)(1-\exp(-2a\zeta))\phi(a-\zeta) da. \tag{E.28}$$

To facilitate the exposition, we introduce short-hand notation

$$f(a) \coloneqq \sqrt{\frac{\chi(a)}{1+\chi(a)}\phi(a-\zeta)}\mathbb{1}_{\{a\ge 0\}}, \quad g(a) = (1-e^{-2a\zeta})f(a), \quad \widetilde{g}_q(a) = \widetilde{\psi}_q(a,\lambda)\frac{1+\chi(a)}{\chi(a)} f(a).$$

Comparing the definitions of f and g with the right-hand sides of (E.26), (E.27), and (E.28), we obtain, for $q \in \{1,2\}$,

$$\mathrm{Corr}(\psi_i, \widetilde{\psi}_{q,i}(\lambda)) = (1+o(1))\theta_q(\lambda), \quad \theta_q(\lambda) \coloneqq \frac{\langle g, \widetilde{g}_q\rangle}{\|g\|\|\widetilde{g}_q\|}, \tag{E.29}$$

where $\langle g, \widetilde{g}_q\rangle \coloneqq \int g(a)\widetilde{g}_q(a) da$ and $\|g\| \coloneqq \sqrt{\langle g, g\rangle}$ stand for the L^2-inner product and L^2-norm. Moreover, we note that $1 - \theta_q(\lambda) = \frac{1}{2}\frac{1}{\|g\|^2}\left|g - \frac{\|g\|}{\|\widetilde{g}_q\|} \times \widetilde{g}_q\right|^2$. Let d^* be the scalar that minimizes $\|g - d \times \widetilde{g}_q\|$. It obviously holds that, for $q \in \{1,2\}$,

$$1 - \theta_q(\lambda) \ge \frac{1}{2\|g\|^2}\|g - d^* \times \widetilde{g}_q\|^2. \tag{E.30}$$

Moreover, we have that, if $\theta_q(\lambda) \geq 0$ and for $q \in \{1,2\}$,

$$1 - \theta_q(\lambda) \leq 1 - \theta_q(\lambda)^2 = \frac{1}{\|g\|^2}\|g - d^* \times \widetilde{g}_q\|^2. \tag{E.31}$$

Given (E.29), (E.30), and (E.31), the proof will focus on evaluating $\|g - d^* \times \widetilde{g}_q\|$. In addition, given the definition of $\widetilde{\psi}(a,\lambda)$, we can write, for $q \in \{1,2\}$,

$$\|g - d^* \times \widetilde{g}_q\| = \int (1 - e^{-2a\zeta} - d^* \times h_q(a))^2 f(a)^2 da, \quad \text{with} \quad h_q(a) = (e^{(a^*-a)\zeta} + 1)\widetilde{\psi}_q(a,\lambda). \tag{E.32}$$

Further, to characterize the magnitude of $\|g\|$, we consider, resepectively, case (ia) where $\zeta \to 0$ (under which $\zeta - \zeta^* \lesssim -\sqrt{\log N}$ always holds, as $\zeta^* \gtrsim \sqrt{\log N}$ by the assumption $\rho \leq N^{-d}$ with fixed $d > 0$), case (ib) where $\zeta \gtrsim 1$ and $\zeta - \zeta^* \lesssim -\sqrt{\log N}$, and case (ii) where $\zeta - \zeta^* \geq -c_N\sqrt{\log N}$. Here and below we adopt the notation introduced in the statement of Lemma E1. Since B^* is simply $\|g\|^2$ here, it follows from Lemma E1 that

$$\|g\|^2 \asymp \begin{cases} (1 \wedge \zeta^2)\iota\Phi(a^* - 2\zeta), & \text{cases (ia) and (ib);} \\ \Phi(\zeta - a^*), & \text{case (ii).} \end{cases} \tag{E.33}$$

By the classic subsequence argument, we can establish the lemma as long as we prove it under each of cases (ia), (ib), and (ii).

Step 2. We start with demonstrating that, for each $q \in \{0,1\}$, $\theta_q(\lambda) \to 1$ holds in case (ia) if and only if we choose $\lambda \to 0$. Suppose $\lambda \to 0$ and let $\bar{d}^* = 2\zeta e^{-a^*\zeta}$. We obtain that, for each $q \in \{1,2\}$ and for all $a \geq 0$,

$$\begin{aligned} |1 - e^{-2a\zeta} - \bar{d}^* \times h_q(a)| &= |1 - e^{-2a\zeta} - 2\zeta\psi_q(a,\lambda)(e^{-a\zeta} + e^{-a^*\zeta})| \\ &\leq |1 - e^{-2a\zeta} - 2\zeta a| + 2\zeta\lambda + 2\zeta a(|1 - e^{-a\zeta}| + e^{-a^*\zeta}) \\ &\leq 4\zeta^2 a^2 + 2\zeta\lambda + 2e^{-a^*\zeta}\zeta a. \end{aligned} \tag{E.34}$$

The second inequality comes from that $|\psi_q(a,\lambda)| \leq a$ and $|\psi_q(a,\lambda) - a| \leq \lambda$ for all $a \geq 0$. The last inequality comes from that $-\frac{1}{2}x^2 \leq 1 - e^{-x} - x \leq 0$ and $0 \leq 1 - e^{-x} \leq x$ for all $x \geq 0$. As a result, we have, for each $q \in \{1,2\}$,

$$\begin{aligned} \|g - d^* \times \widetilde{g}_q\|^2 &\leq \|g - \bar{d}^* \times \widetilde{g}_q\|^2 \lesssim \iota \int_0^\infty (\zeta^4 a^4 + \zeta^2\lambda^2 + e^{-2a^*\zeta}\zeta^2 a^2)\phi(a - 2\zeta)da \\ &\lesssim \iota(\zeta^4 + \zeta^2\lambda^2 + e^{-2a^*\zeta}\zeta^2) = o(\iota\zeta^2). \end{aligned} \tag{E.35}$$

Here the first inequality holds by the definition of d^*. The second inequality comes from (E.32), (E.34), and $f(a)^2 \leq \chi(a)\phi(a - \zeta) = \iota\phi(a - 2\zeta)$. The third inequality comes from that

$\int a^j\phi(a-2\zeta)da = \int(a+2\zeta)^j\phi(a)da \lesssim \int(a^j+\zeta^j)\phi(a)da \lesssim 1$ for $j \in \{2,4\}$ as $\zeta \to 0$. The last inequality comes from $\zeta \to 0$, $a^*\zeta \to \infty$, and $\lambda \to 0$.

Now suppose $\lambda \gtrsim 1$. Then it holds that, for each $q \in \{1,2\}$,

$$\|g - d^* \times \widetilde{g}_q\|^2 \geq \int_0^{\lambda\wedge 1} (1-e^{-2a\zeta})^2 f(a)^2 da \gtrsim \iota\zeta^2 \int_0^{\lambda\wedge 1} a^2\phi(a-2\zeta) \gtrsim \iota\zeta^2. \tag{E.36}$$

The first inequality comes from $\widetilde{g}_q(a) = 0$ for all $a \in (0,\lambda)$. The second inequality comes from $\inf_{a\leq 1} f(a)^2 \geq \frac{1}{2}\chi(a)\phi(a-\zeta)$. According to (E.33), we have $\|g\|^2 \eqsim \iota\zeta^2$ as $\Phi(a^*-2\zeta) \eqsim 1$ (by the first part of (E.1)) when $\zeta \to 0$. Given (E.29), we prove that for each $q \in \{1,2\}$, $\mathrm{Corr}(\psi_i, \widetilde{\psi}_{q,i}(\lambda)) \to 1$ if (by (E.31) and (E.35)) and only if (by (E.30) and (E.36)) $\lambda \to 0$ when we are in case (ia), i.e., when we have $\zeta \to 0$. We hence establish the first part of (E.24).

