Chilean Journal of Statistics
Volume 15, Number 2, December 2024, Pages 126–146
DOI: 10.32372/ChJS.15-02-02

Statistical Modeling
Research Article

A method of multi-dimensional variable selection for additive partial linear models

Munaf Y. Hmood[1,*] and Hayder Raaid Talib[2]
[1]Department of Statistics, University of Baghdad, Baghdad, Iraq
[2]Department of Statistics, University of Sumer, Dhi Qar, Iraq

(Received: 30 May 2023 · Accepted in final form: 08 May 2024)

Abstract

In high-dimensional semiparametric regression, balancing accuracy and interpretability often requires combining dimension reduction with variable selection. This study introduces two novel methods for dimension reduction in additive partial linear models: (i) minimum average variance estimation (MAVE) combined with the adaptive least absolute shrinkage and selection operator (MAVE-ALASSO) and (ii) MAVE with smoothly clipped absolute deviation (MAVE-SCAD). These methods leverage the flexibility of MAVE for sufficient dimension reduction while incorporating adaptive penalties to ensure sparse and interpretable models. The performance of both methods is evaluated through simulations using the mean squared error and variable selection criteria, assessing the correct detection of zero coefficients and the false omission of nonzero coefficients. A practical application involving financial data from the Baghdad Soft Drinks Company demonstrates their utility in identifying key predictors of stock market value. The results indicate that MAVE-SCAD performs well in high-dimensional and complex scenarios, whereas MAVE-ALASSO is better suited to small samples, producing more parsimonious models. These results highlight the effectiveness of these two methods in addressing key challenges in semiparametric modeling.

Keywords: Adaptive least absolute shrinkage and selection operator · Dimension reduction · LASSO · Mean squared error · Minimum average variance estimation · Smoothly clipped absolute deviation

Mathematics Subject Classification: Primary 62G08 · Secondary 62J99.

1. Introduction

In modern data analysis, balancing interpretability and flexibility in high-dimensional semiparametric models is a critical challenge. Additive partial linear models (APLMs) are powerful tools in this context, as they generalize multiple linear regression models while allowing nonlinear relationships for selected explanatory variables (covariates). APLMs combine the interpretability of parametric components with the flexibility of nonparametric ones, making them particularly suitable for scenarios where the response variable depends linearly on some covariates but nonlinearly on others (Lian et al., 2014; Xinyu and Wendun, 2019).

*Corresponding author. Email: hayder.raaid@uos.edu.iq (H.R.Talib)

Compared to purely linear models, APLMs offer greater flexibility, and to fully nonparametric models, they mitigate the curse of dimensionality, which is a common limitation in high-dimensional data analysis (Alkenani and Yu, 2013; Huda and Hmood, 2022). However, effectively modeling APLMs in high dimensions often requires sufficient dimension reduction (SDR) techniques, which project the original p-dimensional covariate space onto a d-dimensional subspace ($d < p$), preserving essential regression information without fully specifying the model or error distribution (Leng et al., 2009; Hua et al., 2012).

A limitation in the literature lies in the disconnection between dimension reduction and variable selection. While minimum average variance estimation (MAVE) is a widely recognized SDR technique (Xia et al., 2014), it lacks mechanisms for selecting relevant variables, which are critical for achieving sparse and interpretable models. Adaptive penalties, such as the adaptive least absolute shrinkage and selection operator —ALASSO— (Zou, 2006) and smoothly clipped absolute deviation —SCAD— (Fan and Li, 2001), have demonstrated strong capabilities in variable selection and achieving oracle properties. However, their integration with MAVE in the context of APLMs remains underexplored (Hua et al., 2012).

This study seeks to address this gap by introducing two novel methods: minimum average variance estimation (MAVE) with adaptive least absolute shrinkage and selection operator (MAVE-ALASSO) and (ii) MAVE with smoothly SCAD (MAVE-SCAD). These approaches extend existing SDR techniques by incorporating adaptive penalties to achieve sparse estimation while preserving model accuracy. Specifically, MAVE-ALASSO combines MAVE with the adaptive LASSO, offering strong sparsity and consistency properties, while MAVE-SCAD integrates SCAD with MAVE, reducing shrinkage bias for large coefficients. Our contributions are twofold. First, we formalize these integrated approaches for variable selection and dimension reduction within the APLM framework. Second, we compare the performance of MAVE-ALASSO and MAVE-SCAD using mean squared error (MSE) and additional variable selection accuracy criteria, "C" (correct identification of zero coefficients) and "I" (incorrect omission of nonzero coefficients). The effectiveness of these methods is demonstrated through extensive simulations and a real-data application.

