# Variance Estimation with Dependence and Heterogeneous Means

Luther Yap*

July 16, 2026

**Abstract**

This paper develops a framework for variance estimation under dependence and heterogeneous means. This paper shows that consistent estimation of the variance target is impossible in general, and characterizes necessary and sufficient conditions for conservative variance estimation using dual cones. To choose among the valid estimators, this paper formulates three criteria — minimal correction, pointwise level estimand, and pointwise MSE — and shows how an eigenvalue truncation solution is optimal under all three criteria. This characterization and solution allow us to assess if existing variance estimators are valid and optimal in their respective settings, and construct the first optimal variance estimator that is simultaneously robust to heterogeneous means and cross-cluster serial correlation.

**Keywords:** variance estimation, heterogeneous means, two-way clustering, HAC, conic duality, design-based inference.
**JEL codes:** C12, C13, C23, C33.

## 1 Introduction

Suppose $Y_n \in \mathbb{R}^n$ is a vector of random variables with $E[Y_n] = \mu \in \mathcal{M}_n$ and $Var(Y_n) = \Sigma \in \mathcal{S}_n$. We are interested in a variance target $\theta_n = \text{tr}(W_n\Sigma)$ for some known weight matrix $W_n$. The object $\theta_n$ can be interpreted as the oracle variance for $\sum_i Y_{n,i}$, which is used for inference after applying a central limit theorem. As an example, with independent observations, $W_n = I_n$, but $W_n$ also encodes dependent observations such as one/ two- way clustering and serial correlation. To obtain the variance target, we rely on a variance estimator that takes a quadratic form $Y_n' A_n Y_n$,

*Email: lyap@nus.edu.sg. Department of Economics, National University of Singapore. The latest version is available at https://lutheryap.github.io/files/twclus_het_means_wp.pdf for some $A_n$. Observe that $E[Y_n' A_n Y_n] = \mathrm{tr}(A_n \Sigma) + \mu' A_n \mu$. When the mean is known or consistently estimable, the second term $\mu' A_n \mu$ is innocuous and can be removed by demeaning, and the first term can recover the target. The difficulty confronted in this paper is when means can be arbitrarily heterogeneous, so the means are no longer consistently estimable. Considering heterogeneous means with a general dependence structure is the novel problem addressed in this paper.

Heterogeneous means arise in several settings. Under design-based uncertainty, where potential outcomes are fixed and cluster dependence reflects correlated assignment or sampling (Abadie et al., 2023), the score for each observation is heterogeneous and not consistently estimable by construction. In asset pricing, when we condition on the set of assets and time observed, any omitted priced factor leaves the regression score with a mean that varies across assets and time. In both cases the individual means are not consistently estimable even though the grand mean is, which is the regime in which a central limit theorem governs the sum while the summands' means remain unidentified and the variance of the sum becomes difficult to estimate.

This paper makes three contributions.

First, this paper establishes the limits of identification in the problem in Section 3. With dependence and arbitrary mean heterogeneity, this paper shows that consistent estimation of the variance target $\theta_n = \mathrm{tr}(W_n \Sigma)$ is impossible in general (Theorem 1), because the second moment alone is uninformative of the split between the covariance component and the mean component. The obstruction is one of identification, not of finite samples. Since consistency is unattainable, we can only pursue the second best goal, a conservative variance estimand, one that is guaranteed not to fall below the target, so that the associated statistical test remains valid (i.e., with size guarantee). Seemingly disparate estimators are quadratic forms governed by the same two objects: a linear subspace $\mathcal{M}_n$ of admissible mean vectors and a closed convex cone $\mathcal{S}_n$ of admissible covariance matrices. Necessary and sufficient conditions for conservativeness in terms of these spaces are derived (Theorem 2): $Y_n' A_n Y_n$ is uniformly conservative for $\theta_n$ if and only if $A_n - W_n$ lies in the dual cone $\mathcal{S}_n^*$ and $A_n$ is positive semidefinite on $\mathcal{M}_n$.

In light of how there are many variance estimators that are valid, the second contribution of the paper is to propose an optimal variance estimator in Section 4. Optimality is defined relative to criteria, so this paper proposes three natural criteria — minimal correction (which respects the dependence structure), minimum level of the variance estimand (in the interest of having the smallest variance used in the t-statistic and hence the most powerful test), and mean-squared error (MSE) (which is a common benchmark that assesses the quality of an estimator). The solution proposed in this paper is to set $A_n$ to be an eigenvalue truncation (ET) of $W_n$. Under some conditions, this solution is optimal under all three criteria. (Theorem 3) If these conditions do not hold, then these criteria can nonetheless be transformed into constrained conic optimization problems, so an optimal solution can be obtained numerically. (Proposition 3)

Given the framework and conditions for validity and optimality, the third contribution of the paper is to assess if existing variance estimators in economically relevant settings with dependence are valid and optimal under heterogeneity, and if they are not, obtain the optimal conservative variance estimator. Section 5 considers several existing examples. The plug-in one-way cluster robust variance estimator is valid under mean heterogeneity (a result known from Abadie et al. (2023)), but now we also certify that it is optimal. In two-way clustering, the plug-in variance estimator from Cameron et al. (2011) (CGM) is not valid under mean heterogeneity, while the CGM2 modification from Davezies et al. (2021) is valid — these results are known from Xu and Yap (2024), and are corollaries of our setting. While CGM2 is valid, the new result is that CGM2 is not optimal as it does not match the ET solution — the difference is that CGM2 adds observation-specific variances to CGM, while ET adds the observation-specific within-transformed variances, which results in a smaller variance estimand. Section 6 extends the two-way cluster dependence to allow for cross-cluster weak dependence, a dependence structure considered in Chiang et al. (2024) (CHS). The plug-in variance estimator from CHS is invalid under mean heterogeneity, so a novel optimal conservative variance estimator is proposed as a corollary of this paper's theory. This section also develops the $\psi$-dependence framework, the central limit theorem, and consistency. Simulations and an empirical application close the section.

**Related literature.** This paper relates to three strands of literature.

The first is variance estimation in dependent arrays. Cameron et al. (2011) and Chiang et al. (2024) provide variance estimators under the two-way clustering and two-way clustering with cross cluster correlations, but both assume mean homogeneity. This paper extends their variance estimators to be robust to mean heterogeneity. Beyond variance estimation in the setting with two-way clustering with cross cluster correlations, this paper also provides a central limit theorem that is weaker than the one used in CHS, as this paper relies on the $\psi$-dependence limit theory from Kojevnikov et al. (2021) (KMS) that does not rely on an Aldous-Hoover representation, thereby accommodating data-generating processes outside the representation, similar to Yap (2025).

The second is distributional and mean heterogeneity. In time series, Chan (2022) and Casini (2023) permit time-varying means and second moments but require enough regularity to estimate the mean function consistently, whereas we impose no regularity on the mean sequence and instead modify the estimand to guarantee conservativeness. The interaction between dependence and mean heterogeneity has been rather disparate in design-based settings: settings with independence (Abadie et al., 2020) or one-way clustering (Abadie et al., 2023) have conservative plug-in variances, while two-way clustering does not (Xu and Yap, 2024). This paper's framework unifies these dependence structures, and the conservativeness criterion gives a general and practical way to check if a dependence-robust plug-in is conservative. Further, this paper extends conservative variance estimation under mean heterogeneity beyond these settings to the CHS dependence structure, and additionally derives optimal variances.

Third, a long tradition in time-series and clustered variance estimation is concerned that a quadratic form $Y_n' W_n Y_n$ can return a negative number in finite samples when $W_n$ is not PSD. Newey and West (1987) address this ex ante by adopting the Bartlett kernel; Andrews (1991) characterizes the class of kernels delivering positive-semidefinite estimates and identifies the quadratic-spectral kernel as minimizing asymptotic mean-squared error within it. A complementary, ex post tradition takes a possibly non-PSD estimate and projects it onto the PSD cone by eigenvalue correction, the operation which appears in clustered inference as the eigenvalue correction of Chiang et al. (2024). Our framework differs in two respects. First, the non-PSD-ness we confront is not a sampling-noise artifact of a particular realization but a structural feature of the population weight matrix that the dependence design dictates so it does not vanish as the sample grows and cannot be diagnosed from any single estimate. Second, and consequently, our optimality is stated for the estimand, not the estimate: where the classical corrections enforce a threshold on a realization or minimize the distance of an estimate to the positive-semidefinite cone, ours minimizes the mean-induced bias of the targeted variance subject to conservativeness.

# 2 Framework and Quadratic Representation

## 2.1 Statistical model and parameter

For each $n \geq 1$, let $Y_n = (Y_{n,1}, \dots, Y_{n,n})' \in \mathbb{R}^n$ be observed, distributed according to $P_n$. Define

$$\mu(P_n) := E_{P_n}[Y_n], \qquad \Sigma(P_n) := \mathrm{Var}_{P_n}(Y_n).$$

This paper will drop the indexing on $P_n$ when it is without ambiguity.

The researcher commits to two structural objects:

- a linear subspace $\mathcal{M}_n \subseteq \mathbb{R}^n$ of admissible mean vectors;
- a closed convex cone $\mathcal{S}_n \subseteq \mathrm{Sym}_n^+$ of admissible covariance matrices, where $\mathrm{Sym}_n^+$ denotes the positive semidefinite cone in the space of symmetric $n \times n$ matrices $\mathrm{Sym}_n$.

The model is

$$\mathcal{P}_n(\mathcal{M}_n, \mathcal{S}_n) := \big\{ P_n : E_{P_n}\|Y_n\|^2 < \infty,\ \mu(P_n) \in \mathcal{M}_n,\ \Sigma(P_n) \in \mathcal{S}_n \big\}.$$

Let $W_n \in \mathrm{Sym}_n$ be a known symmetric weight matrix. The parameter of interest is

$$\theta_n(P_n) := \mathrm{tr}\big(W_n \Sigma(P_n)\big), \qquad P_n \in \mathcal{P}_n(\mathcal{M}_n, \mathcal{S}_n).$$

An important context where $\theta_n$ is interesting is whenever some central limit theorem is applied while allowing for dependence across observations and mean heterogeneity. The central limit theorem yields $\left(\sum_{i=1}^n Y_{n,i} - E[Y_{n,i}]\right)/\sqrt{\mathrm{Var}\left(\sum_{i=1}^n Y_{n,i}\right)} \xrightarrow{d} N(0,1)$. While $\frac{1}{n}\sum_{i=1}^n E[Y_{n,i}]$ is consistently estimable from the theorem, the individual means may not be when there is sufficient heterogeneity, which makes the estimation of $\mathrm{Var}\left(\sum_{i=1}^n Y_{n,i}\right)$ tricky when conducting statistical inference. When $W_n$ corresponds to an adjacency matrix and encodes several dependence structures, this parameter of interest $\theta_n$ converges to $\mathrm{Var}\left(\sum_{i=1}^n Y_{n,i}\right)$ under some conditions, even when dependence across $Y_{n,i}$ is permitted. For instance, $W_n = I_n$ corresponds to independent observations. In time series settings with a kernel adjustment for the variance, $W_n$ would reflect these kernel adjustments.

The pair $(\mathcal{M}_n, \mathcal{S}_n)$ encodes the researcher's structural commitments. Larger $\mathcal{M}_n$ and $\mathcal{S}_n$ correspond to weaker assumptions and stronger demands for robustness. Two extreme cases are instructive. At one end, $\mathcal{S}_n = \mathrm{Sym}_n^+$ imposes no structure on the covariance and yields the maximally dependence-robust posture: the researcher seeks conservativeness against any covariance compatible with $\Sigma \succeq 0$. At the other end, $\mathcal{S}_n = \{\sigma^2 I_n : \sigma^2 \geq 0\}$ encodes iid sampling and admits the classical demeaned sample variance. Intermediate cones encode block independence (clusters), bandedness ($m$-dependence), or diagonal structure (without correlation), each yielding its own optimal estimator.

**Examples of $\mathcal{S}_n$.**

- $\mathcal{S}_n^{\mathrm{iid}} := \{\sigma^2 I_n : \sigma^2 \geq 0\}$: iid sampling.
- $\mathcal{S}_n^{\mathrm{diag}} := \{\Sigma \succeq 0 : \Sigma \text{ diagonal}\}$: heterogeneous variances but uncorrelated.
- $\mathcal{S}_n^{\mathrm{block}} := \{\Sigma \succeq 0 : \Sigma_{ij} = 0 \text{ if } g(i) \neq g(j)\}$: cluster independence.
- $\mathcal{S}_n^{\mathrm{band}(b)} := \{\Sigma \succeq 0 : \Sigma_{ij} = 0 \text{ if } |i-j| > b\}$: $b$-dependence.

Similarly, $\mathcal{M}_n$ can encode the known structure of the problem. Fully unrestricted means corresponds to $\mathcal{M}_n = \mathbb{R}^n$, while constant means corresponds to $\mathcal{M}_n = \mathrm{span}\{1_n\}$, as two extreme cases. If we have heterogeneous means but are able to consistently estimate the mean of all elements in $\mu$, then after demeaning by the grand mean, we may restrict our attention to $\mathcal{M}_n = \{\mu : 1_n'\mu = 0\} = 1_n^{\perp}$.

## 2.2 Quadratic-form estimators and dual cones

A quadratic-form estimator takes the form $\widehat{V}_{A,n}(Y_n) := Y_n' A_n Y_n$ for some $A_n \in \mathrm{Sym}_n$, with expectation

$$V_{A,n}(P_n) := E_{P_n}[Y_n' A_n Y_n] = \mathrm{tr}\big(A_n \Sigma(P_n)\big) + \mu(P_n)' A_n \mu(P_n). \tag{1}$$

$V_{A,n}(P_n)$ is an estimand, which is not necessarily the target $\theta_n(P_n)$.

The canonical plug-in estimator for the sum of $Y_n$ (i.e., $1_n' Y_n$) corresponds to $A_n = W_n$, with bias

$$V_{W,n}(P_n) - \theta_n(P_n) = \mu(P_n)' W_n \mu(P_n).$$

This bias can be negative when $W_n$ has negative directions inside $\mathcal{M}_n$, arising in anticonservativeness in inference.

To analyze conservativeness across the full pair $(\mathcal{M}_n, \mathcal{S}_n)$, we use cone duality. The *dual cone* of $\mathcal{S}_n$ is

$$\mathcal{S}_n^* := \{B \in \mathrm{Sym}_n : \mathrm{tr}(B\Sigma) \geq 0 \text{ for all } \Sigma \in \mathcal{S}_n\}. \tag{2}$$

Since $\mathcal{S}_n \subseteq \mathrm{Sym}_n^+$ and $\mathrm{Sym}_n^+$ is self-dual, $\mathcal{S}_n^* \supseteq \mathrm{Sym}_n^+$, with equality when $\mathcal{S}_n = \mathrm{Sym}_n^+$. For strict subcones, $\mathcal{S}_n^*$ contains symmetric matrices that are not positive semidefinite, and this larger admissible region is precisely what licenses non-PSD corrections such as demeaning.

Examples of dual cones, all easily verified from (2):

$$(\mathcal{S}_n^{\mathrm{iid}})^* = \{A : \mathrm{tr}(A) \geq 0\}, \quad (\mathcal{S}_n^{\mathrm{diag}})^* = \{A : \mathrm{diag}(A) \geq 0\}.$$

For $\mathcal{S}_n^{\mathrm{block}}$, $A \in \mathcal{S}_n^{\mathrm{block},*}$ if and only if each within-cluster diagonal block of $A$ is positive semidefinite, with arbitrary off-block entries. The covariance cones sit in a partial order: $\mathcal{S}_n^{\mathrm{iid}} \subset \mathcal{S}_n^{\mathrm{diag}} \subset \mathcal{S}_n^{\mathrm{block}} \subset \mathrm{Sym}_n^+$. Dual cones reverse the order.

# 3 Impossibility and Conservativeness

## 3.1 When consistent estimation is impossible

The same second moment matrix may admit different decompositions into a mean component and a covariance component:

$$E[Y_n Y_n'] = \Sigma(P_n) + \mu(P_n)\mu(P_n)'.$$

The target $\theta_n(P_n) = \mathrm{tr}(W_n \Sigma(P_n))$, however, depends only on the covariance component. Hence, if the model permits a rank-one term $\mu\mu'$ to be shifted between covariance and mean while preserving admissibility, then the target changes but the expectation of every quadratic estimator does not. This yields an impossibility result for uniform consistency over the quadratic class.

**Theorem 1** (Impossibility of consistency)**.** *Fix $n$ and consider the model $\mathcal{P}_n(\mathcal{M}_n, \mathcal{S}_n)$. Suppose there exist $\Sigma_0 \in \mathcal{S}_n$ and $v_n \in \mathcal{M}_n$ with $\Sigma_0 + v_n v_n' \in \mathcal{S}_n$. Then, for any $A_n \in \mathrm{Sym}_n$,*

$$\sup_{P_n \in \mathcal{P}_n(\mathcal{M}_n, \mathcal{S}_n)} \left|\mathbb{E}_{P_n}[\widehat{V}_{A,n}] - \theta_n(P_n)\right| \geq \frac{1}{2}|v_n' W_n v_n|.$$

To gain some intuition for the result, suppose $\mathcal{S}_n$ is the set of diagonal matrices, and $v_n = (1, 0, \cdots, 0)' \in \mathcal{M}_n$. Then, the second moment alone cannot distinguish $\mu_1 = 1$ and $\Sigma_{11} = 1$ from $\mu_1 = 0$ and $\Sigma_{11} = 2$.

As a corollary, write $\lambda_n \asymp \theta_n$ for the scale of the target. If $v_n' W_n v_n \neq o(\lambda_n)$, then the relative worst-case bias of every quadratic estimator is bounded below,

$$\frac{1}{\lambda_n} \sup_{P_n \in \mathcal{P}_n(\mathcal{M}_n, \mathcal{S}_n)} \left| \mathbb{E}_{P_n}[\widehat{V}_{A,n}] - \theta_n(P_n) \right| \;\geq\; \frac{|v_n' W_n v_n|}{2\lambda_n} \;\not\to\; 0.$$

Consequently no quadratic estimator is consistent in that $\widehat{V}_{A,n}/\theta_n \xrightarrow{p} 1$, uniformly over the model: at the two Gaussian points $P_n, Q_n$ of Theorem 1 the normalized forms $\widehat{V}_{A,n}/\theta_n$ are $L^2$-bounded along the sequence, hence uniformly integrable, so consistency would force relative asymptotic unbiasedness $\mathbb{E}_{P_n}[\widehat{V}_{A,n}]/\theta_n \to 1$ at each point, contradicting the display.

The proof rests on a two-point moment-matching argument: two distributions $P_n, Q_n \in \mathcal{P}_n$ can have the same expectation of every quadratic functional yet different values of $\theta_n$, rendering them indistinguishable in expectation to the quadratic class. The factor reflects that any estimator's expectation is a single number that cannot simultaneously match two admissible distributions with the same quadratic-form expectation but targets differing by $v_n' W_n v_n$, so its bias against the worse of the two is at least half that gap.

**Example 1** (Demeaned sample variance)**.** The following example does not satisfy the conditions of the impossibility result. Let $W_n = I_n$, $\mathcal{M}_n = \mathrm{span}(\mathbf{1}_n)$, and $\mathcal{S}_n = \mathcal{S}_n^{\mathrm{iid}}$, so $\theta_n = n\sigma^2$. Then $\mathcal{S}_n^* = \{A : \mathrm{tr}(A) \geq 0\}$, and

$$A_n = \frac{n}{n-1} M_{\mathbf{1}_n}, \qquad M_{\mathbf{1}_n} := I_n - \frac{1}{n}\mathbf{1}_n \mathbf{1}_n'.$$

The estimator $Y_n' A_n Y_n = \frac{n}{n-1} \sum_i (Y_{n,i} - \bar{Y}_n)^2$ is exactly unbiased for $n\sigma^2$ on the model.

The impossibility in Theorem 1 requires the cone $\mathcal{S}_n$ to admit a perturbation along the mean direction: there must exist $\Sigma_0 \in \mathcal{S}_n$ and $v_n \in \mathcal{M}_n$ with $\Sigma_0 + v_n v_n' \in \mathcal{S}_n$ and $v_n' W_n v_n \neq 0$. In the iid case this feasibility condition fails. Any $v_n \in \mathcal{M}_n = \mathrm{span}\{1_n\}$ has the form $v_n = \alpha 1_n$, so $v_n v_n' = \alpha^2 1_n 1_n'$, a rank-one matrix with nonzero off-diagonal entries. For any $\Sigma_0 = \sigma_0^2 I_n \in \mathcal{S}_n^{\mathrm{iid}}$, the perturbed covariance $\Sigma_0 + \alpha^2 1_n 1_n'$ is not a scalar multiple of $I_n$ and therefore lies outside $\mathcal{S}_n^{\mathrm{iid}}$ whenever $\alpha \neq 0$. The cone is too narrow for the adversary to hide mean variation inside a covariance perturbation, and consistent estimation of $\theta_n = n\sigma^2$ is restored.

## 3.2 Characterization of conservative estimands

In light of how consistent variance estimation may be impossible, we may use conservative variance estimators to avoid being oversized in inference. The following result provides necessary and sufficient conditions for a variance estimand to be conservative using the dual cone representation.

**Theorem 2** (Conservativeness via dual cones)**.** *The estimand* $V_{A,n}(P_n) = E_{P_n}[Y_n' A_n Y_n]$ *satisfies* $V_{A,n}(P_n) \geq \theta_n(P_n)$ *for every* $P_n \in \mathcal{P}_n(\mathcal{M}_n, \mathcal{S}_n)$ *if and only if*

$$(C1) \quad A_n - W_n \in \mathcal{S}_n^*, \tag{3}$$

$$(C2) \quad \mu' A_n \mu \geq 0 \text{ for all } \mu \in \mathcal{M}_n. \tag{4}$$

When $\mathcal{S}_n = \mathrm{Sym}_n^+$, self-duality gives $\mathcal{S}_n^* = \mathrm{Sym}_n^+$ and (C1) reduces to $A_n - W_n \succeq 0$ on $\mathbb{R}^n$, since $\mathrm{tr}(B\Sigma) \geq 0$ for every PSD $\Sigma$ if and only if $B$ is itself PSD. For strict subcones, (C1) is weaker, admitting estimators that exploit covariance structure to sharpen the correction. A corollary of this lemma is that the plug-in variance estimator with $W_n$ is conservative whenever $W_n$ is positive semidefinite.

However, if $A_n = W_n$ is not positive semidefinite, (C1) can still be satisfied, and (C2) is still satisfied if $\mathcal{M}_n$ is sufficiently restricted (e.g., $\mathcal{M}_n = \{0\}$). Nonetheless, if $\mathcal{M}_n = \mathbb{R}^n$, then (C2) is equivalent to positive semidefiniteness of $A_n = W_n$, so $W_n$ not being positive semidefinite implies that $V_{A,n}(P_n) < \theta_n(P_n)$ for some $\mu \in \mathcal{M}_n$. This result will be evident in CGM and $m$-dependence.