Step 3. Next, we show that, for each $q \in \{1,2\}$, $\theta_q(\lambda) \to 1$ does not hold for any λ sequence in case (ib). Note that we always have $\zeta \gtrsim 1$ and $\zeta - a^* \lesssim 1$ in case (ib), due to $\zeta - a^* \leq \zeta - \zeta^*$ by the first part of (E.1). We introduce a set sequence $A := ((2\zeta)\wedge a^* - 4\Delta, (2\zeta)\wedge a^*)$, with $\Delta := \frac{1}{1+(2\zeta-a^*)_+} \wedge (\zeta/2)$. Because $\phi(x)' = -x\phi(x)$, it holds that, for any real sequence x and any fixed positive constant C,

$$\phi(x) \lesssim \inf_{a:|a-x|\leq \frac{C}{1+|x|}} \phi(a) \leq \sup_{a:|a-x|\leq \frac{C}{1+|x|}} \phi(a) \lesssim \phi(x). \tag{E.37}$$

As a result, we have $\inf_{a\in A} f(a)^2 \gtrsim \iota\phi((2\zeta)\wedge a^* - 2\zeta)$ (by (E.37), the relation right after (E.6), the definition of f, and $A \subset (0,\infty)$), we have from (E.32) that, for $q \in \{1,2\}$,

$$\|g - d^* \times \widetilde{g}_q\|^2 \gtrsim \iota\phi((2\zeta)\wedge a^* - 2\zeta) \int_A (1 - e^{-2a\zeta} - d^* \times h_q(a))^2 da. \tag{E.38}$$

According to (E.33) and (E.15), $\|g\|^2 \eqsim \iota\phi((2\zeta)\wedge a^* - 2\zeta)\Delta$ in case (ib). Then, combining (E.30) and (E.38), we obtain that, if $\theta(\lambda) \to 1$, then, for $q \in \{1,2\}$,

$$\int_A (1 - e^{-2a\zeta} - d^* \times h_q(a))^2 da = o(\Delta). \tag{E.39}$$

It follows from Lemma E2 and (E.39) that, for each $q \in \{1,2\}$, if $\theta_q(\lambda) \to 1$, there exists a set sequence $\bar{A}$ such that,

$$\int_{A-\bar{A}} da = o(\Delta) \quad \text{and} \quad \sup_{a\in\bar{A}} |d^* \times h_q(a) - (1-e^{-2a\zeta})| = o(1). \tag{E.40}$$

It is almost obvious that (E.40) can not hold for any $q \in \{1,2\}$ and for any λ, given the drastic difference between the definition of $h_q(a)$ and $1 - e^{-2a\zeta}$. To see this more clearly,

we first consider the situation where $\zeta \asymp 1$. We note that, under $\zeta \asymp 1$ and uniformly over $a \in A$ (which becomes $(2\zeta - 4\Delta, 2\zeta)$ for large N), $h_q(a) = (1 + o(1))e^{a^*\zeta}\widetilde{\psi}_q(a, \lambda)$. Apply the triangular inequality to (E.40) leads to $\sup_{a\in\bar{A}} |d^* e^{a^*\zeta} \times \widetilde{\psi}_q(a, \lambda) - (1 - e^{-2a\zeta})| = o(1)$, which can not hold for any $q \in \{1, 2\}$ and for any λ sequence, because of the nonlinearity of $1 - e^{-2a\zeta}$ in a as $\zeta \asymp 1$.

Then we consider the situation where $\zeta \to \infty$. It is trivial from (E.39) and that $\widetilde{\psi}_q(a, \lambda) = 0$ for $a \le \lambda$ that $\lambda \le (2\zeta) \wedge a^* - 7\Delta/2$ for large N if $\theta_q(\lambda) \to 1$, for each $q \in \{1, 2\}$. As a direct result, we have, by the definition of $h_q(a)$, that, for each $q \in \{1, 2\}$, if $\theta_q(\lambda) \to 1$ and $\zeta \to \infty$,

$$\frac{\inf_{a\in A^+} h_q(a)}{\sup_{a\in A^-} h_q(a)} \gtrsim \exp(\zeta\Delta) \to \infty, \tag{E.41}$$

where $A^+ := ((2\zeta) \wedge a^* - 3\Delta, (2\zeta) \wedge a^* - 2\Delta)$ and $A^- := ((2\zeta) \wedge a^* - \Delta, (2\zeta) \wedge a^*)$. (E.41) and (E.40) clearly contradict as, for large N, any $\bar{A}$ satisfying the first part of (E.40) would overlap with both A^+ and A^-, and hence would violate the second part of (E.40) according to (E.41). In other words, $\theta_q(\lambda) \to 1$ does not hold under any λ sequence and for any $q \in \{1, 2\}$ if we are in case (ib).

Step 4. Now we demonstrate that, for each $q \in \{1, 2\}$, $\theta_q(\lambda) \to 1$ does not hold for any λ sequence if we are in case (ii) and $\Delta^{-1}(a^* - \lambda) \to \infty$. We introduce a set sequences $A^* := (a^* \vee \zeta + \Delta, a^* \vee \zeta + 2\Delta)$ with $\Delta := \frac{1}{1+(a^*-\zeta)_+}$. We note that $\inf_{a\in A} \phi(a-\zeta) \gtrsim \phi((a^*-\zeta)_+)$ by (E.37). Then, from (E.32) it follows that, for $q \in \{1, 2\}$,

$$\|g - d^* \times \widetilde{g}_q\|^2 \gtrsim \phi((a^* - \zeta)_+) \int_{A^*} (1 - e^{-2a\zeta} - d^* \times h_q(a))^2 da. \tag{E.42}$$

According to (E.33) and (E.15), $\|g\|^2 \asymp \phi((a^* - \zeta)_+)\Delta$. Then, combining (E.30) and (E.42), we obtain that, for $q \in \{1, 2\}$, if $\theta_q(\lambda) \to 1$,

$$\int_{A^*} (1 - e^{-2a\zeta} - d^* \times h_1(a))^2 da = o(\Delta). \tag{E.43}$$

Moreover, it follows from Lemma E2 and (E.43) that, for $q \in \{1, 2\}$, if $\theta_q(\lambda) \to 1$, there exists a set sequence $\bar{A}$ such that

$$\int_{A^*-\bar{A}} da = o(\Delta) \quad \text{and} \quad \sup_{a\in\bar{A}} |d^* \times h_q(a) - 1| = o(1). \tag{E.44}$$

Here we also use $e^{-a^*\zeta} \to 0$ (by the first part of (E.1) and $\zeta^* \asymp \sqrt{\log N} \to \infty$ due to the assumption on ρ). Now we prove by contradiction that, for each $q \in \{1, 2\}$, $\theta_q(\lambda) \to 1$ can not hold if $\Delta^{-1}(a^*-\lambda) \to \infty$. For this purpose, we introduce a set sequence $A' := (a^*-2\Delta, a^*-\Delta)$.

