

Article

Adaptive Wrapped Robust Canonical Correlation Analysis in High-Dimensional Data

Hasan Bulut ^{1,*} , Müjgan Zobu ² and Vedat Sağlam ¹¹ Department of Statistics, Faculty of Science, Ondokuz Mayıs University, Samsun 55139, Türkiye; vsaglam@omu.edu.tr² Department of Statistics, Faculty of Science and Letters, Amasya University, Amasya 05100, Türkiye; mujgan.zobu@amasya.edu.tr

* Correspondence: hasan.bulut@omu.edu.tr

Abstract

Classical canonical correlation analysis becomes numerically unstable when the number of variables is large relative to the sample size and is sensitive to contamination in observations or individual cells. This study develops an integrated robust and regularized procedure that combines bounded cellwise wrapping, shrinkage estimation of the joint correlation matrix, and robust reweighting in a low-dimensional canonical score space. The resulting observation weights enter a second regularized canonical correlation fit, so the final estimator remains well defined when the combined number of variables exceeds the sample size. The simulation study shows that relative estimation accuracy depends on the signal strength, contamination mechanism, and dimensional configuration. The proposed estimator is competitive in several moderate-signal settings and has a clear computational advantage, whereas the minimum regularized covariance determinant plug-in estimator provides lower estimation error in many high-signal configurations. An additional ultra-high-dimensional experiment demonstrates numerical feasibility with modest memory use but also reveals substantial attenuation, identifying a limitation of the present dense estimator. The results therefore support a regime-dependent interpretation rather than a claim of uniform superiority. The complete reproducible simulation workflow is provided.

Keywords: robust canonical correlation analysis; robust statistics; cellwise contamination; high-dimensional data; shrinkage correlation; wrapping transformation; minimum regularized covariance determinant

MSC: 62H20; 62H12; 62F35; 62J07



Academic Editor: Jiancang Zhuang

Received: 15 June 2026

Revised: 19 July 2026

Accepted: 24 July 2026

Published: 27 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Canonical Correlation Analysis (CCA), introduced by Hotelling [1], is a widely used multivariate method for studying linear associations between two sets of variables. It constructs linear combinations of the variables in each block and identifies the combinations with the largest correlation. By identifying linear combinations of variables in each set that achieve maximal correlation, CCA provides a lower-dimensional summary that captures the strongest dependencies between complex data sets [2,3]. Despite its wide applicability in fields such as genomics, chemometrics, and finance, the effectiveness of classical CCA is severely challenged by two defining characteristics of modern data: high dimensionality ($p > n$) and outlier contamination [4,5].

In high-dimensional settings where the number of variables exceeds the sample size, the sample covariance matrix becomes singular, rendering the matrix inversions required for classical CCA impossible [6,7]. To mitigate this, regularized and sparse extensions have been proposed. For instance, the Penalized Matrix Decomposition (PMD) by Witten et al. [8] utilizes ℓ_1 penalties to induce sparsity and handle high dimensions. However, these methods are non-robust and can be easily distorted by even a few atypical observations.

Robust statistics traditionally focused on casewise (rowwise) outliers, where an entire observation vector deviates from the bulk of the data. A landmark development for high-dimensional scatter estimation is the Minimum Regularized Covariance Determinant (MRCD) estimator proposed by Boudt et al. [9]. MRCD extends the classic MCD approach [10] by employing a ridge-type regularization, ensuring the estimated matrix is well-conditioned and invertible even when $p > n$. A common strategy in current literature is the plug-in approach, where MRCD estimates are substituted into the classical CCA framework to achieve robustness [11,12].

However, a more recent paradigm shift has highlighted the danger of cellwise (elementwise) outliers, where individual entries within a data matrix are corrupted. In high dimensions, cellwise contamination is particularly problematic due to “outlier propagation”: even a small fraction of contaminated cells can affect a vast majority of rows, causing traditional casewise robust methods like MRCD to break down [13,14]. To tackle the simultaneous need for robustness and sparsity, Pfeiffer et al. [11] proposed the ccaMM algorithm. This method utilizes a flexible optimization framework based on the Method of Multipliers (MM) and gradient descent, allowing for the direct computation of sparse canonical vectors. While ccaMM is a powerful tool, its overall robustness depends heavily on the initial plug-in covariance estimator used, such as MRCD or OGK [11].

In this study, we develop an integrated robust and regularized canonical correlation procedure for data sets that may simultaneously exhibit high dimensionality and cellwise or casewise contamination. The procedure combines three established statistical ideas: bounded wrapping of individual observations, shrinkage regularization of the joint correlation matrix, and robust reweighting in a low-dimensional canonical score space. The individual components are not presented as new. The methodological contribution lies in their coordinated use within a single estimator and, in particular, in using the score-space robust weights in a second shrinkage-regularized canonical correlation fit. This construction remains computationally feasible in the regime $p + q > n$, where an unregularized covariance matrix is singular.

The empirical study is therefore designed to characterize regime-dependent statistical and computational trade-offs rather than to establish uniform superiority over existing robust canonical correlation procedures.

Accordingly, the objectives of this study are to: (i) formulate the integrated estimator and clearly describe the role of each component; (ii) investigate its finite-sample behaviour under clean, casewise-contaminated, and cellwise-contaminated settings; (iii) examine its computational feasibility in high- and ultra-high-dimensional settings; and (iv) provide a reproducible software implementation. We do not claim that wrapping, shrinkage, or robust score estimation is individually novel; instead, the contribution concerns the estimator construction, the resulting robust high-dimensional workflow, and its systematic evaluation against relevant competing procedures.

For clarity, the abbreviations used throughout the manuscript are as follows: canonical correlation analysis (CCA), Adaptive Wrapped Robust Canonical Correlation Analysis (AWRcca), minimum covariance determinant (MCD), minimum regularized covariance determinant (MRCD), positive semidefinite (PSD), root mean squared error (RMSE).

The remainder of this article is organized as follows: Section 2 provides a comprehensive review of existing methodologies for robust and high-dimensional CCA, focusing on the ccaMM framework and MRCD-based plug-in approaches [9,11]. Section 3 introduces the proposed AWRcca algorithm, detailing its technical components such as robust wrapping transformations, consistency corrections, and analytical shrinkage. Section 4 presents a simulation study comparing AWRcca with the selected robust competing methods under clean, casewise-contaminated, cellwise-contaminated, and mixed-contamination settings. Section 5 demonstrates the practical utility of the proposed method through an application to a real-world high-dimensional data set [11,15]. Section 6 introduces a dedicated R function developed to implement the AWRcca algorithm, providing users with a computationally efficient tool for robust multivariate analysis [16]. Finally, Section 7 concludes the paper with a summary of the main contributions and potential directions for future research.

2. Existing Robust CCA Approaches

In this section, we briefly review the robust canonical correlation analysis (CCA) methods that serve as benchmarks in our simulation study. Specifically, we focus on two widely used approaches for handling high-dimensional and contaminated data: (i) the MRCD-based plug-in CCA method, and (ii) the ccaMM framework based on constrained optimization and sparse robust estimation.

2.1. Canonical Correlation Analysis Based on MRCD Plug-In

A natural robust extension of classical CCA is obtained by replacing the sample covariance matrix with a robust scatter estimator. In high-dimensional settings ($p + q > n$), the sample covariance matrix is singular and cannot be inverted, which makes classical CCA infeasible. To address this issue, Boudt et al. [9] proposed the Minimum Regularized Covariance Determinant (MRCD) estimator, which extends the classical MCD estimator [10] through ridge-type regularization.

Let $Z = [X, Y] \in \mathbb{R}^{n \times (p+q)}$ denote the joint data matrix. The MRCD estimator computes a robust location vector $\hat{\mu}_{\text{MRCD}}$ and a positive definite scatter matrix $\hat{\Sigma}_{\text{MRCD}}$ of the form

$$\hat{\Sigma}_{\text{MRCD}} = (1 - \lambda)\hat{\Sigma}_{\text{MCD}} + \lambda I_{p+q}, \quad (1)$$

where $\hat{\Sigma}_{\text{MCD}}$ is the classical MCD scatter matrix computed on a high-breakdown subset of the data, $\lambda \in (0, 1)$ is a regularization parameter, and I_{p+q} denotes the identity matrix. This regularization guarantees positive definiteness and numerical stability even when $p + q > n$.

In the MRCD plug-in CCA approach, the robust covariance estimator $\hat{\Sigma}_{\text{MRCD}}$ is transformed into a robust correlation matrix by standardizing with its diagonal elements. The resulting block correlation matrix is then partitioned into submatrices corresponding to X and Y , and canonical correlations are obtained by applying the classical CCA eigenvalue or singular value decomposition to these robust blocks. This approach has been widely adopted in recent robust multivariate literature due to its conceptual simplicity and strong protection against casewise outliers [11].

However, since MRCD is fundamentally a casewise robust estimator, its robustness deteriorates under cellwise contamination, where even a small fraction of contaminated cells may propagate across rows and compromise the estimation accuracy [13,14]. This limitation motivates the comparison with more flexible optimization-based approaches.

2.2. ccaMM: Robust and Sparse Canonical Correlation via MM Optimization

To simultaneously address robustness, sparsity, and high dimensionality, Pfeiffer et al. [11] proposed the ccaMM algorithm, which formulates CCA as a constrained optimization problem solved using the Method of Multipliers (MM). Unlike plug-in approaches, ccaMM directly estimates the canonical vectors by maximizing an association measure subject to normalization and sparsity constraints.

Let $a \in \mathbb{R}^p$ and $b \in \mathbb{R}^q$ denote the canonical vectors. The ccaMM framework seeks to maximize a robust association criterion of the form

$$\max_{a,b} a^\top \hat{\Sigma}_{xy} b \quad \text{subject to} \quad a^\top \hat{\Sigma}_{xx} a = 1, \quad b^\top \hat{\Sigma}_{yy} b = 1, \quad (2)$$

where $\hat{\Sigma}_{xx}$, $\hat{\Sigma}_{yy}$, and $\hat{\Sigma}_{xy}$ are robust covariance or correlation estimates. Sparsity is induced through ℓ_1 -type penalties on a and b , controlled by user-specified penalty parameters.

The optimization problem is solved iteratively using gradient-based updates embedded in an MM scheme, which ensures monotone convergence of the objective function. Robustness is introduced through the choice of the initial scatter estimator, such as MRCD or OGK, and through optional nearest positive definite (nearPD) corrections that stabilize the optimization in high-dimensional settings [11].

While ccaMM provides a highly flexible and general framework for sparse robust CCA, its performance and robustness properties depend critically on the underlying scatter estimator and on several tuning parameters, including learning rates, penalty weights, and stopping criteria. Moreover, the iterative nature of the algorithm leads to substantially higher computational costs compared to plug-in approaches, particularly in large-scale simulation studies.

2.3. Positioning of the Proposed Method

The methods considered in this study represent different strategies for robust canonical correlation analysis. Plug-in procedures replace the classical covariance matrix with a robust covariance estimator, whereas optimization-based procedures incorporate a robust loss directly into the canonical correlation objective. The proposed procedure follows a third strategy: it limits the effect of individual cells before covariance estimation, stabilizes the resulting correlation matrix by shrinkage, and subsequently refines the fit using weights estimated in the low-dimensional canonical score space.

The proposed method should therefore be interpreted as an integrated transformation–regularization–reweighting procedure rather than as a universally superior replacement for existing robust canonical correlation methods. The empirical comparisons are intended to identify the contamination and dimensional regimes in which the different procedures are effective, as well as their computational and statistical trade-offs.

3. Proposed Adaptive Wrapped Robust Canonical Correlation Analysis

We construct the Adaptive Wrapped Robust Canonical Correlation Analysis procedure from established robust and regularized components. The procedure first applies a bounded wrapping transformation to robustly standardized variables, then estimates a shrinkage-regularized joint correlation matrix, and finally uses robust distances in the initial canonical score space to define observation weights. These weights enter a second regularized canonical correlation fit, so that the final estimator remains well defined even when the combined dimension exceeds the sample size. Algorithm 1 summarizes this estimation sequence. The starting point of AWRcca is the observation that, in high-dimensional settings, even a small fraction of contaminated cells may propagate across observations and severely distort covariance and correlation estimates. To prevent this phenomenon,

AWRcca first applies a columnwise wrapping transformation based on robust location and scale estimates. By shrinking moderate outliers and neutralizing extreme values at the variable level, this transformation suppresses the influence of cellwise contamination without discarding entire observations, thereby preserving valuable information in partially contaminated rows.

More specifically, let $\mathbf{X} = (x_{ij})$ denote the data matrix. Each variable is first standardized using robust location and scale estimators, namely the median and the median absolute deviation (MAD):

$$z_{ij} = \frac{x_{ij} - \text{med}(x_j)}{\text{MAD}(x_j)}. \quad (3)$$

Following Raymaekers and Rousseeuw [17], the wrapping transformation is defined as

$$\psi_{b,c}(z) = \begin{cases} z, & |z| \leq b, \\ q_1 \tanh(q_2(c - |z|)) \text{sign}(z), & b < |z| < c, \\ 0, & |z| \geq c, \end{cases} \quad (4)$$

where $0 < b < c$ are cutoff constants and q_1, q_2 are scaling parameters ensuring smoothness. In practice, we set $b = 1.5$ and $c = 4$.

Since the wrapping transformation shrinks extreme values, it alters the scale of the variables. To restore consistency under the Gaussian model, a scaling factor is introduced. Specifically, let $Z \sim \mathcal{N}(0, 1)$ and define a constant ξ such that

$$\text{Var}(\xi \psi_{b,c}(Z)) = 1, \quad Z \sim \mathcal{N}(0, 1). \quad (5)$$

Because the wrapping function in Equation (4) is odd and the standard normal distribution is symmetric,

$$\mathbb{E}[\psi_{b,c}(Z)] = 0.$$

Consequently, the consistency condition in Equation (5) gives

$$\xi = \left\{ \mathbb{E}[\psi_{b,c}(Z)^2] \right\}^{-1/2}.$$

Writing $\phi(\cdot)$ for the standard normal density and using the piecewise definition of $\psi_{b,c}$, this constant can be evaluated deterministically as

$$\xi = \left[2 \int_0^b z^2 \phi(z) \, dz + 2q_1^2 \int_b^c \tanh^2\{q_2(c - z)\} \phi(z) \, dz \right]^{-1/2}.$$

For the constants used throughout this study,

$$b = 1.5, \quad c = 4, \quad q_1 = 1.540793, \quad q_2 = 0.8622731,$$

numerical integration gives

$$\mathbb{E}[\psi_{b,c}(Z)^2] = 0.7532528310$$

and therefore

$$\xi = 1.152204624.$$

This deterministic value was used in the simulation study and can be reproduced by direct numerical evaluation of the preceding one-dimensional integrals. Throughout the remainder of the manuscript, ξ denotes this Gaussian consistency correction factor.

The final transformed observations are obtained as

$$\tilde{x}_{ij} = \zeta \psi_{b,c}(z_{ij}) \cdot \text{MAD}(x_j) + \text{med}(x_j). \quad (6)$$

This transformation preserves statistical efficiency under Gaussian assumptions while limiting the influence of cellwise contamination.