The characterization of conservativeness appears similar to the classical positive-semidefiniteness corrections in the variance-estimation literature. Proposition 1 states the connection between the estimator and the estimand: if $W_n$ is positive semidefinite, then both the estimator is always positive semidefinite, and the estimand will be valid.

**Proposition 1** (Correction of Estimand vs Estimate)**.** *Consider the plug-in* $\widehat{V}_{W,n} = Y_n' W_n Y_n$ *with target* $\theta_n(P_n) = \mathrm{tr}(W_n \Sigma(P_n))$*. The following are equivalent:*

*(i)* $\widehat{V}_{W,n} \geq 0$ *for every realization* $Y_n \in \mathbb{R}^n$ *(realized non-negativity);*

*(ii)* $E_{P_n}\left[\widehat{V}_{W,n}\right] \geq \theta_n$ *on* $\mathcal{P}_n(\mathbb{R}^n, \mathcal{S}_n)$ *for every closed convex cone* $\mathcal{S}_n \subseteq \mathrm{Sym}_n^+$*;*

*(iii)* $W_n \succeq 0$.

Beyond this similarity, the purpose and objective of the conservativeness are different. Corrections of the variance estimator are motivated by realized non-negativity: a quadratic form $Y_n' W_n Y_n$ with $W_n \not\succeq 0$ can return a negative number in a given sample, and the standard remedy enforces $\widehat{V} \geq 0$ directly. A realized-non-negativity fix acts on the *estimate*: it adjusts the random variable $\widehat{V}_{W,n}$ so that no draw falls below zero, but it does not change which estimand $V_{W,n}(P_n) = \mathrm{tr}(W_n\Sigma) + \mu' W_n \mu$ is being targeted. When the plug-in is anticonservative, the problem is that $V_{W,n}(P_n) < \theta_n(P_n)$, i.e. $\mu' W_n \mu < 0$, even though $V_{W,n}(P_n)$ itself remains positive because the covariance term $\mathrm{tr}(W_n\Sigma) = \theta_n$ dominates; a floor at zero leaves the gap $\mu' W_n \mu$ untouched. The correction developed in this paper acts instead on the *estimand*: replacing $W_n$ with an $A_n$ that satisfies the conditions of Theorem 2 changes the targeted quantity.

# 4 Optimal Conservative Variance Estimation

This section develops criteria that define an optimal variance estimator. Section 4.1 introduces these optimization criteria while Section 4.2 shows that an eigenvalue truncation (ET) solution $A_n^\star$ solves all problems under some conditions. When the conditions for ET optimality fail, these criteria can nonetheless be formulated as conic programs that can be solved numerically — the topic of Section 4.3.

## 4.1 Three optimization problems

By Theorem 2, every uniformly conservative quadratic estimator on $\mathcal{P}_n(\mathcal{M}_n, \mathcal{S}_n)$ corresponds to a matrix $A_n \in \mathrm{Sym}_n$ satisfying (C1)–(C2). Among such estimators we seek one that is "best". Three natural criteria arise: the first is a minimal correction criterion that ensures $A_n$ departs from $W_n$ as little as possible; the second is a level criterion that minimizes the variance estimand subject to validity, in the interest of the most powerful tests; and the third is the MSE, the common benchmark for the quality of an estimator.

The target $\theta_n = \mathrm{tr}(W_n\Sigma)$ is a known function of $W_n$, and the plug-in $A_n = W_n$ estimates it with mean bias $\mu' W_n \mu$. The plug-in is anticonservative precisely when $W_n \not\succeq 0$ (Theorem 2); when it is conservative, no correction is warranted. Where a correction is needed, the natural object is the one that repairs anticonservativeness while disturbing the plug-in as little as possible. We measure the disturbance by the worst-case upward adjustment of $A_n$ relative to the plug-in $W_n$ over admissible mean directions,

$$\mathcal{R}_n(A_n) := \sup_{\mu\in\mathcal{M}_n,\ \|\mu\|=1} \mu'(A_n - W_n)\mu. \tag{5}$$

**Criterion 1: Minimal Correction.** Let $P_{\mathcal{M}_n}$ denote the orthogonal projector onto $\mathcal{M}_n$. The minimal-correction problem is

$$\min_{A_n\in\mathrm{Sym}_n} \mathcal{R}_n(A_n) \qquad \text{subject to} \quad A_n - W_n \in \mathcal{S}_n^*, \quad P_{\mathcal{M}_n} A_n P_{\mathcal{M}_n} \succeq 0, \tag{6}$$

and it depends on $(\mathcal{M}_n, W_n)$ alone, so its solution applies uniformly across the model class. In the optimization problem (6) (C2) is imposed in the equivalent single-matrix form $P_{\mathcal{M}_n} A_n P_{\mathcal{M}_n} \succeq 0$. The equivalence is immediate: for every $\mu \in \mathcal{M}_n$ one has $P_{\mathcal{M}_n}\mu = \mu$, so $\mu' A_n \mu = \mu' P_{\mathcal{M}_n} A_n P_{\mathcal{M}_n} \mu$, and ranging over $\mu \in \mathcal{M}_n$ shows (C2) holds if and only if the restriction of $A_n$ to $\mathcal{M}_n$ is positive semidefinite. The projected form is the usable one for computation: it is a single linear matrix inequality rather than an infinite family of scalar constraints.

**Criterion 2: Level.** The level of the estimand targeted by $A_n$ is its expectation

$$V_{A,n}(\mu, \Sigma) = \mathbb{E}_{P_n}[Y_n' A_n Y_n] = \mathrm{tr}(A_n\Sigma) + \mu' A_n \mu = \mathrm{tr}\big(A_n(\Sigma + \mu\mu')\big),$$

and among conservative corrections one would like the one whose estimand is smallest, so as not to over-inflate the variance and lose power. The pointwise level problem at a given admissible $(\mu, \Sigma)$ is

$$\min_{A_n \in \mathrm{Sym}_n} \operatorname{tr}\big(A_n(\Sigma + \mu\mu')\big) \qquad \text{s.t.} \qquad A_n - W_n \in \mathcal{S}_n^*, \quad P_{\mathcal{M}_n} A_n P_{\mathcal{M}_n} \succeq 0. \tag{7}$$

**Criterion 3: MSE.** With $\mathrm{MSE}_n(A_n; \mu, \Sigma) := \mathbb{E}_{P_n}\big[(Y_n' A_n Y_n - \theta_n)^2\big]$, the pointwise MSE problem at a given admissible $(\mu, \Sigma)$ is

$$\min_{A_n \in \mathrm{Sym}_n} \mathrm{MSE}_n(A_n; \mu, \Sigma) \qquad \text{s.t.} \qquad A_n - W_n \in \mathcal{S}_n^*, \quad P_{\mathcal{M}_n} A_n P_{\mathcal{M}_n} \succeq 0. \tag{8}$$

Both (7) and (8) reference the unknown $(\mu, \Sigma)$, and—unlike minimal correction—neither admits a worst-case formulation over the model: along $\Sigma = c\Sigma_0$ the MSE variance term grows like $c^2$ and the level term $\operatorname{tr}(A_n \Sigma)$ grows like $c$, so the supremum over the scale-free cone $\mathcal{S}_n$ diverges for every feasible $A_n$ and cannot rank candidates. Section 4.3 resolves this by replacing the worst case with an average case.

**Remark 1** (The bias and level criteria coincide)**.** A natural objective to consider is to minimize the bias $|V_{A,n}(\mu, \Sigma) - \theta_n|$, but this problem coincides with the level problem. The bias of $A_n$ decomposes as $V_{A,n}(\mu, \Sigma) - \theta_n = \operatorname{tr}\big(A_n(\Sigma + \mu\mu')\big) - \operatorname{tr}(W_n \Sigma)$, and $\operatorname{tr}(W_n\Sigma)$ does not depend on $A_n$. Hence, minimizing the bias and minimizing the level are therefore the *same* optimization problem. This result is also intuitive: validity implies $V_{A,n}(\mu, \Sigma) \geq \theta_n$, so for a given $(\mu, \Sigma)$, the bias is simply $V_{A,n}(\mu, \Sigma) - \theta_n$, and minimizing that object is the same as minimizing $V_{A,n}(\mu, \Sigma)$, subject to validity.

**Remark 2** (A degenerate minimizer)**.** Since (6) reads $A_n$ only through its $\mathcal{M}_n$-block, it cannot exclude estimators that are cheap on $\mathcal{M}_n$ but wild off it. Take $\mathcal{S}_n^{\mathrm{diag}}$, $\mathcal{M}_n = \mathbf{1}_n^{\perp}$, and $W_n = I_n$. Because $W_n = I_n$ acts as the identity on $\mathcal{M}_n$, every feasible $A_n$ satisfies

$$\mathcal{R}_n(A_n) = \sup_{\mu \in \mathcal{M}_n,\ \|\mu\|=1} \mu'(A_n - I_n)\mu = \Big(\sup_{\mu \in \mathcal{M}_n,\ \|\mu\|=1} \mu' A_n \mu\Big) - 1 \ \geq\ -1,$$

the inequality by (C2), with equality if and only if $\mu' A_n \mu = 0$ for every $\mu \in \mathcal{M}_n$. The value $-1$ is therefore the minimum of (6), and it is attained by $A_n = \mathbf{1}_n \mathbf{1}_n'$, which is feasible — (C1) holds since $\operatorname{diag}(\mathbf{1}_n\mathbf{1}_n' - I_n) = 0$, and (C2) since $P_{\mathcal{M}_n}\mathbf{1}_n = 0$ — and which yields the trivial estimator $\widehat{V}_{A,n} = (\sum_i Y_{n,i})^2$. The minimal-correction criterion thus selects an estimator with no sampling content at all, strictly preferring it to the truncation; it is insufficient for selecting an estimator, and the variance discipline of the MSE problem (8) is what rules such minimizers out.

The MSE criterion controls sampling variance directly, eliminating the degeneracy illustrated by $(\sum_i Y_{n,i})^2$. The cost is that the objective involves the third- and fourth-moment tensors. Specifying these requires distributional assumptions beyond the second-moment structure encoded by $(\mathcal{M}_n, \mathcal{S}_n)$.

## 4.2 The eigenvalue-truncation solution

The weight matrix $W_n$ has the following spectral decomposition

$$W_n = U_n \Lambda_n U_n', \qquad \Lambda_n = \operatorname{diag}(\lambda_{n,1}, \dots, \lambda_{n,n}).$$

Write $\lambda^+ := \max\{\lambda, 0\}$ and $\lambda^- := \max\{-\lambda, 0\}$, set $\Lambda_n^\pm := \operatorname{diag}(\lambda_{n,1}^\pm, \dots, \lambda_{n,n}^\pm)$, and define

$$W_n^+ := U_n \Lambda_n^+ U_n', \qquad W_n^- := U_n \Lambda_n^- U_n', \qquad \lambda_n^{\min} := \min_j \lambda_{n,j} = \lambda_{\min}(W_n),$$

so $W_n = W_n^+ - W_n^-$, both terms are positive semidefinite, and $\lambda_n^{\min}$ is the smallest eigenvalue of $W_n$. The *eigenvalue-truncation* (ET) weight matrix is:

$$A_n^\star := W_n^+ = U_n \Lambda_n^+ U_n'. \tag{9}$$

It is identical to $W_n$ except that it truncates the negative eigenvalues at zero. $A_n^\star$ depends only on the known $W_n$; it never references $\mathcal{S}_n$ and never requires an estimate of $\Sigma$. Due to Theorem 2, $A_n^\star$ is valid for any $(\mathcal{M}_n, \mathcal{S}_n)$, since $W_n^+$ is positive semidefinite.

The following definitions and assumptions are used to characterize optimality.

**Definition 1** (Spectral class and aligned cone)**.** Fix an orthonormal eigenbasis $\{q_{n,j}\}_{j=1}^n$ of $W_n$, so that $W_n = \sum_{j=1}^n \lambda_{n,j}\, q_{n,j} q_{n,j}'$ (where $q_{n,j} = U_{n,\cdot j}$), and write

$$\Delta(a) := \sum_{j=1}^n a_j\, q_{n,j} q_{n,j}', \qquad a \in \mathbb{R}^n.$$

The *spectral class* is

$$\mathcal{A}_n^{\mathrm{sp}} := \big\{ A_n(a) := W_n + \Delta(a) \ : \ a \in \mathbb{R}^n \big\},$$

and the *aligned cone* is

$$\mathcal{S}_n^{W\text{-diag}} := \big\{ \Delta(s) : s \in \mathbb{R}_+^n \big\},$$

the covariances commuting with $W_n$.

$\mathcal{S}_n^{W\text{-diag}}$ is generated by the eigenprojectors $q_{n,j} q_{n,j}'$, so

$$\mathcal{S}_n^{W\text{-diag}} \subseteq \mathcal{S}_n \iff q_{n,j} q_{n,j}' \in \mathcal{S}_n \ \text{ for every } j. \tag{10}$$

The containment (10) is the operative restriction on the cone in (ii) and (iii) below. It is a demand on $\mathcal{S}_n$ from below, and is inherited by every larger cone.

The MSE criterion is a fourth-order object, so it cannot be evaluated from $(\mathcal{M}_n, \mathcal{S}_n)$ alone. We impose the following shape restriction, which makes the MSE a function of $(\mu, \Sigma)$ only.

**Assumption 1** (No skewness, Gaussian fourth-cumulant structure)**.** For every $P_n \in \mathcal{P}_n(\mathcal{M}_n, \mathcal{S}_n)$ the vector $Y_n$ has finite fourth moments, and $Z := Y_n - \mu$ satisfies

$$\mathbb{E}_{P_n}[Z_i Z_j Z_k] = 0 \quad \text{for all } i, j, k, \qquad \kappa_{ijkl}(P_n) = 0 \quad \text{for all } i, j, k, l,$$

where $\kappa_{ijkl}(P_n) = \mathbb{E}_{P_n}[Z_i Z_j Z_k Z_l] - \mathbb{E}[Z_i Z_j]\mathbb{E}[Z_k Z_l] - \mathbb{E}[Z_i Z_k]\mathbb{E}[Z_j Z_l] - \mathbb{E}[Z_i Z_l]\mathbb{E}[Z_j Z_k]$.

The no-skewness condition is implied by symmetry of $Z$ about zero, $Z \overset{d}{=} -Z$; the Gaussian fourth-cumulant condition is implied by Gaussianity, and for elliptical distributions $\kappa^{(4)}$ is proportional to the Gaussian tensor scaled by a kurtosis parameter. Assumption 1 restricts the *shape* of $Z$, not its location $\mu$ or scale $\Sigma$, both of which remain arbitrary within $\mathcal{M}_n \times \mathcal{S}_n$.

**Theorem 3** (Eigenvalue truncation: validity and optimality)**.** *Let $A_n^\star$ be as in (9), and assume $\mathcal{M}_n = \mathbb{R}^n$.*

*(i)* ***Minimal correction.*** *Suppose $\mathcal{S}_n \neq \{0\}$. Then, $A_n^\star = W_n^+$ solves (6) with optimal value $\mathcal{R}_n(A_n^\star) = (-\lambda_n^{\min})_+$.*

*Parts (ii)(a)–(b) and (iii) assume in addition the containment (10).*

*(ii)* ***Level.***

- *(a) For every $\Psi = \Sigma + \mu\mu' \succeq 0$, $A_n^\star$ solves the level problem (7) over the uniformly conservative members of $\mathcal{A}_n^{\mathrm{sp}}$. This solution is unique if and only if $\psi_j := q_{n,j}' \Psi q_{n,j} > 0$ for every $j$.*
- *(b) If in addition $\Psi \in \mathcal{S}_n^{W\text{-diag}}$, then $A_n^\star$ minimizes $\mathrm{tr}(A_n \Psi)$ over* all *uniformly conservative $A_n \in \mathrm{Sym}_n$, spectral or not.*
- *(c) Let $\mathcal{S}_n$ be closed. If $q_{n,k} q_{n,k}' \notin \mathcal{S}_n$ for some $k$ with $\lambda_{n,k} > 0$, then some uniformly conservative $A_n \in \mathrm{Sym}_n$ satisfies $\mathrm{tr}(A_n \Psi) < \mathrm{tr}(A_n^\star \Psi)$ at $\Psi = q_{n,k} q_{n,k}' \succeq 0$.*

*(iii)* ***MSE.*** *Suppose Assumption 1 holds. Then for every $\mu \in \mathbb{R}^n$ and every $\Sigma \in \mathcal{S}_n^{W\text{-diag}}$,*

$$\mathrm{MSE}_n(A_n^\star; \mu, \Sigma) \leq \mathrm{MSE}_n(A_n; \mu, \Sigma) \qquad \text{for every uniformly conservative } A_n \in \mathcal{A}_n^{\mathrm{sp}},$$

*so $A_n^\star$ solves (8) pointwise; the minimizer is unique if $s_j := q_{n,j}' \Sigma q_{n,j} > 0$ for every $j$.*

In (ii) and (iii), the spectral class $\mathcal{A}_n^{\mathrm{sp}}$ is the set of weight matrices built from the known $W_n$ by adjusting its eigenvalues but not its eigenvectors: every estimator in Section 5 lies in it, so it is not restrictive. Within it, ET is the Loewner-least element. By (ii)(b), at any $\Psi$ diagonal in the eigenbasis of $W_n$, no uniformly conservative $A_n$ undercuts ET. What a non-spectral $A_n$ can do is beat ET at *particular* $(\mu, \Sigma)$;

capturing that gain means orienting weight toward the unknown second moment and, for MSE, toward its split into $\mu\mu'$ and $\Sigma$, which Theorem 1 forbids learning. Parts (ii)(a) and (iii) differ only in that (iii) claims pointwise optimality in $\mathcal{S}_n^{\mathrm{W-diag}}$ while (ii) claims pointwise optimality for every $\Psi \succeq 0$.

**Remark 3** (The conditions constrain the cone)**.** The constraint (10) asks the cone to place mass on each eigendirection of $W_n$, which can be restrictive. Let the covariance model be a zero pattern, $\mathcal{S}_n = \{\Sigma \succeq 0 : \Sigma_{ij} = 0$ whenever units $i$ and $j$ are not permitted to covary$\}$. By (10), what matters is whether each $q_{n,j}q'_{n,j}$ vanishes at every forbidden cell. Under two-way clustering, units are indexed by $(g, h)$ with $g \le G$ and $h \le H$, $G, H \ge 2$; two units may covary exactly when they share $g$ or share $h$. The top eigenvector is $q = \mathbf{1}_n/\sqrt{n}$, and its projector $\mathbf{1}_n\mathbf{1}'_n/n$ is nonzero at *every* cell, including the pairs sharing neither index. Hence $qq' \notin \mathcal{S}_n$, (10) fails, and by Theorem 3(ii)(c) some uniformly conservative $A_n$ strictly undercuts ET at $\Psi = qq'$. Under a bare two-way cone, then, ET remains uniformly conservative and minimal-correction optimal, but is not level- or MSE-optimal.

A researcher unwilling to buy optimality with those covariance restrictions can instead enlarge the cone until the containment holds. Call a convex cone $\mathcal{T} \subseteq \mathrm{Sym}_n^+$ an *admissible enlargement* of $\mathcal{S}_n$ if $\mathcal{S}_n \subseteq \mathcal{T}$ and $\mathcal{T}$ satisfies (10). Adding the aligned cone produces the smallest enlargement, formalized in the next proposition.

**Proposition 2** (The minimal aligned enlargement)**.** *Let $\mathcal{M}_n = \mathbb{R}^n$, let $\mathcal{S}_n \subseteq \mathrm{Sym}_n^+$ be a convex cone, and let*

$$\mathcal{S}_n^{\oplus} := \mathcal{S}_n + \mathcal{S}_n^{W\text{-diag}} = \{\Sigma_1 + \Sigma_2 : \Sigma_1 \in \mathcal{S}_n,\ \Sigma_2 \in \mathcal{S}_n^{W\text{-diag}}\}$$

*denote their Minkowski sum.*

*(i)* ***Minimality.*** *$\mathcal{S}_n^{\oplus}$ is an admissible enlargement of $\mathcal{S}_n$, and $\mathcal{S}_n^{\oplus} \subseteq \mathcal{T}$ for every admissible enlargement $\mathcal{T}$.*

*(ii)* ***The dual price.*** *$(\mathcal{S}_n^{\oplus})^* = \mathcal{S}_n^* \cap (\mathcal{S}_n^{W\text{-diag}})^*$, where*

$$(\mathcal{S}_n^{W\text{-diag}})^* = \big\{D \in \mathrm{Sym}_n :\ q'_{n,j} D q_{n,j} \ge 0 \ \ \textit{for every } j\big\},$$

*so $A_n$ is uniformly conservative on $\mathcal{P}_n(\mathbb{R}^n, \mathcal{S}_n^{\oplus})$ if and only if it is uniformly conservative on $\mathcal{P}_n(\mathbb{R}^n, \mathcal{S}_n)$ and $q'_{n,j}(A_n - W_n)q_{n,j} \ge 0$ for every $j$. Moreover $\mathcal{T}^* \subseteq (\mathcal{S}_n^{\oplus})^*$ for every admissible enlargement $\mathcal{T}$.*

Optimality is then a corollary. Since $\mathcal{S}_n^{\oplus}$ satisfies (10), Theorem 3(ii)(a)–(b) and (iii) apply to it verbatim, so $A_n^\star = W_n^+$ solves the level problem at every $\Psi \succeq 0$ within $\mathcal{A}_n^{\mathrm{sp}}$, and, under Assumption 1, the MSE problem at every $\mu \in \mathbb{R}^n$ and every $\Sigma \in \mathcal{S}_n^{W\text{-diag}}$. The enlargement changes the constraint set, not the estimator: by (9), $A_n^\star$ never references $\mathcal{S}_n$, so any convergence result for $Y'_n A_n^\star Y_n$ established under the dependence environment that generates $\mathcal{S}_n$ continues to hold unchanged.

**Remark 4** (Optimality over a restricted dual cone)**.** Validity is unaffected. What the enlargement changes is the feasible set of (6), (7) and (8): by Proposition 2(ii) the constraint (C1) tightens from $A_n - W_n \in \mathcal{S}_n^*$ to

$$A_n - W_n \in \mathcal{S}_n^* \quad \text{and} \quad q_{n,j}'(A_n - W_n)q_{n,j} \geq 0 \ \text{ for every } j,$$

which discards exactly the competitors that undercut ET along eigendirections Theorem 3(ii)(c) constructs. Optimality of ET is therefore optimality *over a restricted dual cone*, and by Proposition 2 the restriction is the smallest one that delivers it: any other admissible enlargement discards more competitors.