Suppose $\Delta^{-1}(a^* - \lambda) \to \infty$ holds. Then we obtain that, for $q \in \{1, 2\}$, if $\theta_q(\lambda) \to 1$ and for all $a \in A^*$ and $a' \in A'$,

$$h_q(a') = (1 + o(1))\frac{\exp((a^* - a')\zeta) + 1}{\exp((a^* - a)\zeta) + 1} \times h_q(a) = (1 + o(1))\chi(a')^{-1} \times h_q(a). \tag{E.45}$$

The first equality comes from $\Delta^{-1}(a^* - \lambda) \to \infty$ and the definition of $\widetilde{\psi}_q$. The last equality comes from the definitions of A^* and A' and that $\exp(\zeta\Delta) \to \infty$, which in turn is a result of that in case (ii) we have (a) $\zeta \geq \zeta^* - c_N\sqrt{\log N} \gtrsim \sqrt{\log N}$ and (b) $\Delta^{-1} \lesssim c_N\sqrt{\log N}$ due to $a^* - \zeta \asymp \zeta^* - \zeta \geq -c_N\sqrt{\log N}$. Combining the second part of (E.44) and (E.45), we obtain that, for $q \in \{1, 2\}$, if $\theta_q(\lambda) \to 1$ and for all $a \in A'$,

$$d^* \times h_q(a) = (1 + o(1))\chi(a)^{-1}. \tag{E.46}$$

Since $\inf_{a\in A'} \chi(a)^{-1} \to \infty$ due to $\exp(\zeta\Delta) \to \infty$, it follows from (E.46) that, for $q \in \{1, 2\}$,

$$\begin{aligned} \|g - d^* \times \widetilde{g}_q\|^2 &\geq \int_{A'} (1 - e^{-2a\zeta} - d^* \times h_q(a))^2 f(a)^2 da \\ &\gtrsim \int_{A'} \chi(a)^{-1}\phi(a - \zeta)da \gtrsim \Delta \inf_{a\in A'} \phi(a - \zeta) \gtrsim \|g\|^2, \end{aligned}$$

where we note that $\inf_{a\in A'} \chi(a)^{-1} \to \infty$ (by $\chi(a^*) = 1$ and $\exp(\zeta\Delta) \to \infty$) and $\inf_{a\in A'} \phi(a - \zeta) \gtrsim \phi((a^* - \zeta)_+)$, and we recall $\|g\|^2 \asymp \phi((a^* - \zeta)_+)\Delta$. This contradicts (E.30) and proves that, for $q \in \{1, 2\}$, $\theta_q(\lambda) \to 1$ can not hold if we are in case (ii) and $\Delta^{-1}(a^* - \lambda) \to \infty$.

Step 5. This step completes the proof regarding condition (E.23) for $q = 1$, and prove the second part of (E.24). Given (E.29) and the privious analysis for cases (ia) and (ib), to show condition (E.23) for $q = 1$ we only need to demonstrate that, in case (ii), there exists some sequence λ such that $\theta_1(\lambda) \to 1$ if and only if $\zeta - a^* \to \infty$.

We first show the "only if". By definition of A^* (see above (E.42)), it holds that $\sup_{a\in A^*} \exp((a^* - a)\zeta) \leq \exp(-\zeta\Delta) \to 0$. Substituting this result into the definition of h_1 given in (E.32) and that $\psi_1(a, \lambda) = (a - \lambda)_+$ for $a \geq 0$, we obtain from the second part of (E.44) that $\sup_{a\in\bar{A}}(a - \lambda)_+ = (1 + o(1)) \inf_{a\in\bar{A}}(a - \lambda)_+$. This results, given the first part of (E.44) and the definition of A^*, translates into $(a^* \vee \zeta + 2\Delta - \lambda)_+ = (1 + o(1))(a^* \vee \zeta + \Delta - \lambda)_+$. Also, it is trivial from (E.43) that $\lambda \leq a^* \vee \zeta + 3\Delta/2$ for large N. Therefore, we have that, if $\theta_1(\lambda) \to 1$,

$$\Delta^{-1}(a^* \vee \zeta - \lambda) \to \infty. \tag{E.47}$$

However, according to Step 4 and the subsequence argument, $\theta_1(\lambda) \to 1$ also leads to $\Delta^{-1}(a^* - \lambda) \lesssim 1$. This bound and (E.47) can simultaneously hold only if $\zeta - a^* \to \infty$.

Now we show the "if" part. We note

$$\|f-g\|^2 \le \int_0^\infty e^{-2a\zeta}\phi(a-\zeta)da = \Phi(-\zeta) \to 0, \tag{E.48}$$

where the first inequality comes from $f(a)^2 \le \phi(a-\zeta)$. Moreover, letting $\bar{d}^* = (\zeta-\lambda)^{-1}$, we obtain

$$\begin{aligned}
&\|f-\bar{d}^*\times\widetilde{g}_1\|^2 \\
=\;& \int_0^\lambda f(a)^2da + \int_\lambda^\infty \left(1-(e^{(a^*-a)\zeta}+1)\times\frac{a-\lambda}{\zeta-\lambda}\right)^2 f(a)^2da \\
\le\;& \int_0^\lambda f(a)^2da + 2\int_\lambda^\infty \left(\frac{\zeta-a}{\zeta-\lambda}\right)^2 f(a)^2da + 2\int_\lambda^\infty e^{2(a^*-a)\zeta}\left(\frac{a-\lambda}{\zeta-\lambda}\right)^2 f(a)^2da \\
\le\;& \Phi(\lambda-\zeta) + \frac{2}{(\zeta-a^*)^2} + \frac{4}{(\zeta-\lambda)^2\rho}\int_\lambda^\infty \phi(a)(a-\lambda)^2da.
\end{aligned} \tag{E.49}$$

Here the equality comes from (E.32) and the first inequality is obvious. For the second inequality we use $f(a)^2 \le \phi(a-\zeta)$, $\int(x-\zeta)^2\phi(x-\zeta)=1$, and $e^{2(a^*-a)\zeta}f(a)^2 \le \chi(a)^{-1}\phi(a-\zeta) = \frac{2}{\rho}\phi(a)$. We also note $\Phi(\lambda-\zeta)\to 0$ if $\lambda \le a^*$, as $\zeta - a^* \to \infty$. Hence, given (E.49), and noting $\int_\lambda^\infty \phi(a)(a-\lambda)^2da \eqsim \frac{1}{1+\lambda^2}\phi(\lambda)$, we have that $\|f-\bar{d}^*\times\widetilde{g}_1\|\to 0$, when λ satisfies the second part of (E.24). (Such λ always exists, because $\lambda = a^*$ satisfy the condition as $\zeta - a^* \to \infty$ and $\rho^{-1}\phi(a^*) = \phi(a^*-\zeta)/2 \to 0$.) Given (E.48), and noting $\|g\|^2 \eqsim 1$ from (E.33) (note $\zeta - a^* \to \infty$ can only occur in case (ii) by the first part of (E.1)), we obtain $\|g-\bar{d}^*\times\widetilde{g}_1\| = o(\|g\|)$. Since $\|g-d^*\times\widetilde{g}_1\| \le \|g-\bar{d}^*\times\widetilde{g}_1\|$ by definition of d^*, the "if" part follows from (E.31). We hence establish condition (E.23) for $q=1$.

Now we prove the second part of (E.24). The paragraph after (E.49) already proves the "if" part of the second part of (E.24).

We now demonstrate the "only if" part. When $\zeta-\lambda \lesssim 1$, it holds that, for all d,

$$\|f-d\times\widetilde{g}_1\|^2 \ge \int_{a^*}^\lambda f(a)^2da \ge \frac{1}{2}(\Phi(\lambda-\zeta)-\Phi(a^*-\zeta)) \gtrsim 1. \tag{E.50}$$

The first inequality comes from $\lambda \ge a^*$ as $\zeta - a^* \to \infty$, the second inequality comes from that $f(a)^2 \ge \frac{1}{2}\phi(a-\zeta)$ for all $a \ge a^*$, and the last inequality comes from $\zeta-\lambda \lesssim 1$ and $\zeta - a^* \to \infty$. Combining (E.48) and (E.50), and recalling (E.31) and that $\|g\|^2 \eqsim 1$ from (E.33) under $\zeta - a^* \to \infty$, we obtain that the condition $\zeta - \lambda \to \infty$ is necessary.