Let $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ denote the transformed data matrices. The sample covariance and cross-covariance matrices used in AWRcca are computed as

$$\mathbf{S}_{XX} = \frac{1}{n-1} (\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}})^\top (\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}}), \quad (7)$$

$$\mathbf{S}_{YY} = \frac{1}{n-1} (\tilde{\mathbf{Y}} - \bar{\tilde{\mathbf{Y}}})^\top (\tilde{\mathbf{Y}} - \bar{\tilde{\mathbf{Y}}}), \quad (8)$$

$$\mathbf{S}_{XY} = \frac{1}{n-1} (\tilde{\mathbf{X}} - \bar{\tilde{\mathbf{X}}})^\top (\tilde{\mathbf{Y}} - \bar{\tilde{\mathbf{Y}}}), \quad (9)$$

where

$$\bar{\tilde{\mathbf{X}}} = \frac{1}{n} \sum_{i=1}^n \tilde{\mathbf{x}}_i, \quad \bar{\tilde{\mathbf{Y}}} = \frac{1}{n} \sum_{i=1}^n \tilde{\mathbf{y}}_i$$

are the column-wise sample mean vectors. These covariance matrices inherit robustness from the wrapping transformation by limiting the influence of cellwise contamination, while retaining consistency and efficiency under Gaussian assumptions.

After transformation, the robustly standardized data are used to construct a correlation matrix. In high-dimensional regimes where the total number of variables exceeds the sample size, this matrix may be ill-conditioned or numerically indefinite. To guarantee a valid correlation structure, the matrix is projected onto the space of positive semi-definite correlation matrices whenever necessary, ensuring numerical stability for subsequent computations.

To further stabilize estimation in the $p + q > n$ setting, AWRcca incorporates an analytical shrinkage regularization of the correlation matrix. The shrinkage intensity is estimated directly from the data using a Ledoit–Wolf type rule, yielding a well-conditioned matrix that balances variance reduction and bias without requiring cross-validation [18].

Canonical correlations are then obtained via a singular value decomposition of a whitened cross-covariance matrix. This closed-form solution provides an initial robust estimate of the canonical correlations and vectors.

Although the preceding steps effectively address cellwise contamination and numerical instability, structural casewise outliers may still persist. To detect such observations, AWRcca projects the data onto a low-dimensional canonical score space and computes robust distances using a Minimum Covariance Determinant (MCD) estimator. Observations with large robust distances are downweighted, yielding weights that reflect their conformity to the dominant canonical structure.

In the final stage, the score-space weights are incorporated into the estimation of a weighted joint correlation matrix. The weighted matrix is regularized using the corresponding shrinkage estimate, and the final canonical correlations and directions are obtained from a second regularized canonical decomposition. This construction avoids the dimensional restriction of classical CCA fitted only to the retained observations and remains computationally feasible when $p + q > n$.

The complete estimation sequence is summarized in Algorithm 1. The algorithm reports the statistical operations required to reproduce the estimator, whereas input validation, zero-scale handling, and numerical matrix safeguards are implemented at the software level.

Algorithm 1. Adaptive Wrapped Robust Canonical Correlation Analysis (AWRcca)**Input:** $\mathbf{X} \in \mathbb{R}^{n \times p}$, $\mathbf{Y} \in \mathbb{R}^{n \times q}$; wrapping constants $0 < b < c$; cutoff probability α .**Output:** $\hat{\rho}, \hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{w}}, \hat{\lambda}^{(0)}, \hat{\lambda}^{(w)}$.**Robust wrapping**1: $\hat{\mu}_x \leftarrow \text{med}(\mathbf{X}), \hat{\sigma}_x \leftarrow \text{MAD}(\mathbf{X}); \quad \hat{\mu}_y \leftarrow \text{med}(\mathbf{Y}), \hat{\sigma}_y \leftarrow \text{MAD}(\mathbf{Y}).$ 2: $\mathbf{Z}_x \leftarrow (\mathbf{X} - \mathbf{1}\hat{\mu}_x^\top) \text{diag}(\hat{\sigma}_x)^{-1}; \quad \mathbf{Z}_y \leftarrow (\mathbf{Y} - \mathbf{1}\hat{\mu}_y^\top) \text{diag}(\hat{\sigma}_y)^{-1}.$ 3: $\mathbf{W}_x \leftarrow \xi\psi_{b,c}(\mathbf{Z}_x); \quad \mathbf{W}_y \leftarrow \xi\psi_{b,c}(\mathbf{Z}_y); \quad \mathbf{W} \leftarrow [\mathbf{W}_x, \mathbf{W}_y].$ **Initial regularized solution**4: $\hat{\mathbf{R}}^{(0)} \leftarrow \text{Cor}(\mathbf{W}); \quad \hat{\lambda}^{(0)} \leftarrow \text{LW}(\mathbf{W}).$ 5: $\hat{\mathbf{R}}_{\text{reg}}^{(0)} \leftarrow (1 - \hat{\lambda}^{(0)})\hat{\mathbf{R}}^{(0)} + \hat{\lambda}^{(0)}\mathbf{I}_{p+q}.$ 6: $(\hat{\rho}^{(0)}, \hat{\mathbf{A}}^{(0)}, \hat{\mathbf{B}}^{(0)}) \leftarrow \mathcal{C}(\hat{\mathbf{R}}_{\text{reg}}^{(0)}).$ **Score-space reweighting**7: $m \leftarrow \min\{5, \text{rank}(\hat{\mathbf{R}}_{xy}^{(0)})\}.$ 8: $\mathbf{S} \leftarrow [\mathbf{W}_x \hat{\mathbf{A}}_{(:,1:m)}^{(0)}, \mathbf{W}_y \hat{\mathbf{B}}_{(:,1:m)}^{(0)}].$ 9: $(\hat{\mu}_S, \hat{\Sigma}_S) \leftarrow \text{MCD}(\mathbf{S}).$ 10: **for** $i = 1, \dots, n$ **do**11: $RD_i^2 \leftarrow (\mathbf{s}_i - \hat{\mu}_S)^\top \hat{\Sigma}_S^{-1} (\mathbf{s}_i - \hat{\mu}_S).$ 12: $\hat{w}_i \leftarrow \mathbb{I}\{RD_i^2 \leq \chi_{2m,\alpha}^2\}.$ 13: **end for****Final weighted regularized solution**14: $\hat{\mathbf{R}}^{(w)} \leftarrow \text{Cor}_{\hat{\mathbf{w}}}(\mathbf{W}); \quad \hat{\lambda}^{(w)} \leftarrow \text{LW}_{\hat{\mathbf{w}}}(\mathbf{W}).$ 15: $\hat{\mathbf{R}}_{\text{reg}}^{(w)} \leftarrow (1 - \hat{\lambda}^{(w)})\hat{\mathbf{R}}^{(w)} + \hat{\lambda}^{(w)}\mathbf{I}_{p+q}.$ 16: $(\hat{\rho}, \hat{\mathbf{A}}, \hat{\mathbf{B}}) \leftarrow \mathcal{C}(\hat{\mathbf{R}}_{\text{reg}}^{(w)}).$ 17: **return** $\hat{\rho}, \hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{w}}, \hat{\lambda}^{(0)}, \hat{\lambda}^{(w)}.$

In Algorithm 1, $\text{Cor}_{\mathbf{w}}(\cdot)$ denotes a weighted correlation operator and $\text{LW}_{\mathbf{w}}(\cdot)$ denotes its corresponding shrinkage rule. For a partitioned regularized correlation matrix $\mathbf{R}$, the operator $\mathcal{C}(\mathbf{R})$ represents the canonical decomposition

$$\mathbf{T} = \mathbf{R}_{xx}^{-1/2} \mathbf{R}_{xy} \mathbf{R}_{yy}^{-1/2} = \mathbf{U} \mathbf{D} \mathbf{V}^\top,$$

with

$$\hat{\rho} = \text{diag}(\mathbf{D}), \quad \hat{\mathbf{A}} = \mathbf{R}_{xx}^{-1/2} \mathbf{U}, \quad \hat{\mathbf{B}} = \mathbf{R}_{yy}^{-1/2} \mathbf{V}.$$

Unlike an unregularized reweighting step, the final stage does not require the number of retained observations to exceed the combined dimension. The weighted correlation matrix is regularized before the final canonical decomposition. Therefore, the final estimation step remains well defined in the high-dimensional regime $p + q > n$, subject to the usual numerical conditions required for computing the weighted correlations.

The robustness of AWRcca is produced by two complementary mechanisms. The wrapping transformation limits the contribution of individual cells before joint correlation estimation, whereas score-space reweighting limits the contribution of observations that are atypical with respect to the estimated canonical structure. These mechanisms can be clarified through the boundedness properties of the corresponding transformations.

Let

$$\kappa_{b,c} = \sup_{z \in \mathbb{R}} |\psi_{b,c}(z)|.$$

For the wrapping function in Equation (4), $\kappa_{b,c} < \infty$, and one may use

$$\kappa_{b,c} \leq \max\{b, q_1\},$$

because the transformation equals z when $|z| \leq b$, is bounded by q_1 when $b < |z| < c$, and is zero when $|z| \geq c$. It follows that every consistency-corrected wrapped cell satisfies

$$|\xi \psi_{b,c}(z_{ij})| \leq \xi \kappa_{b,c}.$$

Conditional on positive and finite marginal scale estimates, the absolute contribution of a transformed observation to an unnormalized covariance entry is therefore bounded by

$$\xi^2 \kappa_{b,c}^2.$$

Hence, an arbitrarily large value in an individual cell cannot make its transformed covariance contribution arbitrarily large. Moreover, because $\psi_{b,c}(z) = 0$ for $|z| \geq c$, the transformation is redescending: sufficiently extreme standardized cells have zero direct contribution after wrapping.

The casewise protection mechanism operates in the low-dimensional canonical score space. For observation i , AWRcca assigns the weight

$$\hat{w}_i = \mathbb{I}\{RD_i^2 \leq \chi_{2m,\alpha}^2\}.$$

An observation exceeding the robust distance cutoff has $\hat{w}_i = 0$ and consequently does not contribute to the weighted correlation matrix used in the final regularized fit. This step limits the effect of observations that are jointly atypical even when their individual cells are not sufficiently extreme to be removed by the wrapping transformation. Its effectiveness is conditional on the initial wrapped fit providing an informative low-dimensional score representation.

Regularization provides a complementary numerical property. Let $\hat{\mathbf{R}}$ denote either the initial or weighted positive semidefinite correlation estimate. For

$$\hat{\mathbf{R}}_{\text{reg}} = (1 - \hat{\lambda})\hat{\mathbf{R}} + \hat{\lambda}\mathbf{I}_{p+q}, \quad 0 < \hat{\lambda} \leq 1,$$

positive semidefiniteness of $\hat{\mathbf{R}}$ gives

$$\lambda_{\min}(\hat{\mathbf{R}}_{\text{reg}}) \geq \hat{\lambda} > 0.$$

Thus, the regularized matrix is positive definite and invertible, including when $p + q > n$. This result applies to both the initial and final weighted fits and prevents singularity of the matrices required for the canonical decomposition.

These properties explain how AWRcca bounds the direct effect of contaminated cells, excludes observations identified as atypical in the canonical score space, and preserves numerical well-posedness in high dimensions. They do not constitute a complete influence-function characterization or establish a universal finite-sample breakdown point for the full adaptive estimator. Such properties also depend on the marginal location and scale estimates, the initial score representation, the shrinkage rule, and the contamination configuration. The simulation evidence in Section 4 should therefore be interpreted as a complement to these theoretical robustness mechanisms.

The computational complexity of AWRcca can be described in terms of the sample size n and the combined dimension $d = p + q$. Columnwise robust standardization and wrapping require $O(nd)$ operations. Forming the dense initial joint correlation matrix requires $O(nd^2)$ operations and $O(d^2)$ storage. The positive semidefinite correction, matrix

whitening, and dense spectral decompositions have worst-case complexity $O(d^3)$. More explicitly, the canonical decomposition requires operations of order

$$O(p^3 + q^3 + pq \min\{p, q\}),$$

which is bounded above by $O(d^3)$.

The score-space reweighting step is performed in dimension $2m$, where

$$m \leq 5.$$

If the number of MCD initialization and refinement steps is treated as fixed, constructing the score matrix and computing its robust distances require approximately $O(ndm + nm^2)$ operations and $O(nm)$ additional storage. Because m is bounded by five, this part is low-dimensional and is not the dominant operation in the considered high-dimensional settings. The final weighted correlation matrix requires a second $O(nd^2)$ operation, followed by another dense regularized canonical decomposition with worst-case complexity $O(d^3)$.

Consequently, the overall worst-case time complexity of the dense AWRcca implementation is

$$O(nd^2 + d^3),$$

and its memory requirement is

$$O(nd + d^2).$$

When d is moderate relative to n , construction of the two correlation matrices can account for a substantial part of the computation. When d is very large relative to n , the dense spectral operations become dominant. The approximately 30-min runtime observed in the $n = 100$, $p = 5000$, $q = 60$ experiment is consistent with this dense-matrix complexity. The corresponding peak memory use below 1.75 GiB is also consistent with storage dominated by a limited number of dense $d \times d$ matrices and their numerical workspaces.

This complexity analysis concerns the present dense implementation and does not by itself establish a universal runtime ordering between methods. The empirical runtime comparisons in Section 4 additionally reflect implementation details, convergence behavior, software libraries, and hardware. Sparse or iterative matrix methods could reduce the cost of future AWRcca extensions when only a small number of canonical directions is required.

4. Simulation Study

This section evaluates the finite-sample accuracy, robustness, numerical stability, and computational cost of AWRcca relative to the MRCD plug-in estimator and ccaMM. The revised design distinguishes genuinely high-dimensional configurations from lower-dimensional settings, uses a data generating mechanism whose first population canonical correlation is known exactly, and reports both estimation error and method failures. Detailed configuration-specific results for the revised simulation study, including RMSE and computation-time summaries, are provided in Appendix A.

4.1. Simulation Design

For each replication, a common latent variable $Z_i \sim \mathcal{N}(0, 1)$ was generated together with independent noise vectors. Let $\mathbf{a} \in \mathbb{R}^p$ and $\mathbf{b} \in \mathbb{R}^q$ denote loading vectors obtained from independent uniform draws on $(0.8, 1.2)$ and normalized to unit Euclidean length. The two data blocks were generated as

$$\mathbf{X} = \sqrt{\rho} \mathbf{Z} \mathbf{a}^\top + \sqrt{1 - \rho} \mathbf{E}_x, \quad \mathbf{Y} = \sqrt{\rho} \mathbf{Z} \mathbf{b}^\top + \sqrt{1 - \rho} \mathbf{E}_y. \quad (10)$$

Because the loading vectors have unit norm and the noise variables have unit variance, the first population canonical correlation under Equation (10) is exactly ρ . In the Gaussian setting, the entries of the noise matrices were independent standard normal variables. In the heavy-tailed setting, the noise variables followed a Student distribution with three degrees of freedom and were divided by $\sqrt{3}$ to obtain unit variance.