An $A_n$ beating ET under $\mathcal{S}_n$ places weight on the cells $\mathcal{S}_n$ *declares* to be exactly zero. The extra dual constraint of Proposition 2(ii), $q_{n,j}'(A_n - W_n)q_{n,j} \geq 0$ for every $j$, is precisely the inequality it violates. Such an estimator draws its advantage from being valid if the zero pattern holds exactly, but fragile to misspecification of covariances. Declining that advantage is what $\mathcal{S}_n^{\oplus}$ formalizes.

The ET variance estimator is shown to converge to the intended variance estimand under some primitive conditions in Lemma 2 in the appendix. Hence, when coupled with a central limit theorem, the variance estimator can be used for robust and optimal inference. These limit theorems are often derived under particular dependence structures, so the formal statement is deferred to Section 6. We fix $A_n^\star = W_n^+$ as the closed-form benchmark.

### 4.3 Average-case criteria

Outside the regime where optimality is established, no correction dominates pointwise: the MSE- and level-minimizing $A_n$ move with $(\mu, \Sigma)$, orienting weight toward the mass of $\Sigma$ in a way Theorem 1 forbids learning. Similarly, the minimal correction problem may not be minimized at ET when $\mathcal{M}_n \neq \mathbb{R}^n$. A worst-case selection is unavailable as the entire identification problem is that the second moment $\mathbb{E}[Y_n Y_n'] = \Sigma + \mu\mu'$ does not reveal how its mass splits between $\mu\mu'$ and $\Sigma$. There is no canonical adversary, and selecting a unique estimator requires committing to *how* the residual uncertainty in $(\mu, \Sigma)$ is weighted.

We make that weighting explicit. Let $\pi$ be a researcher-declared probability measure on $\mathcal{M}_n \times \mathcal{S}_n$ with finite fourth moments, and define the $\pi$-averaged MSE and the $\pi$-averaged level,

$$\overline{\mathrm{MSE}}_n(A_n; \pi) := \int \mathrm{MSE}_n(A_n; \mu, \Sigma)\, d\pi(\mu, \Sigma), \tag{11}$$

$$\overline{V}_n(A_n; \pi) := \int \mathrm{tr}\big(A_n(\Sigma + \mu\mu')\big)\, d\pi(\mu, \Sigma) = \mathrm{tr}(A_n \Psi_\pi), \tag{12}$$

where $\Psi_\pi := \mathbb{E}_\pi[\Sigma + \mu\mu'] \succeq 0$ is the prior second moment of $Y_n$. The associated programs are

$$\min_{A_n \in \mathrm{Sym}_n} \overline{\mathrm{MSE}}_n(A_n; \pi) \qquad \text{s.t.} \qquad A_n - W_n \in \mathcal{S}_n^*, \quad P_{\mathcal{M}_n} A_n P_{\mathcal{M}_n} \succeq 0, \tag{13}$$

$$\min_{A_n \in \mathrm{Sym}_n} \mathrm{tr}(A_n \Psi_\pi) \qquad \text{s.t.} \qquad A_n - W_n \in \mathcal{S}_n^*, \quad P_{\mathcal{M}_n} A_n P_{\mathcal{M}_n} \succeq 0. \tag{14}$$

The measure $\pi$ requires no estimation of $\mu$ or $\Sigma$ and never invokes the extreme rays of $\mathcal{S}_n$. Averaging names no adversary, so both (13) and (14) remain finite. Notably, the choice of $\pi$ is irrelevant for validity, because the constraints are satisfied by the $A_n$ that solves the programs even if the $\pi$ is misspecified.

The minimal-correction criterion does not require a declared $\pi$, because its objective $\mathcal{R}_n(A_n) = \sup_{\mu \in \mathcal{M}_n, \|\mu\|=1} \mu'(A_n - W_n)\mu$ depends only on $(\mathcal{M}_n, W_n)$, both known, and its worst case is taken over the compact unit sphere of $\mathcal{M}_n$. The optimization problem for this criterion directly is also conic.

**Proposition 3** (Conicity of the optimization problems)**.** *Suppose $\mathcal{S}_n$ is a closed convex cone and $\mathcal{M}_n$ is a linear subspace.*

*(i) The minimal-correction problem (6) is a convex conic program in $A_n$.*

*(ii) Let $\pi$ be a probability measure on $\mathcal{M}_n \times \mathcal{S}_n$ with finite fourth moments, and suppose that for $\pi$-almost every $(\mu, \Sigma)$ the admissible moment tensors $(m^{(3)}, \kappa^{(4)})$ are such that $\mathrm{MSE}_n(A_n; \mu, \Sigma)$ is convex in $A_n$. Then the $\pi$-averaged MSE program (13) is a convex conic program in $A_n$.*

*(iii) Let $\pi$ be a probability measure on $\mathcal{M}_n \times \mathcal{S}_n$ with finite second moments. Then the $\pi$-averaged level program (14) is a* linear *conic program in $A_n$.*

Conicity is practically attractive: any specific instance of these three programs can be solved with interior-point solvers once the problem data $(W_n, \mathcal{M}_n, \mathcal{S}_n)$ and the moment-tensor class are specified. The conic formulation is the fallback when no closed form is available, say, when the conditions of Theorem 3 do not hold, and restores a computable optimum.

Under the conditions of Theorem 3, ET is also the optimal solution for these optimization problems regardless of the choice of $\pi$. Since ET is pointwise optimal for *every* $\pi$, it continues to be optimal after integrating over $\pi$. The average-case criterion is thus a single criterion across regimes.

However, unlike the ET solution, it is in general more difficult to show that the variance estimators obtained from solving numerical conic programs converge to the intended estimand.

# 5 Examples and Optimal Corrections

This section explains how the framework's optimization problem applies to standard and new variance estimators. Each subsection states the weight matrix $W_n$, the admissible cone $\mathcal{S}_n$, and the optimal correction $A_n^\star$.

### 5.1 One-way cluster-robust estimators

With observations partitioned into clusters $G_1, \dots, G_{C_n}$ and indicator vectors $B_c \in \mathbb{R}^n$, the one-way cluster-robust estimator is

$$\widehat{V}_{\text{clu},n} = \sum_{c=1}^{C_n} \left( \sum_{i \in G_c} Y_{n,i} \right)^2 = Y_n' W_n^{\text{clu}} Y_n, \qquad W_n^{\text{clu}} = \sum_{c=1}^{C_n} B_c B_c'.$$

Each $B_c B_c'$ is positive semidefinite, so $W_n^{\text{clu}}$ is globally PSD, and the plug-in is uniformly conservative for every admissible mean space. Hence $A_n^\star = W_n^{\text{clu}}$, and heterogeneous means cannot induce anticonservativeness when the weight matrix is globally positive semidefinite. This result shows how the plug-in one-way cluster robust variance estimator is not only conservative in design-based settings (as established in Abadie et al. (2023)), it is also optimal.

### 5.2 Time series: HAC estimators

A generic HAC estimator is the quadratic form

$$\widehat{V}_{\text{HAC},n} = \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{i-j}{b_n}\right) Y_{n,i} Y_{n,j} = Y_n' W_n^{\text{HAC}} Y_n, \qquad (W_n^{\text{HAC}})_{ij} = K\left(\frac{i-j}{b_n}\right),$$

with $W_n^{\text{HAC}}$ symmetric Toeplitz. Whether the plug-in $A_T^\star = W_n^{\text{HAC}}$ is already optimal turns entirely on the *spectral window* of the kernel. For a symmetric banded Toeplitz matrix $B$ with value $c_k$ on its $k$-th diagonal, the spectral window is the generating function

$$f(\omega) := c_0 + 2 \sum_{k \geq 1} c_k \cos(k\omega), \qquad \omega \in [-\pi, \pi],$$

the Fourier transform of the kernel weights. Two facts connect $f$ to the eigenvalues of $B$: every eigenvalue lies in the range of $f$ (each Rayleigh quotient is a weighted average of $f$ values), and by Grenander–Szegő the empirical eigenvalue distribution converges to that of $f$. The sign of $f$ therefore governs whether the plug-in is conservative. We contrast two kernels.

#### 5.2.1 Naive m-dependence plug-in

Consider a scalar time series $\{Y_t\}_{t=1}^T$ that is $m$-dependent, so $\mathrm{Cov}(Y_t, Y_{t+k}) = 0$ for $|k| > m$. The variance of the sum is

$$\mathrm{Var}\left(\sum_{t=1}^{T} Y_t\right) = \sum_t \mathrm{Var}(Y_t) + \sum_{k=1}^{m} \sum_t \big[\mathrm{Cov}(Y_t, Y_{t+k}) + \mathrm{Cov}(Y_{t+k}, Y_t)\big].$$

The naive plug-in replaces each covariance with the product $Y_t Y_{t+k}$ and weights all admissible lags equally,

$$\widehat{V}_T^{\text{naive}} = \sum_t Y_t^2 + \sum_{k=1}^{m} \sum_t \big(Y_t Y_{t+k} + Y_{t+k} Y_t\big) = Y' B_m Y,$$

where $B_m \in \mathrm{Sym}_T$ is the symmetric banded Toeplitz matrix with ones on bands $0, 1, \ldots, m$. Its spectral window is the Dirichlet kernel of order $m$,

$$f_{\mathrm{Dir}}(\omega) = 1 + 2 \sum_{k=1}^{m} \cos(k\omega) = \frac{\sin\big((m + \frac{1}{2})\omega\big)}{\sin(\omega/2)},$$

which oscillates in sign, vanishing at $\omega = 2\pi j/(2m + 1)$, $j = 1, \ldots, m$, and taking negative values between successive zeros. A positive fraction of the eigenvalues of $B_m$ are negative, so $B_m$ is indefinite. Two consequences follow. First, the realized $Y' B_m Y$ can be negative in a given sample regardless of the mean—the classical defect that motivates a positive-semidefinite kernel. Second, under the agnostic mean space $\mathcal{M}_n = \mathbb{R}^T$ the negative eigendirections lie inside $\mathcal{M}_n$, so the plug-in estimand is anticonservative for means loading on them. Due to Theorem 2, the variance estimand can be anticonservative — an example is provided in Appendix A. The eigenvalue truncation $A_T^\star = B_m^+$ addresses both issues.

The optimal correction is the positive-part operator

$$A_T^\star = B_m^+ = U \max(\Lambda, 0) U',$$

where $B_m = U \Lambda U'$ is the spectral decomposition. In the frequency domain, the optimal estimator has spectral window $f^\star(\omega) = \max\{f_{\mathrm{Dir}}(\omega), 0\}$, modifying $B_m$ only on the frequency band where $f_{\mathrm{Dir}} < 0$.

#### 5.2.2 Newey–West with Bartlett kernel

The Newey–West estimator replaces uniform weights by linearly declining Bartlett weights $\omega(k, M) = 1 - k/(M + 1)$ for $|k| \leq M$,

$$\widehat{V}_T^{\mathrm{NW}} = Y' C_M Y, \qquad (C_M)_{ts} = \omega(|t - s|, M),$$

with $C_M \in \mathrm{Sym}_T$ banded Toeplitz. Its spectral window is the Fejér kernel,

$$f_{\mathrm{Fej}}(\omega) = \frac{1}{M+1}\left(\frac{\sin((M+1)\omega/2)}{\sin(\omega/2)}\right)^2 \geq 0,$$

the squared modulus of a Dirichlet kernel, hence nonnegative everywhere. Because every eigenvalue of a symmetric Toeplitz matrix lies in the range of its spectral window, nonnegativity of $f_{\mathrm{Fej}}$ forces $C_M \succeq 0$ exactly, in every finite sample—the positive-semidefiniteness for which the Bartlett kernel was introduced (Newey and West, 1987). No truncation is ever required: $A_T^\star = C_M$, and the identification problem of this paper does not bind in the Bartlett–HAC setting.

### 5.3 Two-way clustering

We work in a balanced panel with $G$ groups and $T$ periods, one observation per cell, so $n = GT$, and identify $\mathbb{R}^n \cong \mathbb{R}^T \otimes \mathbb{R}^G$. With row clusters indexed by $g$ and column clusters by $t$, the Cameron–Gelbach–Miller (CGM) weight matrix is

$$W_n^{\text{2-way}} = W_n^G + W_n^T - W_n^{G\cap T}, \qquad W_n^G = J_T \otimes I_G, \quad W_n^T = I_T \otimes J_G, \quad W_n^{G\cap T} = I_n,$$

where $J_G = \mathbf{1}_G\mathbf{1}_G'$ and $J_T = \mathbf{1}_T\mathbf{1}_T'$. Each component is positive semidefinite, but the subtraction renders $W_n^{\text{2-way}}$ indefinite, so under heterogeneous means the plug-in is anticonservative (Xu and Yap, 2024)—a direct instance of the converse in Theorem 2 at $\mathcal{M}_n = \mathbb{R}^n$. Let

$$P_{00} := I_n - \frac{W_n^G}{T} - \frac{W_n^T}{G} + \frac{J_n}{n}, \qquad J_n = \mathbf{1}_n\mathbf{1}_n',$$

be the orthogonal projector implementing two-way demeaning, $(P_{00}\, y)_{gt} = y_{gt} - \bar{y}_{\cdot t} - \bar{y}_{g\cdot} + \bar{y}_{\cdot\cdot}$, and the associated subspace $\mathcal{S}_{00} := \mathrm{range}(P_{00})$.

**Corollary 1** (Optimal two-way correction)**.** *Under $\mathcal{M}_n = \mathbb{R}^n$, the eigenvalue-truncation estimator of Theorem 3 is*

$$A_n^\star = (W_n^{\text{2-way}})^+ = W_n^{\text{2-way}} + P_{00},$$

*equivalently* $Y_n' A_n^\star Y_n = Y_n' W_n^{\text{2-way}} Y_n + \sum_{g,t}\left(Y_{gt} - \bar{Y}_{\cdot t} - \bar{Y}_{g\cdot} + \bar{Y}_{\cdot\cdot}\right)^2$.

The CGM matrix subtracts $W_n^{G\cap T} = I_n$ to undo the double-counting of cells entering both the row-cluster and column-cluster sums. Consequently, on the two-way within subspace (vectors $v$ orthogonal to every row and column mean), we have $W_n^{\text{2-way}} v = -v$, resulting in a negative eigenvalue of $-1$. Truncating its eigenvalue to zero adds exactly the projector $P_{00}$ onto that subspace. Since the within subspace is precisely where heterogeneous means generate anticonservativeness (Theorem 2), the correction places nonnegative mass only where the plug-in was deficient.

CGM2 uses weight matrix $W_n^G + W_n^T = A_n^\star + (I_n - P_{00})$. Since $I_n - P_{00} \succeq 0$, CGM2 is uniformly conservative but non-optimal: it overshoots $A_n^\star$ on the between-group and between-time directions, where no correction is needed. This overshoot is the same order as the correction when means carry between-cluster heterogeneity, so there is a finite-sample effect, visible in the simulations of Section 6.5, where ET delivers tighter intervals than the conservative C2NW (Appendix A.3) while retaining size control.

### 5.4 Two-way clustering with serial correlation

We now allow cross-cluster serial correlation on the time dimension. In the balanced panel ($n = GT$), the Chiang–Hansen–Sasaki (CHS) plug-in is

$$W_n^{\mathrm{CHS}} = J_T \otimes I_G + C_M \otimes (J_G - I_G),$$

where $C_M \in \mathbb{R}^{T\times T}$ is the Bartlett kernel matrix, $(C_M)_{ts} = \omega(|t-s|, M)$ with $\omega(m, M) = \max\{1 - m/(M+1), 0\}$. This result is stated in Lemma 1 of Appendix A. Let $P_G := J_G/G$ and $Q_G := I_G - P_G$ be the group-mean and group-demeaning projectors, and define the indefinite $T \times T$ block

$$B := J_T - C_M, \qquad B^- := (C_M - J_T)_+ \succeq 0,$$

the negative eigenmass of $B$. Under $\mathcal{M}_n = \mathbb{R}^n$, $A_n^\star = (W_n^{\mathrm{CHS}})^+$ by Theorem 3. Due to Theorem 2, the naive CHS variance estimator can be anticonservative, as exemplified in Appendix A.

**Corollary 2** (Optimal correction under serial correlation)**.**

$$A_n^\star = (W_n^{\mathrm{CHS}})^+ = W_n^{\mathrm{CHS}} + B^- \otimes Q_G,$$

*equivalently, with $\check{Y}_{g,t} := Y_{g,t} - \bar{Y}_{\cdot t}$ the group-demeaned (cross-sectionally centered) score,*

$$Y_n' A_n^\star Y_n = Y_n' W_n^{\mathrm{CHS}} Y_n + \sum_{g=1}^{G} \check{Y}_{g,\cdot}'\, B^-\, \check{Y}_{g,\cdot}.$$

Split the cross-section into the group-mean direction $P_G$ and its complement $Q_G$; the CHS weight acts block-diagonally across them. On the group-symmetric block the weight is $J_T + (G-1)C_M$, a nonnegative combination of two PSD matrices, hence already conservative—no correction. On the group-demeaned block the weight is $J_T - C_M$: the rank-one temporal aggregator $J_T$, concentrated on the constant-in-time direction, is partly canceled by the Fejér kernel $C_M$, producing negative eigenvalues. Only this block needs repair, and the truncation adds back precisely its negative eigenmass $B^-$, supported on the group-demeaned scores. The correction is thus a single $T \times T$ eigendecomposition applied within clusters. At $M = 0$, $C_M = I_T$ and

$B^- = I_T - J_T/T$ is the time-demeaning projector, so $B^- \otimes Q_G$ reduces to two-way demeaning and we recover Corollary 1.

**Remark 5** (EVC floors the estimate; ET repairs the estimand)**.** Chiang et al. (2024) render their plug-in non-negative through an eigenvalue correction (EVC) that replaces the negative eigenvalues of the *realized* meat $\widehat{\Omega}_{NT}$ by zero. In the notation here this acts on the data-space realization $Y_n' W_n^{\mathrm{CHS}} Y_n$ (scalar case: $\max\{Y_n' W_n^{\mathrm{CHS}} Y_n, 0\}$), addressing (i) of Proposition 1. It does not act on $W_n^{\mathrm{CHS}}$, so it cannot address (ii). At the plug-in covariance the bias-free part $\mathrm{tr}(W_n^{\mathrm{CHS}}\Sigma_n) > 0$, so the estimand $V_{\mathrm{CHS}} = \mathrm{tr}(W_n^{\mathrm{CHS}}\Sigma_n) + \mu_n' W_n^{\mathrm{CHS}} \mu_n$ is positive and $\widehat{\Omega}_{NT} \to V_{\mathrm{CHS}} > 0$, leaving the correction asymptotically inactive—yet $V_{\mathrm{CHS}} < V_{\mathrm{true}}$ whenever $\mu_n' W_n^{\mathrm{CHS}} \mu_n < 0$ (the counterexample of Appendix A).

# 6 Inference under Panel Dependence

This section develops the inferential machinery needed to implement variance estimators in this setting of Section 5.4.

## 6.1 Dependence framework

Consider a two-way cluster dependent setting, where one dimension exhibits cross-cluster weak dependence. To keep the exposition concrete, suppose we have a panel with both a cross sectional and a time series component. The time series component contains time indices $t \in \{1, 2, \cdots, T\}$ for every observation, and observations within the same time period may be arbitrarily correlated. The cross sectional component can be partitioned on some dimension $G$, with cluster indices $g \in \{1, 2, \cdots, G\}$ such that every observation within the same $g$ cluster may be arbitrarily correlated. Let $i \in \{1, 2, \cdots, n\} =: N_n$ index observations, $g(i)$ denote its cross sectional cluster, and $t(i)$ denote its time period. For example, in a canonical balanced panel with each cross sectional unit observed in every time period, we have $i = (g, t)$. While the setting in this paper permits unbalanced panels, the balanced panel with $G \asymp T$ is used as a running example.

There is a triangular array of multivariate random vectors $\{Y_{n,i}\}, n \geq 1, Y_{n,i} \in \mathbb{R}^v$, where $n$ denotes the number of observations. Define the moment of this random vector as $\|Y_{n,i}\|_p = (E[\|Y_{n,i}\|^p])^{1/p}$. All observations $i$ may be serially correlated. The distance $d_n(i, j)$ between any two observations $i$ and $j$ is 0 if they share a cluster on any dimension, and the difference in their time index otherwise.

Let $N_t^T$ denote the set of observations in time $t$, $N_g^G$ denote the set of observations in cluster $g$, and $N_{t,g}^{T\cap G} := N_t^T \cap N_g^G$ the set of observations in both time $t$ and cluster $g$. Define the following objects to measure how concentrated observations are across clusters and time:

$$\delta_n^{\partial}(s;k) := \frac{1}{n}\sum_{i\in N_n}\left(|N^T_{t(i)+s}| + |N^T_{t(i)-s}| + \mathbf{1}\{s=0\}(|N^G_{g(i)}| - |N^T_{t(i)}|)\right)^k; \tag{15}$$

$$\Delta_n(s,m;k) := \frac{1}{n}\sum_{i\in N_n}\left(\sum_{h=s}^{m}(|N^T_{t(i)+h}| + |N^T_{t(i)-h}| + \mathbf{1}\{h=0\}(|N^G_{g(i)}| - |N^T_{t(i)}|))\right)^k \tag{16}$$

$$c_n(s,m;k) := \inf_{\alpha>1}[\Delta_n(s,m;k\alpha)]^{1/\alpha}[\delta_n^{\partial}(s;\tfrac{\alpha}{\alpha-1})]^{1-1/\alpha}. \tag{17}$$

These objects can be interpreted in the following manner. $\delta_n^{\partial}(s;k)$ tells us the average number of observations that are $s$ periods away, while $\Delta_n(s,m;k)$ tells us the number of observations between $m$ and $s$ periods away (inclusive), both raised to the power of $k$. Finally, $c_n(s,m;k)$ gives us an upper bound on the convolution of $\delta_n^{\partial}(s;k)$ and $\Delta_n(s,m;k)$ when Hölder's inequality is applied, which is interpretable as a measure of how quickly neighborhoods grow and overlap. In the running example of a balanced panel, $c_n(s,m;k) \asymp (m-s)^k G^k$.