Now suppose $\zeta-\lambda\to\infty$ holds but $\rho^{-1}\phi(\lambda) \gtrsim (1+\lambda^2)(\zeta-\lambda)^2$. This indicates $\rho^{-1}\phi(\lambda)\to\infty$. Since $\rho^{-1}\phi(a^*)\to 0$ (see after (E.49)), we have $(a^*-\lambda)/(1+\lambda)\to\infty$. Therefore, letting $l(a) = \mathbb{1}_{\{\lambda\le a\le a^*\}}$, we have, for all d,

$$\|\widetilde{g}_1 \times l\|^2 = \int_\lambda^{a^*} (e^{(a^*-a)\zeta}+1)^2(a-\lambda)^2 f(a)^2 da \geq \frac{2}{\rho}\int_\lambda^{a^*}(a-\lambda)^2\phi(a)da \gtrsim \frac{\phi(\lambda)}{\rho(1+\lambda^2)} \gtrsim (\zeta-\lambda)^2. \tag{E.51}$$

The second inequality comes from $(e^{(a^*-a)\zeta}+1)^2 f(a)^2 = (1+\chi(a)^{-1})\phi(a-\zeta)$ and $\chi(a) = \frac{\rho}{2}\phi(a-\zeta)/\phi(a)$ by definition. The third inequality comes from $(a^*-\lambda)/(1+\lambda) \to \infty$. The last inequality comes from $\rho^{-1}\phi(\lambda) \gtrsim (1+\lambda^2)(\zeta-\lambda)^2$ that we suppose. Since $\|f \times l\|^2 = \int_\lambda^{a^*} f(a)^2 da \leq \Phi(a^*-\zeta) \to 0$, given (E.48) and (E.51), we obtain that, for all d,

$$\|g - d \times \widetilde{g}_1\| = \|f - d \times \widetilde{g}_1\| + o(1) \geq \|(f - d \times \widetilde{g}_1) \times l\| + o(1) \gtrsim d \times (\zeta - \lambda) + o(1). \tag{E.52}$$

Given (E.52), and recalling $\|g\|^2 \eqsim 1$ from (E.33), $\|g - d \times \widetilde{g}_1\| = o(1)$ can hold only if $d = o((\zeta-\lambda)^{-1})$. However, for all $d \leq \frac{1}{4}(\zeta-\lambda)^{-1}$, it holds that

$$\|f - d\times\widetilde{g}_1\|^2 \geq \int_{a^*}^{\zeta}(1-d\times(e^{(a^*-a)\zeta}+1)(a-\lambda))^2 f(a)^2 da \geq \frac{1}{4}\int_{a^*}^{\zeta} f(a)^2 da \gtrsim \frac{1}{2} - \Phi(a^*-\zeta) \gtrsim 1. \tag{E.53}$$

The second inequality comes from that $e^{(a^*-a)\zeta}+1 \leq 2$ and $a-\lambda \leq \zeta-\lambda$ for all $a \in (a^*, \zeta)$, the third inequality comes from that $f(a)^2 \geq \frac{1}{2}\phi(a-\zeta)$ for all $a \geq a^*$, and the last inequality comes from $\zeta - a^* \to \infty$. Given (E.48), (E.53) contradicts $\|g - d \times \widetilde{g}_1\| = o(1)$. Therefore, recalling (E.31), we conclude that $\rho^{-1}\phi(\lambda) = o((1+\lambda^2)(\zeta-\lambda)^2)$ is indeed necessary. The second part of (E.24) has been proved.

Step 6. In this step we finish the proof regarding condition (E.23) for $q = 2$. Given (E.29) and the privious analysis for cases (ia) and (ib), we only need to show that, in case (ii), $\theta_2(\lambda) \to 1$ for some sequence λ. Under $\lambda = a^*$ and $\bar{d}^* = \zeta^{-1}$, we obtain

$$\begin{aligned}
\|f - \bar{d}^* \times \widetilde{g}_2\|^2 &= \int_0^{a^*} f(a)^2 da + \int_{a^*}^{\infty}\left(1 - (e^{(a^*-a)\zeta}+1)\times\frac{a}{\zeta}\right)^2 f(a)^2 da \\
&\leq \int_0^{a^*} f(a)^2 da + 2\int_{a^*}^{\infty}\left(\frac{a-\zeta}{\zeta}\right)^2 f(a)^2 da + 2\int_{a^*}^{\infty} e^{2(a^*-a)\zeta}\frac{a^2}{\zeta^2} f(a)^2 da \\
&\leq \iota\Phi(2\zeta - a^*) + \frac{2}{\zeta^2}\int_{a^*-\zeta}^{\infty} a^2\phi(a)da + \frac{2}{\zeta^2}e^{2a^*\zeta}\int_{a^*}^{\infty}(a+\zeta)^2\phi(a+\zeta)da \\
&\lesssim \iota\Phi(2\zeta - a^*) + \frac{2}{\zeta^2}(1+(a^*-\zeta)_+)^2\Phi(\zeta - a^*) \\
&\quad + \frac{2}{\zeta^2}(1+(a^*+\zeta))^2 e^{2a^*\zeta}\Phi(-a^*-\zeta).
\end{aligned} \tag{E.54}$$

Here the equality comes from (E.32) and the first inequality is obvious. For the second inequality we use $f(a)^2 \leq \chi(a)\phi(a-\zeta) = \iota\phi(a-2\zeta)$, $f(a)^2 \leq \phi(a-\zeta)$, and $e^{-2a\zeta}\phi(a-\zeta) = \phi(a+\zeta)$. The last inequality comes from that $\int_x^\infty a^2\phi(a)da \eqsim (1+x_+)^2\Phi(-x)$ uniformly

over x. Moreover, we note $(1+(a^*+\zeta))e^{2a^*\zeta}\Phi(-a^*-\zeta) \approx e^{2a^*\zeta}\phi(a^*+\zeta) = \phi(a^*-\zeta) \approx (1+(a^*-\zeta)_+)\Phi(\zeta-a^*)$ using (E.15). Furthermore, since we are in case (ii), it holds that $\iota\Phi(2\zeta-a^*) = o(\Phi(\zeta-a^*))$ according to (E.17) and (E.21), and that $1+(a^*-\zeta)_+ = o(\zeta)$ (by the first part of (E.1) and $\zeta^* \approx \sqrt{\log N}$ due to the assumption on ρ). Substituting these results into (E.54), we have that, under $\lambda = a^*$ and $\bar{d}^* = \zeta^{-1}$,

$$\|f - \bar{d}^* \times \widetilde{g}_2\|^2 = o(\Phi(\zeta - a^*)). \tag{E.55}$$

On the other hand, we note $\|f-g\|^2 \leq \Phi(-\zeta)$ (see below (E.49)) and $\Phi(-\zeta) = o(\Phi(\zeta-a^*))$ due to $1+(a^*-\zeta)_+ = o(\zeta)$ (see above (E.55)). Combining these results with (E.55), and noting $\|g\|^2 \approx \Phi(\zeta-a^*)$ in case (ii) according to (E.33), we finally obtain that, under $\lambda = a^*$ and $\bar{d}^* = \zeta^{-1}$,

$$\|g - \bar{d}^* \times \widetilde{g}_2\| = o(\|g\|). \tag{E.56}$$

Since $\|g - d^* \times \widetilde{g}_1\| \leq \|g - \bar{d}^* \times \widetilde{g}_1\|$ by definition of d^*, we prove $\theta_2(\lambda) \to 1$ in case (ii) for some sequence λ. (E.23) is completely established.

Next, it follows that, under $\lambda \leq a^*$ and letting $\bar{d}^* = \zeta^{-1}$,

$$\begin{aligned} & \|g - d^* \times \widetilde{g}_2\|^2 && \text{(E.57)} \\ \leq\ & \|g - \bar{d}^* \times \widetilde{g}_2\|^2 \leq o(\|g\|^2) + 2\zeta^{-2}\int_\lambda^{a^*} a^2 \frac{1+\chi(a)}{\chi(a)}\phi(a-\zeta)da \\ \leq\ & o(\|g\|^2) + \frac{8}{\zeta^2\rho}\int_\lambda^{a^*} a^2\phi(a)da \leq o(\|g\|^2) + \frac{8a^{*2}}{\zeta^2\rho}\Phi(-\lambda) \leq o(\|g\|^2) + O(\rho^{-1}\Phi(-\lambda)). && \text{(E.58)} \end{aligned}$$

The second inequality comes from (E.56), the triangle inequality, the definition of $\widetilde{g}_2(a)$ given after (E.28), and that $\widetilde{\psi}_2(a,\lambda) - \widetilde{\psi}_2(a,a^*) = a\mathbb{1}_{\{\lambda \leq a \leq a^*\}}$ for all $a \geq 0$. The third inequality comes from $\frac{1+\chi(a)}{\chi(a)}\phi(a-\zeta) \leq \frac{2}{\chi(a)}\phi(a-\zeta) = \frac{4}{\rho}\phi(a)$ for $a \leq a^*$. The fourth inequality is obvious and the last comes from $a^* \lesssim \zeta$ as we are in case (ii). (E.58), (E.29), and (E.31) together proves (E.25). ∎

Lemma E4. *Suppose the same assumptions as in Lemma E1. Then, for any deterministic positive λ sequence satisfying $\rho N\Phi(\zeta-\lambda) \to \infty$, it holds that, for $q \in \{1,2\}$,*

$$\|\widetilde{\psi}_q(\lambda)\|^2 = (1+o_{\mathrm{P}}(1))N\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2).$$

Here $\widetilde{\psi}_q(\lambda)$ stands for the N-dimensional vector whose components are $\widetilde{\psi}_{q,i}(\lambda)$.