The main simulation crossed sample sizes $n \in \{30, 60, 100\}$, dimensions $p \in \{50, 100\}$ and $q \in \{60, 120\}$, population correlations $\rho \in \{0.5, 0.8\}$, and Gaussian and heavy-tailed distributions. Casewise and cellwise contamination proportions were independently varied over $\{0, 0.10, 0.25\}$.

This produced 432 configurations. Each configuration was independently replicated 100 times, giving 43,200 generated data sets and 129,600 fitted method–data set combinations.

For casewise contamination, the selected rows of the two blocks were replaced by independent Gaussian observations shifted by 8 in the first block and by -8 in the second block. For cellwise contamination, the selected entries of the combined matrix were multiplied by 10. Mixed contamination was created by applying both mechanisms. The four reporting regimes are therefore clean data, casewise contamination only, cellwise contamination only, and simultaneous casewise and cellwise contamination.

The fixed AWRcca constants were $b = 1.5$, $c = 4$, $q_1 = 1.540793$, $q_2 = 0.8622731$, and a score-space cutoff probability of 0.975. At most five canonical score pairs were used in the reweighting step. The MRCD plug-in estimator used a coverage fraction of 0.75 and an inverse ridge value of 0.2. The ccaMM implementation used 2000 epochs, a learning rate of 10^{-3} , a convergence tolerance of 10^{-5} , and penalty and mixing parameters of 0.05 and 0.5, respectively. These settings were fixed before examining the simulation results.

For an estimate $\hat{\rho}_r$ from replication r , performance was assessed using bias, root mean squared error, and mean absolute error:

$$\text{Bias} = \frac{1}{B} \sum_{r=1}^B (\hat{\rho}_r - \rho), \quad \text{RMSE} = \left\{ \frac{1}{B} \sum_{r=1}^B (\hat{\rho}_r - \rho)^2 \right\}^{1/2}, \quad (11)$$

$$\text{MAE} = \frac{1}{B} \sum_{r=1}^B |\hat{\rho}_r - \rho|. \quad (12)$$

Runtime and failure rate were also recorded. A fit was classified as failed when it returned an error or a non-finite estimate. All computations were performed using independent, explicitly recorded random number seeds.

4.2. Overall Results

All three methods completed every requested fit. Thus, the observed failure rate was zero for AWRcca, MRCD, and ccaMM across all 129,600 main simulation fits. Table 1 summarizes performance after pooling within each contamination regime. Because this aggregation combines two population-correlation levels, the regime-specific values should be interpreted together with the stratified results in Figures 1–4.

Because all three methods were applied to the same generated data sets, their estimation errors and computation times were compared using paired analyses. For each of the 432 simulation configurations, the 100 replication-level squared errors and computation times were first averaged within method. Uncertainty in the aggregate RMSE differences was then evaluated using a cluster bootstrap that resampled complete configurations within each contamination regime. This procedure preserves the pairing of the methods and keeps all replications belonging to the same configuration together. A total of 10,000 bootstrap samples were generated using a fixed random number seed.

Table 1. Aggregate simulation performance by contamination regime. Each row pools the relevant dimensional, distributional, contamination-intensity, and population-correlation configurations. Runtime is reported in seconds.

Regime	Method	Mean Estimate	Bias	RMSE	MAE	Mean Time
Clean	AWRcca	0.384	−0.266	0.315	0.270	0.133
	MRCD	0.894	0.244	0.283	0.244	1.619
	ccaMM	0.229	−0.421	0.457	0.422	24.372
Casewise	AWRcca	0.463	−0.187	0.263	0.216	0.088
	MRCD	0.888	0.238	0.278	0.239	1.655
	ccaMM	0.228	−0.422	0.456	0.423	49.611
Cellwise	AWRcca	0.313	−0.337	0.379	0.338	0.132
	MRCD	0.874	0.224	0.269	0.224	1.559
	ccaMM	0.182	−0.468	0.510	0.473	32.092
Mixed	AWRcca	0.412	−0.238	0.334	0.288	0.104
	MRCD	0.859	0.209	0.260	0.214	1.546
	ccaMM	0.189	−0.461	0.512	0.471	51.222

For each comparison in Table 2,

$$\Delta_{\text{RMSE}} = \text{RMSE}_{\text{first}} - \text{RMSE}_{\text{second}},$$

so a negative value favors the first method. The reported time ratio is

$$R_{\text{time}} = \frac{\bar{T}_{\text{first}}}{\bar{T}_{\text{second}}},$$

so a value below one indicates that the first method was faster. The confidence intervals were obtained from the 2.5th and 97.5th percentiles of the cluster-bootstrap distributions. Paired Wilcoxon signed-rank tests were applied to the configuration-level mean squared errors and log-transformed mean computation times. The resulting p -values were adjusted by Holm's method separately within each regime and outcome type.

Table 2. Paired comparisons of estimation error and computation time. The RMSE difference is defined as the first method minus the second method. The time ratio is the mean runtime of the first method divided by that of the second method. Values in parentheses are 95% cluster-bootstrap confidence intervals. The reported p -values are Holm-adjusted paired Wilcoxon signed-rank test results.

Regime	Comparison	Δ_{RMSE} (95% CI)	p_{adj}	Time Ratio (95% CI)
Clean	AWRcca–MRCD	0.032 (−0.020, 0.085)	0.118	0.082 (0.075, 0.092)
	AWRcca–ccaMM	−0.142 (−0.186, −0.100)	<0.001	0.006 (0.005, 0.006)
	MRCD–ccaMM	−0.174 (−0.243, −0.105)	0.004	0.066 (0.058, 0.075)
Casewise	AWRcca–MRCD	−0.016 (−0.057, 0.025)	0.419	0.053 (0.049, 0.058)
	AWRcca–ccaMM	−0.194 (−0.223, −0.166)	<0.001	0.002 (0.002, 0.002)
	MRCD–ccaMM	−0.178 (−0.225, −0.132)	<0.001	0.033 (0.030, 0.037)
Cellwise	AWRcca–MRCD	0.110 (0.067, 0.153)	<0.001	0.085 (0.079, 0.091)
	AWRcca–ccaMM	−0.131 (−0.159, −0.102)	<0.001	0.004 (0.004, 0.004)
	MRCD–ccaMM	−0.241 (−0.290, −0.193)	<0.001	0.049 (0.043, 0.054)
Mixed	AWRcca–MRCD	0.074 (0.042, 0.106)	<0.001	0.067 (0.064, 0.071)
	AWRcca–ccaMM	−0.178 (−0.198, −0.158)	<0.001	0.002 (0.002, 0.002)
	MRCD–ccaMM	−0.252 (−0.287, −0.217)	<0.001	0.030 (0.028, 0.033)

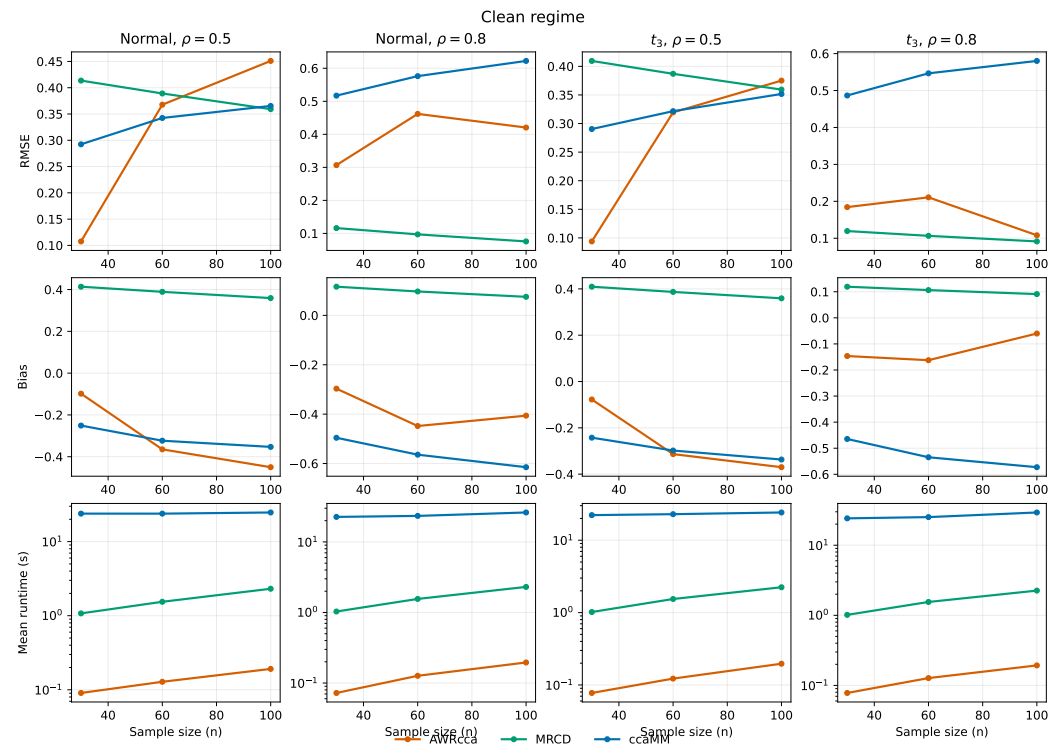


Figure 1. RMSE, bias, and mean runtime under clean data. Values are averaged over the considered values of p and q . Runtime is displayed on a logarithmic scale.

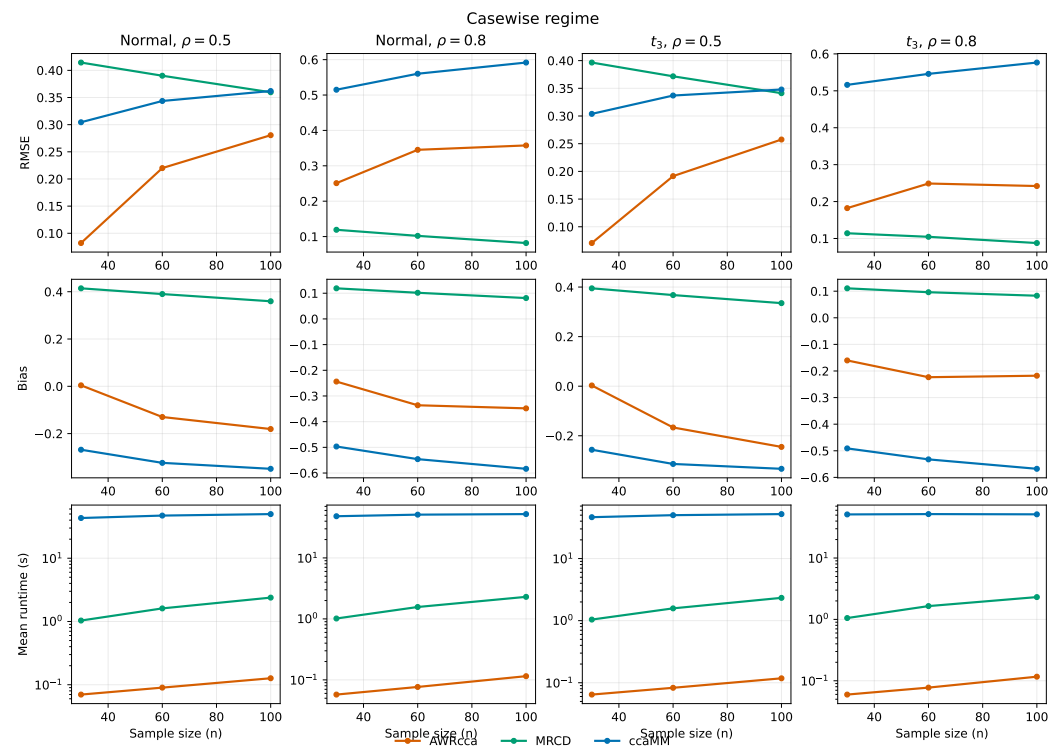


Figure 2. RMSE, bias, and mean runtime under casewise contamination. Values are averaged over p , q , and the nonzero casewise contamination proportions. Runtime is displayed on a logarithmic scale.

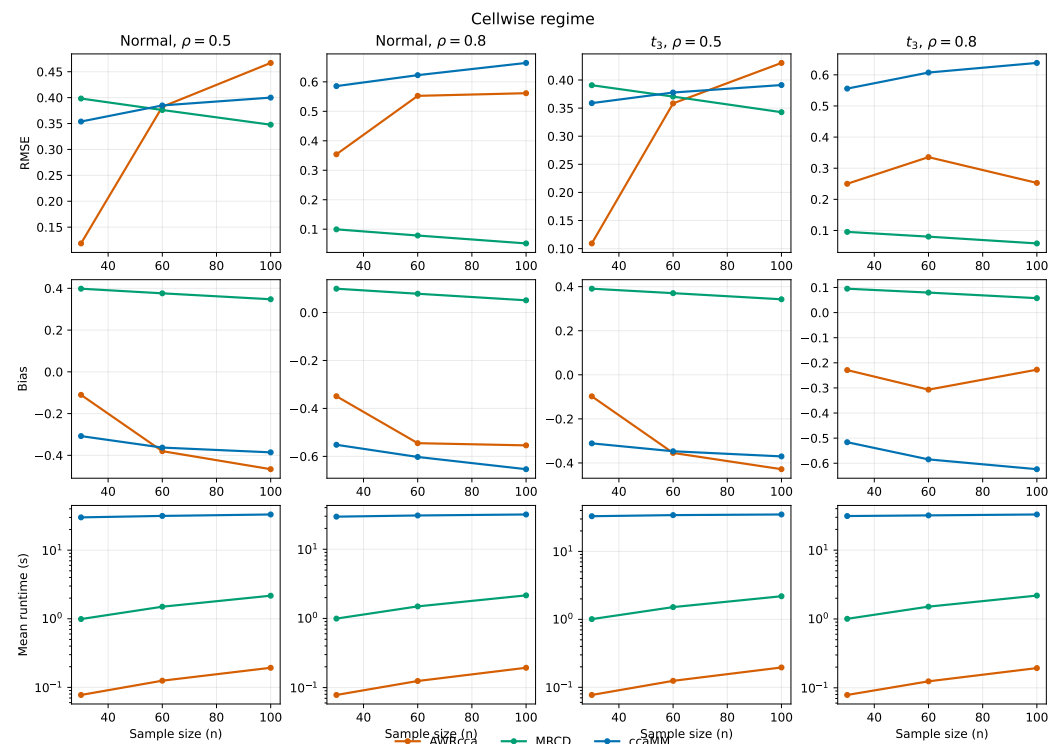


Figure 3. RMSE, bias, and mean runtime under cellwise contamination. Values are averaged over p , q , and the nonzero cellwise contamination proportions. Runtime is displayed on a logarithmic scale.