The rest of this section characterizes $\psi$-dependence, which is the sense in which observations are weakly dependent across clusters and across time. Let $\mathcal{L}_{v,a}$ denote the collection of bounded Lipschitz real functions on $\mathbb{R}^{v\times a}$:

$$\mathcal{L}_{v,a} = \{f : \mathbb{R}^{v\times a} \to \mathbb{R} : \|f\|_\infty < \infty, Lip(f) < \infty\}, \tag{18}$$

where $Lip(f)$ is the Lipschitz constant of $f$ and $\|\cdot\|_\infty$ is the sup-norm. For any positive integers $a,b,s$, let $\mathcal{Q}_n(a,b;s)$ denote the collection of all pairs of observation sets of sizes $a$ and $b$ at distance at least $s$:

$$\mathcal{Q}_n(a,b;s) = \{(A,B) : A,B \subset N_n, |A| = a, |B| = b, \text{and } d_n(A,B) \geq s\}, \tag{19}$$

where $d_n(A,B) = \min_{i\in A}\min_{j\in B} d_n(i,j)$.

Inference is conditional on a $\sigma$-field $\mathcal{C}_n$ that holds fixed the information common to the array—for instance the cluster- and time-level common factors, or, in the design-based reading, the fixed potential outcomes. All expectations, variances, and covariances in this section and Appendix A.4 are taken conditional on $\mathcal{C}_n$, and the dependence coefficients $\theta_n$ below are $\mathcal{C}_n$-measurable. Write $\mu_n := \mathbb{E}[Y_n \mid \mathcal{C}_n]$ and $\Sigma_n := \mathrm{Var}(Y_n \mid \mathcal{C}_n)$ for the conditional first and second moments; these are the $\mathcal{C}_n$-measurable analogues of the $\mu, \Sigma$ of Section 2. The dependence coefficients $\{\theta_{n,s}\}_{s\geq 0}$ below (notation following KMS) are unrelated to the scalar variance target $\theta_n = \mathrm{tr}(W_n\Sigma)$ of Section 2, which does not recur in this section.

**Definition 2** ($\psi$-dependence)**.** A triangular array $\{Y_{n,i}\}, n \geq 1, Y_{n,i} \in \mathbb{R}^v$ is conditionally $\psi$-dependent if for each $n \in \mathbb{N}$, there exist a measurable sequence $\theta_n = \{\theta_{n,s}\}_{s\geq 0}, \theta_{n,0} = 1$, and a collection of nonrandom functions $(\psi_{a,b})_{a,b\in\mathbb{N}}, \psi_{a,b} : \mathcal{L}_{v,a} \times$

$\mathcal{L}_{v,b} \to [0,\infty)$, such that for all $(A,B) \in \mathcal{Q}_n(a,b;s)$ with $s > 0$ and all $f \in \mathcal{L}_{v,a}$ and $g \in \mathcal{L}_{v,b}$,

$$|\mathrm{Cov}(f(Y_{n,A}), g(Y_{n,B}) \mid \mathcal{C}_n)| \le \psi_{a,b}(f,g)\, \theta_{n,s} \quad \text{a.s.} \tag{20}$$

where $Y_{n,A} := (Y_{n,i})_{i \in A}$. We call the sequence $\theta_n$ the dependence coefficients of $\{Y_{n,i}\}$.

This $\psi$-dependence setting generalizes the more familiar strong-mixing processes used in the two-way cluster dependence panel literature with serial correlation.

The setup here generalizes the CHS setting that requires a representation where $Y_{gt} = f(\alpha_g, \gamma_t, \varepsilon_{gt})$, for iid $\alpha_g, \varepsilon_{gt}$ and $\gamma_t$ strictly stationary, which generalizes a standard Aldous–Hoover representation that additionally requires $\gamma_t$ iid. The following example is a DGP that satisfies the conditions of this paper and has homogeneous means but does not permit the above representation.

**Example 2.** Suppose every firm $g$ is observed for $T$ time periods, each indexed by $t$. Data is generated from the following process. There is an innovation sequence $(\nu_{gt})_{t\in\mathbb{Z}}$ iid $N(0,1)$ over $g,t$. For some $|\rho| \in (0,1)$, $D_{gt} = \rho D_{gt-1} + \nu_{gt}$, $t \in \mathbb{Z}$, with $D_{gt}$ stationary and processes independent across $g$. Hence, within a firm, $\{D_{gt}\}_t$ is an AR(1) Markov chain with parameter $\rho$, and across firms, $\{D_{gt}\}_g$ are independent copies of this AR(1) process.

Such a process does not permit the CHS representation. Suppose to a contradiction that $D_{gt}$ admits a representation where $D_{gt} = f(\alpha_g, \gamma_t, \varepsilon_{gt})$ for iid $\alpha_g, \varepsilon_{gt}$, and strictly stationary $\gamma_t$. In the DGP, $\gamma_t = 0$, because if $\gamma_t$ is stochastic, then there must be dependence across $g$ units, which is ruled out in this DGP. With the resulting representation $D_{gt} = f(\alpha_g, 0, \varepsilon_{gt})$, it follows that $\mathrm{Cov}(D_{gt}, D_{gs_1}) = \mathrm{Cov}(D_{gt}, D_{gs_2})$ for $t \ne s_1 \ne s_2 \ne t$. However, this is untrue in the DGP as $\mathrm{Cov}(D_{g,t}, D_{g,t-1}) = \rho/(1-\rho^2) \ne \rho^2/(1-\rho^2) = \mathrm{Cov}(D_{g,t}, D_{g,t-2})$.

## 6.2 Central limit theorem

The results are stated in a setting where $\sum_{i\in N_n} E[Y_{n,i}] = 0$, and we are interested in constructing the variance of $\sum_{i\in N_n} Y_{n,i}$. With $V_{\text{true}} := Var(\sum_{i\in N_n} Y_{n,i})$, let $\lambda_n := \lambda_{\min}(V_{\text{true}})$ denote its smallest eigenvalue.

**Assumption 2.** The triangular array $\{Y_{n,i}\}$ is $\psi$-dependent with the dependence coefficients $\{\theta_n\}$ satisfying: (a) For some constant $C > 0$, $\psi_{a,b}(f,g) \le C \times ab(\|f\|_\infty + Lip(f))(\|g\|_\infty + Lip(g))$; (b) $\sup_{n\ge1} \max_{s\ge1} \theta_{n,s} < \infty$; (c) For some $p > 4$, $\sup_{n\ge1} \max_{i\in N_n} \|Y_{n,i}\|_p < \infty$ a.s.

**Assumption 3.** There exists a positive sequence $m_n \to \infty$ such that for $k = 1, 2$, and the $p$ in Assumption 2,

(a) $\frac{n}{\lambda_n^{1+k/2}} \sum_{s\ge0} c_n(s, m_n; k) \theta_{n,s}^{1-(2+k)/p} \to 0$; and

(b) $\frac{n^2 \theta_{n,m_n}^{1-(1/p)}}{\lambda_n^{1/2}} \to 0$.

These conditions balance the dependence strength, neighborhood growth, and variance growth. In an AR(1) process with positive coefficient $\rho$, $\lambda_n \asymp T$, and both conditions are satisfied with $m_n = T^{1/3}$ as $T \to \infty$. In the running balanced panel example with common random effects within clusters and an AR(1) process in the common time effects, $\lambda_n \asymp T^2 G$, and again $m_n = T^{1/3}$ suffices.

**Theorem 4.** *Suppose Assumptions 2 and 3 hold a.s. Let $\nu$ be a nonstochastic vector with $\dim(\nu) = \dim(Y_{n,i})$ and $\|\nu\| = 1$. Then, for $S_n := \sum_{i \in N_n} \nu'(Y_{n,i} - E[Y_{n,i}])$ and $\sigma_n^2 := Var(S_n)$,*

$$\sup_{t \in \mathbb{R}} \left| \Pr\left\{ \frac{S_n}{\sigma_n} \le t \right\} - \Phi(t) \right| \xrightarrow{a.s.} 0, \tag{21}$$

*where $\Phi$ denotes the distribution function of $N(0,1)$.*

This result suffices for a uniform multivariate central limit theorem due to the Cramer–Wold device.

## 6.3 Consistency

By Corollary 2, the optimal variance estimator is the exact eigenvalue truncation $(W_n^{\text{CHS}})^+ = W_n^{\text{CHS}} + B^- \otimes Q_G$, with $B^- = (C_M - J_T)_+$ and $Q_G = I_G - J_G/G$. In estimator form,

$$\widehat{V}^{\star}_{\text{ET}} := \widehat{V}_{\text{CHS}} + \sum_{g=1}^{G} \check{Y}'_{n,(g,\cdot)} \, B^- \, \check{Y}_{n,(g,\cdot)}, \qquad \check{Y}_{n,(g,t)} := Y_{n,(g,t)} - \bar{Y}_{\cdot t}, \tag{22}$$

where $\check{Y}_{n,(g,t)}$ is the score demeaned across groups within period $t$ and $\check{Y}_{n,(g,\cdot)} := (\check{Y}_{n,(g,t)})_{t=1}^{T}$. Equivalently $\widehat{V}^{\star}_{\text{ET}} = Y_n'(W_n^{\text{CHS}} + B^- \otimes Q_G)Y_n$. The correction is the within-group HAC of the cross-sectionally centered scores, using the exact negative eigenmass $B^-$ of the indefinite temporal block $B = J_T - C_M$ as kernel; computing it is a single $T \times T$ eigendecomposition applied within clusters. The corresponding estimand is

$$V^{\star}_{\text{ET}} := \mathbb{E}[\widehat{V}^{\star}_{\text{ET}} \mid \mathcal{C}_n] = V_{\text{CHS}} + \sum_{g=1}^{G} \mathbb{E}\big[\check{Y}'_{n,(g,\cdot)} \, B^- \, \check{Y}_{n,(g,\cdot)} \mid \mathcal{C}_n\big]. \tag{23}$$

Using $y_t := \sum_{i \in N_t^T} Y_{n,i}$, we compare the estimand to the kernel-adjusted variance

$$\begin{aligned} V_{\text{adj}} := &\sum_{i \in N_n} \sum_{j \in N^G_{g(i)}} \text{Cov}(Y_{n,i}, Y_{n,j}) + \sum_{i \in N_n} \sum_{j \in N^T_{t(i)}} \text{Cov}(Y_{n,i}, Y_{n,j}) - \sum_{i \in N_n} \sum_{j \in N^{T \cap G}_{t(i),g(i)}} \text{Cov}(Y_{n,i}, Y_{n,j}) \\ &+ \sum_{m=1}^{M} \sum_{t=1}^{T-m} \omega(m, M) \left[\text{Cov}(y_t, y_{t+m}) + \text{Cov}(y_{t+m}, y_t)\right], \end{aligned} \tag{24}$$

which will be shown to be asymptotically equivalent to $V_{\text{true}}$. This $V_{\text{adj}}$ is $\operatorname{tr}(W_n^{\text{CHS}}\Sigma_n)$, the truncated kernel covariance: the variance part of the CHS plug-in ($\mathbb{E}[\widehat{V}_{\text{CHS}}] = V_{\text{adj}} + \mu_n' W_n^{\text{CHS}}\mu_n$), and the quantity $\widehat{V}_{\text{ET}}^{\star}$ targets.

We state the assumptions for the consistency theorem.

**Assumption 4.** $\frac{n}{\lambda_n^2}\sum_{s\geq 0} c_n(s, M; 2)\, \theta_{n,s}^{1-4/p} \to 0$.

This assumption mimics Assumption 4.1(iii) in KMS. In the balanced panel the left-hand side is $\asymp M^2/T^2$, so the condition reduces to $M = o(T)$; the choice $M \asymp T^{1/3}$ used in the simulations and application satisfies it.

**Assumption 5.** The following hold almost surely:

(a) $\frac{n}{\lambda_n}\sum_{m=1}^{M} |\omega(m, M) - 1| \delta_n^{\partial}(m; 1)\theta_{n,m}^{1-2/p} \to 0$;

(b) $\frac{1}{\lambda_n}\sum_{m\geq 1}\sum_{t=1}^{T-m}\sum_{g=1}^{G} |N_{t,g}^{T\cap G}||N_{t+m,g}^{T\cap G}|\theta_{n,m}^{1-2/p} \to 0$;

(c) $\frac{1}{\lambda_n}\sum_{m=M+1}^{T-1}\sum_{t=1}^{T-m} |N_t^T||N_{t+m}^T|\theta_{n,m}^{1-2/p} \to 0$.

In the running AR(1) examples with $\rho \in (0, 1)$ and the triangular kernel, all three conditions are implied by $\rho^m$ decay and the balanced panel structure (where $|N_{t,g}^{T\cap G}| = 1$).

**Theorem 5** (Consistency of the eigenvalue-truncation estimator)**.** *The following hold.*

*(a) $V_{\text{ET}}^{\star} - V_{\text{adj}}$ is positive semidefinite.*

*(b) Under Assumptions 2 and 4, and provided $\lambda_{\min}(V_{\text{ET}}^{\star})/\lambda_n \geq 1+o(1)$ and $Mn/\lambda_n \to 0$,*

$$(V_{\text{ET}}^{\star})^{-1}\, \widehat{V}_{\text{ET}}^{\star} \xrightarrow{p} I_v.$$

*(c) Under Assumptions 2 and 5, $V_{\text{true}}^{-1}V_{\text{adj}} \to I_v$.*

Observe that (c) implies $\lambda_{\min}(V_{\text{adj}})/\lambda_n = 1 + o(1)$, so by (a) $\lambda_{\min}(V_{\text{ET}}^{\star})/\lambda_n \geq 1 + o(1)$. The condition $Mn/\lambda_n \to 0$ in (b) is not an additional restriction: in the balanced panel $n = GT$ and $\lambda_n \asymp T^2 G$, so $Mn/\lambda_n = M/T$, which is exactly what Assumption 4 reduces to ($M = o(T)$, satisfied by $M \asymp T^{1/3}$). Both auxiliary conditions in (b) are therefore automatically satisfied under the assumptions.

**Corollary 3** (Asymptotically valid inference)**.** *Let $\nu \in \mathbb{R}^v$ with $\|\nu\| = 1$, and suppose the conditions of Theorem 4 and of Theorem 5 hold. Then, for $S_n = \sum_{i\in N_n} \nu'(Y_{n,i} - \mathbb{E}[Y_{n,i}])$ and any $\alpha \in (0, 1)$,*

$$\limsup_{n\to\infty}\ \Pr\left\{\frac{|S_n|}{\sqrt{\nu'\widehat{V}_{\text{ET}}^{\star}\nu}} > z_{1-\alpha/2}\right\} \ \leq\ \alpha,$$

*with equality if and only if $\nu'(V_{\text{ET}}^{\star} - V_{\text{true}})\nu = o\big(\nu' V_{\text{true}}\nu\big)$.*

## 6.4 Extension to regression

While the theoretical exposition focuses on means, the extension to the method of moments (which includes standard regressions) is straightforward. With a regression equation $Y_i = X_i'\beta + u_i$, and array $\{(X_i', u_i)\}$ satisfying $\psi$-dependence, as long as $\frac{1}{n}\sum_{i \in N_n} X_i X_i'$ has a minimum eigenvalue bounded away from zero, and $\{X_i u_i\}$ and $\{\text{vec}(X_i X_i')\}$ satisfy the conditions of Theorem 4, the OLS estimator $\widehat{\beta}$ is asymptotically normal and consistent. The variance of the coefficient estimator is computed by applying $\widehat{V}^{\star}_{\text{ET}}$ to the estimated score $X_i \widehat{u}_i$ in place of $Y_{n,i}$. Standard OLS arguments then apply when limit theorems hold. Writing $X_i \widehat{u}_i = X_i u_i - X_i X_i'(\widehat{\beta} - \beta)$, the feasible estimator differs from its infeasible counterpart $\sum_{i,j}(A_n^{\star})_{ij}(X_i u_i)(X_j u_j)'$ only through terms linear and quadratic in $\widehat{\beta} - \beta$, which the conditions of Theorem 4 render negligible provided $\max_i \|X_i\|^2 \|\widehat{\beta} - \beta\|$ is negligible at the CLT rate.

In the numerical results that follow, we additionally augment the methods with "C2NW", which is a natural extension of CGM2 to cross-cluster serial correlation. This method is conservative, but is not necessarily optimal — its details are relegated to Appendix A.

## 6.5 Simulation

The simulations show how the variance estimators that are robust to heterogeneous means compare with existing methods, and isolate the regime in which the optimal correction delivers a concrete advantage. Data are generated from the linear model

$$Y_{gt} = \beta_0 + (\beta_1 + \eta_{gt})\, X_{gt} + \delta_{gt} + U_{gt}, \tag{25}$$

where the regressor and disturbance follow the CHS variance-component structure,

$$\begin{aligned} X_{gt} &= w_\alpha \alpha_g^x + w_\gamma \gamma_t^x + w_\varepsilon \varepsilon_{gt}^x, \\ U_{gt} &= w_\alpha \alpha_g^u + w_\gamma \gamma_t^u + w_\varepsilon \varepsilon_{gt}^u, \end{aligned} \tag{26}$$

with $(\beta_0, \beta_1) = (0.1, 0.1)$ and $(w_\alpha, w_\gamma, w_\varepsilon) = (0.15, 0.20, 0.15)$. The latent components $(\alpha_g^x, \alpha_g^u, \varepsilon_{gt}^x, \varepsilon_{gt}^u)$ are mutually independent $N(0,1)$, and the common time effects $(\gamma_t^x, \gamma_t^u)$ follow AR(1) processes with coefficient $\rho$.

The two heterogeneity terms are rank-one cluster-by-time interactions

$$\eta_{gt} = \kappa\, c_g^\eta\, d_t^\eta, \qquad \delta_{gt} = \delta\, c_g^\delta\, d_t^\delta,$$

where the loading vectors $c_g^\bullet$ and $d_t^\bullet$ are independent, mean-zero draws (the $d_t^\bullet$ are serially uncorrelated). Centering both factors places each interaction in the two-way demeaned subspace $\mathcal{S}_{00}$—precisely the subspace on which the two-way clustering weight matrix carries its negative eigenvalue (Section 5). The term $\eta_{gt}$ is a mean-zero random *slope*: because it is mean-zero and independent of $X$, the OLS estimand for $\beta_1$ remains exactly 0.1, so the test is correctly centered; but the score $X_{gt}u_{gt}$ inherits the component $\eta_{gt}X_{gt}^2$, which has a heterogeneous *mean* on $\mathcal{S}_{00}$. The term $\delta_{gt}$ is an additive outcome-level component omitted from the regression, contributing heterogeneous score *variance* on the same subspace. We fix $\kappa = 0.8$ and report results with the level heterogeneity switched off ($\delta = 0$) and on ($\delta = 0.8$). We hold $G = 50$, and use $T = 120$ for most designs, comparable to the dimensions of the empirical application in Section 6.6, and vary $\rho \in \{0.25, 0.50, 0.75\}$.

In each of 1,000 draws we regress $Y_{gt}$ on $X_{gt}$ and a constant, compute standard errors by each method, and test $H_0 : \beta_1 = 0.1$ at the 5% level. Table 1 reports the empirical rejection rates. Across every design the Monte Carlo mean of $\widehat{\beta}_1$ remains within sampling error of 0.1, confirming the test is correctly centered and that the reported rates are size, not bias.

Table 1: Rejection rates of nominal 5% tests of $H_0 : \beta_1 = 0.1$. $G = 50$; 1,000 Monte Carlo draws. "Level Het." indicates whether the additive cluster-by-time level heterogeneity is present ($\delta = 0.8$) or absent ($\delta = 0$); the random-slope heterogeneity $\kappa = 0.8$ is present throughout.

| row | T | rho | Level Het. | EHW | CRg | CRt | CGM | CHS | C2NW | ET |
|---|---|---|---|---|---|---|---|---|---|---|
| (I) | 120 | 0.25 | No | 0.477 | 0.220 | 0.157 | 0.075 | 0.071 | 0.010 | 0.055 |
| (II) | 120 | 0.50 | No | 0.521 | 0.254 | 0.204 | 0.121 | 0.091 | 0.016 | 0.074 |
| (III) | 120 | 0.75 | No | 0.631 | 0.364 | 0.343 | 0.215 | 0.120 | 0.041 | 0.107 |
| (IV) | 120 | 0.25 | Yes | 0.201 | 0.130 | 0.115 | 0.078 | 0.075 | 0.000 | 0.026 |
| (V) | 120 | 0.50 | Yes | 0.245 | 0.173 | 0.140 | 0.098 | 0.080 | 0.001 | 0.036 |
| (VI) | 120 | 0.75 | Yes | 0.364 | 0.274 | 0.267 | 0.200 | 0.098 | 0.012 | 0.068 |
| (VII) | 1000 | 0.75 | No | 0.734 | 0.133 | 0.510 | 0.104 | 0.058 | 0.031 | 0.057 |
| (VIII) | 1000 | 0.75 | Yes | 0.551 | 0.137 | 0.457 | 0.125 | 0.064 | 0.030 | 0.056 |

EHW denotes Eicker–Huber–White; CRg and CRt denote one-way cluster-robust within $g$ and within $t$; CGM denotes Cameron–Gelbach–Miller; CHS denotes Chiang–Hansen–Sasaki (as written in their paper, without subtracting within-cluster lag); C2NW denotes the conservative estimator $\widehat{V}_{\mathrm{C2NW}}$; ET denotes the eigenvalue-truncation estimator $\widehat{V}_{\mathrm{ET}}^{\star}$.

Three patterns emerge. First, the estimators that ignore dependence (EHW, CRg, CRt) over-reject grossly throughout, confirming that some form of dependence robustness is essential. Second, among the dependence-robust estimators, CGM and CHS over-reject substantially when mean heterogeneity is present, and the distortion grows with the persistence $\rho$. ET controls size far better in these designs, because its correction targets exactly the directions on which the heterogeneity loads. The conservative estimator C2NW controls size in all designs, at the cost of being noticeably conservative (rejection rates well below nominal).

Third, ET uniformly has lower rejection rates than CHS, including in the designs without level heterogeneity ($\delta = 0$, rows I–III). This reflects the within-cluster correction embedded in $\widehat{V}^{\star}_{\mathrm{ET}}$. The gap in size distortion is modest without level heterogeneity and widens once heterogeneity is present.

There are two qualifications. The cluster-by-time heterogeneity here is serially *uncorrelated*; this is what prevents the CHS HAC lag estimator from absorbing it and exposes the bias in the CHS estimand. Under serially correlated heterogeneity the CHS lag term partially captures the structure and CHS is correspondingly more robust in finite samples. While ET controls size far better than CGM and CHS under heterogeneity, there is some over-rejection in finite-sample designs, but this over-rejection disappears when we increase $T$ to 1000, even for the hardest case with high $\rho$.

## 6.6 Empirical Application

The new variance estimator is applied to a panel of industry portfolios across time. The Fama–French three-factor model (Fama and French, 1993) can be written as:

$$R_{gt} - R_{ft} = \beta_1(R_{Mt} - R_{ft}) + \beta_2 SMB_{gt} + \beta_3 HML_{gt} + e_{gt}, \tag{27}$$

where $R_{gt} - R_{ft}$ is the excess return as $R_{gt}$ denotes the total return of portfolio $g$ in month $t$, and $R_{ft}$ denotes the risk-free rate.