Proof. Throughout the proof λ is an arbitrary sequence as described in the statement of the lemma. The strategy is to calculate the magnitude of $\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2)$ and $\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^4)$, and then establish the probability limit using Chebyshev's inequality. It holds by the definition of $\widetilde{\psi}_{q,i}$

and (E.2) (the condition $\rho \to 0$ it relies on is assumed here too) that, for $q \in \{1, 2\}$,

$$\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^{2q}) = 2(1 + o(1)) \int_0^\infty \widetilde{\psi}_q(a, \lambda)^{2q} \left(\phi(a) + \frac{\rho}{2} \phi(a - \zeta) \right) da. \tag{E.59}$$

On the other hand, we can calculate, for $j \in \{2, 4\}$,

$$\begin{aligned} \int_0^\infty \widetilde{\psi}_1(a, \lambda)^j \phi(a) da &= \int_0^\infty a^j \phi(\lambda + a) da \asymp \iota_{1,1}^j \Phi(-\lambda), & \text{(E.60)} \\ \int_0^\infty \widetilde{\psi}_1(a, \lambda)^j \phi(a - \zeta) da &= \int_0^\infty a^j \phi(\lambda - \zeta + a) da \asymp \iota_{1,2}^j \Phi(\zeta - \lambda), & \text{(E.61)} \end{aligned}$$

where we use short-hand notation $\iota_{1,1} \coloneqq (1 + \lambda)^{-1}$ and $\iota_{1,2} \coloneqq \frac{1+(\zeta-\lambda)_+}{1+(\lambda-\zeta)_+}$. Similarly, we have, for $j \in \{2, 4\}$,

$$\begin{aligned} \int_\lambda^\infty \widetilde{\psi}_2(a, \lambda)^j \phi(a) da &= \int_0^\infty (\lambda + a)^j \phi(\lambda + a) da \asymp \iota_{2,1}^j \Phi(-\lambda), & \text{(E.62)} \\ \int_\lambda^\infty \widetilde{\psi}_2(a, \lambda)^j \phi(a - \zeta) da &= \int_0^\infty (\lambda + a)^j \phi(\lambda - \zeta + a) da \asymp \iota_{2,2}^j \Phi(\zeta - \lambda). & \text{(E.63)} \end{aligned}$$

where $\iota_{2,1} = \lambda + (1 + \lambda)^{-1}$ and $\iota_{2,2} = \lambda + \frac{1+(\zeta-\lambda)_+}{1+(\lambda-\zeta)_+}$. Then it holds that, for $q \in \{1, 2\}$,

$$\begin{aligned} \frac{\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^4)}{\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2)^2} &\lesssim \frac{\iota_{q,1}^4 \Phi(-\lambda) + \iota_{q,2}^4 \rho \Phi(\zeta - \lambda)}{\iota_{q,1}^4 \Phi(-\lambda)^2 + \iota_{q,2}^4 \rho^2 \Phi(\zeta - \lambda)^2} \leq \iota_{q,2}^2 \frac{\iota_{q,1}^2 \Phi(-\lambda) + \iota_{q,2}^2 \rho \Phi(\zeta - \lambda)}{\iota_{q,1}^4 \Phi(-\lambda)^2 + \iota_{q,2}^4 \rho^2 \Phi(\zeta - \lambda)^2} \\ &\leq \iota_{q,2}^2 \frac{2}{\iota_{q,1}^2 \Phi(-\lambda) + \iota_{q,2}^2 \rho \Phi(\zeta - \lambda)} \leq \frac{2}{\rho \Phi(\zeta - \lambda)}. \end{aligned}$$

Here we obtain the first inequality by substituting (E.60), (E.61), (E.62), and (E.63) into (E.59). The second inequality comes from that $\iota_{q,1} \leq \iota_{q,2}$ for $q \in \{1, 2\}$. Since $\rho N \Phi(\zeta - \lambda) \to \infty$ by assumption, we have, for $q \in \{1, 2\}$,

$$\frac{\mathrm{Var}(\|\widetilde{\psi}_q\|^2)}{(\mathrm{E}(\|\widetilde{\psi}_q\|^2))^2} \leq \frac{\mathrm{E}(\widetilde{\psi}_{q,i}^4)}{N \mathrm{E}(\widetilde{\psi}_{q,i}^2)^2} \lesssim \frac{1}{N \rho \Phi(\zeta - \lambda)} \to 0. \tag{E.64}$$

Here we suppress the argument λ of $\widetilde{\psi}_{q,i}$ for simplicity and the first inequality comes from that $\widetilde{\psi}_{q,i}$ is i.i.d. across i. The current lemma directly follows from (E.64) and Chebyshev's inequality. ∎

Lemma E5. *Suppose the same assumptions as in Lemma E1. Let $\psi_{\Delta,1}(a, \lambda, b) \coloneqq b \mathbb{1}_{\{|a|-\lambda \geq -b\}}$ and $\psi_{\Delta,2}(a, \lambda, b) \coloneqq \lambda \mathbb{1}_{\{||a|-\lambda| \leq b\}}$. Also suppose that λ_N and b_N are two deterministic positive sequences satisfying $\rho N \Phi(\zeta - \lambda_N) \to \infty$ and $b_N = o\left(1/\sqrt{\log N}\right)$. Then it holds that, for*

$q \in \{1,2\}$,

$$\mathrm{E}(|\psi_{\Delta,q}(\check{z}_i, \lambda_N, b_N)|^2) = o(\mathrm{E}(\widetilde{\psi}_q(\check{z}_i, \lambda_N)^2)).$$

Proof. Throughout the proof λ_N and b_N are two arbitrary sequences as described in the statement of the lemma. For simplicity, we omit the subscript N of λ_N and b_N and also omit the last two arguments of $\psi_{\Delta,q}$. Using (E.2) (the condition $\rho \to 0$ it relies on is assumed here too), we have, for $q \in \{1,2\}$,

$$\mathrm{E}(\psi_{\Delta,q}(\check{z}_i)^2) = 2(1+o(1)) \int_0^\infty \psi_{\Delta,q}(a)^2 \left(\phi(a) + \frac{\rho}{2}\phi(a-\zeta)\right) da. \tag{E.65}$$

To evaluate the right-hand side of (E.65), we write

$$\begin{aligned}
\int \psi_{\Delta,1}(a)^2 \phi(a) da &= b^2 \int_{\lambda-b}^\infty \phi(a) da = b^2 \Phi(b-\lambda), && \text{(E.66)}\\
\int \psi_{\Delta,1}(a)^2 \phi(a-\zeta) da &= b^2 \int_{\lambda-b}^\infty \phi(a-\zeta) da = b^2 \Phi(\zeta+b-\lambda), && \text{(E.67)}\\
\int \psi_{\Delta,2}(a)^2 \phi(a) da &= \lambda^2 \int_{\lambda-b}^{\lambda+b} \phi(a) da \le 2b\lambda^2 \phi((\lambda-b)_+), && \text{(E.68)}\\
\int \psi_{\Delta,2}(a)^2 \phi(a-\zeta) da &= \lambda^2 \int_{\lambda-b}^{\lambda+b} \phi(a-\zeta) da \le 2b\lambda^2 \phi((\lambda-b-\zeta)_+). && \text{(E.69)}
\end{aligned}$$