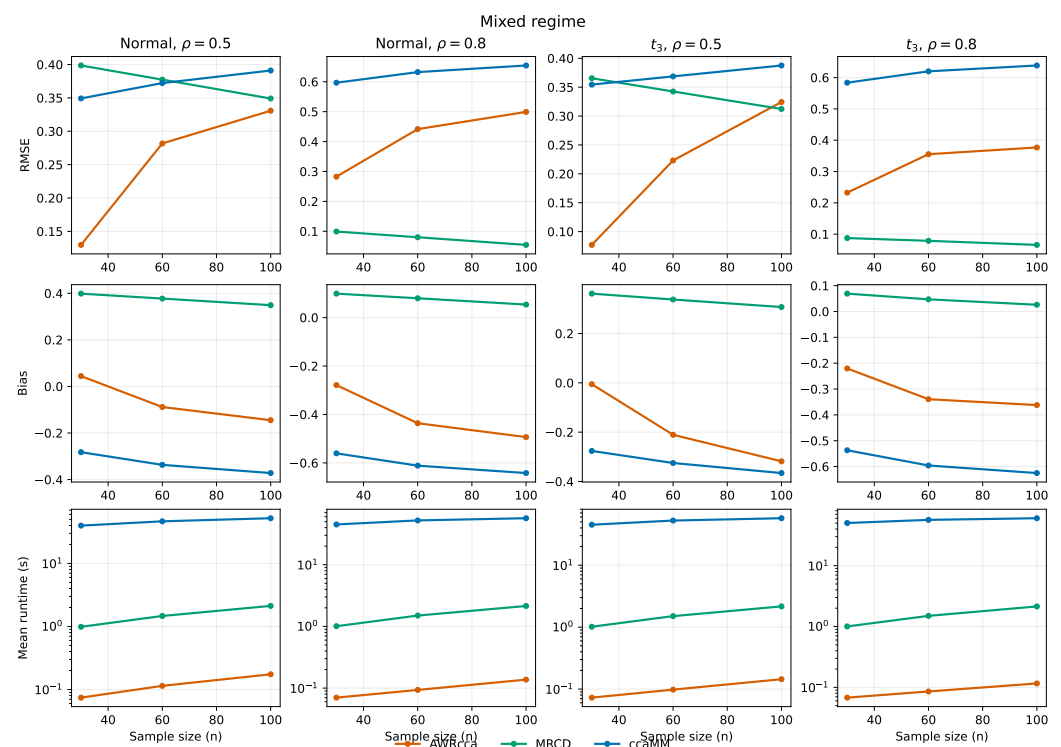


Figure 4. RMSE, bias, and mean runtime under simultaneous casewise and cellwise contamination. Values are averaged over p , q , and the nonzero contamination proportions. Runtime is displayed on a logarithmic scale.

The paired comparisons refine the aggregate rankings in Table 1. Under clean data and casewise contamination, the RMSE differences between AWRcca and MRCD were not statistically distinguishable after Holm adjustment, with adjusted p -values of 0.118

and 0.419, respectively, and bootstrap confidence intervals containing zero. Under cell-wise and mixed contamination, MRCD had significantly smaller aggregate RMSE than AWRcca. Both AWRcca and MRCD had significantly smaller RMSE than ccaMM in every contamination regime.

The computation-time differences were statistically clear in every pairwise comparison. All Holm-adjusted runtime p -values were below 0.001, and every bootstrap confidence interval for the time ratio was entirely below one for the method listed first. In particular, the AWRcca-to-MRCD time ratios ranged from 0.053 to 0.085, while the AWRcca-to-ccaMM ratios ranged from 0.002 to 0.006. These results support a computational advantage for AWRcca while confirming that its estimation accuracy advantage over MRCD is regime dependent rather than uniform.

These inferential comparisons summarize uncertainty across the finite grid of simulation configurations considered in this study. They should not be interpreted as population-level tests covering every possible dimensional, distributional, or contamination setting.

4.3. Clean Data

Under clean data, MRCD had the smallest pooled RMSE, followed by AWRcca and ccaMM. The aggregate values conceal a clear dependence on signal strength. When $\rho = 0.5$, AWRcca had RMSEs of 0.343 and 0.291 under Gaussian and t_3 noise, respectively, compared with 0.388 and 0.386 for MRCD. When $\rho = 0.8$, MRCD had the smallest RMSE, whereas AWRcca exhibited negative bias, particularly under Gaussian noise. Thus, MRCD's tendency to estimate values near 0.89 was disadvantageous at the moderate target correlation but advantageous when the target was 0.8.

The absence of uniform dominance is also visible in the configuration identified by the reviewer. For clean data with $p = 50$, $q = 60$, and $\rho = 0.8$, the AWRcca and MRCD RMSEs under Gaussian noise were, respectively, 0.327 and 0.107 for $n = 30$, 0.408 and 0.082 for $n = 60$, and 0.295 and 0.054 for $n = 100$. Thus, MRCD was more accurate in all three Gaussian configurations, and the AWRcca RMSE was not monotone in n . Under t_3 noise, however, the corresponding AWRcca RMSE decreased from 0.188 to 0.139 and 0.064, whereas the MRCD RMSEs were 0.113, 0.095, and 0.075. AWRcca was therefore slightly more accurate at $n = 100$ in the heavy-tailed configuration. These results show that changes in RMSE with sample size depend on the distribution and the interaction between wrapping, shrinkage, and signal strength; they do not support a general monotonicity or uniform-superiority claim.

AWRcca was considerably faster: its mean runtime was 0.133 s, compared with 1.619 s for MRCD and 24.372 s for ccaMM. Figure 1 shows the dependence of RMSE, bias, and runtime on sample size, distribution, and population correlation.

4.4. Casewise Contamination

Across the casewise contamination configurations, AWRcca had the smallest pooled RMSE (0.263), followed by MRCD (0.278) and ccaMM (0.456). As in the clean-data setting, relative performance depended on the target correlation. For $\rho = 0.5$, AWRcca produced smaller RMSE than both MRCD and ccaMM under Gaussian and heavy-tailed noise. For $\rho = 0.8$, MRCD produced the smallest RMSE, while AWRcca retained a negative bias.

The score-space reweighting step was used in every AWRcca fit, with an average of 41.8 retained observations across the heterogeneous sample sizes and contamination levels. Mean runtimes were 0.088 s for AWRcca, 1.655 s for MRCD, and 49.611 s for ccaMM. The detailed trends are shown in Figure 2.

4.5. Cellwise Contamination

Under cellwise contamination, MRCD had the smallest pooled RMSE (0.269), followed by AWRcca (0.379) and ccaMM (0.510). The ranking again changed with signal strength. At $\rho = 0.5$, AWRcca had smaller RMSE than MRCD under both distributions, although the difference under Gaussian noise was modest. At $\rho = 0.8$, MRCD was substantially closer to the target. AWRcca's negative bias increased under Gaussian cellwise contamination, indicating attenuation associated with wrapping and strong correlation shrinkage in these configurations.

The average AWRcca runtime was 0.132 s, compared with 1.559 s for MRCD and 32.092 s for ccaMM. Figure 3 provides the stratified results.

4.6. Mixed Contamination

When casewise and cellwise contamination occurred simultaneously, MRCD had the smallest pooled RMSE (0.260), followed by AWRcca (0.334) and ccaMM (0.512). At the moderate target correlation, AWRcca was more accurate than MRCD: its RMSEs were 0.289 under Gaussian noise and 0.252 under heavy-tailed noise, compared with 0.376 and 0.342 for MRCD. At the high target correlation, MRCD was more accurate. These results reinforce that no method dominated over both signal levels.

AWRcca required 0.104 s on average, whereas MRCD and ccaMM required 1.546 s and 51.222 s, respectively. Figure 4 summarizes the stratified mixed-contamination results.

4.7. Ultra-High-Dimensional Feasibility Experiment

To address the requested regime in which the number of features is much larger than the sample size, an additional experiment was conducted with $n = 100$, $p = 5000$, $q = 60$, and $\rho = 0.8$ under Gaussian noise. Clean data, 10% casewise contamination, and 10% cellwise contamination were considered, with 20 independent replications per scenario. Because the purpose of this experiment was to assess the scalability of the proposed estimator at the requested dimension, only AWRcca was run. The estimation accuracy, computational time, memory use, and completion rates for the three scenarios are summarized in Table 3.

Table 3. AWRcca results for the ultra-high-dimensional experiment with $n = 100$, $p = 5000$, $q = 60$, and target correlation $\rho = 0.8$. Runtime is reported in minutes and memory in GiB.

Scenario	Success	Mean (SD)	Bias	RMSE	Time	Peak Memory
Clean	20/20	0.253 (0.023)	−0.547	0.548	30.12	1.64
Casewise 10%	20/20	0.291 (0.028)	−0.509	0.510	30.11	1.74
Cellwise 10%	20/20	0.245 (0.016)	−0.555	0.556	30.08	1.74

As reported in Table 3, all 60 fits completed without numerical failure. The mean runtime was approximately 30 min per replication, and the maximum memory use remained below 1.75 GiB. Figure 5 shows the mean estimated first canonical correlation, its variability across replications, and the corresponding mean runtime for each scenario. The estimates remained well below the population target of $\rho = 0.8$ in all three scenarios.

The experiment therefore establishes computational feasibility at the requested dimension. It does not, however, demonstrate statistical accuracy, because the estimates were strongly attenuated relative to the target of 0.8. The mean final shrinkage intensities ranged from 0.973 to 0.977, providing a plausible explanation for this attenuation. Accordingly, the ultra-high-dimensional experiment is interpreted as a scalability result and as evidence of the limitations of the present dense estimator when the dimension is extremely large relative to the sample size.

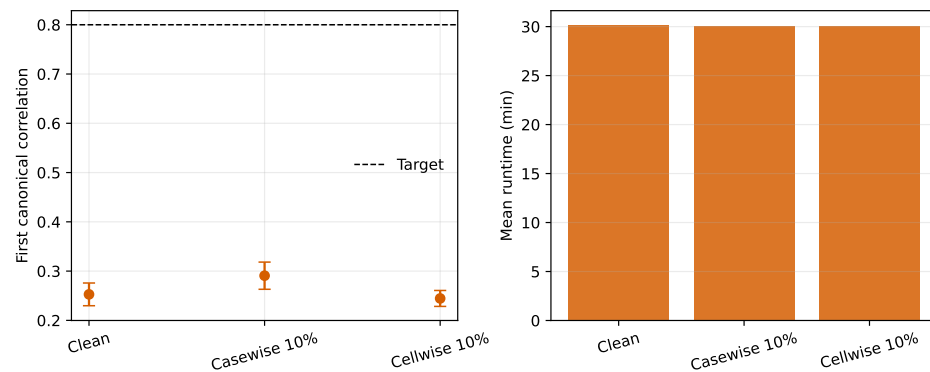


Figure 5. Mean first canonical correlation (with one standard deviation error bars) and mean runtime in the ultra-high-dimensional experiment. The dashed line denotes the population target $\rho = 0.8$.

4.8. Summary of the Simulation Evidence

The revised simulation does not support a claim of uniform statistical superiority. MRCD had the smallest RMSE in 260 of the 432 configurations, AWRcca in 149, and ccaMM in 23. AWRcca was most competitive at the moderate population correlation, particularly under casewise and mixed contamination, whereas MRCD was favored at the higher population correlation because its estimates were concentrated near the upper end of the considered target range. Across all four regimes, AWRcca was approximately 12–19 times faster than MRCD and 183–565 times faster than ccaMM. Thus, the main empirical advantage of AWRcca is computational efficiency combined with competitive accuracy in specific moderate-signal regimes, rather than universal dominance. The pronounced attenuation in the $p = 5000$ experiment identifies dense ultra-high-dimensional estimation as an important limitation and motivates future work on sparse or less aggressively shrunk extensions.

5. Real Data Example

The practical performance of the proposed Adaptive Wrapped Robust Canonical Correlation Analysis (AWRcca) is investigated using the Nutrimouse data set, which is widely regarded as a benchmark example for high-dimensional canonical correlation analysis. This data set has been widely used in previous studies on robust and high-dimensional CCA [7,11].

The Nutrimouse data consist of $n = 40$ mice subjected to different dietary regimes. Two multivariate blocks are observed: a gene expression matrix $\mathbf{X} \in \mathbb{R}^{40 \times 120}$ containing hepatic expression levels of 120 genes, and a lipid concentration matrix $\mathbf{Y} \in \mathbb{R}^{40 \times 21}$ representing 21 fatty acids. Consequently, the total number of variables ($p + q = 141$) greatly exceeds the sample size, rendering classical CCA infeasible and making the data particularly suitable for evaluating regularized and robust CCA methods. Prior to analysis, a mild log-transformation is applied to the gene expression data when appropriate, followed by standardization based on training samples only. This transformation is commonly used in gene expression analysis to reduce skewness and stabilize variances, while preserving relative expression patterns.

Before performing predictive comparisons, the original, unmodified data are examined in order to assess their intrinsic outlier structure. Figure 6 presents an outlier map based on robust Mahalanobis distances computed from a Minimum Regularized Covariance Determinant (MRCD) estimator applied to the concatenated data matrix $[\mathbf{X}, \mathbf{Y}]$.

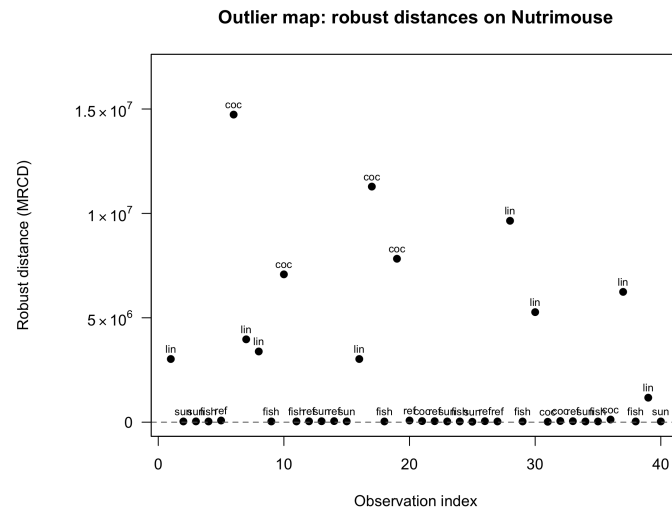


Figure 6. Outlier map based on robust Mahalanobis distances computed using the MRCD estimator on the Nutrimouse data. Observations exceeding the reference threshold indicate potential rowwise outliers.

Several observations exhibit substantially larger robust distances than the bulk of the data, indicating the presence of potential rowwise outliers. These deviations may reflect biological variability or experimental effects associated with specific dietary regimes.

While rowwise outlier detection captures global deviations at the observation level, it does not reveal whether contamination is localized within individual cells. To investigate this aspect, Figure 7 displays a heatmap of robust standardized scores for the gene expression block.

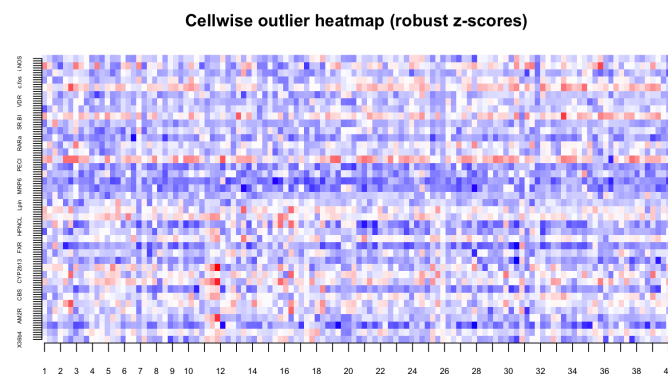


Figure 7. Cellwise outlier heatmap based on robust standardized scores for the gene expression block of the Nutrimouse data. Red and blue regions indicate extreme positive and negative deviations, respectively, highlighting localized cellwise contamination.