I use the industry portfolio dataset from CHS containing 44 industry portfolios across 119 months. Following CHS, I use a within-transformation for both $Y$ and $X = (R_M - R_f, SMB, HML)$ to remove additive fixed effects. Inference on coefficients relies on a CLT applied to $\sum_{g,t} \widetilde{X}_{gt}\widetilde{e}_{gt}$.

Allowing for heterogeneous means is natural here when the three-factor model is a linear approximation of the return process, and inference conditions on the portfolios and sample months observed. Under correct specification the score $\widetilde{X}_{gt}\widetilde{e}_{gt}$ would be conditionally mean-zero, and hence no mean heterogeneity. But suppose a priced factor is omitted—momentum, profitability, investment, or another of the many documented since Fama and French (1993). Conditional on the observed portfolios and months, the omitted portfolio-by-time specific factor is not a random shock but a fixed feature of cell $(g, t)$: it enters the *mean* of the score $\widetilde{X}_{gt}\widetilde{e}_{gt}$. We reserve variance for what remains genuinely stochastic given the conditioning—the idiosyncratic innovations—and treat the systematic, omitted-factor contribution as a heterogeneous conditional mean rather than as excess score variance. The within-transformation we apply, following CHS, removes any additive portfolio-level and time-level component of this mean, so a constant industry alpha or a common-to-all-portfolios mispricing in a given month is swept out. What survives is the *non-separable* interaction between portfolio and time which is precisely the two-way demeaned component on which the two-way clustering weight matrix carries its negative eigenvalue (Section 5). The researcher need not estimate the omitted factor: the role of the eigenvalue truncation is instead to modify the estimand so that inference remains conservative against whatever non-separable mean heterogeneity the omitted structure induces.

Table 2: Standard Errors for Various Methods: 44 Industry Portfolios, 119 Periods

| | coef | EHW | CRg | CRt | CGM | CHS | C2NW | ET |
|---|---|---|---|---|---|---|---|---|
| MKT | 0.959 | 0.022 | 0.055 | 0.031 | 0.059 | 0.059 | 0.076 | 0.062 |
| SMB | 0.076 | 0.029 | 0.035 | 0.041 | 0.045 | 0.053 | 0.083 | 0.059 |
| HML | 0.358 | 0.030 | 0.066 | 0.049 | 0.076 | 0.080 | 0.109 | 0.084 |

Estimates of the factor coefficients with various standard error estimates. EHW denotes Eicker–Huber–White; CRg denotes one-way cluster-robust within $g$; CRt denotes one-way cluster-robust within $t$; CGM denotes Cameron–Gelbach–Miller; CHS denotes Chiang–Hansen–Sasaki; C2NW denotes the conservative estimator $\widehat{V}_{\mathrm{C2NW}}$; ET denotes the optimal estimator $\widehat{V}^{\star}_{\mathrm{ET}}$.

From Table 2, standard errors estimated using ET are higher than the standard errors estimated using CHS. However, the standard errors are not unreasonably large: we retain the statistical significance of the coefficient on HML at the 5% level, while the statistical significance of SMB can now be called into question, as CHS already show. A further takeaway is that cross-cluster serial correlation is empirically important: if serial correlation were negligible, we would expect the CHS standard errors to be similar to CGM, which is not the case for two of the coefficients. We verify that ET meaningfully decreases the standard errors relative to the more conservative C2NW, as expected due to how ET is optimal.

# 7 Conclusion

This paper developed a unified theory of dependence-robust variance estimation under heterogeneous means. Every quadratic estimator $Y_n' A_n Y_n$ has expectation $\mathrm{tr}(A_n\Sigma) + \mu' A_n \mu$, so its behavior is governed by two structural commitments—a subspace $\mathcal{M}_n$ of admissible means and a closed convex cone $\mathcal{S}_n$ of admissible covariances. Cone duality turns this decomposition into an exact conservativeness criterion (Theorem 2), and the eigenvalue truncation $W_n^+$ minimizes the adjustment to the plug-in and, in the aligned regime, the mean-squared error and estimand level (Theorem 3). The same construction doubles as a diagnostic: one-way cluster-robust estimators, and Bartlett-kernel HAC emerge as optimal, whereas two-way clustering is anticonservative under heterogeneous means and is repaired by an explicit, minimal correction. Specializing the theory resolves an open problem in the panel literature— a variance estimator simultaneously robust to heterogeneous means and cross-cluster serial correlation, with consistency and conservativeness established under $\psi$-dependence and borne out in simulations and a Fama–French application.

# A Supplementary Results

## A.1 Details on Optimization Problems

In the minimal correction problem, the constraint $\|\mu\| = 1$ excludes $\mu = 0$. When $\mathcal{M}_n = \{0\}$—the mean is known exactly, so that the demeaned mean vanishes—the supremum in (5) is over the empty set; we adopt the convention $\mathcal{R}_n(A_n) := 0$, reflecting that (C2) is vacuous and no mean-induced correction is warranted, so the plug-in $A_n = W_n$ is feasible and optimal. The criterion reads $A_n$ only through its $\mathcal{M}_n$-block $P_{\mathcal{M}_n} A_n P_{\mathcal{M}_n}$ and is silent about the off-$\mathcal{M}_n$ block.

Under finite fourth moments, expanding the squared error gives

$$\mathrm{MSE}_n(A_n; P_n) = \left[\mathrm{tr}((A_n - W_n)\Sigma) + \mu' A_n \mu\right]^2 + 2\,\mathrm{tr}((A_n\Sigma)^2) + 4\mu' A_n \Sigma A_n \mu + \mathcal{T}_3(A_n; P_n) + \mathcal{K}_4(A_n; P_n), \tag{28}$$

where, writing $Z = Y_n - \mu$, $m^{(3)}_{ijk}(P_n) = \mathbb{E}_{P_n}[Z_i Z_j Z_k]$, and $\kappa_{ijkl}(P_n) = \mathbb{E}_{P_n}[Z_i Z_j Z_k Z_l] - \mathbb{E}[Z_i Z_j]\mathbb{E}[Z_k Z_l] - \mathbb{E}[Z_i Z_k]\mathbb{E}[Z_j Z_l] - \mathbb{E}[Z_i Z_l]\mathbb{E}[Z_j Z_k]$,

$$\mathcal{T}_3(A_n; P_n) := 4 \sum_{i,j,k} (A_n)_{ij} (A_n\mu)_k\, m^{(3)}_{ijk}(P_n), \qquad \mathcal{K}_4(A_n; P_n) := \sum_{i,j,k,l} (A_n)_{ij} (A_n)_{kl}\, \kappa_{ijkl}(P_n).$$

Under Assumption 1, MSE reduces to

$$\mathrm{MSE}_n(A_n; \mu, \Sigma) = \left[\mathrm{tr}((A_n - W_n)\Sigma) + \mu' A_n \mu\right]^2 + 2\,\mathrm{tr}((A_n\Sigma)^2) + 4\mu' A_n \Sigma A_n \mu, \tag{29}$$

depending on $P_n$ only through $(\mu, \Sigma)$.

**Remark 6** (Both restrictions in Theorem 3(iii) are essential)**.** The optimality in Theorem 3(iii) holds over the *spectral* class under a *commuting* cone, and fails if either restriction is dropped. $A_n^\star$ is in particular *not* globally MSE-minimal, and feasible Loewner-refinements $A_n \succeq A_n^\star$ can strictly improve on it.

*The cone must commute with $W_n$.* Take $\mathcal{M}_n = \mathbb{R}^2$, $\mathcal{S}_n = \mathrm{Sym}_2^+$, $Y_n$ Gaussian, $W_n = \mathrm{diag}(1, -1)$, so $A_n^\star = W_n^+ = \mathrm{diag}(1, 0)$. The feasible correction $A_n = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 1/2 \end{smallmatrix}\right)$ satisfies $A_n \succeq 0$, $A_n - W_n \succeq 0$, and at $\mu = 0$, $\Sigma = \left(\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix}\right)$ attains $\mathrm{MSE}_n(A_n) = 3/4 < 3 = \mathrm{MSE}_n(A_n^\star)$: a correction not sharing the eigenstructure of $W_n$ places less weight where $\Sigma$ concentrates. Here $\Sigma$ does not commute with $W_n$.

*The correction must be spectral, even when the cone commutes.* Take $\mathcal{M}_n = \mathbb{R}^2$, $Y_n$ Gaussian, $\mathcal{S}_n = \mathcal{S}_2^{\mathrm{diag}}$ (so condition (c)'s commuting holds), and

$$W_n = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad A_n^\star = W_n^+ = W_n.$$

Fix $\mu = (1, 1)'$, $\Sigma = \mathrm{diag}(1, \sigma)$ with $0 < \sigma < 1/9$, and move into the *non-spectral* feasible direction $D = vv'$, $v = (1, -3)'$, i.e. $A_n = A_n^\star + tD$ for small $t > 0$. Then $A_n - A_n^\star = tD \succeq 0$ and $\mathrm{diag}(A_n - W_n) = t(1, 9) \geq 0$, so $A_n$ is feasible and Loewnerdominates $A_n^\star$. Yet, using $\mathrm{Cov}(Y_n' A_n^\star Y_n, Y_n' D Y_n) = 2\,\mathrm{tr}(A_n^\star \Sigma D \Sigma) + 4\mu' A_n^\star \Sigma D \mu = 2 - 8 = -6$ and $\mathbb{E}[Y_n' A_n^\star Y_n] - \theta_n = \mu' A_n^\star \mu = 1$, $\mathbb{E}[Y_n' D Y_n] = \mathrm{tr}(D\Sigma) + \mu' D \mu = 5 + 9\sigma$,

$$\frac{d}{dt}\,\mathrm{MSE}_n(A_n^\star + tD; \mu, \Sigma)\Big|_{t=0} = 2\,\mathbb{E}\big[(Y_n' A_n^\star Y_n - \theta_n)\, Y_n' D Y_n\big] = 2\big(-6 + (5 + 9\sigma)\big) = 2(9\sigma - 1) < 0.$$

So $\mathrm{MSE}_n(A_n) < \mathrm{MSE}_n(A_n^\star)$ for small $t > 0$: a feasible, conservative, Loewner-dominating refinement strictly beats the truncation. The obstruction is the off-eigenbasis tilt of $D$, which lowers the mean–variance cross term $4\mu' A_n \Sigma A_n \mu$ faster than it raises bias; confining the search to $\mathcal{A}_n^{\mathrm{sp}}$ removes exactly this freedom.

## A.2 Examples for Anticonservativeness

We provide two illustrative examples. The first example shows how the plug-in estimator is anticonservative with m-dependence. The second example shows that the CHS estimand is anticonservative under heterogeneous means.

**Example 3.** Suppose we have an $m$-dependent series, where $m = 1$, and 3 time periods. This example can be extended to the $T \to \infty$ case for a sequence of heterogeneous means. The variance for the sum of a scalar random variable $y_t$ is:

$$\mathrm{Var}\left(\sum_{t=1}^{3} y_t\right) = \sum_{t=1}^{3} \mathrm{Var}\left(y_t\right) + 2\sum_{t=1}^{2} \mathrm{Cov}\left(y_t, y_{t+1}\right). \tag{30}$$

When using a standard heteroskedasticity and autocorrelation robust (HAR) variance estimator, we target $\sum_{t=1}^{3} E\left[y_t^2\right] + 2\sum_{t=1}^{2} E\left[y_t y_{t+1}\right]$, but

$$\begin{aligned} D_1 &:= \sum_{t=1}^{3} E\left[y_t^2\right] + 2\sum_{t=1}^{2} E\left[y_t y_{t+1}\right] - \mathrm{Var}\left(\sum_{t=1}^{3} y_t\right) \\ &= \sum_{t=1}^{3} E\left[y_t\right]^2 + 2\sum_{t=1}^{2} E\left[y_t\right] E\left[y_{t+1}\right]. \end{aligned} \tag{31}$$

The expression $D_1$ above can be negative while having $\sum_{t=1}^{3} E\left[y_t\right] = 0$ when $E[y_1] = 0.5, E[y_2] = -1, E[y_3] = 0.5$. Then, $\sum_{t=1}^{3} E\left[y_t\right]^2 + 2\sum_{t=1}^{2} E\left[y_t\right] E\left[y_{t+1}\right] = 1.5 - 2 = -0.5$. To restore validity, we may add (a scaled version of) $\sum_{t=1}^{3} y_t^2$ to the variance estimator, so that we target $2\sum_{t=1}^{3} E\left[y_t^2\right] + 2\sum_{t=1}^{2} E\left[y_t y_{t+1}\right]$ instead. Using this expression,

$$D_2 := 2\sum_{t=1}^{3} E\left[y_t^2\right] + 2\sum_{t=1}^{2} E\left[y_t y_{t+1}\right] - \mathrm{Var}\left(\sum_{t=1}^{3} y_t\right)$$

$$= 2\sum_{t=1}^{3} E\left[y_t\right]^2 + 2\sum_{t=1}^{2} E\left[y_t\right] E\left[y_{t+1}\right] + \sum_{t=1}^{3} \mathrm{Var}\left(y_t\right)$$

$$\geq \sum_{t=1}^{2} E\left[y_t\right]^2 + 2\sum_{t=1}^{2} E\left[y_t\right] E\left[y_{t+1}\right] + \sum_{t=1}^{2} E\left[y_{t+1}\right]^2 + \sum_{t=1}^{3} \mathrm{Var}\left(y_t\right) \geq 0, \tag{32}$$

so the variance estimand is conservative and robust to mean heterogeneity. When the means are homogeneous, we would overestimate the true variance by $D_2 = \sum_{t=1}^{3} \mathrm{Var}\left(y_t\right)$.

**Example 4.** Suppose we have one observation in each $(g, t)$ intersection, and we have $m = 1$ dependence. We can compare the oracle CHS variance estimand (where $\omega(m, M) = 1$ for all $m \leq 1$ and 0 otherwise) with the true variance. As before, this example can be extended to $T \to \infty$ and $G \to \infty$. The scalar $E\left[Y_{n,i}\right]$ values are in Table 3.

Table 3: Values of scalar $E[Y_{n,i}]$

| | $t=1$ | $t=2$ | $t=3$ | $t=4$ |
|---|---|---|---|---|
| $g=1$ | $-1$ | $-1$ | $1$ | $1$ |
| $g=2$ | $1$ | $1$ | $-1$ | $-1$ |

Then, observe that $\sum_{i\in N_n}\sum_{j\in N^G_{g(i)}} E\left[Y_{n,i}\right] E\left[Y_{n,j}\right] = 0$, $\sum_{i\in N_n}\sum_{j\in N^T_{t(i)}} E\left[Y_{n,i}\right] E\left[Y_{n,j}\right] = 0$, and $\sum_{i\in N_n}\sum_{j\in N^{T\cap G}_{t(i),g(i)}} E\left[Y_{n,i}\right] E\left[Y_{n,j}\right] = 8$. In the serially correlated components, setting $m = 1$, $\sum_{t=1}^{T-m} E\left[y_t\right] E\left[y'_{t+m}\right] = 0$ and $\sum_{t=1}^{T-m}\sum_{g=1}^{G}\sum_{i\in N^{T\cap G}_{t,g}}\sum_{j\in N^{T\cap G}_{t+m,g}} E\left[Y_{n,i}\right] E\left[Y_{n,j}\right] = 2$. Then, on aggregate,

$$\sum_{i\in N_n}\sum_{j\in N^G_{g(i)}} E\left[Y_{n,i}\right] E\left[Y_{n,j}\right] + \sum_{i\in N_n}\sum_{j\in N^T_{t(i)}} E\left[Y_{n,i}\right] E\left[Y_{n,j}\right] - \sum_{i\in N_n}\sum_{j\in N^{T\cap G}_{t(i),g(i)}} E\left[Y_{n,i}\right] E\left[Y_{n,j}\right]$$

$$+ 2\sum_{t=1}^{T-m} E\left[y_t\right] E\left[y_{t+m}\right] - 2\sum_{t=1}^{T-m}\sum_{g=1}^{G}\sum_{i\in N^{T\cap G}_{t,g}}\sum_{j\in N^{T\cap G}_{t+m,g}} E\left[Y_{n,i}\right] E\left[Y_{n,j}\right] = -12 < 0, \tag{33}$$

so the standard CHS estimand is anticonservative.

## A.3 Variance estimators

We now define both variance estimators and their target estimands. The CHS variance estimator (excluding their EVC operation) is:

$$\widehat{V}_{\mathrm{CHS}} = \sum_{i\in N_n}\sum_{j\in N^G_{g(i)}} Y_{n,i} Y'_{n,j} + \sum_{i\in N_n}\sum_{j\in N^T_{t(i)}} Y_{n,i} Y'_{n,j} - \sum_{i\in N_n}\sum_{j\in N^{T\cap G}_{t(i),g(i)}} Y_{n,i} Y'_{n,j}$$

$$
+\sum_{m=1}^{M}\omega(m,M)\left(\sum_{t=1}^{T-m}y_t y_{t+m}' + \sum_{t=1}^{T-m}y_{t+m}y_t'\right)
$$

$$
-\sum_{m=1}^{M}\omega(m,M)\left(\sum_{t=1}^{T-m}\sum_{g=1}^{G}\sum_{i\in N_{t,g}^{T\cap G}}\sum_{j\in N_{t+m,g}^{T\cap G}}Y_{n,i}Y_{n,j}' + \sum_{t=1}^{T-m}\sum_{g=1}^{G}\sum_{i\in N_{t,g}^{T\cap G}}\sum_{j\in N_{t+m,g}^{T\cap G}}Y_{n,j}Y_{n,i}'\right). \tag{34}
$$

The first line is the CGM estimator, the second accounts for serial correlation across clusters, and the third adjusts for double counting. The CHS estimand $V_{\text{CHS}}$ replaces each empirical product with its expectation.

**Lemma 1** (Matrix form of the CHS estimator)**.** *In the balanced panel with one observation per cell, $i = (g, t)$ and $\mathbb{R}^n \cong \mathbb{R}^T \otimes \mathbb{R}^G$, the CHS variance estimator (34) admits the quadratic-form representation*

$$
\widehat{V}_{\text{CHS}} = Y_n' W_n^{\text{CHS}} Y_n, \qquad W_n^{\text{CHS}} = J_T \otimes I_G + C_M \otimes (J_G - I_G).
$$

A simpler conservative estimator extends Davezies et al. (2021)'s CGM2 construction to the serially correlated setting. CGM2 adds $W_n^T$ to the CGM weight matrix to restore uniform conservativeness in the standard two-way setting (Section 5.3); its analog here adds a tractable PSD term $2\sum_t y_t y_t'$ in the kernel-weighted block:

$$
\widehat{V}_{\text{C2NW}} := \sum_{i\in N_n}\sum_{j\in N_{g(i)}^G} Y_{n,i}Y_{n,j}' + \sum_{i\in N_n}\sum_{j\in N_{t(i)}^T} Y_{n,i}Y_{n,j}'
$$

$$
+\sum_{m=1}^{M}\omega(m,M)\left(\sum_{t=1}^{T-m}y_t y_{t+m}' + \sum_{t=1}^{T-m}y_{t+m}y_t' + 2\sum_{t=1}^{T}y_t y_t'\right). \tag{35}
$$

$V_{\text{C2NW}}$ is defined analogously by replacing empirical products with expectations. The estimator omits the $\sum_{t,g,i,j}$ double-counting line of $\widehat{V}_{\text{CHS}}$, matching CHS's own implementation choice.

By construction, $\widehat{V}_{\text{C2NW}}$ adds weight on the group-symmetric subspace ($v \otimes \mathbf{1}_G$ directions), while $\widehat{V}_{\text{ET}}^{\star}$ adds weight precisely on the group-demeaned subspace where the negative eigenvalues live. Both are uniformly conservative, but the optimal correction targets the pathology directly while the conservative correction overshoots.

For implementation, $M$ is selected using the bandwidth selection procedure of Andrews (1991) and $\omega(\cdot)$ is the triangular kernel. This choice follows CHS and is used here for comparability, but we note that the Andrews (1991) rule is derived under stationarity for a single time series and its optimality has not been established for the $\psi$-dependent two-way array considered here; Assumption 4 requires only $M = o(T)$, which the resulting bandwidths satisfy.

## A.4 Consistency Details

**Lemma 2** (Quadratic-form law of large numbers)**.** *Let $\{Y_{n,i}\}$ be a triangular array of scalar random variables that is conditionally $\psi$-dependent given $\{\mathcal{C}_n\}$, with dependence coefficients $\{\theta_{n,s}\}$ satisfying Assumption 2(b). Let $A_n \in \mathrm{Sym}_n$ be a (possibly random but $\mathcal{C}_n$-measurable) symmetric weight matrix, and define*

$$\hat{V}_{A,n} := Y_n' A_n Y_n, \qquad V_{A,n} := \mathbb{E}[Y_n' A_n Y_n \mid \mathcal{C}_n].$$

*Suppose the following primitive conditions hold for some $p > 4$:*

*(i) (Moment bound)* $\sup_{n\geq 1} \max_{i\in N_n} \|Y_{n,i}\|_{\mathcal{C}_n,p} < \infty$ *a.s.*

*(ii) (Weight bound) There is a nonnegative sequence $\{\bar{a}_n\}$ with $\max_{i,j\in N_n} |(A_n)_{ij}| \leq \bar{a}_n$ a.s. for every $n$.*

*(iii) (Support bandwidth) There exists a sequence $b_n \geq 0$ such that $(A_n)_{ij} = 0$ whenever $d_n(i,j) > b_n$.*

*(iv) (Variance summability)*

$$\bar{a}_n^2 \cdot n \sum_{s\geq 0} c_n(s, b_n; 2)\, \theta_{n,s}^{1-4/p} \to 0 \quad a.s.$$

*Then*

$$\hat{V}_{A,n} - V_{A,n} \xrightarrow{p} 0.$$

The lemma is applied to weight matrices that have already been normalized by $\lambda_n$: for an unnormalized $A_n$ with $\max_{i,j} |(A_n)_{ij}| \leq a_n$, taking $A_n/\lambda_n$ in the lemma gives $\bar{a}_n = a_n/\lambda_n$ and recovers $\lambda_n^{-1}(\hat{V}_{A,n} - V_{A,n}) \xrightarrow{p} 0$ under $\frac{a_n^2}{\lambda_n^2}\, n \sum_{s\geq 0} c_n(s, b_n; 2)\theta_{n,s}^{1-4/p} \to 0$. Condition (ii) alone does not deliver (iv): the sum $n \sum_{s\geq 0} c_n(s, b_n; 2)\theta_{n,s}^{1-4/p}$ diverges, so the rate at which $\bar{a}_n$ vanishes is what does the work.