On the other hand, it holds that, for all sequence $(a, \bar{a})$ satisfying $|\bar{a} - a| \le b$,

$$\begin{aligned}
b^2\Phi(-\bar{a}) &\le \frac{c_N}{(1+a_+)\log N}\phi(\bar{a}_+) \le c_N(1+a_+)^{-3}(\phi(a_+) + c_N N^{-1})\\
&\le c_N(1+a_+)^{-2}(\Phi(-a) + c_N N^{-1}). && \text{(E.70)}
\end{aligned}$$

The first inequality comes from $b = o\left(1/\sqrt{\log N}\right)$ and (E.15). The second comes from (D.18) of Lemma D4. For the last inequality we use (E.15), too. Further, we note that $(1+\lambda_+)^{-1} \le \iota_{1,1} \le \iota_{1,2}$ and $(1+(\lambda-\zeta)_+)^{-1} \le \iota_{1,2}$ (introduced after (E.61)). Then, using (E.70) (choose $(a,\bar{a}) = (\lambda, \lambda-b)$ and $(a,\bar{a}) = (\lambda-\zeta, \lambda-\zeta-b)$ respectively), and noting $\rho \to 0$, $\iota_{1,1} \le \iota_{1,2}$, and $\rho\Phi(\zeta-\lambda) \gtrsim N^{-1}$ by assumption, we obtain

$$b^2\Phi(b-\lambda) + b^2\rho\Phi(\zeta+b-\lambda) = o(\iota_{1,1}^2\Phi(-\lambda) + \iota_{1,2}^2\rho\Phi(\zeta-\lambda)). \tag{E.71}$$

Given (E.65), (E.66), (E.67), (E.59), (E.60), and (E.61), we obtain from (E.71) that $\mathrm{E}(\psi_{\Delta,1}(\check{z}_i)^2) = o(\mathrm{E}(\widetilde{\psi}_{1,i}(\lambda)^2))$, which proves the lemma for $q=1$.

Next, it holds that, for all sequence $(a, \bar{a})$ satisfying $|\bar{a} - a| \le b$,

$$b\phi(\bar{a}_+) \le \frac{c_N}{\sqrt{\log N}}\phi(\bar{a}_+) \le c_N(1+a_+)^{-1}\phi(a_+) + c_N N^{-1} \le c_N\Phi(-a) + c_N N^{-1}. \tag{E.72}$$

We recall $x_+ = \max\{x, 0\}$. The first inequality comes from $b = o\left(1/\sqrt{\log N}\right)$. The second comes from (D.18) of Lemma D4. For the last inequality we use (E.15), too. Further, we note that $\lambda \leq \iota_{2,1}$ and $\lambda \leq \iota_{2,2}$ (introduced after (E.61)). Then, using (E.72) (choose $(a, \bar{a}) = (\lambda, \lambda - b)$ and $(a, \bar{a}) = (\lambda - \zeta, \lambda - \zeta - b)$ respectively), and noting $\rho \to 0$, $\iota_{2,1} \leq \iota_{2,2}$, and $\rho\Phi(\zeta - \lambda) \gtrsim N^{-1}$ by assumption, we obtain

$$b\lambda^2\Phi(b - \lambda) + b\lambda^2\rho\Phi(\zeta + b - \lambda) = o(\iota_{1,1}^2\Phi(-\lambda) + \iota_{1,2}^2\rho\Phi(\zeta - \lambda)). \tag{E.73}$$

Given (E.65), (E.68), (E.69), (E.59), (E.62), and (E.63), we obtain from (E.73) that $\mathrm{E}(\psi_{\Delta,2}(\check{z}_i)^2) = o(\mathrm{E}(\widetilde{\psi}_{2,i}(\lambda)^2))$, which proves the lemma for $q = 2$. ∎

Lemma E6. *Suppose the same assumptions as in Proposition B2. Suppose λ is a positive deterministic sequence satisfying $\rho N\Phi(\mu - \lambda) \to \infty$. Also suppose λ' is a $\mathcal{G}$-measurable sequence λ' that satisfies $|\lambda' - \lambda| = o_{\mathrm{P}}\left(1/\sqrt{\log N}\right)$. Then it holds that $\widehat{S}'_q(\lambda') = S_q(\lambda) + o_{\mathrm{P}}(S^{\mathrm{OPT}} + 1)$ for $q = \{1, 2\}$.*

Proof. Throughout the proof λ and λ' are arbitrary sequences as described in the statement of the lemma. Every result holds for $q \in \{1, 2\}$.

We first provide a bound on $\|\widehat{\psi}_q(\lambda') - \widetilde{\psi}_q(\lambda)\|^2$. Since $\check{z}_i = \sqrt{T}(s_i + \bar{\varepsilon}_i)$ and we have $\max_i |\bar{\varepsilon}_i| \lesssim_{\mathrm{P}} \sqrt{(\log N)/T}$ by uniform bound on i.i.d. normal variables and $|s_i| \lesssim N^{-d}$ by assumption, we obtain $\sqrt{(\log N)/T}\max_i |\check{z}_i| = o_{\mathrm{P}}\left(1/\sqrt{\log N}\right)$ as $(\log N)^3/T \to 0$ by assumption. Combining this result with the first part of (D.12) of Lemma D2, and noting $T = o(N)$ by assumption, we obtain

$$\sup_{i \leq N} |\widehat{z}_i - \check{z}_i| = o_{\mathrm{P}}\left(1/\sqrt{\log N}\right). \tag{E.74}$$

Further, by definition it holds that, for all a, $b \geq 0$, and $\lambda \geq 0$,

$$\sup_{(a',\lambda'):|a'-a|+|\lambda'-\lambda| \leq b} |\widetilde{\psi}_q(a', \lambda') - \widetilde{\psi}_q(a, \lambda)| \leq \psi_{\Delta,q}(a, \lambda, b) + \mathbb{1}_{\{q=2\}}\psi_{\Delta,1}(a, \lambda, b), \tag{E.75}$$

where $\psi_{\Delta,q}$ is introduced in the statement of Lemma E5. Substituting (E.74) and $|\lambda' - \lambda| = o_{\mathrm{P}}\left(1/\sqrt{\log N}\right)$ (by assumption) into (E.75) (choose $(a', a) = (\widehat{z}_i, \check{z}_i)$), applying Lemma E5, noting $|\widetilde{\psi}_1(a, \lambda)| \leq |\widetilde{\psi}_2(a, \lambda)|$ by definition, and using Chebyshev's inequality, we obtain

$$\|\widehat{\psi}_q(\lambda') - \widetilde{\psi}_q(\lambda)\|^2 = o(N\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2)). \tag{E.76}$$

Using $\max_{i \leq N} |\widehat{\sigma}_i/\sigma_i - 1| \lesssim_{\mathrm{P}} c_N$ by (D.1) of Lemma D1 and the triangular inequality, we further have

$$\|\sigma\widehat{w}'_q(\lambda') - \widetilde{\psi}_q(\lambda)\|^2 = o(N\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2)). \tag{E.77}$$

Next, we write

$$|\sigma\widehat{w}'_q(\lambda')^\intercal\psi - \widetilde{\psi}_q(\lambda)^\intercal\psi| \leq \|\sigma\widehat{w}'_q(\lambda')^\intercal - \widetilde{\psi}_q(\lambda)\|\|\psi\| = o_{\mathrm{P}}\left(\sqrt{N\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2)}(1+S^{\mathrm{OPT}})\right). \tag{E.78}$$