The heatmap reveals scattered extreme values across different genes and mice, suggesting the presence of cellwise contamination rather than purely casewise aberrations. Taken together, the robust distance plot and the cellwise heatmap indicate that the Nutrimouse data exhibit a mixed contamination structure, involving both rowwise heterogeneity and localized cellwise anomalies. This observation provides strong motivation for the use of wrapping-based robust procedures, such as AWRcca, which are specifically designed to mitigate the impact of cellwise contamination.

Since the true canonical correlations are unknown for real data, method comparison is conducted within a predictive framework. The data are repeatedly split into training (70%)

and test (30%) sets. Canonical vectors ($\hat{\mathbf{a}}, \hat{\mathbf{b}}$) are estimated on the training data and then applied to the test data. Performance is evaluated using the test-set canonical correlation

$$\text{MCC} = \text{cor}(\mathbf{X}_{\text{test}}\hat{\mathbf{a}}, \mathbf{Y}_{\text{test}}\hat{\mathbf{b}}), \quad (13)$$

which quantifies the strength of association captured on unseen observations. The MCC criterion is particularly suitable in this context, as it directly evaluates the generalizability of the estimated canonical relationship to unseen data.

To further assess robustness, semi-synthetic contamination is introduced in the test data. Two contamination schemes are considered: cellwise contamination, where 10% of randomly selected cells are perturbed, and rowwise contamination, where 10% of randomly selected observations are shifted. Contamination is applied exclusively to the test set, allowing direct evaluation of out-of-sample robustness. The proposed AWRcca is compared with an MRCD-based plug-in CCA approach and the ccaMM method based on MRCD estimation. The predictive performance and computational cost of the three methods under the clean, cellwise-contaminated, and rowwise-contaminated test conditions are summarized in Table 4.

Table 4. Nutrimouse predictive performance over $R = 500$ random splits (70% train/30% test). MCC is computed as in (13). Reported are mean (SD) and median [Q1, Q3] for MCC, and mean runtime in seconds.

Scenario	Method	Mean MCC (SD)	Median [Q1, Q3]	Mean Time (s)
clean	AWRcca	0.7518 (0.1976)	0.8008 [0.7366, 0.8645]	0.020
	MRCD plug-in	0.2907 (0.2173)	0.2224 [0.0897, 0.4502]	0.367
	ccaMM (MRCD)	0.3077 (0.1890)	0.2833 [0.1442, 0.4574]	30.256
cell110	AWRcca	0.5654 (0.1457)	0.5745 [0.4453, 0.6864]	0.018
	MRCD plug-in	0.3232 (0.1774)	0.2795 [0.1954, 0.4773]	0.352
	ccaMM (MRCD)	0.1957 (0.1674)	0.1468 [0.0689, 0.3065]	30.224
row10	AWRcca	0.5984 (0.2481)	0.6718 [0.4887, 0.7993]	0.019
	MRCD plug-in	0.2724 (0.2331)	0.1814 [0.1254, 0.3543]	0.360
	ccaMM (MRCD)	0.2123 (0.2035)	0.1525 [0.0714, 0.2728]	30.254

Within this Nutrimouse resampling design, AWRcca yielded the largest mean and median test-set MCC in the clean, cellwise-contaminated, and rowwise-contaminated scenarios. The differences were especially pronounced for the clean data and for the cellwise-contamination experiment. AWRcca also required approximately 0.02 s per split, compared with approximately 0.35–0.37 s for MRCD plug-in CCA and 30.2 s for ccaMM. These application-specific findings indicate that the integrated wrapping and reweighting strategy can provide favorable predictive performance at low computational cost for this data set. They complement the simulation results, in which the relative estimation accuracy depended on the signal strength and contamination regime, and should not be interpreted as evidence of universal superiority.

6. Software Implementation and Reproducibility

An R interface for Adaptive Wrapped Robust Canonical Correlation Analysis is available through the open-source MVTests package [16]. The development repository and source code are publicly available at <https://github.com/hsnbulut/MVTests> (accessed on 1 June 2026). The complete estimator and computational workflow used for the revised numerical study in Section 4 are provided with the supplementary reproducibility materials, together with the scenario definitions, random number seeds, and aggregation scripts.

The principal user interface is the AWRcca() function. It accepts two numeric matrices with the same number of observations and allows the wrapping thresholds, robust distance

cutoff probability, and upper bound on shrinkage intensity to be specified. The reference values used in this study are $b = 1.5$, $c = 4$, $\alpha = 0.975$, and $\lambda_{\max} = 0.5$. For these wrapping constants, the Gaussian consistency factor is evaluated deterministically from the one-dimensional integral given in Section 3, yielding $\xi = 1.152204624$.

The fitted object is a list containing five components: `cor`, the estimated canonical correlations; `shrink_used`, the shrinkage intensity used by the fitted procedure; `weights`, the binary observation weights obtained from robust score-space distances; and `u1` and `v1`, the first pair of canonical scores from the initial regularized solution. The number of retained observations can be obtained by summing the binary weights.

A minimal reproducible example is presented in Listing 1. The two simulated data blocks share a common latent component, and their combined dimension exceeds the sample size. The example reports only quantities returned by the publicly available implementation; no numerical result is hard-coded.

Listing 1. Minimal reproducible example of the AWRcca implementation for two simulated high-dimensional data blocks sharing a common latent component.

```

1 library(MVTests)
2
3 set.seed(123)
4 n <- 60; p <- 40; q <- 50
5
6 z <- rnorm(n)
7 a <- seq(0.80, 0.30, length.out = p)
8 b <- seq(0.75, 0.25, length.out = q)
9
10 X <- outer(z, a) + matrix(rnorm(n * p), n, p)
11 Y <- outer(z, b) + matrix(rnorm(n * q), n, q)
12
13 stopifnot(nrow(X) == nrow(Y), p + q > n)
14
15 fit <- AWRcca(X, Y)
16
17 c(
18   first_correlation = fit$cor[1],
19   shrinkage = fit$shrink_used,
20   retained = sum(fit$weights),
21   score_correlation = cor(fit$u1, fit$v1)
22 )

```

The binary weights can be used to identify observations flagged by the score-space reweighting step, whereas `u1` and `v1` can be used to construct diagnostic score plots. The complete simulation workflow, including scenario definitions, random number seeds, replication-level outputs, aggregation scripts, tables, figures, and session information, is provided separately as supplementary reproducibility material. This separates the concise user-level example from the full computational workflow used in the numerical study.

7. Conclusions

This study developed and evaluated an integrated robust and regularized canonical correlation procedure for contaminated high-dimensional data. The individual ingredients are established statistical tools; the contribution lies in coordinating bounded cellwise wrapping, shrinkage correlation estimation, and score-space robust reweighting within a single estimator. In particular, the score-space observation weights enter a second regular-

ized canonical correlation fit, thereby preserving numerical feasibility when the combined dimension exceeds the sample size.

The simulation study comprised 432 configurations, 43,200 generated data sets, and 129,600 method fits. It did not support a claim of uniform statistical superiority. MRCD plug-in CCA attained the smallest RMSE in 260 configurations, AWRcca in 149, and ccaMM in 23. AWRcca was most competitive at the moderate target correlation, particularly under casewise and mixed contamination, whereas MRCD was generally favored at the higher target correlation. Across the four principal regimes, AWRcca was approximately 12–19 times faster than MRCD and 183–565 times faster than ccaMM. Thus, its main empirical advantage is computational efficiency combined with competitive accuracy in specific moderate-signal regimes.

The additional experiment with $n = 100$, $p = 5000$, and $q = 60$ established computational feasibility: all 60 fits completed, mean runtime was approximately 30 min per replication, and peak memory use remained below 1.75 GiB. However, the estimated first canonical correlations were strongly attenuated relative to the target value of 0.8. This result limits the scope of the statistical claims and indicates that the present dense estimator may require sparse regularization or less aggressive shrinkage when the dimension is extremely large relative to the sample size.

In the Nutrimouse application, AWRcca produced the largest mean and median test-set canonical correlations in each evaluated resampling scenario while requiring substantially less computation than ccaMM. These findings demonstrate practical utility for this particular high-dimensional data set, but they are interpreted together with the regime-dependent simulation evidence rather than as proof of universal dominance.

Future research should investigate statistical properties of the integrated estimator and extend it to sparse canonical directions, multi-block and multi-view data, and functional or longitudinal settings. Particular attention should be given to reducing shrinkage-induced attenuation in ultra-high-dimensional regimes and to evaluating sensitivity to the wrapping constants and robust distance cutoff.

Author Contributions: Conceptualization, H.B.; methodology, H.B.; software, H.B.; validation, H.B., M.Z. and V.S.; formal analysis, H.B.; investigation, H.B., M.Z. and V.S.; writing—original draft preparation, H.B.; writing—review and editing, H.B., M.Z. and V.S.; visualization, H.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The Nutrimouse data set analyzed in this study is available through the public R package CCA. The implementation of the proposed AWRcca method is available in the MVTests R package, as described in the Software Availability section.

Acknowledgments: During the preparation of this study, the authors used OpenAI ChatGPT (version 5.3) for the purposes of proofreading. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Simulation Tables

This appendix reports the detailed results of the revised simulation study described in Section 4. All entries were computed from 100 independent replications per configuration. The tables report root mean squared error and mean computation time for AWRcca, MRCD

plug-in CCA, and ccaMM under clean, casewise-contaminated, cellwise-contaminated, and mixed-contamination regimes.

Table A1. Simulation results for clean data ($\epsilon_{\text{row}} = \epsilon_{\text{cell}} = 0$). Average RMSE and mean computation time (in seconds).

<i>n</i>	<i>p</i>	<i>q</i>	ρ	dist	RMSE			Time (s)		
					AWRcca	MRCD	ccaMM	AWRcca	MRCD	ccaMM
30	50	60	0.5	normal	0.177	0.403	0.291	0.103	0.620	20.628
30	50	60	0.5	t_3	0.144	0.400	0.290	0.057	0.460	19.805
30	50	60	0.8	normal	0.327	0.107	0.519	0.057	0.452	19.694
30	50	60	0.8	t_3	0.188	0.113	0.488	0.056	0.451	18.550
30	50	120	0.5	normal	0.103	0.413	0.265	0.071	1.030	19.376
30	50	120	0.5	t_3	0.094	0.408	0.274	0.071	1.014	19.758
30	50	120	0.8	normal	0.310	0.116	0.514	0.070	1.014	17.425
30	50	120	0.8	t_3	0.187	0.119	0.469	0.073	1.008	19.614
30	100	60	0.5	normal	0.105	0.414	0.297	0.095	0.896	19.768
30	100	60	0.5	t_3	0.098	0.410	0.279	0.082	0.891	19.603
30	100	60	0.8	normal	0.315	0.116	0.496	0.071	0.897	19.639
30	100	60	0.8	t_3	0.191	0.120	0.492	0.081	0.899	21.629
30	100	120	0.5	normal	0.044	0.424	0.317	0.093	1.752	35.648
30	100	120	0.5	t_3	0.040	0.419	0.319	0.101	1.710	29.693
30	100	120	0.8	normal	0.275	0.125	0.540	0.092	1.764	33.678
30	100	120	0.8	t_3	0.172	0.127	0.497	0.102	1.704	36.909
60	50	60	0.5	normal	0.389	0.371	0.336	0.106	0.666	22.906
60	50	60	0.5	t_3	0.321	0.370	0.317	0.104	0.680	22.145
60	50	60	0.8	normal	0.408	0.082	0.574	0.107	0.702	21.287
60	50	60	0.8	t_3	0.139	0.095	0.525	0.104	0.684	20.253
60	50	120	0.5	normal	0.372	0.390	0.325	0.118	1.522	23.835
60	50	120	0.5	t_3	0.326	0.388	0.313	0.118	1.520	22.491
60	50	120	0.8	normal	0.475	0.099	0.574	0.120	1.535	21.121
60	50	120	0.8	t_3	0.225	0.108	0.552	0.121	1.548	26.123
60	100	60	0.5	normal	0.373	0.390	0.347	0.123	1.361	21.285
60	100	60	0.5	t_3	0.323	0.388	0.316	0.119	1.366	21.671
60	100	60	0.8	normal	0.463	0.098	0.563	0.123	1.377	21.363
60	100	60	0.8	t_3	0.204	0.107	0.547	0.126	1.356	22.545
60	100	120	0.5	normal	0.338	0.405	0.362	0.167	2.613	27.327
60	100	120	0.5	t_3	0.308	0.402	0.342	0.148	2.598	25.284
60	100	120	0.8	normal	0.502	0.109	0.594	0.157	2.608	29.692
60	100	120	0.8	t_3	0.276	0.116	0.562	0.158	2.614	31.818
100	50	60	0.5	normal	0.446	0.329	0.366	0.175	1.062	24.069
100	50	60	0.5	t_3	0.329	0.333	0.334	0.182	0.980	23.524
100	50	60	0.8	normal	0.295	0.054	0.595	0.180	0.984	22.538
100	50	60	0.8	t_3	0.064	0.075	0.543	0.179	0.981	20.992
100	50	120	0.5	normal	0.456	0.363	0.358	0.185	2.270	24.043
100	50	120	0.5	t_3	0.387	0.363	0.353	0.186	2.225	23.599
100	50	120	0.8	normal	0.449	0.078	0.628	0.188	2.280	26.561
100	50	120	0.8	t_3	0.126	0.093	0.597	0.187	2.240	31.748
100	100	60	0.5	normal	0.454	0.361	0.360	0.189	2.056	23.774
100	100	60	0.5	t_3	0.380	0.361	0.352	0.191	1.992	24.095
100	100	60	0.8	normal	0.422	0.078	0.631	0.190	2.049	26.050
100	100	60	0.8	t_3	0.090	0.093	0.580	0.190	1.995	27.977
100	100	120	0.5	normal	0.448	0.384	0.377	0.218	3.852	26.980
100	100	120	0.5	t_3	0.404	0.381	0.368	0.228	3.760	25.461
100	100	120	0.8	normal	0.517	0.093	0.635	0.224	3.897	29.344
100	100	120	0.8	t_3	0.151	0.105	0.601	0.215	3.792	36.541

Table A2. Simulation results under casewise contamination ($\epsilon_{\text{cell}} = 0$). Average RMSE and mean computation time (in seconds) for varying $(n, p, q, \rho, \text{dist}, \epsilon_{\text{row}})$.