**Remark 7.** Conditions (i), (ii), and (iv) are direct adaptations of Assumption 4.1 of KMS to matrix-form weights. Condition (iii) holds whenever $A_n^\star$ inherits the support pattern of the underlying plug-in $W_n$, as for the block and banded cones of Section 2. It fails for the exact truncation of Corollary 2, whose correction $B^- \otimes Q_G$ is globally supported in time; that case is handled in Proposition 5 below by applying the lemma to the band-limited proxy and bounding the difference in operator norm.

We establish consistency of the eigenvalue-truncation estimator $\widehat{V}_{\mathrm{ET}}^\star$ of (22) and of the conservative estimator $\widehat{V}_{\mathrm{C2NW}}$ of Appendix A.3. The argument has three layers. First, two band-limited quadratic forms—the conservative $\widehat{V}_{\mathrm{C2NW}}$ and a proxy $\widehat{V}_{\mathrm{band}}$ for the truncation—obey a law of large numbers directly through Lemma 2; the C2NW branch additionally invokes Assumption 6, since its within-period block is not entrywise bounded. Second, the exact truncation $\widehat{V}^{\star}_{\text{ET}}$, which is not band-limited, inherits consistency from the proxy $\widehat{V}_{\text{band}}$ because the two differ by an operator of vanishing relative size. Third, the common target $V_{\text{adj}}$ is asymptotically equivalent to $V_{\text{true}}$.

Throughout, $\widehat{V}_{\text{band}} := Y_n' A_n^{\text{band}} Y_n$ for

$$A_n^{\text{band}} := W_n^{\text{CHS}} + C_M \otimes Q_G = J_T \otimes I_G + \tfrac{G-1}{G}\, C_M \otimes J_G, \qquad V_{\text{band}} := \mathbb{E}[\widehat{V}_{\text{band}} \mid \mathcal{C}_n],$$

the band-limited proxy; it plays no role beyond the consistency proof.

**Proposition 4** (Conservativeness). *$V^{\star}_{\text{ET}} - V_{\text{adj}}$ and $V_{\text{C2NW}} - V_{\text{adj}}$ are both positive semidefinite.*

The weight matrix of $\widehat{V}_{\text{C2NW}}$, unlike those of $\widehat{V}_{\text{CHS}}$ and $\widehat{V}^{\star}_{\text{ET}}$, is not entrywise bounded: the conservative term $2\sum_t y_t y_t'$ sits inside the bracket summed over $m$, so every lag deposits on the same within-period pairs, which accumulate weight $2\sum_{m=1}^{M} \omega(m, M) = M$. The lag terms are innocuous by contrast: $\omega(m, M)\sum_t y_t y_{t+m}'$ deposits on pairs at temporal distance *exactly* $m$, so distinct lags occupy disjoint entries and each receives a single kernel weight — which is why $W_n^{\text{CHS}}$ has entries in $[0, 2]$ irrespective of $M$. Since the variance bound of Lemma 2 scales with the square of the largest entry, $\widehat{V}_{\text{C2NW}}$ requires an additional condition.

**Assumption 6.** $\frac{M^2 n}{\lambda_n^2} \sum_{s\ge 0} c_n(s, 0; 2)\, \theta_{n,s}^{1-4/p} \to 0.$

The $O(M)$ entries are confined to pairs at distance 0, so Assumption 6 evaluates $c_n(\cdot)$ at bandwidth 0 rather than $M$; since $c_n(s, 0; 2) = 0$ for $s \ge 1$, the sum collapses to its $s = 0$ term. In the balanced panel $c_n(0, 0; 2) \asymp (G + T)^2$ and $\lambda_n \asymp T^2 G$, so the left-hand side is $\asymp M^2/T^2$ — the same order as that of Assumption 4, and likewise reducing to $M = o(T)$. The two estimators $\widehat{V}_{\text{C2NW}}$ and $\widehat{V}^{\star}_{\text{ET}}$ therefore require the same bandwidth condition in the balanced panel, and the case for $\widehat{V}^{\star}_{\text{ET}}$ rests on the efficiency ordering of Remark 8, not on weaker regularity.

**Lemma 3** (Consistency of the band-limited estimators). *Suppose Assumption 2 holds and $\lambda_{\min}(V_\bullet)/\lambda_n \ge 1 + o(1)$. Then*

$$V_\bullet^{-1}\widehat{V}_\bullet \xrightarrow{p} I_v,$$

*for $\bullet = \text{band}$ under Assumption 4, and for $\bullet = \text{C2NW}$ under Assumptions 4 and 6.*

**Proposition 5** (Consistency of the exact eigenvalue-truncation estimator). *Suppose Assumptions 2 and 4 hold, $\lambda_{\min}(V^{\star}_{\text{ET}})/\lambda_n \ge 1 + o(1)$, and $Mn/\lambda_n \to 0$ (equivalently $M/T \to 0$ in the balanced panel, where $\lambda_n \asymp T^2 G$). Then*

$$(V^{\star}_{\text{ET}})^{-1}\, \widehat{V}^{\star}_{\text{ET}} \xrightarrow{p} I_v.$$

**Proposition 6.** *Under Assumptions 2 and 5,* $V_{\text{true}}^{-1}V_{\text{adj}} \to I_v$.

**Remark 8** (Efficiency ordering)**.** $\widehat{V}_{\text{C2NW}}$ overshoots by placing weight on the group-symmetric subspace, where $W_n^{\text{CHS}}$ needs no correction, rather than on the group-demeaned subspace where its negative eigenvalues live. Since $2\sum_{m=1}^{M}\omega(m,M) = M$ for the triangular kernel, the within-period pairs of $\widehat{V}_{\text{C2NW}}$ carry weight $M$, so

$$A_n^{\text{C2NW}} = J_T \otimes I_G + (C_M + M\, I_T) \otimes J_G = \big[J_T + G\, C_M + GM\, I_T\big] \otimes P_G \;+\; J_T \otimes Q_G,$$

while $A_n^{\text{band}} = \big[J_T + (G-1)C_M\big] \otimes P_G + J_T \otimes Q_G$ from the proof of Corollary 2. The two coincide on the $Q_G$ block, and

$$A_n^{\text{C2NW}} - A_n^{\text{band}} = \big(C_M + GM\, I_T\big) \otimes P_G \;\succeq\; 0,$$

positive definite on the group-symmetric block, so $V_{\text{C2NW}} - V_{\text{band}} \succeq 0$. Since $A_n^{\text{band}} - A_n^{\star} = (C_M - B^{-}) \otimes Q_G$ has operator norm at most $M+1$, we have $V_{\text{band}} - V_{\text{ET}}^{\star} = O(Mn) = o(\lambda_n)$, and $V_{\text{C2NW}} \succeq V_{\text{ET}}^{\star}$ up to that order.

Under the conditions of Proposition 6, the auxiliary condition $\lambda_{\min}(V_\bullet)/\lambda_n \geq 1 + o(1)$ holds for $\bullet \in \{\text{C2NW}, \text{ET}^\star\}$, since $\lambda_{\min}(V_\bullet)/\lambda_{\min}(V_{\text{adj}}) \geq 1$ by Proposition 4 and $\lambda_{\min}(V_{\text{adj}})/\lambda_n = 1 + o(1)$ by Proposition 6. In the running AR(1) examples with $\rho \in (0,1)$ and the triangular kernel, all three parts of Assumption 5 are implied by $\rho^m$ decay and the balanced panel structure.

# B Proofs

Let $N_n(i; s)$ denote the set of observations that are within distance $s$ of observation $i$, and let $N_n^{\partial}(i; s)$ denote the set of observations that are exactly at distance $s$ from observation $i$:

$$\begin{aligned} N_n(i; s) &:= \{j \in N_n : d_n(i,j) \leq s\}\,; \text{ and} \\ N_n^{\partial}(i; s) &:= \{j \in N_n : d_n(i,j) = s\}\,. \end{aligned} \tag{36}$$

For matrices $A, B$, the notation $A \geq B$ means $A - B$ is positive semidefinite (psd).

## B.1 Proofs of Theorems

*Proof of Theorem 1.* Fix any $A_n \in \text{Sym}_n$ and any $\Sigma_0 \in \mathcal{S}_n$, $v_n \in \mathcal{M}_n$ satisfying the hypotheses. Define $\Sigma_1 := \Sigma_0 + v_n v_n' \in \mathcal{S}_n$ and consider

$$P_n : Y_n \sim N(0, \Sigma_1), \qquad Q_n : Y_n \sim N(v_n, \Sigma_0).$$

Both have finite second moments, mean in $\mathcal{M}_n$ (since $0, v_n \in \mathcal{M}_n$), and covariance in $\mathcal{S}_n$, so both lie in $\mathcal{P}_n(\mathcal{M}_n, \mathcal{S}_n)$. The expectations of the quadratic estimator under each distribution are

$$\mathbb{E}_{P_n}[\widehat{V}_{A,n}] = \operatorname{tr}(A_n\Sigma_1) + 0 = \operatorname{tr}(A_n\Sigma_0) + v_n' A_n v_n,$$
$$\mathbb{E}_{Q_n}[\widehat{V}_{A,n}] = \operatorname{tr}(A_n\Sigma_0) + v_n' A_n v_n,$$

which are identical: $\mathbb{E}_{P_n}[\widehat{V}_{A,n}] = \mathbb{E}_{Q_n}[\widehat{V}_{A,n}]$ for every symmetric $A_n$. However, the target values differ:

$$\theta_n(P_n) - \theta_n(Q_n) = \operatorname{tr}(W_n(\Sigma_1 - \Sigma_0)) = v_n' W_n v_n.$$

Let $\Delta := v_n' W_n v_n$. For any choice of $A_n$, since the estimator expectation is common across $P_n$ and $Q_n$ while the target differs by $\Delta$, the absolute bias must satisfy

$$\big|\mathbb{E}_{P_n}[\widehat{V}_{A,n}] - \theta_n(P_n)\big| + \big|\mathbb{E}_{Q_n}[\widehat{V}_{A,n}] - \theta_n(Q_n)\big| \geq |\Delta|,$$

so $\sup_{P\in\{P_n,Q_n\}} |\mathbb{E}_P[\widehat{V}_{A,n}] - \theta_n(P)| \geq |\Delta|/2$, and hence the supremum over the full model is at least $|\Delta|/2$.

□

*Proof of Theorem 2.* By (1), conservativeness is

$$\operatorname{tr}\big((A_n - W_n)\Sigma\big) + \mu' A_n \mu \geq 0 \quad \text{for all } (\mu, \Sigma) \in \mathcal{M}_n \times \mathcal{S}_n.$$

Since $0 \in \mathcal{M}_n$ as $\mathcal{M}_n$ is a linear subspace of $\mathbb{R}^n$ and $0 \in \mathcal{S}_n$ as $\mathcal{S}_n$ is a closed convex cone, taking $\mu = 0$ and $\Sigma = 0$ separately decouples the joint requirement into (3)–(4). Sufficiency is immediate. □

*Proof of Theorem 3.* Throughout, $\mathcal{M}_n = \mathbb{R}^n$, so (C2) reads $A_n \succeq 0$ and $\mathcal{R}_n(A_n) = \sup_{\|\mu\|=1} \mu'(A_n - W_n)\mu = \lambda_{\max}(A_n - W_n)$.

*(i).* Let $A_n$ be uniformly conservative and set $\Delta := A_n - W_n$. Two lower bounds combine. First, take a unit eigenvector $u$ with $W_n u = \lambda_n^{\min} u$; by (C2), $u' A_n u \geq 0$, so

$$\mathcal{R}_n(A_n) \;\geq\; u'\Delta u \;=\; u' A_n u - \lambda_n^{\min} \;\geq\; -\lambda_n^{\min}.$$

Second, since $\mathcal{S}_n \neq \{0\}$ pick $\Sigma_0 \in \mathcal{S}_n \setminus \{0\}$; it is positive semidefinite and nonzero, so if $\mathcal{R}_n(A_n) < 0$, i.e. $\Delta \prec 0$, then $\operatorname{tr}(\Delta\Sigma_0) < 0$, contradicting $\Delta \in \mathcal{S}_n^*$. Hence $\mathcal{R}_n(A_n) \geq 0$. Combining, $\mathcal{R}_n(A_n) \geq (-\lambda_n^{\min})_+$. The truncation attains this bound and is feasible: $A_n^\star - W_n = W_n^- \succeq 0 \in \mathcal{S}_n^*$ gives (C1), $A_n^\star = W_n^+ \succeq 0$ gives (C2), and $\mathcal{R}_n(A_n^\star) = \lambda_{\max}(W_n^-) = (-\lambda_n^{\min})_+$. So $A_n^\star$ solves (6).

In (ii) and (iii), (10) enters only through the memberships $q_{n,k}q_{n,k}' \in \mathcal{S}_n$, which turn (C1) into the scalar inequalities $q_{n,k}'(A_n - W_n)q_{n,k} \geq 0$.

*The feasible box.* Under (10), the uniformly conservative members of $\mathcal{A}_n^{\mathrm{sp}}$ are exactly

$$\{A_n(a) : a_j \geq \lambda^-_{n,j}\ \forall j\}, \qquad \text{whose Loewner-least element is } A_n(\lambda^-) = W_n^+ = A_n^\star. \tag{37}$$

Indeed, if $a_j \geq \lambda^-_{n,j}$ for all $j$ then $a \geq 0$, so $\Delta(a) \succeq 0 \in \mathrm{Sym}_n^+ \subseteq \mathcal{S}_n^*$, and $A_n(a) = \sum_j(\lambda_{n,j}+a_j)q_{n,j}q'_{n,j} \succeq 0$: both (C1) and (C2) hold. Conversely let $A_n(a)$ be uniformly conservative. If $\lambda_{n,k} \leq 0$, (C2) gives $a_k \geq -\lambda_{n,k} = \lambda^-_{n,k}$. If $\lambda_{n,k} > 0$, then (C1) at $q_{n,k}q'_{n,k} \in \mathcal{S}_n$ gives $a_k = \mathrm{tr}\big(\Delta(a)\, q_{n,k}q'_{n,k}\big) \geq 0 = \lambda^-_{n,k}$.

*(ii)(a).* $\mathrm{tr}(A_n(a)\Psi) = \mathrm{tr}(W_n\Psi) + \sum_j a_j\psi_j$ with $\psi_j \geq 0$ is coordinatewise nondecreasing on the box (37), hence minimized at its corner $a = \lambda^-$; linearity in $a$ makes the minimizer unique exactly when every $\psi_j > 0$.

*(ii)(b).* Let $A_n$ be uniformly conservative and fix $k$. If $\lambda_{n,k} > 0$, (C1) at $q_{n,k}q'_{n,k} \in \mathcal{S}_n$ gives $q'_{n,k}A_nq_{n,k} \geq q'_{n,k}W_nq_{n,k} = \lambda_{n,k} = \lambda^+_{n,k}$; if $\lambda_{n,k} \leq 0$, (C2) gives $q'_{n,k}A_nq_{n,k} \geq 0 = \lambda^+_{n,k}$. Hence, for $\Psi$ diagonal in the eigenbasis, $\mathrm{tr}(A_n\Psi) = \sum_j \psi_j\, q'_{n,j}A_nq_{n,j} \geq \sum_j \psi_j\lambda^+_{n,j} = \mathrm{tr}(A_n^\star\Psi)$.

*(ii)(c).* Recall $A_n(a) = W_n + \Delta(a)$ from Definition 1, so at $\Psi = q_{n,k}q'_{n,k}$ every spectral candidate has level $\mathrm{tr}(A_n(a)\Psi) = \lambda_{n,k} + a_k$, while $\mathrm{tr}(A_n^\star\Psi) = \lambda_{n,k}$. The claim is therefore equivalent to exhibiting a uniformly conservative $A_n(a)$ with $a_k < 0$. By the box (37), no such $a$ exists when (10) holds; the construction below extracts an $a''$ with $a''_k < 0$ when $q_{n,k}q'_{n,k} \notin \mathcal{S}_n$.

Let $\mathcal{C} := \{d(\Sigma) : \Sigma \in \mathcal{S}_n\} \subseteq \mathbb{R}^n_+$, where $d(\Sigma) := \big(q'_{n,1}\Sigma q_{n,1}, \ldots, q'_{n,n}\Sigma q_{n,n}\big)'$; it is a convex cone, being a linear image of one, and $\overline{\mathcal{C}}$ is a closed convex cone. Here $\{e_j\}_{j=1}^n$ is the standard basis of $\mathbb{R}^n$ in the eigen-index $j$, so that $e_k = d(q_{n,k}q'_{n,k})$ is the diagonal image of the $k$-th eigenprojector.

We first claim $e_k \notin \overline{\mathcal{C}}$. If it were, there would be $\Sigma^{(m)} \in \mathcal{S}_n$ with $d(\Sigma^{(m)}) \to e_k$; then $q'_{n,k}\Sigma^{(m)}q_{n,k} \to 1$, so rescaling within the cone we may take $q'_{n,k}\Sigma^{(m)}q_{n,k} = 1$, whence $\mathrm{tr}(\Sigma^{(m)}) = \sum_j q'_{n,j}\Sigma^{(m)}q_{n,j} \to 1$. The sequence is bounded, so along a subsequence $\Sigma^{(m)} \to \Sigma^\star$, and $\Sigma^\star \in \mathcal{S}_n$ because $\mathcal{S}_n$ is closed, with $\mathrm{tr}(\Sigma^\star) = q'_{n,k}\Sigma^\star q_{n,k} = 1$. For $\Sigma \succeq 0$ of unit trace, $q'\Sigma q \leq 1$ with equality only at $\Sigma = qq'$; hence $\Sigma^\star = q_{n,k}q'_{n,k} \in \mathcal{S}_n$, a contradiction.

Separation therefore supplies $a \in \mathbb{R}^n$ with $a'd \geq 0$ for all $d \in \overline{\mathcal{C}}$ and $a_k = a'e_k < 0$. Write $\mathcal{C}^* := \{b : b'd \geq 0\ \forall d \in \overline{\mathcal{C}}\}$, a convex cone containing $\mathbb{R}^n_+$ because $\overline{\mathcal{C}} \subseteq \mathbb{R}^n_+$. Since $\mathrm{tr}\big(\Delta(b)\Sigma\big) = b'd(\Sigma)$ for every $\Sigma$, we have $\Delta(b) \in \mathcal{S}_n^*$ whenever $b \in \mathcal{C}^*$.

The separating $a$ delivers the negative $k$-th coordinate we need but may violate (C2) off coordinate $k$; we correct this without disturbing the sign of $a_k$ by adding a small nonnegative shift in the remaining coordinates. Fix $0 < \varepsilon < \lambda_{n,k}/|a_k|$, set $\delta := \varepsilon \max_j |a_j|$, and let

$$a'' := \varepsilon a + \lambda^- + \delta \textstyle\sum_{j\neq k} e_j \ \in\ \mathcal{C}^*,$$

the membership because $\mathcal{C}^*$ is a convex cone containing $\varepsilon a$ and $\mathbb{R}^n_+$. Then $\Delta(a'') \in \mathcal{S}_n^*$, giving (C1). For (C2): at $j = k$, $\lambda^-_{n,k} = 0$ and $\lambda_{n,k} + a''_k = \lambda_{n,k} + \varepsilon a_k > 0$; for $j \neq k$,

$\lambda_{n,j}+a''_j = \lambda^+_{n,j}+\varepsilon a_j+\delta \geq 0$ since $\delta \geq \varepsilon|a_j|$. So $A_n(a'') \in \mathcal{A}^{\mathrm{sp}}_n$ is uniformly conservative, yet at $\Psi = q_{n,k}q'_{n,k}$,

$$\mathrm{tr}\big(A_n(a'')\Psi\big) = \lambda_{n,k} + \varepsilon a_k < \lambda_{n,k} = \mathrm{tr}(A^\star_n \Psi),$$

which is the assertion of (c).

*(iii).* By (10), $\Sigma \in \mathcal{S}^{W\text{-diag}}_n$ lies in $\mathcal{S}_n$, so any $P_n$ with moments $(\mu, \Sigma)$, $\mu \in \mathbb{R}^n$, lies in $\mathcal{P}_n(\mathbb{R}^n, \mathcal{S}_n)$: Assumption 1 applies and the MSE is given by (29). Write $\Sigma = \Delta(s)$, so $\Sigma q_{n,j} = s_j q_{n,j}$ with $s_j \geq 0$, and set $m_j := q'_{n,j}\mu$ and $c_j := \lambda_{n,j}+a_j$. On the box (37), $a \geq 0$ and $c_j \geq \lambda^+_{n,j} \geq 0$. Diagonality in $\{q_{n,j}\}$ collapses each term of (29):

$$\mathrm{tr}((A_n(a) - W_n)\Sigma) = \sum_j a_j s_j, \quad \mu' A_n(a)\mu = \sum_j c_j m_j^2,$$
$$\mathrm{tr}((A_n(a)\Sigma)^2) = \sum_j c_j^2 s_j^2, \quad \mu' A_n(a)\Sigma A_n(a)\mu = \sum_j c_j^2 s_j m_j^2,$$

so that, with $b(a) := \sum_j a_j s_j + \sum_j c_j m_j^2$,

$$\mathrm{MSE}_n(A_n(a); \mu, \Sigma) = b(a)^2 + 2\sum_j c_j^2 s_j^2 + 4\sum_j c_j^2 s_j m_j^2.$$

The quantity $b(a)$ is the bias, nonnegative by Theorem 2 and, on the box, termwise. For each $\ell$,

$$\frac{\partial\, \mathrm{MSE}_n}{\partial a_\ell} = 2b(a)\,(s_\ell + m_\ell^2) + 2c_\ell\big(2s_\ell^2 + 4s_\ell m_\ell^2\big) \;\geq\; 0,$$

so $\mathrm{MSE}_n$ is coordinatewise nondecreasing on the box and is minimized at its corner $a = \lambda^-$, i.e. at $A^\star_n$. If every $s_j > 0$ then $2s_\ell^2 + 4s_\ell m_\ell^2 > 0$ and $c_\ell > 0$ off the corner, so the derivative is strictly positive there and the minimizer is unique. □

*Proof of Theorem 4.* It suffices to verify the conditions of KMS Theorem 3.2, using notation in this setup. Our Assumption 2 is identical to KMS Assumption 2.1. $Y_{n,i}$ satisfying that condition implies that $\nu' Y_{n,i}$ also satisfies that condition. Assumption 2(c)'s finite moment condition satisfies KMS Assumption 3.3. Observe that, since $\|\nu\| = 1$, we have $\sigma_n^2 \geq \lambda_n$ as the worst case $\sigma_n^2$ loads all weight on a combination of variance components that results in the smallest eigenvalue. Since $\sigma_n^2 \geq \lambda_n$, for some arbitrary constant $C > 0$,

$$\frac{n}{\sigma_n^{2+k}} \sum_{s\geq 0} c_n\,(s, m_n; k)\,\theta_{n,s}^{1-\frac{2+k}{p}} \leq \frac{Cn}{\lambda_n^{1+k/2}} \sum_{s\geq 0} c_n\,(s, m_n; k)\,\theta_{n,s}^{1-\frac{2+k}{p}} \xrightarrow{a.s.} 0, \text{ and} \tag{38}$$

$$\frac{n^2\theta_{n,m_n}^{1-(1/p)}}{\sigma_n} \leq \frac{Cn^2\theta_{n,m_n}^{1-(1/p)}}{\lambda_n^{1/2}} \xrightarrow{a.s.} 0. \tag{39}$$

It then remains to show that the setting in this paper satisfies the corresponding definitions in KMS. First, normalize $N_n(j;s)=\emptyset$ for $s<0$, and observe that, when using $s=0$,

$$\frac{1}{n}\sum_{i\in N_n}\left|N_n^{\partial}(i;0)\right|^k \le \frac{1}{n}\sum_{i\in N_n}\left(\left|N_{g(i)}^G\right|+\left|N_{t(i)}^T\right|\right)^k=\delta_n^{\partial}(0;k)\,. \tag{40}$$

There are two cases to consider: when $m=0$ and when $m\ge 1$. When $m=0$, observe that, for all $s\ge 0$, we obtain

$$\frac{1}{n}\sum_{i\in N_n}\left|N_n(i;0)\right|^k \le \frac{1}{n}\sum_{i\in N_n}\left(\left|N_{g(i)}^G\right|+\left|N_{t(i)}^T\right|\right)^k=\Delta_n(s,0;k)\,, \tag{41}$$

so the $\delta_n^{\partial}(0;k)$ and $\Delta_n(s,0;k)$ defined here place an upper bound on the corresponding definitions in KMS for $m=0$.