The first inequality comes from Cauchy-Schwarz inequality. For the last equality we utilize $\|\psi\| \lesssim_{\mathrm{P}} 1 + S^{\mathrm{OPT}}$ by Corollary 2 and (E.77). On the other hand, we have

$$\mathrm{Var}(\widetilde{\psi}_q(\lambda)^\intercal\psi) \leq N\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2\psi_i^2) \leq N\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2\alpha_i^2) = o(N\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2)). \tag{E.79}$$

The first inequality comes from that both $\widetilde{\psi}_{q,i}(\lambda)$ and ψ_i are i.i.d. across i. The second inequality comes from $\psi_i = \mathrm{E}(\alpha_i|\mathcal{G})$ and that $\widetilde{\psi}_{q,i}(\lambda)$ is $\mathcal{G}$-measurable. The last equality comes from $|\alpha_i| \leq \mu = o(1)$. As a result of (E.79) and Chebyshev's inequality, we have

$$\widetilde{\psi}_q(\lambda)^\intercal\psi = N\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)\psi_i) + o_{\mathrm{P}}\left(\sqrt{N\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2)}\right). \tag{E.80}$$

Combining (E.78) and (E.80), we obtain

$$\sigma\widehat{w}'_q(\lambda')^\intercal\psi = N\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)\psi_i) + o_{\mathrm{P}}\left(\sqrt{N\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2)}(1+S^{\mathrm{OPT}})\right). \tag{E.81}$$

Moreover, it holds that

$$\begin{aligned}\sigma^2\|\widehat{w}'_q(\lambda')\|^2 &= \|\widetilde{\psi}_q(\lambda)\|^2 + O(\|\sigma\widehat{w}'_q(\lambda') - \widetilde{\psi}_q(\lambda)\|^2 + \|\widetilde{\psi}_q(\lambda)\|\|\sigma\widehat{w}'_q(\lambda') - \widetilde{\psi}_q(\lambda)\|) \\ &= (1+o_{\mathrm{P}}(1))N\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2).\end{aligned} \tag{E.82}$$

The first inequality comes from Cauchy-Schwarz inequality. The second equality is a direct result of Lemma E4 and (E.77).

Next, we note that $\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)) = 0$ by symmetry of $\widetilde{\psi}(a,\lambda)$ in a. Because $\widetilde{\psi}_{q,i}(\lambda)$ is i.i.d. across i and independent of β, it follows

$$\mathrm{E}(\|\beta^\intercal\widetilde{\psi}_q(\lambda)\|^2|\beta) \lesssim \mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2)N\|\beta\|_{\mathrm{MAX}}^2 \lesssim_{\mathrm{P}} N\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2). \tag{E.83}$$

Then we have

$$\sigma\|\beta^\intercal\widehat{w}'_q(\lambda')\| \leq \|\beta^\intercal\widetilde{\psi}_q(\lambda)\| + \|\beta^\intercal(\sigma\widehat{w}'_q(\lambda') - \widetilde{\psi}_q(\lambda))\| \lesssim_{\mathrm{P}} \sqrt{N\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2)} + c_N N\sqrt{\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2)}. \tag{E.84}$$

where the last inequality comes from (E.83) and (E.77). Therefore, we obtain

$$\|\mathbb{M}_\beta \widehat{w}_q'(\lambda')\|^2 = \|\widehat{w}_q'(\lambda')\|^2 - \widehat{w}_q'(\lambda')^\intercal \mathbb{P}_\beta \widehat{w}_q'(\lambda') = \sigma^2 \|\widehat{w}_q'(\lambda')\|^2 + o_{\mathrm{P}}(N\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2)), \quad \text{(E.85)}$$

$$\widehat{w}_q'(\lambda')^\intercal \mathbb{M}_\beta \psi = \widehat{w}_q'(\lambda')^\intercal \psi + \widehat{w}_q'(\lambda')^\intercal \mathbb{P}_\beta \psi = \widehat{w}_q'(\lambda')^\intercal \psi + o_{\mathrm{P}}\left(\sqrt{N\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2)}\right). \quad \text{(E.86)}$$

The last equalities of both results come from (E.84) and $\lambda_{\min}(\beta^\intercal\beta) \gtrsim_{\mathrm{P}} N$. For the last equality of (E.86), we also use $\|\beta^\intercal\psi\| \lesssim_{\mathrm{P}} N^{1/2}\mathrm{E}(\alpha_i^2)^{1/2} \lesssim_{\mathrm{P}} c_N N^{1/2}$. We now conclude that

$$\begin{aligned} \frac{\widehat{w}_q'(\lambda')^\intercal \mathbb{M}_\beta \psi}{\|\mathbb{M}_\beta \widehat{w}_q'(\lambda')\|} &= \frac{(1+o_{\mathrm{P}}(1))\widehat{w}_q'(\lambda')^\intercal \psi}{\sqrt{N\mathrm{E}(\widetilde{\psi}_{q,i}(\lambda)^2)}} + o_{\mathrm{P}}(1) \\ &= (1+o_{\mathrm{P}}(1))S_q(\lambda) + o_{\mathrm{P}}(1+S^{\mathrm{OPT}}) = S_q(\lambda) + o_{\mathrm{P}}(S^{\mathrm{OPT}}+1). \quad \text{(E.87)} \end{aligned}$$

The first equality comes from (E.85), (E.86), and (E.82). The second comes from (E.81). The last equality comes from $S_q(\lambda) \leq N^{1/2}\mathrm{E}(\psi_i^2)^{1/2} = S^{\mathrm{OPT}}$, in which the inequality is just Cauchy-Schwarz and the equality holds by definition.

Finally, (C.5) and (C.6) establishes that, under Assumption 1 and for all $\mathcal{G}$-measurable w, $w^\intercal(\alpha - \mathrm{E}(\alpha|\mathcal{G}))/\|w\| = o_{\mathrm{P}}(1)$. Choosing $w = \widehat{w}_q'(\lambda')^\intercal\mathbb{M}_\beta$, and noting $\psi = \sigma^{-1}\mathrm{E}(\alpha|\mathcal{G})$, we obtain

$$\widehat{S}_q'(\lambda') = \frac{\widehat{w}_q'(\lambda')^\intercal \mathbb{M}_\beta \alpha}{\sigma\|\mathbb{M}_\beta \widehat{w}_q'(\lambda')\|} = \frac{\widehat{w}_q'(\lambda')^\intercal \mathbb{M}_\beta \psi}{\|\mathbb{M}_\beta \widehat{w}_q'(\lambda')\|} + o_{\mathrm{P}}(1). \quad \text{(E.88)}$$

Given (E.87) and (E.88), we prove the lemma. ∎

Lemma E7. *Suppose the same assumptions as in Proposition B2. Then it holds that* $\widehat{S}_q'(\lambda) = o_{\mathrm{P}}(1)$*, for all* $\mathcal{G}$*-measurable positive sequence* λ *satisfying* $\mu^2\rho N\Phi(\zeta - \lambda) = o_{\mathrm{P}}(1)$ *and for* $q \in \{1, 2\}$.

Proof. Throughout the proof λ is an arbitrary sequence as described in the statement of the lemma. Every result holds for $q \in \{1, 2\}$. We let λ^* be a deterministic sequence such that $\lambda \geq \lambda^*$ in probability and $\mu^2\rho N\Phi(\zeta - \lambda^*) \to 0$, which is apparently always feasible.

When $\Phi(\zeta - \lambda^*) \gtrsim 1$, we have $\mu^2\rho N \to 0$ and thereby $(S^{\mathrm{OPT}})^2 = N\mathrm{E}(\psi_i^2) \leq N\mathrm{E}(s_i^2) \leq \rho N\mu^2 \to 0$. Since $\widehat{S}_q'(\lambda)$ is the Sharpe ratio generated by $\widehat{w}_q'(\lambda)^\intercal\mathbb{M}_\beta$, which is $\mathcal{G}$-measurable and satisfy $\widehat{w}_q'(\lambda)^\intercal\mathbb{M}_\beta\beta = 0$. Then we obtain $\widehat{S}_q'(\lambda) \leq S^{\mathrm{OPT}} + o_{\mathrm{P}}(1) = o_{\mathrm{P}}(1)$ from the second part of Theorem 2 and Corollary 2 (the assumptions of the lemma obviously guarantee the prerequisites of both). Since $-\widehat{S}_q'(\lambda)$ is the Sharpe ratio generated by $-\widehat{w}_q'(\lambda)^\intercal\mathbb{M}_\beta$, we obtain $-\widehat{S}_q'(\lambda) \leq o_{\mathrm{P}}(1)$, too.