n	p	q	ρ	ϵ_{row}	dist	RMSE			Time (s)		
						AWRcca	MRCd	ccaMM	AWRcca	MRCd	ccaMM
30	50	60	0.5	0.10	normal	0.142	0.403	0.288	0.047	0.463	42.867
30	50	60	0.5	0.10	t_3	0.117	0.398	0.273	0.047	0.457	44.535
30	50	60	0.5	0.25	normal	0.056	0.405	0.284	0.045	0.484	45.942
30	50	60	0.5	0.25	t_3	0.059	0.378	0.287	0.049	0.488	49.713
30	50	60	0.8	0.10	normal	0.328	0.109	0.487	0.047	0.446	47.816
30	50	60	0.8	0.10	t_3	0.202	0.113	0.514	0.046	0.616	48.542
30	50	60	0.8	0.25	normal	0.275	0.113	0.495	0.033	0.473	52.337
30	50	60	0.8	0.25	t_3	0.204	0.105	0.479	0.031	0.447	56.772
30	50	120	0.5	0.10	normal	0.066	0.413	0.281	0.069	1.017	43.643
30	50	120	0.5	0.10	t_3	0.061	0.407	0.313	0.064	0.999	44.839
30	50	120	0.5	0.25	normal	0.074	0.414	0.288	0.070	1.067	46.666
30	50	120	0.5	0.25	t_3	0.056	0.383	0.319	0.064	1.062	52.693
30	50	120	0.8	0.10	normal	0.288	0.117	0.517	0.061	0.991	48.436
30	50	120	0.8	0.10	t_3	0.184	0.120	0.499	0.066	1.010	49.153
30	50	120	0.8	0.25	normal	0.215	0.121	0.516	0.047	1.027	52.173
30	50	120	0.8	0.25	t_3	0.187	0.108	0.509	0.050	1.017	57.177
30	100	60	0.5	0.10	normal	0.067	0.415	0.323	0.082	0.898	41.537
30	100	60	0.5	0.10	t_3	0.055	0.409	0.309	0.074	0.874	42.819
30	100	60	0.5	0.25	normal	0.074	0.416	0.330	0.068	0.956	44.201
30	100	60	0.5	0.25	t_3	0.057	0.383	0.282	0.064	0.910	48.105
30	100	60	0.8	0.10	normal	0.289	0.118	0.532	0.063	0.872	46.294
30	100	60	0.8	0.10	t_3	0.187	0.120	0.514	0.069	1.007	47.097
30	100	60	0.8	0.25	normal	0.220	0.121	0.495	0.049	0.908	50.365
30	100	60	0.8	0.25	t_3	0.187	0.110	0.517	0.060	0.884	55.506
30	100	120	0.5	0.10	normal	0.036	0.425	0.324	0.089	1.693	39.895
30	100	120	0.5	0.10	t_3	0.041	0.419	0.336	0.083	1.790	41.280
30	100	120	0.5	0.25	normal	0.141	0.425	0.317	0.089	1.700	41.018
30	100	120	0.5	0.25	t_3	0.119	0.395	0.311	0.071	1.744	46.950
30	100	120	0.8	0.10	normal	0.242	0.126	0.559	0.091	1.699	43.032
30	100	120	0.8	0.10	t_3	0.167	0.126	0.551	0.088	1.719	44.307
30	100	120	0.8	0.25	normal	0.150	0.129	0.519	0.069	1.705	45.617
30	100	120	0.8	0.25	t_3	0.141	0.112	0.546	0.066	1.715	52.249
60	50	60	0.5	0.10	normal	0.369	0.371	0.342	0.076	0.718	45.598
60	50	60	0.5	0.10	t_3	0.332	0.369	0.338	0.077	0.669	47.071
60	50	60	0.5	0.25	normal	0.075	0.373	0.340	0.057	0.713	50.841
60	50	60	0.5	0.25	t_3	0.124	0.340	0.325	0.051	0.731	54.221
60	50	60	0.8	0.10	normal	0.440	0.084	0.556	0.078	0.677	50.815
60	50	60	0.8	0.10	t_3	0.178	0.097	0.517	0.079	1.459	47.210
60	50	60	0.8	0.25	normal	0.280	0.093	0.545	0.039	0.700	54.983
60	50	60	0.8	0.25	t_3	0.280	0.087	0.488	0.038	0.684	57.687
60	50	120	0.5	0.10	normal	0.337	0.391	0.345	0.093	1.547	47.238
60	50	120	0.5	0.10	t_3	0.314	0.388	0.336	0.093	1.517	48.295
60	50	120	0.5	0.25	normal	0.086	0.391	0.336	0.083	1.659	52.320
60	50	120	0.5	0.25	t_3	0.058	0.360	0.340	0.054	1.557	56.671
60	50	120	0.8	0.10	normal	0.462	0.101	0.579	0.094	1.532	50.689
60	50	120	0.8	0.10	t_3	0.253	0.109	0.556	0.093	1.508	48.956
60	50	120	0.8	0.25	normal	0.221	0.107	0.552	0.052	1.605	55.600
60	50	120	0.8	0.25	t_3	0.240	0.116	0.552	0.053	1.525	58.105
60	100	60	0.5	0.10	normal	0.338	0.390	0.338	0.087	1.382	44.568
60	100	60	0.5	0.10	t_3	0.306	0.388	0.320	0.090	1.339	45.377
60	100	60	0.5	0.25	normal	0.094	0.392	0.345	0.078	1.434	49.142
60	100	60	0.5	0.25	t_3	0.067	0.359	0.338	0.081	1.419	53.741
60	100	60	0.8	0.10	normal	0.485	0.100	0.575	0.096	1.352	48.732
60	100	60	0.8	0.10	t_3	0.267	0.107	0.552	0.093	1.512	47.851
60	100	60	0.8	0.25	normal	0.221	0.106	0.549	0.055	1.402	53.265

Table A2. Cont.

n	p	q	ρ	ε_{row}	dist	RMSE			Time (s)		
						AWRcca	MRCd	ccaMM	AWRcca	MRCd	ccaMM
60	100	60	0.8	0.25	t_3	0.242	0.102	0.544	0.050	1.355	55.988
60	100	120	0.5	0.10	normal	0.297	0.406	0.357	0.130	2.620	43.058
60	100	120	0.5	0.10	t_3	0.274	0.402	0.354	0.132	2.736	43.375
60	100	120	0.5	0.25	normal	0.165	0.406	0.346	0.115	2.806	45.107
60	100	120	0.5	0.25	t_3	0.057	0.368	0.345	0.085	2.650	50.998
60	100	120	0.8	0.10	normal	0.493	0.111	0.566	0.123	2.542	46.209
60	100	120	0.8	0.10	t_3	0.308	0.116	0.581	0.126	2.558	45.695
60	100	120	0.8	0.25	normal	0.159	0.115	0.559	0.076	2.726	49.485
60	100	120	0.8	0.25	t_3	0.223	0.103	0.578	0.086	2.577	54.462
100	50	60	0.5	0.10	normal	0.458	0.330	0.359	0.127	1.054	47.558
100	50	60	0.5	0.10	t_3	0.376	0.330	0.345	0.134	0.987	47.637
100	50	60	0.5	0.25	normal	0.071	0.332	0.354	0.076	1.031	53.686
100	50	60	0.5	0.25	t_3	0.175	0.297	0.328	0.069	1.070	56.063
100	50	60	0.8	0.10	normal	0.401	0.055	0.588	0.146	0.998	50.990
100	50	60	0.8	0.10	t_3	0.108	0.076	0.540	0.139	1.631	46.075
100	50	60	0.8	0.25	normal	0.309	0.066	0.562	0.059	1.036	54.467
100	50	60	0.8	0.25	t_3	0.305	0.071	0.546	0.062	0.994	55.409
100	50	120	0.5	0.10	normal	0.448	0.362	0.368	0.142	2.411	49.030
100	50	120	0.5	0.10	t_3	0.403	0.362	0.330	0.141	2.205	49.343
100	50	120	0.5	0.25	normal	0.103	0.364	0.361	0.100	2.359	54.862
100	50	120	0.5	0.25	t_3	0.110	0.321	0.351	0.071	2.333	59.696
100	50	120	0.8	0.10	normal	0.495	0.083	0.587	0.141	2.251	51.882
100	50	120	0.8	0.10	t_3	0.185	0.095	0.577	0.144	2.217	48.079
100	50	120	0.8	0.25	normal	0.233	0.089	0.581	0.066	2.350	55.565
100	50	120	0.8	0.25	t_3	0.296	0.085	0.585	0.069	2.252	57.694
100	100	60	0.5	0.10	normal	0.444	0.360	0.364	0.137	2.108	47.203
100	100	60	0.5	0.10	t_3	0.396	0.360	0.339	0.143	1.960	47.363
100	100	60	0.5	0.25	normal	0.104	0.362	0.362	0.094	2.085	52.413
100	100	60	0.5	0.25	t_3	0.127	0.334	0.352	0.089	2.097	56.674
100	100	60	0.8	0.10	normal	0.469	0.080	0.595	0.140	1.986	50.518
100	100	60	0.8	0.10	t_3	0.198	0.092	0.587	0.142	1.955	46.415
100	100	60	0.8	0.25	normal	0.234	0.087	0.570	0.064	2.094	55.037
100	100	60	0.8	0.25	t_3	0.292	0.082	0.576	0.066	1.997	56.352
100	100	120	0.5	0.10	normal	0.438	0.383	0.374	0.181	4.024	45.225
100	100	120	0.5	0.10	t_3	0.408	0.380	0.369	0.187	4.121	46.153
100	100	120	0.5	0.25	normal	0.179	0.384	0.354	0.156	3.998	48.917
100	100	120	0.5	0.25	t_3	0.065	0.345	0.368	0.108	3.857	53.106
100	100	120	0.8	0.10	normal	0.553	0.094	0.631	0.191	3.743	48.171
100	100	120	0.8	0.10	t_3	0.285	0.103	0.589	0.197	3.668	47.166
100	100	120	0.8	0.25	normal	0.167	0.100	0.620	0.111	3.930	51.873
100	100	120	0.8	0.25	t_3	0.268	0.099	0.613	0.114	3.761	55.149

Table A3. Simulation results under cellwise contamination ($\varepsilon_{\text{row}} = 0$). Average RMSE and mean computation time (in seconds) for varying ($n, p, q, \rho, \text{dist}, \varepsilon_{\text{cell}}$).

n	p	q	ρ	$\varepsilon_{\text{cell}}$	dist	RMSE			Time (s)		
						AWRcca	MRCd	ccaMM	AWRcca	MRCd	ccaMM
30	50	60	0.5	0.10	normal	0.189	0.378	0.334	0.054	0.442	25.402
30	50	60	0.5	0.10	t_3	0.156	0.373	0.345	0.056	0.452	29.404
30	50	60	0.5	0.25	normal	0.195	0.391	0.349	0.058	0.450	33.095
30	50	60	0.5	0.25	t_3	0.188	0.382	0.346	0.055	0.442	33.563
30	50	60	0.8	0.10	normal	0.367	0.080	0.603	0.053	0.428	25.486
30	50	60	0.8	0.10	t_3	0.202	0.081	0.529	0.056	0.455	26.925
30	50	60	0.8	0.25	normal	0.412	0.093	0.582	0.056	0.449	31.945
30	50	60	0.8	0.25	t_3	0.288	0.090	0.546	0.055	0.446	32.364
30	50	120	0.5	0.10	normal	0.111	0.393	0.361	0.071	0.966	27.132

Table A3. Cont.

<i>n</i>	<i>p</i>	<i>q</i>	ρ	$\varepsilon_{\text{cell}}$	dist	RMSE			Time (s)		
						AWRcca	MRCD	ccaMM	AWRcca	MRCD	ccaMM
30	50	120	0.5	0.10	t_3	0.103	0.385	0.370	0.070	1.017	31.290
30	50	120	0.5	0.25	normal	0.118	0.402	0.350	0.074	0.981	35.377
30	50	120	0.5	0.25	t_3	0.114	0.393	0.351	0.072	0.995	36.437
30	50	120	0.8	0.10	normal	0.352	0.094	0.590	0.075	1.001	26.103
30	50	120	0.8	0.10	t_3	0.227	0.090	0.541	0.074	1.004	29.816
30	50	120	0.8	0.25	normal	0.380	0.104	0.545	0.071	0.990	34.337
30	50	120	0.8	0.25	t_3	0.291	0.098	0.576	0.073	0.993	34.623
30	100	60	0.5	0.10	normal	0.113	0.394	0.358	0.085	0.857	25.849
30	100	60	0.5	0.10	t_3	0.107	0.388	0.366	0.092	0.889	30.880
30	100	60	0.5	0.25	normal	0.126	0.405	0.347	0.089	0.866	33.586
30	100	60	0.5	0.25	t_3	0.119	0.396	0.340	0.085	0.864	34.260
30	100	60	0.8	0.10	normal	0.337	0.095	0.585	0.089	0.859	25.503
30	100	60	0.8	0.10	t_3	0.242	0.091	0.536	0.095	0.887	29.027
30	100	60	0.8	0.25	normal	0.379	0.105	0.585	0.089	0.864	32.969
30	100	60	0.8	0.25	t_3	0.293	0.101	0.555	0.085	0.867	32.922
30	100	120	0.5	0.10	normal	0.046	0.408	0.374	0.094	1.669	26.845
30	100	120	0.5	0.10	t_3	0.041	0.401	0.377	0.095	1.720	32.267
30	100	120	0.5	0.25	normal	0.051	0.416	0.358	0.096	1.691	33.154
30	100	120	0.5	0.25	t_3	0.047	0.408	0.377	0.093	1.683	35.335
30	100	120	0.8	0.10	normal	0.289	0.108	0.603	0.092	1.684	26.863
30	100	120	0.8	0.10	t_3	0.201	0.103	0.588	0.103	1.711	30.338
30	100	120	0.8	0.25	normal	0.317	0.116	0.592	0.101	1.677	33.485
30	100	120	0.8	0.25	t_3	0.255	0.110	0.575	0.090	1.679	34.123
60	50	60	0.5	0.10	normal	0.406	0.349	0.398	0.106	0.668	27.256
60	50	60	0.5	0.10	t_3	0.352	0.348	0.375	0.105	0.668	31.946
60	50	60	0.5	0.25	normal	0.421	0.361	0.385	0.105	0.677	33.644
60	50	60	0.5	0.25	t_3	0.401	0.355	0.356	0.109	0.671	34.789
60	50	60	0.8	0.10	normal	0.480	0.055	0.612	0.105	0.648	26.884
60	50	60	0.8	0.10	t_3	0.206	0.061	0.588	0.106	0.674	29.197
60	50	60	0.8	0.25	normal	0.593	0.062	0.611	0.105	0.665	33.148
60	50	60	0.8	0.25	t_3	0.340	0.068	0.612	0.104	0.665	32.230
60	50	120	0.5	0.10	normal	0.368	0.373	0.376	0.122	1.518	29.077
60	50	120	0.5	0.10	t_3	0.345	0.367	0.381	0.121	1.532	33.394
60	50	120	0.5	0.25	normal	0.391	0.381	0.376	0.120	1.479	35.646
60	50	120	0.5	0.25	t_3	0.370	0.375	0.378	0.119	1.485	37.022
60	50	120	0.8	0.10	normal	0.535	0.077	0.612	0.116	1.460	27.753
60	50	120	0.8	0.10	t_3	0.292	0.079	0.602	0.120	1.525	30.978
60	50	120	0.8	0.25	normal	0.597	0.085	0.606	0.123	1.501	34.716
60	50	120	0.8	0.25	t_3	0.409	0.085	0.575	0.120	1.492	34.407
60	100	60	0.5	0.10	normal	0.376	0.374	0.385	0.120	1.322	28.199
60	100	60	0.5	0.10	t_3	0.345	0.367	0.376	0.121	1.341	32.733
60	100	60	0.5	0.25	normal	0.392	0.382	0.369	0.121	1.317	34.462
60	100	60	0.5	0.25	t_3	0.383	0.375	0.366	0.122	1.307	35.402
60	100	60	0.8	0.10	normal	0.522	0.074	0.637	0.116	1.307	26.609
60	100	60	0.8	0.10	t_3	0.273	0.077	0.603	0.118	1.337	29.013
60	100	60	0.8	0.25	normal	0.598	0.083	0.641	0.116	1.308	34.390
60	100	60	0.8	0.25	t_3	0.397	0.085	0.635	0.118	1.324	33.679
60	100	120	0.5	0.10	normal	0.348	0.391	0.408	0.151	2.493	29.132
60	100	120	0.5	0.10	t_3	0.329	0.385	0.403	0.151	2.570	32.259
60	100	120	0.5	0.25	normal	0.356	0.398	0.383	0.156	2.515	35.059
60	100	120	0.5	0.25	t_3	0.342	0.392	0.386	0.148	2.539	35.944
60	100	120	0.8	0.10	normal	0.518	0.093	0.623	0.160	2.520	28.182
60	100	120	0.8	0.10	t_3	0.317	0.089	0.631	0.156	2.537	31.525
60	100	120	0.8	0.25	normal	0.577	0.100	0.640	0.155	2.528	34.436
60	100	120	0.8	0.25	t_3	0.451	0.098	0.613	0.152	2.527	34.628
100	50	60	0.5	0.10	normal	0.465	0.313	0.394	0.176	0.934	28.976
100	50	60	0.5	0.10	t_3	0.398	0.311	0.387	0.181	0.969	32.278
100	50	60	0.5	0.25	normal	0.479	0.322	0.386	0.176	0.944	34.590
100	50	60	0.5	0.25	t_3	0.442	0.315	0.396	0.180	0.941	35.963