Now, consider the case of $m\ge 1$. The case of $s=0$ is derived in Equation (40). When $s\ge 1$, the corresponding definitions are also satisfied as

$$\frac{1}{n}\sum_{i\in N_n}\left|N_n^{\partial}(i;s)\right|^k=\frac{1}{n}\sum_{i\in N_n}\left(\left|N_{t(i)+s}^T\right|+\left|N_{t(i)-s}^T\right|\right)^k=\delta_n^{\partial}(s;k)\,,\ \text{and} \tag{42}$$

$$\begin{aligned}
&\frac{1}{n}\sum_{i\in N_n}\max_{j\in N_n^{\partial}(i;s)}\left|N_n(i;m)\setminus N_n(j;s-1)\right|^k\\
&=\frac{1}{n}\sum_{i\in N_n}\max_{j\in N_n^{\partial}(i;s)}\left|\sum_{h=0}^{m}\left|N_{t(i)+h}^T\right|+\sum_{h=0}^{m}\left|N_{t(i)-h}^T\right|-\sum_{h=0}^{s-1}\left|N_{t(i)+h}^T\right|-\sum_{h=0}^{s-1}\left|N_{t(i)-h}^T\right|\right|^k\\
&=\frac{1}{n}\sum_{i\in N_n}\left(\sum_{h=0}^{m}\left|N_{t(i)+h}^T\right|+\sum_{h=0}^{m}\left|N_{t(i)-h}^T\right|-\sum_{h=0}^{s-1}\left|N_{t(i)+h}^T\right|-\sum_{h=0}^{s-1}\left|N_{t(i)-h}^T\right|\right)^k\\
&=\frac{1}{n}\sum_{i\in N_n}\left(\sum_{h=s}^{m}\left|N_{t(i)+h}^T\right|+\sum_{h=s}^{m}\left|N_{t(i)-h}^T\right|\right)^k=\Delta_n(s,m;k)\,.
\end{aligned} \tag{43}$$

The expression $c_n(s,m;k)=\inf_{\alpha>1}\left[\Delta_n(s,m;k\alpha)\right]^{1/\alpha}\left[\delta_n^{\partial}\left(s;\frac{\alpha}{\alpha-1}\right)\right]^{1-\frac{1}{\alpha}}$ is identical to KMS. Consequently, the conditions of KMS Theorem 3.2 are satisfied under the Assumptions 2 and 3. $\square$

*Proof of Theorem 5.* (a) By Corollary 2, $(W_n^{\text{CHS}})^+-W_n^{\text{CHS}}=B^-\otimes Q_G\succeq 0$, and $(W_n^{\text{CHS}})^+\succeq 0$ on $\mathcal{M}_n=\mathbb{R}^n$. Hence the bias decomposition

$$V_{\text{ET}}^{\star}-V_{\text{adj}}=\operatorname{tr}\big((B^-\otimes Q_G)\,\Sigma_n\big)+\mu_n'(W_n^{\text{CHS}})^+\mu_n$$

has both terms nonnegative: the first is a trace of positive semidefinite matrices, the second is nonnegative by positive semidefiniteness of $(W_n^{\text{CHS}})^+$.

(b) This is Proposition 5 in Appendix A.4. The argument writes $\widehat{V}_{\text{ET}}^{\star} = \widehat{V}_{\text{band}} - Y_n' D Y_n$, where $\widehat{V}_{\text{band}} = Y_n'(W_n^{\text{CHS}} + C_M \otimes Q_G)Y_n$ is a band-limited proxy consistent by the quadratic LLN of Lemma 2, and $D = (C_M - B^-) \otimes Q_G$ has $\|D\|_{\text{op}} \leq M + 1$, so $Y_n' D Y_n = O_p(Mn) = o_p(\lambda_n)$.

(c) This is Proposition 6 in Appendix A.4: under Assumption 5 the discarded long-lag and boundary terms vanish relative to $\lambda_n$. □

## B.2 Proofs of Propositions

*Proof of Proposition 1.* (i)⇔(iii) is the definition of positive semidefiniteness. For (ii)⇔(iii): the plug-in $A_n = W_n$ satisfies (C1) trivially, since $A_n - W_n = 0 \in \mathcal{S}_n^*$ for every cone. By Theorem 2, conservativeness on $\mathcal{P}_n(\mathcal{M}_n, \mathcal{S}_n)$ is therefore equivalent to (C2) alone, $\mu' W_n \mu \geq 0$ for all $\mu \in \mathcal{M}_n$, independently of $\mathcal{S}_n$. At $\mathcal{M}_n = \mathbb{R}^n$ this reads $W_n \succeq 0$. □

*Proof of Proposition 2. (i).* Let $\Sigma_i = \Sigma_i^{(1)} + \Sigma_i^{(2)}$ with $\Sigma_i^{(1)} \in \mathcal{S}_n$ and $\Sigma_i^{(2)} \in \mathcal{S}_n^{W\text{-diag}}$, $i = 1, 2$, and let $t \in [0, 1]$. Then

$$t\Sigma_1 + (1-t)\Sigma_2 = \underbrace{\left[t\Sigma_1^{(1)} + (1-t)\Sigma_2^{(1)}\right]}_{\in\ \mathcal{S}_n} + \underbrace{\left[t\Sigma_1^{(2)} + (1-t)\Sigma_2^{(2)}\right]}_{\in\ \mathcal{S}_n^{W\text{-diag}}} \ \in\ \mathcal{S}_n^{\oplus},$$

each bracket lying in its own cone by the convexity of that cone; likewise $c(\Sigma^{(1)} + \Sigma^{(2)}) = c\Sigma^{(1)} + c\Sigma^{(2)} \in \mathcal{S}_n^{\oplus}$ for $c \geq 0$. Convexity of the summands is doing the work that the union of the two cones, being nonconvex in general, cannot do. Also $\mathcal{S}_n^{\oplus} \subseteq \text{Sym}_n^+$ because $\text{Sym}_n^+$ is closed under addition, and 0 belongs to each summand, so $\mathcal{S}_n^{\oplus}$ contains both $\mathcal{S}_n$ and $\mathcal{S}_n^{W\text{-diag}}$: it is admissible. For minimality, let $\mathcal{T}$ be admissible, and take $\Sigma_1 \in \mathcal{S}_n \subseteq \mathcal{T}$ and $\Sigma_2 \in \mathcal{S}_n^{W\text{-diag}} \subseteq \mathcal{T}$. Convexity of $\mathcal{T}$ gives $\frac{1}{2}\Sigma_1 + \frac{1}{2}\Sigma_2 \in \mathcal{T}$, and the cone property gives $\Sigma_1 + \Sigma_2 \in \mathcal{T}$. Hence $\mathcal{S}_n^{\oplus} \subseteq \mathcal{T}$.

*(ii).* If $D \in \mathcal{S}_n^* \cap (\mathcal{S}_n^{W\text{-diag}})^*$ then $\operatorname{tr}\big(D(\Sigma_1 + \Sigma_2)\big) = \operatorname{tr}(D\Sigma_1) + \operatorname{tr}(D\Sigma_2) \geq 0$; conversely, taking $\Sigma_2 = 0$ and then $\Sigma_1 = 0$ in $D \in (\mathcal{S}_n^{\oplus})^*$ delivers the two memberships. The stated form of $(\mathcal{S}_n^{W\text{-diag}})^*$ follows from $\operatorname{tr}\big(D\Delta(s)\big) = \sum_j s_j\, q_{n,j}' D q_{n,j}$, which is nonnegative for every $s \geq 0$ if and only if every coefficient is. The characterization of uniform conservativeness follows because (C2) does not involve the cone. Finally, if $\mathcal{T}$ is admissible then $\mathcal{S}_n^{\oplus} \subseteq \mathcal{T}$ by (i), and dualization reverses inclusion, so $\mathcal{T}^* \subseteq (\mathcal{S}_n^{\oplus})^*$; since (C1) requires $A_n - W_n \in \mathcal{T}^*$, the feasible set under $\mathcal{T}$ is contained in the feasible set under $\mathcal{S}_n^{\oplus}$. □

*Proof of Proposition 3.* In all three parts the feasible set is fixed: $A_n - W_n \in \mathcal{S}_n^*$ is a conic constraint (the dual cone $\mathcal{S}_n^*$ is closed convex), and $P_{\mathcal{M}_n} A_n P_{\mathcal{M}_n} \succeq 0$ is a linear matrix inequality, hence conic. It remains to treat the objectives.

*(i).* Introduce $t \in \mathbb{R}$ and rewrite (6) as $\min t$ subject to $\mu'(A_n - W_n)\mu \leq t$ for all $\mu \in \mathcal{M}_n$ with $\|\mu\| = 1$, equivalently the linear matrix inequality $t\, P_{\mathcal{M}_n} \succeq$

$P_{\mathcal{M}_n}(A_n - W_n)P_{\mathcal{M}_n}$. This is a linear objective subject to an linear matrix inequality (LMI) and the conic feasibility constraints, hence a convex conic program.

*(ii).* For $\pi$-almost every $(\mu, \Sigma)$ the integrand $\mathrm{MSE}_n(\,\cdot\,; \mu, \Sigma)$ is convex in $A_n$ by hypothesis. The averaged objective

$$\overline{\mathrm{MSE}}_n(A_n; \pi) = \int \mathrm{MSE}_n(A_n; \mu, \Sigma)\, d\pi(\mu, \Sigma)$$

is therefore an integral of convex functions, hence convex in $A_n$; the finite-fourth-moment assumption guarantees the integral is finite for every $A_n \in \mathrm{Sym}_n$. Minimizing a convex objective over the convex conic feasible set is a convex conic program.

*(iii).* By (1), for each $(\mu, \Sigma)$ the level $V_{A,n}(\mu, \Sigma) = \mathrm{tr}\big(A_n(\Sigma + \mu\mu')\big)$ is a linear functional of $A_n$. Integrating against $\pi$ and using linearity of the trace,

$$\overline{V}_n(A_n; \pi) = \int \mathrm{tr}\big(A_n(\Sigma + \mu\mu')\big)\, d\pi(\mu, \Sigma) = \mathrm{tr}\big(A_n\, \Psi_\pi\big), \qquad \Psi_\pi := \mathbb{E}_\pi[\Sigma + \mu\mu'],$$

finite since $\pi$ has finite second moments, and linear in $A_n$. Minimizing a linear objective over the conic feasible set is a linear conic program. $\square$

*Proof of Proposition 4.* Throughout write $a_i := \mathbb{E}[Y_{n,i} \mid \mathcal{C}_n]$ and $\bar{y}_t := \mathbb{E}[y_t \mid \mathcal{C}_n] = \sum_{i \in N_t^T} a_i$.

$\widehat{V}^\star_{\mathrm{ET}}$. By Corollary 2, $(W_n^{\mathrm{CHS}})^+ - W_n^{\mathrm{CHS}} = B^- \otimes Q_G \succeq 0$ and $(W_n^{\mathrm{CHS}})^+ \succeq 0$, so

$$V^\star_{\mathrm{ET}} - V_{\mathrm{adj}} = \mathrm{tr}\big((B^- \otimes Q_G)\, \Sigma_n\big) + \mu_n'(W_n^{\mathrm{CHS}})^+\mu_n \ \geq\ 0,$$

the first term a trace of a product of positive semidefinite matrices, the second non-negative by $(W_n^{\mathrm{CHS}})^+ \succeq 0$.

$\widehat{V}_{\mathrm{C2NW}}$. Replacing each empirical product in $\widehat{V}_{\mathrm{C2NW}}$ by its conditional expectation and subtracting $V_{\mathrm{adj}}$ of (24), the covariance parts of the two cluster double-sums and of the lag terms cancel against $V_{\mathrm{adj}}$, leaving the within-cell covariance, the group/time *mean* outer products, and the mean part of the kernel-weighted block:

$$\begin{aligned} V_{\mathrm{C2NW}} - V_{\mathrm{adj}} = {} & \underbrace{\sum_i \sum_{j \in N^G_{g(i)}} a_i a_j'}_{\text{(I) within-group means}} + \underbrace{\sum_i \sum_{j \in N^T_{t(i)}} a_i a_j'}_{\text{(II) within-time means}} + \underbrace{\sum_i \sum_{j \in N^{T\cap G}_{t(i),g(i)}} \mathrm{Cov}(Y_{n,i}, Y_{n,j} \mid \mathcal{C}_n)}_{\text{(III) within-cell covariance}} \\ & + \underbrace{\sum_{m=1}^{M} \omega(m, M) \sum_{t=1}^{T-m} \big(\bar{y}_t \bar{y}_{t+m}' + \bar{y}_{t+m}\bar{y}_t'\big) \ + \ 2\sum_{m=1}^{M} \omega(m, M) \sum_{t=1}^{T} \bar{y}_t \bar{y}_t'}_{\text{(IV) kernel-weighted mean block}}. \end{aligned}$$

Terms (I)–(III) are positive semidefinite: (I) telescopes to $\sum_g \big(\sum_{i \in N_g^G} a_i\big)\big(\sum_{i \in N_g^G} a_i\big)'$ and (II) likewise to $\sum_t \big(\sum_{i \in N_t^T} a_i\big)\big(\sum_{i \in N_t^T} a_i\big)'$, each a sum of outer products; (III) equals $\sum_{g,t} \operatorname{Var}\big( \sum_{i \in N_{t,g}^{T} \cap G} Y_{n,i} \mid \mathcal{C}_n \big) \succeq 0$. For (IV), fix $m$ and complete the square lag by lag. For each $t \le T - m$,

$$\bar{y}_t \bar{y}_{t+m}' + \bar{y}_{t+m} \bar{y}_t' = (\bar{y}_t + \bar{y}_{t+m})(\bar{y}_t + \bar{y}_{t+m})' - \bar{y}_t \bar{y}_t' - \bar{y}_{t+m} \bar{y}_{t+m}',$$

so the two cross terms are a square minus the two diagonal squares. The doubled within-period term supplies those diagonal squares. For each $m$,

$$2 \sum_{t=1}^{T} \bar{y}_t \bar{y}_t' - \sum_{t=1}^{T-m} \big( \bar{y}_t \bar{y}_t' + \bar{y}_{t+m} \bar{y}_{t+m}' \big) = \sum_{t=1}^{T} c_{t,m}\, \bar{y}_t \bar{y}_t', \qquad c_{t,m} \in \{0, 1, 2\},$$

a nonnegative-weighted sum of outer products, since each period is used as an endpoint at most twice. Therefore

$$(\mathrm{IV}) = \sum_{m=1}^{M} \omega(m, M) \left[ \sum_{t=1}^{T-m} (\bar{y}_t + \bar{y}_{t+m})(\bar{y}_t + \bar{y}_{t+m})' + \sum_{t=1}^{T} c_{t,m}\, \bar{y}_t \bar{y}_t' \right] \ \succeq \ 0,$$

every summand a positive semidefinite outer product and $\omega(m, M) \ge 0$ for the triangular kernel. Adding (I)–(IV) gives $V_{\mathrm{C2NW}} - V_{\mathrm{adj}} \succeq 0$. □

*Proof of Proposition 5.* By the Cramér–Wold device and the minimum-eigenvalue normalization, as in the proof of Lemma 3, it suffices to show $\frac{1}{\lambda_n} \nu' (\widehat{V}_{\mathrm{ET}}^{\star} - V_{\mathrm{ET}}^{\star}) \nu \xrightarrow{p} 0$ for every unit vector $\nu$; set $\tilde{Y}_{n,i} := \nu' Y_{n,i}$, a scalar $\psi$-dependent array with the dependence coefficients of $Y_n$.

*Decomposition.* The proxy and the exact estimator differ by $D := (C_M - B^-) \otimes Q_G$, since $\widehat{V}_{\mathrm{band}} - \widehat{V}_{\mathrm{ET}}^{\star} = Y_n' \big( (C_M - B^-) \otimes Q_G \big) Y_n$. Writing $\widehat{\Delta} := \tilde{Y}_n' D \tilde{Y}_n$ and $\Delta := \mathbb{E}[\widehat{\Delta} \mid \mathcal{C}_n]$,

$$\nu' (\widehat{V}_{\mathrm{ET}}^{\star} - V_{\mathrm{ET}}^{\star}) \nu = \nu' (\widehat{V}_{\mathrm{band}} - V_{\mathrm{band}}) \nu - (\widehat{\Delta} - \Delta).$$

By Lemma 3 the first term is $o_p(\lambda_n)$, so it remains to bound $\widehat{\Delta} - \Delta$.

*Operator-norm bound on $D$.* Both $C_M$ and $B^- = (C_M - J_T)_+$ are positive semidefinite. Since the spectral window of $C_M$ is the Fejér kernel, $\lambda_{\max}(C_M) \le f_{\mathrm{Fej}}(0) = M+1$; and as $-J_T \preceq 0$, $\lambda_{\max}(B^-) = \big( \lambda_{\max}(C_M - J_T) \big)_+ \le \lambda_{\max}(C_M) \le M + 1$. Hence $\lambda_{\max}(C_M - B^-) \le \lambda_{\max}(C_M) \le M+1$ and $\lambda_{\min}(C_M - B^-) \ge -\lambda_{\max}(B^-) \ge -(M+1)$, so $\|C_M - B^-\|_{\mathrm{op}} \le M + 1$. As $\|Q_G\|_{\mathrm{op}} = 1$,

$$\|D\|_{\mathrm{op}} = \|C_M - B^-\|_{\mathrm{op}}\, \|Q_G\|_{\mathrm{op}} \le M + 1.$$

*Conclusion.* By the variational characterization of the operator norm,

$$|\widehat{\Delta}| = |\tilde{Y}_n' D \tilde{Y}_n| \le \|D\|_{\mathrm{op}}\, \|\tilde{Y}_n\|^2 \le (M + 1) \sum_{i \in N_n} \tilde{Y}_{n,i}^2.$$

By Assumption 2(c), $\mathbb{E}[\tilde{Y}_{n,i}^2 \mid \mathcal{C}_n] \le \bar{c}$ a.s. uniformly in $i, n$, so $\mathbb{E}[|\widehat{\Delta}| \mid \mathcal{C}_n] \le (M+1)\bar{c}\,n = O(Mn)$ a.s. This gives both $|\Delta| \le \mathbb{E}[|\widehat{\Delta}| \mid \mathcal{C}_n] = O(Mn)$ and, by the conditional Markov inequality, $\widehat{\Delta} = O_p(Mn)$. Therefore

$$\frac{|\widehat{\Delta} - \Delta|}{\lambda_n} \le \frac{|\widehat{\Delta}| + |\Delta|}{\lambda_n} = O_p\Big(\frac{Mn}{\lambda_n}\Big) = o_p(1),$$

using $Mn/\lambda_n \to 0$. Combining the two displays gives $\frac{1}{\lambda_n}\nu'(\widehat{V}^{\star}_{\mathrm{ET}} - V^{\star}_{\mathrm{ET}})\nu \xrightarrow{p} 0$ for each fixed unit $\nu$, the unconditional convergence following by dominated convergence as in Lemma 2. Exactly as in the proof of Lemma 3, in fixed dimension $v$ this per-direction bound yields $\lambda_n^{-1}\|\widehat{V}^{\star}_{\mathrm{ET}} - V^{\star}_{\mathrm{ET}}\|_{\mathrm{op}} \xrightarrow{p} 0$, whence $(V^{\star}_{\mathrm{ET}})^{-1}\widehat{V}^{\star}_{\mathrm{ET}} \xrightarrow{p} I_v$ using $\|(V^{\star}_{\mathrm{ET}})^{-1}\|_{\mathrm{op}} \le \lambda_n^{-1}(1 + o(1))$ from the hypothesis $\lambda_{\min}(V^{\star}_{\mathrm{ET}})/\lambda_n \ge 1 + o(1)$. □

*Proof of Proposition 6.* It suffices to show $\frac{1}{\lambda_n}\nu'(V_{\mathrm{true}} - V_{\mathrm{adj}})\nu = o(1)$ for any unit vector $\nu$; setting $\tilde{Y}_{n,i} = \nu' Y_{n,i}$, this is Lemma 4, whose conditions are immediate from Assumption 5. □

### B.3 Statements and Proofs of Lemmas

*Proof of Lemma 1.* Read $\widehat{V}_{\mathrm{CHS}} = \sum_{i,j}(W_n^{\mathrm{CHS}})_{ij} Y_{n,i} Y_{n,j}'$, so that each line of (34) is the quadratic form of a weight matrix on the index pair $\big((t,g),(s,h)\big)$. With one observation per cell, $N^G_{g(i)} = \{(g,s)\}_s$, $N^T_{t(i)} = \{(h,t)\}_h$, $N^{T\cap G}_{t(i),g(i)} = \{(g,t)\}$, and $y_t = \sum_h Y_{(h,t)}$, so the five lines map to

$$\begin{array}{rlcl}
\text{within-group} & \sum_i \sum_{j \in N^G_{g(i)}} & \longmapsto & J_T \otimes I_G, \\
\text{within-time} & \sum_i \sum_{j \in N^T_{t(i)}} & \longmapsto & I_T \otimes J_G, \\
\text{within-cell} & -\sum_i \sum_{j \in N^{T\cap G}_{t(i),g(i)}} & \longmapsto & -\, I_T \otimes I_G, \\
\text{aggregated lag} & \sum_{m \ge 1} \omega(m,M) \sum_t \big(y_t y_{t+m}' + y_{t+m} y_t'\big) & \longmapsto & (C_M - I_T) \otimes J_G, \\
\text{double-count} & -\sum_{m \ge 1} \omega(m,M) \sum_t \sum_g \big(Y_{(g,t)} Y_{(g,t+m)}' + \mathrm{tr.}\big) & \longmapsto & -\,(C_M - I_T) \otimes I_G,
\end{array}$$

using $(C_M)_{ts} = \omega(|t-s|, M)$ and $\omega(0,M) = 1$. Summing the five contributions and collecting the $\otimes J_G$ and $\otimes I_G$ blocks,

$$\underbrace{\big[I_T + (C_M - I_T)\big]}_{=C_M} \otimes J_G \;+\; \underbrace{\big[J_T - I_T - (C_M - I_T)\big]}_{=J_T - C_M} \otimes I_G \;=\; C_M \otimes J_G + (J_T - C_M) \otimes I_G,$$

which rearranges to $J_T \otimes I_G + C_M \otimes (J_G - I_G) = W_n^{\mathrm{CHS}}$. □

*Proof of Lemma 2.* We adapt the proof of Proposition 4.1 of Kojevnikov et al. (2021), translating from the HAC weight notation $\omega_n(s)$ used there to the matrix entries $(A_n)_{ij}$ used here. Throughout, all probabilistic statements are conditional on $\mathcal{C}_n$; the unconditional statement follows by the dominated convergence theorem applied to the conditional probabilities, since they are bounded by 1 and converge to 0 a.s.