When $\Phi(\zeta - \lambda^*) \to 0$, we have $\Phi(-\lambda^*) \leq \Phi(\zeta - \lambda^*) \to 0$. Moreover, it holds that, for all sequence $(a, \bar{a})$ satisfying $|\bar{a} - a| = o\left(1/\sqrt{\log N}\right)$,

$$\Phi(-\bar{a}) \lesssim \frac{1}{1+a_+}\phi(\bar{a}_+) \lesssim \frac{1}{1+a_+}(\phi(a_+) + c_N N^{-2}) \lesssim \Phi(-a) + c_N N^{-2}. \quad \text{(E.89)}$$

The first inequality comes from $|\bar{a} - a| = o\left(1/\sqrt{\log N}\right)$ and (E.15). The second comes from (D.18) of Lemma D4. For the last inequality we use (E.15), too. Then, using (E.89) (choose $a = \lambda^*$ and $a = \lambda^* - \zeta$ respectively), we have that, for all $b_N = o\left(1/\sqrt{\log N}\right)$,

$$\begin{cases} N\Phi(b_N - \lambda^*) \lesssim N\Phi(-\lambda^*) + c_N N^{-1} = o(N), \\ \rho N\Phi(\zeta + b_N - \lambda^*) \lesssim \rho N\Phi(\zeta - \lambda^*) + c_N \rho N^{-1} \lesssim o(\mu^{-2}). \end{cases} \tag{E.90}$$

Here the last inequality comes from $\mu^2 \rho N\Phi(\zeta - \lambda^*) = o(1)$ and $\mu = o(1)$ by assumption. Hence, it holds that, for some $b_N = o\left(1/\sqrt{\log N}\right)$,

$$\begin{aligned} \sum_{i \le N} \mathbb{1}_{\{\widehat{w}'_{q,i}(\lambda) \neq 0\}} &\lesssim_{\mathrm{P}} \sum_{i \le N} \mathbb{1}_{\{|\check{z}_i| \ge \lambda^* - b_N\}} \\ &\lesssim_{\mathrm{P}} N\Phi(b_N - \lambda^*) + \rho N\Phi(\zeta + b_N - \lambda^*) = o(N), \qquad \text{(E.91)} \\ \sum_{i \le N} \mathbb{1}_{\{\widehat{w}'_{q,i}(\lambda) \neq 0, \alpha_i \neq 0\}} &\lesssim_{\mathrm{P}} \sum_{i \le N} \mathbb{1}_{\{|\check{z}_i| \ge \lambda^* - b_N, \alpha_i \neq 0\}} \lesssim_{\mathrm{P}} \rho N\Phi(\zeta + b_N - \lambda^*) \lesssim o(\mu^{-2}). \qquad \text{(E.92)} \end{aligned}$$

For both results, the first inequality comes from (E.74), the second inequality comes from the density of $\check{z}_i$ and Chebyshev's inequality, the last inequality comes from (E.90). As a result, we have

$$\|\beta^\intercal \widehat{w}'_q(\lambda)\| \lesssim \|\beta\|_{\mathrm{MAX}} \|\widehat{w}'_q(\lambda)\| \sqrt{\sum_{i \le N} \mathbb{1}_{\{\widehat{w}'_{q,i}(\lambda) \neq 0\}}} = o_{\mathrm{P}}\left(\sqrt{N}\|\widehat{w}'_q(\lambda)\|\right). \tag{E.93}$$

Here the first inequality comes from Cauchy-Schwarz inequality, and the last equality comes from (E.91). Then we obtain

$$\|\mathbb{M}_\beta \widehat{w}'_q(\lambda)\|^2 = \|\widehat{w}'_q(\lambda)\|^2 - \widehat{w}'_q(\lambda)^\intercal \mathbb{P}_\beta \widehat{w}'_q(\lambda) = (1 + o_{\mathrm{P}}(1))\|\widehat{w}'_q(\lambda)\|^2. \tag{E.94}$$

For the last equality, we use (E.93) and $\lambda_{\min}(\beta^\intercal\beta) \gtrsim_{\mathrm{P}} N$. Similarly, using (E.93), $\lambda_{\min}(\beta^\intercal\beta) \gtrsim_{\mathrm{P}} N$ and $\|\beta^\intercal\alpha\| \lesssim_{\mathrm{P}} N^{1/2}\mathrm{E}(\alpha_i^2)^{1/2} \lesssim_{\mathrm{P}} c_N N^{1/2}$, we obtain

$$\widehat{w}'_q(\lambda)^\intercal \mathbb{M}_\beta \alpha = \widehat{w}'_q(\lambda)^\intercal \alpha + \widehat{w}'_q(\lambda)^\intercal \mathbb{P}_\beta \alpha = \widehat{w}'_q(\lambda)^\intercal \alpha + o_{\mathrm{P}}(\|\widehat{w}'_q(\lambda)\|). \tag{E.95}$$

Combining (E.85) and (E.86), we now conclude that

$$\widehat{S}'_q(\lambda) = \frac{\widehat{w}'_q(\lambda)^\intercal \mathbb{M}_\beta \alpha}{\|\mathbb{M}_\beta \widehat{w}'_q(\lambda)\|} = (1 + o_{\mathrm{P}}(1)) \frac{\widehat{w}'_q(\lambda)^\intercal \alpha}{\|\widehat{w}'_q(\lambda)\|} + o_{\mathrm{P}}(1). \tag{E.96}$$

On the other hand, it holds that

$$\frac{|\widehat{w}_q'(\lambda)^\intercal \alpha|}{\|\widehat{w}_q'(\lambda)\|} \leq \mu \frac{\sum_{i\leq N} |\widehat{w}_{q,i}'(\lambda)| \mathbb{1}_{\{\widetilde{w}_{q,i}(\lambda)\neq 0, \alpha_i \neq 0\}}}{\|\widehat{w}_q'(\lambda)\|} \leq \mu \sqrt{\sum_{i\leq N} \mathbb{1}_{\{\widehat{w}_{q,i}'(\lambda)\neq 0, \alpha_i \neq 0\}}} = o_{\mathrm{P}}(1). \tag{E.97}$$

The first inequality holds by $|s_i| = \frac{\mu}{\sigma} \mathbb{1}_{\{\alpha_i \neq 0\}}$, the second by Cauchy-Schwarz inequality, and the last equality holds by (E.92). The lemma follows from (E.96) and (E.97). ∎

References

Andrews, D. W. and X. Cheng (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica 80*(5), 2153–2211.

Brown, L. D. and E. Greenshtein (2009). Nonparametric empirical Bayes and compound decision approaches to estimation of a high-dimensional vector of normal means. *The Annals of Statistics 37*(4), 1685 – 1704.

Efron, B. (2011). Tweedie's formula and selection bias. *Journal of the American Statistical Association 106*(496), 1602–1614.

Folland, G. (2009). *A Guide to Advanced Real Analysis*. Number 37. MAA.

Horn, R. A. and C. R. Johnson (2012). *Matrix analysis*. Cambridge university press.

Liu, W. and Q.-M. Shao (2014). Phase transition and regularized bootstrap in large-scale t-tests with false discovery rate control. *The Annals of Statistics 42*(5), 2003–2025.

Robbins, H. (1956). An empirical bayes approach to statistics. *Berkeley Symposium on Mathematical Statistics and Probability 3*, 157–163.