Table A3. Cont.

n	p	q	ρ	$\varepsilon_{\text{cell}}$	dist	RMSE			Time (s)		
						AWRcca	MRCd	ccaMM	AWRcca	MRCd	ccaMM
100	50	60	0.8	0.10	normal	0.400	0.019	0.664	0.172	0.949	27.722
100	50	60	0.8	0.10	t_3	0.095	0.034	0.649	0.180	0.969	30.348
100	50	60	0.8	0.25	normal	0.565	0.027	0.653	0.180	0.951	33.811
100	50	60	0.8	0.25	t_3	0.227	0.038	0.635	0.180	0.943	33.463
100	50	120	0.5	0.10	normal	0.459	0.347	0.400	0.186	2.162	31.330
100	50	120	0.5	0.10	t_3	0.420	0.342	0.389	0.187	2.208	33.941
100	50	120	0.5	0.25	normal	0.473	0.354	0.381	0.188	2.172	35.904
100	50	120	0.5	0.25	t_3	0.453	0.349	0.377	0.187	2.166	37.841
100	50	120	0.8	0.10	normal	0.519	0.052	0.656	0.184	2.138	29.779
100	50	120	0.8	0.10	t_3	0.184	0.058	0.625	0.187	2.198	32.119
100	50	120	0.8	0.25	normal	0.637	0.060	0.663	0.186	2.137	35.331
100	50	120	0.8	0.25	t_3	0.360	0.065	0.634	0.186	2.158	34.776
100	100	60	0.5	0.10	normal	0.459	0.345	0.422	0.183	1.910	30.110
100	100	60	0.5	0.10	t_3	0.413	0.341	0.403	0.188	1.968	32.847
100	100	60	0.5	0.25	normal	0.474	0.354	0.395	0.184	1.927	35.591
100	100	60	0.5	0.25	t_3	0.446	0.348	0.381	0.185	1.918	36.058
100	100	60	0.8	0.10	normal	0.503	0.049	0.668	0.184	1.918	28.223
100	100	60	0.8	0.10	t_3	0.157	0.056	0.626	0.183	1.948	31.251
100	100	60	0.8	0.25	normal	0.634	0.057	0.668	0.194	1.910	33.458
100	100	60	0.8	0.25	t_3	0.341	0.061	0.641	0.184	1.919	34.066
100	100	120	0.5	0.10	normal	0.461	0.371	0.415	0.223	3.677	31.260
100	100	120	0.5	0.10	t_3	0.424	0.365	0.403	0.229	3.704	33.298
100	100	120	0.5	0.25	normal	0.465	0.378	0.408	0.227	3.623	37.207
100	100	120	0.5	0.25	t_3	0.448	0.372	0.392	0.230	3.627	37.197
100	100	120	0.8	0.10	normal	0.564	0.072	0.667	0.220	3.612	29.881
100	100	120	0.8	0.10	t_3	0.243	0.075	0.651	0.224	3.723	32.552
100	100	120	0.8	0.25	normal	0.669	0.079	0.675	0.227	3.613	36.213
100	100	120	0.8	0.25	t_3	0.416	0.081	0.645	0.222	3.637	34.979

Table A4. Simulation results under mixed contamination ($\varepsilon_{\text{row}} > 0$, $\varepsilon_{\text{cell}} > 0$). Average RMSE and mean computation time (in seconds) for varying ($n, p, q, \rho, \text{dist}, \varepsilon_{\text{row}}, \varepsilon_{\text{cell}}$).

n	p	q	ρ	ε_{row}	$\varepsilon_{\text{cell}}$	dist	RMSE			Time (s)		
							AWRcca	MRCd	ccaMM	AWRcca	MRCd	ccaMM
30	50	60	0.5	0.10	0.10	normal	0.144	0.380	0.349	0.048	0.434	37.548
30	50	60	0.5	0.10	0.10	t_3	0.138	0.370	0.357	0.047	0.459	45.056
30	50	60	0.5	0.10	0.25	normal	0.172	0.391	0.325	0.056	0.447	40.408
30	50	60	0.5	0.10	0.25	t_3	0.157	0.380	0.329	0.051	0.434	41.787
30	50	60	0.5	0.25	0.10	normal	0.047	0.380	0.346	0.054	0.433	41.799
30	50	60	0.5	0.25	0.10	t_3	0.064	0.324	0.338	0.044	0.451	50.020
30	50	60	0.5	0.25	0.25	normal	0.203	0.390	0.342	0.056	0.436	47.107
30	50	60	0.5	0.25	0.25	t_3	0.053	0.342	0.340	0.055	0.448	52.855
30	50	60	0.8	0.10	0.10	normal	0.365	0.080	0.629	0.046	0.438	42.186
30	50	60	0.8	0.10	0.10	t_3	0.229	0.080	0.527	0.048	0.443	48.520
30	50	60	0.8	0.10	0.25	normal	0.405	0.093	0.556	0.049	0.448	44.331
30	50	60	0.8	0.10	0.25	t_3	0.288	0.088	0.567	0.047	0.451	48.396
30	50	60	0.8	0.25	0.10	normal	0.292	0.080	0.573	0.037	0.450	48.317
30	50	60	0.8	0.25	0.10	t_3	0.253	0.086	0.552	0.032	0.449	55.764
30	50	60	0.8	0.25	0.25	normal	0.324	0.094	0.602	0.054	0.434	52.805
30	50	60	0.8	0.25	0.25	t_3	0.280	0.087	0.577	0.053	0.451	58.590
30	50	120	0.5	0.10	0.10	normal	0.064	0.394	0.378	0.064	0.971	38.543
30	50	120	0.5	0.10	0.10	t_3	0.062	0.382	0.340	0.069	1.007	45.006
30	50	120	0.5	0.10	0.25	normal	0.084	0.403	0.324	0.075	0.985	40.463
30	50	120	0.5	0.10	0.25	t_3	0.069	0.390	0.362	0.094	1.047	43.803
30	50	120	0.5	0.25	0.10	normal	0.084	0.393	0.353	0.069	0.971	40.639
30	50	120	0.5	0.25	0.10	t_3	0.042	0.332	0.343	0.074	0.998	48.649
30	50	120	0.5	0.25	0.25	normal	0.256	0.402	0.322	0.077	0.989	44.644
30	50	120	0.5	0.25	0.25	t_3	0.067	0.349	0.329	0.079	1.003	49.273
30	50	120	0.8	0.10	0.10	normal	0.306	0.095	0.622	0.072	0.965	42.480
30	50	120	0.8	0.10	0.10	t_3	0.233	0.085	0.585	0.065	0.995	50.269
30	50	120	0.8	0.10	0.25	normal	0.337	0.104	0.576	0.069	0.988	44.548
30	50	120	0.8	0.10	0.25	t_3	0.278	0.095	0.562	0.069	1.004	48.160
30	50	120	0.8	0.25	0.10	normal	0.233	0.092	0.609	0.096	1.103	49.233
30	50	120	0.8	0.25	0.10	t_3	0.214	0.088	0.617	0.054	1.002	53.306

Table A4. Cont.

n	p	q	ρ	ε_{row}	$\varepsilon_{\text{cell}}$	dist	RMSE			Time (s)		
							AWRcca	MRCd	ccaMM	AWRcca	MRCd	ccaMM
30	50	120	0.8	0.25	0.25	normal	0.247	0.103	0.593	0.073	0.982	51.232
30	50	120	0.8	0.25	0.25	t_3	0.235	0.076	0.595	0.083	1.011	55.239
30	100	60	0.5	0.10	0.10	normal	0.076	0.394	0.359	0.068	0.843	36.429
30	100	60	0.5	0.10	0.10	t_3	0.061	0.384	0.372	0.070	0.885	43.333
30	100	60	0.5	0.10	0.25	normal	0.087	0.405	0.345	0.082	0.860	40.358
30	100	60	0.5	0.10	0.25	t_3	0.072	0.392	0.338	0.080	0.916	42.231
30	100	60	0.5	0.25	0.10	normal	0.088	0.395	0.357	0.077	0.850	39.331
30	100	60	0.5	0.25	0.10	t_3	0.046	0.336	0.352	0.069	0.881	46.000
30	100	60	0.5	0.25	0.25	normal	0.254	0.404	0.348	0.086	0.862	43.669
30	100	60	0.5	0.25	0.25	t_3	0.076	0.349	0.361	0.079	0.872	49.029
30	100	60	0.8	0.10	0.10	normal	0.305	0.095	0.591	0.083	0.864	41.605
30	100	60	0.8	0.10	0.10	t_3	0.233	0.088	0.576	0.080	0.867	47.435
30	100	60	0.8	0.10	0.25	normal	0.345	0.105	0.590	0.071	0.867	41.854
30	100	60	0.8	0.10	0.25	t_3	0.279	0.095	0.556	0.081	0.874	47.775
30	100	60	0.8	0.25	0.10	normal	0.238	0.094	0.611	0.056	0.861	45.584
30	100	60	0.8	0.25	0.10	t_3	0.205	0.077	0.561	0.059	0.861	52.395
30	100	60	0.8	0.25	0.25	normal	0.252	0.103	0.609	0.076	0.854	49.826
30	100	60	0.8	0.25	0.25	t_3	0.237	0.082	0.631	0.082	0.872	53.939
30	100	120	0.5	0.10	0.10	normal	0.036	0.407	0.371	0.088	1.675	34.869
30	100	120	0.5	0.10	0.10	t_3	0.038	0.395	0.386	0.083	1.707	40.602
30	100	120	0.5	0.10	0.25	normal	0.033	0.416	0.364	0.093	1.697	37.291
30	100	120	0.5	0.10	0.25	t_3	0.036	0.404	0.375	0.092	1.709	39.611
30	100	120	0.5	0.25	0.10	normal	0.149	0.407	0.360	0.095	1.657	34.896
30	100	120	0.5	0.25	0.10	t_3	0.112	0.349	0.383	0.087	1.718	41.483
30	100	120	0.5	0.25	0.25	normal	0.298	0.416	0.344	0.100	1.684	38.541
30	100	120	0.5	0.25	0.25	t_3	0.139	0.369	0.366	0.091	1.711	42.360
30	100	120	0.8	0.10	0.10	normal	0.253	0.109	0.577	0.085	1.655	38.923
30	100	120	0.8	0.10	0.10	t_3	0.193	0.098	0.614	0.082	1.696	45.534
30	100	120	0.8	0.10	0.25	normal	0.278	0.116	0.605	0.085	1.876	39.333
30	100	120	0.8	0.10	0.25	t_3	0.226	0.107	0.601	0.085	1.695	41.999
30	100	120	0.8	0.25	0.10	normal	0.164	0.108	0.603	0.076	1.704	43.505
30	100	120	0.8	0.25	0.10	t_3	0.155	0.079	0.605	0.073	1.697	48.112
30	100	120	0.8	0.25	0.25	normal	0.179	0.115	0.601	0.095	1.659	43.842
30	100	120	0.8	0.25	0.25	t_3	0.181	0.089	0.611	0.093	1.691	48.794
60	50	60	0.5	0.10	0.10	normal	0.393	0.349	0.376	0.082	0.665	40.947
60	50	60	0.5	0.10	0.10	t_3	0.364	0.346	0.363	0.078	0.668	48.167
60	50	60	0.5	0.10	0.25	normal	0.390	0.361	0.343	0.095	0.661	46.949
60	50	60	0.5	0.10	0.25	t_3	0.386	0.355	0.368	0.080	0.641	49.396
60	50	60	0.5	0.25	0.10	normal	0.076	0.355	0.359	0.092	0.640	49.046
60	50	60	0.5	0.25	0.10	t_3	0.166	0.295	0.361	0.046	0.675	58.538
60	50	60	0.5	0.25	0.25	normal	0.290	0.360	0.354	0.112	0.648	59.113
60	50	60	0.5	0.25	0.25	t_3	0.188	0.310	0.356	0.101	0.660	62.069
60	50	60	0.8	0.10	0.10	normal	0.483	0.055	0.600	0.077	0.644	45.956
60	50	60	0.8	0.10	0.10	t_3	0.283	0.060	0.606	0.079	0.657	52.270
60	50	60	0.8	0.10	0.25	normal	0.587	0.064	0.628	0.079	0.650	51.658
60	50	60	0.8	0.10	0.25	t_3	0.377	0.068	0.612	0.079	0.652	56.338
60	50	60	0.8	0.25	0.10	normal	0.339	0.058	0.631	0.037	0.628	55.400
60	50	60	0.8	0.25	0.10	t_3	0.322	0.064	0.606	0.039	0.663	60.789
60	50	60	0.8	0.25	0.25	normal	0.472	0.067	0.593	0.091	0.649	59.848
60	50	60	0.8	0.25	0.25	t_3	0.414	0.078	0.568	0.066	0.656	62.540
60	50	120	0.5	0.10	0.10	normal	0.342	0.373	0.376	0.094	1.474	40.756
60	50	120	0.5	0.10	0.10	t_3	0.323	0.365	0.364	0.094	1.503	47.781
60	50	120	0.5	0.10	0.25	normal	0.358	0.382	0.383	0.115	1.476	47.362
60	50	120	0.5	0.10	0.25	t_3	0.339	0.373	0.346	0.103	1.513	50.740
60	50	120	0.5	0.25	0.10	normal	0.066	0.376	0.357	0.108	1.454	46.924
60	50	120	0.5	0.25	0.10	t_3	0.086	0.306	0.367	0.065	1.511	55.482
60	50	120	0.5	0.25	0.25	normal	0.343	0.383	0.364	0.123	1.463	57.423
60	50	120	0.5	0.25	0.25	t_3	0.111	0.326	0.350	0.116	1.498	63.010
60	50	120	0.8	0.10	0.10	normal	0.545	0.077	0.639	0.091	1.441	44.906
60	50	120	0.8	0.10	0.10	t_3	0.321	0.076	0.627	0.095	1.507	50.671
60	50	120	0.8	0.10	0.25	normal	0.585	0.084	0.658	0.095	1.671	51.112
60	50	120	0.8	0.10	0.25	t_3	0.434	0.084	0.639	0.095	1.490	54.879
60	50	120	0.8	0.25	0.10	normal	0.260	0.077	0.628	0.092	1.589	57.173
60	50	120	0.8	0.25	0.10	t_3	0.291	0.091	0.582	0.053	1.489	59.897
60	50	120	0.8	0.25	0.25	normal	0.393	0.086	0.612	0.107	1.442	61.423
60	50	120	0.8	0.25	0.25	t_3	0.379	0.082	0.606	0.087	1.488	63.103
60	100	60	0.5	0.10	0.10	normal	0.346	0.374	0.391	0.093	1.311	38.226
60	100	60	0.5	0.10	0.10	t_3	0.326	0.366	0.386	0.098	1.336	46.678
60	100	60	0.5	0.10	0.25	normal	0.362	0.382	0.389	0.113	1.299	45.815
60	100	60	0.5	0.10	0.25	t_3	0.344	0.374	0.370	0.100	1.317	47.188
60	100	60	0.5	0.25	0.10	normal	0.069	0.376	0.368	0.109	1.280	44.911
60	100	60	0.5	0.25	0.10	t_3	0.096	0.308	0.370	0.065	1.334	55.095
60	100	60	0.5	0.25	0.25	normal	0.348	0.383	0.369	0.122	1.288	56.206
60	100	60	0.5	0.25	0.25	t_3	0.130	0.318	0.373	0.115	1.336	62.500
60	100	60	0.8	0.10	0.10	normal	0.526	0.076	0.649	0.089	1.278	43.693
60	100	60	0.8	0.10	0.10	t_3	0.329	0.075	0.654	0.093	1.325	49.697
60	100	60	0.8	0.10	0.25	normal	0.599	0.085	0.650	0.098	1.306	49.290
60	100	60	0.8	0.10	0.25	t_3	0.438	0.084	0.624	0.092	1.303	54.322
60	100	60	0.8	0.25	0.10	normal	0.265	0.079	0.627	0.062	1.297	53.266
60	100	60	0.8	0.25	0.10	t_3	0.302	0.087	0.612	0.051	1.309	59.465