**Step 1: Centering.** Define $z_{ij} := Y_{n,i}Y_{n,j} - \mathbb{E}[Y_{n,i}Y_{n,j} \mid \mathcal{C}_n]$, so that $\mathbb{E}[z_{ij} \mid \mathcal{C}_n] = 0$. Then

$$\hat{V}_{A,n} - V_{A,n} = \sum_{i \in N_n} \sum_{j \in N_n} (A_n)_{ij}\, z_{ij}.$$

By condition (iii), the inner sum is restricted to pairs $(i,j)$ with $d_n(i,j) \le b_n$:

$$\hat{V}_{A,n} - V_{A,n} = \sum_{i \in N_n} \sum_{j \in N_n : d_n(i,j) \le b_n} (A_n)_{ij}\, z_{ij}.$$

**Step 2: Conditional variance bound.** We compute

$$\mathrm{Var}\Big(\hat{V}_{A,n} - V_{A,n} \mid \mathcal{C}_n\Big) = \sum_{(i,j,k,l)} (A_n)_{ij}(A_n)_{kl}\, \mathbb{E}[z_{ij}z_{kl} \mid \mathcal{C}_n],$$

where the sum runs over quadruples $(i,j,k,l) \in N_n^4$ with $d_n(i,j) \le b_n$ and $d_n(k,l) \le b_n$. By condition (ii), $|(A_n)_{ij}(A_n)_{kl}| \le \bar{a}_n^2$, so

$$\mathrm{Var}\Big(\hat{V}_{A,n} - V_{A,n} \mid \mathcal{C}_n\Big) \le \bar{a}_n^2 \sum_{(i,j,k,l)} \big|\mathbb{E}[z_{ij}z_{kl} \mid \mathcal{C}_n]\big|.$$

Group the quadruples by the distance between the pairs $\{i,j\}$ and $\{k,l\}$:

$$\mathcal{H}_n(s, b_n) := \big\{(i,j,k,l) \in N_n^4 : d_n(i,j) \le b_n,\ d_n(k,l) \le b_n,\ d_n(\{i,j\},\{k,l\}) = s\big\},$$

so that

$$\mathrm{Var}\Big(\hat{V}_{A,n} - V_{A,n} \mid \mathcal{C}_n\Big) \le \bar{a}_n^2 \sum_{s \ge 0} \sum_{(i,j,k,l) \in \mathcal{H}_n(s,b_n)} \big|\mathbb{E}[z_{ij}z_{kl} \mid \mathcal{C}_n]\big|.$$

**Step 3: Bounding individual covariances.** The variables $z_{ij}$ do not involve $A_n$, so the bounds here are free of $\bar{a}_n$. For $s \ge 1$, the pairs $\{i,j\}$ and $\{k,l\}$ are at distance $s$, so by KMS Theorem A.1 (combined with Assumption 2(a) and condition (i) above with $p > 4$),

$$\big|\mathbb{E}[z_{ij}z_{kl} \mid \mathcal{C}_n]\big| \le C_1 \cdot \theta_{n,s}^{1-4/p} \quad \text{a.s.},$$

for a constant $C_1 > 0$ depending only on $p$ and the moment bound in (i). For $s = 0$, the bound is uniform: $|\mathbb{E}[z_{ij}z_{kl} \mid \mathcal{C}_n]| \le C_1$ by Cauchy–Schwarz applied to the conditional fourth moment, which is finite by condition (i).

**Step 4: Counting quadruples.** By inequality (A.13) of KMS,

$$|\mathcal{H}_n(s, b_n)| \le 4n\, c_n(s, b_n; 2),$$

where $c_n(s, b_n; 2)$ is defined in equation (11) of the paper.

**Step 5: Combining.** Substituting Steps 3 and 4 into the variance bound from Step 2:

$$\operatorname{Var}\Big(\hat{V}_{A,n} - V_{A,n} \mid \mathcal{C}_n\Big) \le 4C_1\, \bar{a}_n^2 \cdot n \sum_{s\ge 0} c_n(s, b_n; 2)\, \theta_{n,s}^{1-4/p} \xrightarrow{\text{a.s.}} 0$$

by condition (iv).

**Step 6: Conclusion.** By Chebyshev's inequality, for any $\varepsilon > 0$,

$$\mathbb{P}\Big(\big|\hat{V}_{A,n} - V_{A,n}\big| > \varepsilon \,\Big|\, \mathcal{C}_n\Big) \le \frac{1}{\varepsilon^2} \operatorname{Var}\Big(\hat{V}_{A,n} - V_{A,n} \mid \mathcal{C}_n\Big) \xrightarrow{\text{a.s.}} 0.$$

Taking unconditional expectations and applying dominated convergence (the conditional probabilities are bounded by 1), $\mathbb{P}\Big(\big|\hat{V}_{A,n} - V_{A,n}\big| > \varepsilon\Big) \to 0$. ☐

*Proof of Lemma 3.* By Slutsky's lemma and the Cramér–Wold device it suffices to show $\frac{1}{\lambda_n}\nu'(\widehat{V}_\bullet - V_\bullet)\nu = o_p(1)$ for any unit vector $\nu$; set $\tilde{Y}_{n,i} := \nu' Y_{n,i}$, a scalar $\psi$-dependent array with the dependence coefficients of $Y_n$. Using $2\sum_{m=1}^{M} \omega(m, M) = M$, decompose

$$A_n^{\text{C2NW}} = \underbrace{J_T \otimes I_G + C_M \otimes J_G}_{=:A_n^{\text{lag}}} + \underbrace{M\, I_T \otimes J_G}_{=:A_n^0},$$

and apply Lemma 2 to each of $A_n^{\text{band}}/\lambda_n$, $A_n^{\text{lag}}/\lambda_n$, and $A_n^0/\lambda_n$; the conclusion for $\bullet = \text{C2NW}$ follows by summing the two latter limits. We verify the conditions for $\bullet \in \{\text{C2NW}, \text{band}\}$.

(i) Moment bound: from Assumption 2(c).

(ii) Weight bound: the entries of $A_n^{\text{band}} = W_n^{\text{CHS}} + C_M \otimes Q_G$ and of $A_n^{\text{lag}}$ are bounded by a constant, so $\bar{a}_n \asymp \lambda_n^{-1}$; the entries of $A_n^0$ equal $M$, so $\bar{a}_n \asymp M/\lambda_n$.

(iii) Support bandwidth: $b_n = M$ for $A_n^{\text{band}}$ and $A_n^{\text{lag}}$, whose nonzero entries lie on pairs $(i, j)$ with $g(i) = g(j)$, $t(i) = t(j)$, or $|t(i) - t(j)| \le M$; and $b_n = 0$ for $A_n^0$, whose nonzero entries lie only on pairs with $t(i) = t(j)$.

(iv) Variance summability: for $A_n^{\text{band}}$ and $A_n^{\text{lag}}$, condition (iv) of Lemma 2 reads, up to a constant, $\frac{n}{\lambda_n^2}\sum_{s\ge 0} c_n(s, M; 2)\theta_{n,s}^{1-4/p} \to 0$, which is Assumption 4. For $A_n^0$ it reads $\frac{M^2 n}{\lambda_n^2}\sum_{s\ge 0} c_n(s, 0; 2)\theta_{n,s}^{1-4/p} \to 0$, which is Assumption 6.

Lemma 2 then gives, for each fixed unit $\nu$, $\frac{1}{\lambda_n}\nu'(\widehat{V}_\bullet - V_\bullet)\nu \xrightarrow{p} 0$. Because $v$ is fixed, evaluating this at the coordinate directions $e_a$ and the pairwise sums $e_a + e_b$ controls every entry of the symmetric $v \times v$ matrix $\widehat{V}_\bullet - V_\bullet$, so $\lambda_n^{-1}\|\widehat{V}_\bullet - V_\bullet\|_{\text{op}} \xrightarrow{p} 0$. Writing $V_\bullet^{-1}\widehat{V}_\bullet - I_v = V_\bullet^{-1}(\widehat{V}_\bullet - V_\bullet)$ and using $\|V_\bullet^{-1}\|_{\text{op}} = \lambda_{\min}(V_\bullet)^{-1} \le \lambda_n^{-1}(1 + o(1))$ from the hypothesis,

$$\big\|V_\bullet^{-1}\widehat{V}_\bullet - I_v\big\|_{\text{op}} \le \lambda_{\min}(V_\bullet)^{-1}\, \|\widehat{V}_\bullet - V_\bullet\|_{\text{op}} \xrightarrow{p} 0,$$

which is the claim. $\square$

**Lemma 4.** *Suppose Assumptions 2 and 5 hold and $Y_{n,i}$ is scalar. Then,*

$$\frac{1}{\lambda_n}\left(V_{true}-V_{adj}\right)\xrightarrow{p}0. \tag{44}$$

*Proof of Lemma 4.* Let $\vartheta_1,\vartheta_2,\vartheta_3$ denote arbitrary positive and finite constants.

$$\begin{aligned}
V_{adj}-V_{true}=&\sum_{m=1}^{M}\sum_{t=1}^{T-m}\left(\omega\left(m,M\right)-1\right)\left(\mathrm{Cov}\left(y_t,y_{t+m}\right)+\mathrm{Cov}\left(y_{t+m},y_t\right)\right)\\
&+\sum_{m\geq 1}\sum_{t=1}^{T-m}\sum_{g=1}^{G}\sum_{i\in N_{t,g}^{T\cap G}}\sum_{j\in N_{t+m,g}^{T\cap G}}\left(\mathrm{Cov}\left(Y_{n,i},Y_{n,j}\right)+\mathrm{Cov}\left(Y_{n,j},Y_{n,i}\right)\right)\\
&+\sum_{m=M+1}^{T-1}\sum_{t=1}^{T-m}\mathrm{Cov}\left(y_t,y_{t+m}\right)+\sum_{m=M+1}^{T-1}\sum_{t=1}^{T-m}\mathrm{Cov}\left(y_{t+m},y_t\right).
\end{aligned} \tag{45}$$

By applying Theorem A1 from KMS, for $d_n\left(i,j\right)=s\geq 1$, we obtain $\left|\mathrm{Cov}\left(Y_{n,i},Y_{n,j}\right)\right|\leq \vartheta_2\theta_{n,s}^{1-\frac{2}{p}}$, where $\vartheta_2=C\left(\mu\vee 1\right)^2\bar{\theta}$ for some constant $C>0$. Then,

$$\begin{aligned}
&\left|\sum_{m=1}^{M}\sum_{t=1}^{T-m}\left(\omega\left(m,M\right)-1\right)\mathrm{Cov}\left(y_t,y_{t+m}\right)\right|\\
&\leq\sum_{m=1}^{M}\sum_{t=1}^{T-m}\left|\omega\left(m,M\right)-1\right|\sum_{i\in N_t^T}\sum_{j\in N_{t+m}^T}\left|\mathrm{Cov}\left(Y_{n,i},Y_{n,j}\right)\right|\\
&\leq\sum_{m=1}^{M}\left|\omega\left(m,M\right)-1\right|\sum_{t=1}^{T}\sum_{i\in N_t^T}\left|N_{t(i)+m}^T\right|\vartheta_2\theta_{n,m}^{1-\frac{2}{p}}\\
&\leq\sum_{m=1}^{M}\left|\omega\left(m,M\right)-1\right|\sum_{i\in N_n}\left|N_{t(i)+m}^T\right|\vartheta_2\theta_{n,m}^{1-\frac{2}{p}}\\
&\leq\vartheta_2 n\sum_{m=1}^{M}\left|\omega\left(m,M\right)-1\right|\delta_n^{\partial}\left(m;1\right)\theta_{n,m}^{1-\frac{2}{p}},
\end{aligned} \tag{46}$$

so it suffices for this scaled object to converge.

Now, we show that $\frac{1}{\lambda_n}\sum_{m\geq 1}\sum_{t=1}^{T-m}\sum_{g=1}^{G}\sum_{i\in N_{t,g}^{T\cap G}}\sum_{j\in N_{t+m,g}^{T\cap G}}\mathrm{Cov}\left(Y_{n,i},Y_{n,j}\right)$ is asymptotically negligible:

$$\sum_{m\geq 1}\sum_{t=1}^{T-m}\sum_{g=1}^{G}\sum_{i\in N_{t,g}^{T\cap G}}\sum_{j\in N_{t+m,g}^{T\cap G}}\left|\mathrm{Cov}\left(Y_{n,i},Y_{n,j}\right)\right|$$
$$\leq \sum_{m\geq 1}\sum_{t=1}^{T-m}\sum_{g=1}^{G}\left|N_{t,g}^{T\cap G}\right|\left|N_{t+m,g}^{T\cap G}\right|\vartheta_2\theta_{n,m}^{1-\frac{2}{p}}. \tag{47}$$

The object converges to zero after scaling due to the assumption used.

Finally, we show $\frac{1}{\lambda_n}\sum_{m=M+1}^{T-1}\sum_{t=1}^{T-m}\mathrm{Cov}\left(y_t,y_{t+m}\right)$ is asymptotically negligible:

$$\sum_{m=M+1}^{T-1}\sum_{t=1}^{T-m}\left|\mathrm{Cov}\left(y_t,y_{t+m}\right)\right|\leq \sum_{m=M+1}^{T-1}\sum_{t=1}^{T-m}\left|N_t^T\right|\left|N_{t+m}^T\right|\vartheta_3\theta_{n,m}^{1-\frac{2}{p}}. \tag{48}$$

The object converges to zero after scaling due to the assumption used. □

## B.4 Proof of Corollaries

*Proof of Corollary 1.* Write $P_G := J_G/G$, $Q_G := I_G - P_G$ and $P_T, Q_T$ analogously; these are orthogonal idempotents on each factor. Substituting $J_G = GP_G$, $J_T = TP_T$, $I = P + Q$ into $W_n^{\text{2-way}}$ and collecting terms in the commuting blocks $\{P_T, Q_T\}\otimes\{P_G, Q_G\}$,

$$W_n^{\text{2-way}} = (G{+}T{-}1)\,P_T{\otimes}P_G + (G{-}1)\,Q_T{\otimes}P_G + (T{-}1)\,P_T{\otimes}Q_G - Q_T{\otimes}Q_G.$$

The coefficients are the eigenvalues; the only negative one is $-1$, on $Q_T\otimes Q_G$ (nonempty iff $G,T\geq 2$). The positive part zeros this block, so $A_n^\star = W_n^{\text{2-way}}{+}Q_T{\otimes}Q_G$. Finally $Q_T\otimes Q_G = (I_T-P_T)\otimes(I_G-P_G) = I_n - W_n^G/T - W_n^T/G + J_n/n = P_{00}$ by inclusion–exclusion, with the stated quadratic form. □

*Proof of Corollary 2.* Decompose $I_G = P_G + Q_G$ and $J_G - I_G = (G-1)P_G - Q_G$, and substitute into $W_n^{\text{CHS}}$. Since $P_G, Q_G$ are orthogonal idempotent,

$$W_n^{\text{CHS}} = \underbrace{\left[J_T + (G-1)C_M\right]}_{=:A}{\otimes}P_G \;+\; \underbrace{\left[J_T - C_M\right]}_{=B}{\otimes}Q_G,$$

an orthogonal direct sum over $\mathbb{R}^T\otimes\mathrm{range}(P_G)$ and $\mathbb{R}^T\otimes\mathrm{range}(Q_G)$, so the positive part acts blockwise: $(W_n^{\text{CHS}})^+ = A^+\otimes P_G + B^+\otimes Q_G$. Both $J_T$ and $C_M$ are PSD (the latter a Fejér kernel), so $A = J_T + (G-1)C_M \succeq 0$ and $A^+ = A$. For the second block, $B^+ = B + B^-$ with $B^- = (C_M - J_T)_+ \succeq 0$, giving $B^+\otimes Q_G = B\otimes Q_G + B^-\otimes Q_G$. Summing, $A_n^\star = W_n^{\text{CHS}} + B^-\otimes Q_G$. Since $Q_G$ acts on the cluster index to produce $\check{Y}$, the quadratic form is $Y_n' W_n^{\text{CHS}} Y_n + \sum_g \check{Y}_{g,\cdot}' B^- \check{Y}_{g,\cdot}$. □

*Proof of Corollary 3.* Write $\sigma_n^2 := \mathrm{Var}(S_n) = \nu' V_{\text{true}} \nu$ and $\phi_n := \nu' V_{\text{true}} \nu / \nu' V^{\star}_{\text{ET}} \nu$; by Theorem 5(a),(c), $\nu' V^{\star}_{\text{ET}} \nu \geq \nu' V_{\text{adj}} \nu = \nu' V_{\text{true}} \nu \,(1 + o(1))$, so $\phi_n \leq 1 + o(1)$. By Theorem 5(b), $\nu' \widehat{V}^{\star}_{\text{ET}} \nu = \nu' V^{\star}_{\text{ET}} \nu \,(1 + o_p(1))$, hence

$$r_n^2 := \frac{\sigma_n^2}{\nu' \widehat{V}^{\star}_{\text{ET}} \nu} = \phi_n \, \frac{\nu' V^{\star}_{\text{ET}} \nu}{\nu' \widehat{V}^{\star}_{\text{ET}} \nu} = \phi_n \,(1 + o_p(1)) \;\leq\; 1 + o_p(1).$$

Since $|S_n| / \sqrt{\nu' \widehat{V}^{\star}_{\text{ET}} \nu} = (|S_n|/\sigma_n)\, r_n$ and Theorem 4 gives $|S_n|/\sigma_n \xrightarrow{d} |Z|$, $Z \sim N(0,1)$, fix $\varepsilon > 0$. On $\{r_n \leq 1 + \varepsilon\}$, $\{(|S_n|/\sigma_n)\, r_n > z_{1-\alpha/2}\} \subseteq \{|S_n|/\sigma_n > z_{1-\alpha/2}/(1+\varepsilon)\}$, so

$$\Pr\Big\{ \tfrac{|S_n|}{\sqrt{\nu' \widehat{V}^{\star}_{\text{ET}} \nu}} > z_{1-\alpha/2} \Big\} \;\leq\; \Pr\Big\{ \tfrac{|S_n|}{\sigma_n} > \tfrac{z_{1-\alpha/2}}{1+\varepsilon} \Big\} + \Pr\{ r_n > 1 + \varepsilon \}.$$

The second term vanishes since $r_n \leq 1 + o_p(1)$; the first, $z_{1-\alpha/2}/(1+\varepsilon)$ being a continuity point of the law of $|Z|$, converges to $2\Phi\big(-z_{1-\alpha/2}/(1+\varepsilon)\big)$. Taking $\limsup_n$ and then $\varepsilon \downarrow 0$,

$$\limsup_{n \to \infty} \Pr\Big\{ \tfrac{|S_n|}{\sqrt{\nu' \widehat{V}^{\star}_{\text{ET}} \nu}} > z_{1-\alpha/2} \Big\} \;\leq\; 2\Phi(-z_{1-\alpha/2}) = \alpha.$$

For the equality claim, $r_n^2 = \phi_n (1 + o_p(1))$ with $\phi_n \leq 1 + o(1)$ deterministic, so the limsup attains $\alpha$ iff $\phi_n \to 1$; by (a) and (c) this holds iff $\nu'(V^{\star}_{\text{ET}} - V_{\text{true}})\nu = o\big(\nu' V_{\text{true}} \nu\big)$. $\square$

# References


Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2020): "Sampling-based versus design-based uncertainty in regression analysis," *Econometrica*, 88, 265–296.

——— (2023): "When should you adjust standard errors for clustering?" *The Quarterly Journal of Economics*, 138, 1–35.

Andrews, D. W. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica: Journal of the Econometric Society*, 817–858.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011): "Robust inference with multiway clustering," *Journal of Business & Economic Statistics*, 29, 238–249.

Casini, A. (2023): "Theory of evolutionary spectra for heteroskedasticity and autocorrelation robust inference in possibly misspecified and nonstationary models," *Journal of Econometrics*, 235, 372–392.

Chan, K. W. (2022): "Optimal difference-based variance estimators in time series: A general framework," *The Annals of Statistics*, 50, 1376–1400.

Chiang, H. D., B. E. Hansen, and Y. Sasaki (2024): "Standard errors for two-way clustering with serially correlated time effects," *Review of Economics and Statistics*, 1–40.

Davezies, L., X. D'haultfœuille, and Y. Guyonvarch (2021): "Empirical process results for exchangeable arrays," *The Annals of Statistics*, 49, 845–862.

Fama, E. F. and K. R. French (1993): "Common risk factors in the returns on stocks and bonds," *Journal of financial economics*, 33, 3–56.

Kojevnikov, D., V. Marmer, and K. Song (2021): "Limit theorems for network dependent random variables," *Journal of Econometrics*, 222, 882–908.

Newey, W. and K. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Xu, R. and L. Yap (2024): "Clustering with Potential Multidimensionality: Inference and Practice," *arXiv preprint arXiv:2411.13372*.

Yap, L. (2025): "Asymptotic theory for two-way clustering," *Journal of Econometrics*, 249, 106001.