Table A4. Cont.

<i>n</i>	<i>p</i>	<i>q</i>	ρ	ε_{row}	ε_{cell}	dist	RMSE			Time (s)		
							AWRcca	MRCd	ccaMM	AWRcca	MRCd	ccaMM
60	100	60	0.8	0.25	0.25	normal	0.387	0.087	0.609	0.107	1.291	60.840
60	100	60	0.8	0.25	0.25	t_3	0.381	0.067	0.616	0.084	1.322	63.874
60	100	120	0.5	0.10	0.10	normal	0.300	0.392	0.401	0.126	2.456	37.025
60	100	120	0.5	0.10	0.10	t_3	0.289	0.382	0.399	0.126	2.538	41.134
60	100	120	0.5	0.10	0.25	normal	0.313	0.399	0.367	0.146	2.482	41.372
60	100	120	0.5	0.10	0.25	t_3	0.301	0.389	0.370	0.135	2.514	45.648
60	100	120	0.5	0.25	0.10	normal	0.131	0.393	0.392	0.141	2.435	40.307
60	100	120	0.5	0.25	0.10	t_3	0.053	0.327	0.373	0.093	2.537	49.584
60	100	120	0.5	0.25	0.25	normal	0.382	0.399	0.365	0.155	2.435	52.628
60	100	120	0.5	0.25	0.25	t_3	0.068	0.340	0.383	0.152	2.514	58.796
60	100	120	0.8	0.10	0.10	normal	0.517	0.093	0.629	0.125	2.520	39.751
60	100	120	0.8	0.10	0.10	t_3	0.351	0.088	0.646	0.120	2.525	45.546
60	100	120	0.8	0.10	0.25	normal	0.570	0.100	0.635	0.123	2.626	46.255
60	100	120	0.8	0.10	0.25	t_3	0.454	0.097	0.612	0.131	2.501	51.008
60	100	120	0.8	0.25	0.10	normal	0.197	0.094	0.654	0.083	2.497	55.559
60	100	120	0.8	0.25	0.10	t_3	0.259	0.084	0.654	0.083	2.545	58.323
60	100	120	0.8	0.25	0.25	normal	0.341	0.101	0.673	0.141	2.465	61.125
60	100	120	0.8	0.25	0.25	t_3	0.348	0.075	0.656	0.122	2.498	62.467
100	50	60	0.5	0.10	0.10	normal	0.458	0.313	0.396	0.130	0.919	43.507
100	50	60	0.5	0.10	0.10	t_3	0.409	0.308	0.388	0.137	0.947	51.726
100	50	60	0.5	0.10	0.25	normal	0.476	0.322	0.378	0.143	0.933	53.145
100	50	60	0.5	0.10	0.25	t_3	0.458	0.318	0.382	0.138	0.943	55.842
100	50	60	0.5	0.25	0.10	normal	0.093	0.319	0.378	0.163	0.925	54.733
100	50	60	0.5	0.25	0.10	t_3	0.238	0.253	0.355	0.063	0.941	61.674
100	50	60	0.5	0.25	0.25	normal	0.252	0.322	0.383	0.179	0.923	60.858
100	50	60	0.5	0.25	0.25	t_3	0.310	0.270	0.338	0.173	0.938	62.746
100	50	60	0.8	0.10	0.10	normal	0.480	0.020	0.661	0.125	0.924	49.269
100	50	60	0.8	0.10	0.10	t_3	0.195	0.032	0.647	0.135	0.942	56.595
100	50	60	0.8	0.10	0.25	normal	0.625	0.028	0.672	0.132	0.927	58.776
100	50	60	0.8	0.10	0.25	t_3	0.313	0.038	0.643	0.135	0.944	61.508
100	50	60	0.8	0.25	0.10	normal	0.377	0.027	0.653	0.058	0.913	57.644
100	50	60	0.8	0.25	0.10	t_3	0.355	0.071	0.629	0.052	0.932	60.943
100	50	60	0.8	0.25	0.25	normal	0.544	0.033	0.585	0.135	0.912	59.907
100	50	60	0.8	0.25	0.25	t_3	0.488	0.059	0.557	0.069	0.939	62.714
100	50	120	0.5	0.10	0.10	normal	0.458	0.347	0.394	0.141	2.115	44.582
100	50	120	0.5	0.10	0.10	t_3	0.425	0.342	0.394	0.144	2.178	50.534
100	50	120	0.5	0.10	0.25	normal	0.468	0.356	0.398	0.166	2.119	54.505
100	50	120	0.5	0.10	0.25	t_3	0.455	0.347	0.395	0.149	2.174	57.279
100	50	120	0.5	0.25	0.10	normal	0.060	0.349	0.385	0.177	2.126	55.085
100	50	120	0.5	0.25	0.10	t_3	0.162	0.277	0.399	0.071	2.174	61.974
100	50	120	0.5	0.25	0.25	normal	0.311	0.356	0.383	0.187	2.097	62.657
100	50	120	0.5	0.25	0.25	t_3	0.258	0.281	0.386	0.178	2.111	64.984
100	50	120	0.8	0.10	0.10	normal	0.565	0.054	0.640	0.142	2.129	48.910
100	50	120	0.8	0.10	0.10	t_3	0.268	0.060	0.638	0.144	2.168	54.119
100	50	120	0.8	0.10	0.25	normal	0.679	0.060	0.655	0.143	2.333	58.113
100	50	120	0.8	0.10	0.25	t_3	0.427	0.067	0.654	0.142	2.131	60.638
100	50	120	0.8	0.25	0.10	normal	0.286	0.055	0.660	0.091	2.221	62.911
100	50	120	0.8	0.25	0.10	t_3	0.349	0.077	0.653	0.066	2.147	62.088
100	50	120	0.8	0.25	0.25	normal	0.477	0.063	0.618	0.162	2.066	62.314
100	50	120	0.8	0.25	0.25	t_3	0.470	0.086	0.584	0.081	2.108	63.938
100	100	60	0.5	0.10	0.10	normal	0.451	0.347	0.401	0.143	1.885	40.907
100	100	60	0.5	0.10	0.10	t_3	0.426	0.340	0.394	0.146	1.929	49.874
100	100	60	0.5	0.10	0.25	normal	0.474	0.354	0.390	0.168	1.887	51.994
100	100	60	0.5	0.10	0.25	t_3	0.452	0.347	0.397	0.159	1.952	55.277
100	100	60	0.5	0.25	0.10	normal	0.070	0.349	0.396	0.177	1.882	53.420
100	100	60	0.5	0.25	0.10	t_3	0.158	0.275	0.390	0.073	1.916	60.821
100	100	60	0.5	0.25	0.25	normal	0.319	0.354	0.386	0.189	1.844	61.923
100	100	60	0.5	0.25	0.25	t_3	0.253	0.288	0.393	0.180	1.893	64.436
100	100	60	0.8	0.10	0.10	normal	0.547	0.051	0.639	0.145	1.870	47.051
100	100	60	0.8	0.10	0.10	t_3	0.281	0.057	0.649	0.148	1.910	54.928
100	100	60	0.8	0.10	0.25	normal	0.677	0.058	0.676	0.143	1.995	55.631
100	100	60	0.8	0.10	0.25	t_3	0.410	0.066	0.645	0.148	1.890	60.811
100	100	60	0.8	0.25	0.10	normal	0.298	0.054	0.674	0.092	1.919	57.605
100	100	60	0.8	0.25	0.10	t_3	0.353	0.087	0.649	0.066	1.890	61.904
100	100	60	0.8	0.25	0.25	normal	0.481	0.061	0.621	0.162	1.852	62.170
100	100	60	0.8	0.25	0.25	t_3	0.469	0.059	0.612	0.079	1.893	64.583
100	100	120	0.5	0.10	0.10	normal	0.449	0.371	0.396	0.183	3.601	39.521
100	100	120	0.5	0.10	0.10	t_3	0.422	0.361	0.394	0.180	3.640	45.446
100	100	120	0.5	0.10	0.25	normal	0.459	0.376	0.385	0.215	3.577	49.354
100	100	120	0.5	0.10	0.25	t_3	0.443	0.369	0.396	0.186	3.620	52.844
100	100	120	0.5	0.25	0.10	normal	0.134	0.373	0.400	0.206	3.555	48.993
100	100	120	0.5	0.25	0.10	t_3	0.115	0.305	0.394	0.115	3.642	58.831
100	100	120	0.5	0.25	0.25	normal	0.362	0.379	0.408	0.224	3.519	58.922
100	100	120	0.5	0.25	0.25	t_3	0.203	0.316	0.405	0.216	3.584	63.660
100	100	120	0.8	0.10	0.10	normal	0.596	0.073	0.684	0.175	3.541	44.348
100	100	120	0.8	0.10	0.10	t_3	0.361	0.073	0.675	0.183	3.593	50.838
100	100	120	0.8	0.10	0.25	normal	0.693	0.079	0.690	0.179	3.557	54.809
100	100	120	0.8	0.10	0.25	t_3	0.491	0.081	0.695	0.187	3.579	61.335
100	100	120	0.8	0.25	0.10	normal	0.232	0.076	0.680	0.104	3.629	64.225
100	100	120	0.8	0.25	0.10	t_3	0.331	0.077	0.664	0.102	3.609	63.553
100	100	120	0.8	0.25	0.25	normal	0.426	0.081	0.660	0.207	3.539	64.613
100	100	120	0.8	0.25	0.25	t_3	0.466	0.062	0.623	0.119	3.517	63.002

References

- Hotelling, H. Relations between two sets of variates. In *Breakthroughs in Statistics: Methodology and Distribution*; Springer: Berlin/Heidelberg, Germany, 1992; pp. 162–190.
- Johnson, R.A.; Wichern, D.W. *Applied Multivariate Statistical Analysis*; Prentice Hall: Upper Saddle River, NJ, USA, 2002.
- Varmuza, K.; Filzmoser, P. *Introduction to Multivariate Statistical Analysis in Chemometrics*; CRC Press: Boca Raton, FL, USA, 2016.
- Giraud, C. *Introduction to High-Dimensional Statistics*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2021.
- Centofanti, F.; Hubert, M.; Rousseeuw, P.J. Cellwise and Casewise Robust Covariance in High Dimensions. *arXiv* **2025**, arXiv:2505.19925.
- Pourahmadi, M. *High-Dimensional Covariance Estimation: With High-Dimensional Data*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
- Wilms, I.; Croux, C. Sparse canonical correlation analysis from a predictive point of view. *Biom. J.* **2015**, *57*, 834–851. [[CrossRef](#)] [[PubMed](#)]
- Witten, D.M.; Tibshirani, R.; Hastie, T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* **2009**, *10*, 515–534. [[CrossRef](#)] [[PubMed](#)]
- Boudt, K.; Rousseeuw, P.J.; Vanduffel, S.; Verdonck, T. The minimum regularized covariance determinant estimator. *Stat. Comput.* **2020**, *30*, 113–128.
- Rousseeuw, P.J. Least median of squares regression. *J. Am. Stat. Assoc.* **1984**, *79*, 871–880. [[CrossRef](#)]
- Pfeiffer, P.; Alfons, A.; Filzmoser, P. Efficient computation of sparse and robust maximum association estimators. *Comput. Stat. Data Anal.* **2025**, *207*, 108133. [[CrossRef](#)]
- Coraggio, L.; Coretto, P.; Peluso, A. Community Detection In Financial Networks And The Importance Of Being Robust. In *CLADAG-VOC 2025 Book of Abstracts*; Zaccaria Editore: Montese, Italy, 2025; pp. 48–48.
- Alqallaf, F.; Van Aelst, S.; Yohai, V.J.; Zamar, R.H. Propagation of outliers in multivariate data. *Ann. Stat.* **2009**, *37*, 311–331. [[CrossRef](#)]
- Raymaekers, J.; Rousseeuw, P.J. Challenges of cellwise outliers. *Econom. Stat.* **2024**, *38*, 6–25. [[CrossRef](#)]
- González, I.; Déjean, S.; Martin, P.G.; Baccini, A. CCA: An R package to extend canonical correlation analysis. *J. Stat. Softw.* **2008**, *23*, 1–14.
- Bulut, H. An R package for multivariate hypothesis tests: MVTests. *Technol. Appl. Sci.* **2019**, *14*, 132–138. [[CrossRef](#)]
- Raymaekers, J.; Rousseeuw, P.J. Fast robust correlation for high-dimensional data. *Technometrics* **2021**, *63*, 184–198.
- Ledoit, O.; Wolf, M. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* **2004**, *88*, 365–411. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.