The practical relevance of this study is underscored by its potential applications in high-dimensional, nonlinear data settings, such as financial modeling, economic forecasting, and other domains in data science. By addressing the dual challenges of dimension reduction and variable selection, the proposed methods provide a robust and computationally efficient framework for analyzing complex data.

The article is organized as follows. Section 2 introduces the methodological framework Section 3 discusses the practical implementation of MAVE. In Section 4, we focus on the MAVE-ALASSO and MAVE-SCAD procedures. In Section 5, the performance of the proposed methods is evaluated through an extensive simulation study, while Section 6 illustrates their application to a real-world dataset. Section 7 concludes the article.

2. Methodological framework

In this section, we present the methodological framework underlying APLMs and their integration with SDR techniques. We focus on the use of MAVE coupled with adaptive penalization for dimension reduction and variable selection.

2.1 *Additive partial linear model*

APLMs are semiparametric models that combine both a linear component for some and nonparametric components for the remaining ones (Huda and Hmood, 2022; Hua et al., 2012; Da Silva et al., 2011). By partially specifying the model as linear and partially as nonparametric, APLMs allow the inclusion of nonlinear effects while retaining interpretability and reducing complexity relative to fully nonparametric approaches.

Formally, an APLM can be written as

$$Y = \boldsymbol{X}^\top \boldsymbol{\beta} + \sum_{k=1}^{K} g_k(Z_k) + \varepsilon, \tag{2.1}$$

where, for each $i \in \{1, \ldots, n\}$, $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})^\top$ is the vector of linear covariates; $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{ik})^\top$ is the vector of nonlinear covariates; $g_1, \ldots, g_K$ are unknown smooth functions; $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_d)^\top$ is a vector of unknown parameters corresponding to the linear part; and ε_i is the random error, assumed to be independent of $(\boldsymbol{X}_i, \boldsymbol{Z}_i)$ with mean zero and variance σ^2. To ensure identifiability for each g_k, it is customary to impose the constraint $\mathrm{E}[g_k(Z_k)] = 0$, for $k \in \{1, \ldots, K\}$.

An important advantage of APLMs is their capacity to combine the clarity of linear modeling for certain covariates with the ability to represent nonlinear effects through $g_1, \ldots, g_K$. By balancing these components, APLMs address the limitations of fully nonparametric models, which often require substantial sample sizes to reliably estimate high-dimensional structures (Alkenani and Yu, 2013; Huda and Hmood, 2022).

In subsequent sections, we incorporate ideas from SDR into the APLM framework. Specifically, we utilize MAVE coupled with penalization (ALASSO or SCAD) to perform both dimension reduction and variable selection within the APLM.

2.2 *Sufficient dimension reduction*

SDR aims to find a low-dimensional linear subspace of the original covariate space that preserves essential information about the conditional distribution $Y \mid \boldsymbol{X}$. A key concept in SDR is the central mean subspace (CMS), denoted by $S_{\mathrm{E}[Y|\boldsymbol{X}]}$, which is the smallest subspace $\mathcal{S} \subseteq \mathrm{R}^p$ such that $\mathrm{E}[Y \mid \boldsymbol{X}] = \mathrm{E}[Y \mid P_{\mathcal{S}}\boldsymbol{X}]$, where $P_{\mathcal{S}}$ is the orthogonal projection onto the subspace $\mathcal{S}$. Intuitively, $S_{\mathrm{E}[Y|\boldsymbol{X}]}$ spans all linear combinations of $\boldsymbol{X}$ needed to determine $\mathrm{E}[Y \mid \boldsymbol{X}]$ (Hua et al., 2012; Cook and Bing, 2002; Wang et al., 2014; Cook and Forzani, 2009).

Several methods have been proposed to estimate the CMS, including the iterative Hessian transformation and the MAVE approach (Horowitz, 2015). In the context of APLMs, combining MAVE with penalization not only helps to reduce dimensionality but also facilitates variable selection. In the next sections, we present two penalized MAVE procedures: MAVE-ALASSO and MAVE-SCAD. The MAVE method (Huang et al., 2008) is a widely used approach for SDR that employs local linear smoothing. It does not require strong assumptions on the distribution or the functional form of the covariates X, and it adaptively estimates the underlying structure without specifying a parametric link function. Consequently, MAVE is flexible and can effectively capture nonlinear relationships within the CMS.

2.3 *Minimum average variance estimation*

To estimate the projection matrix $\boldsymbol{B}$ in a dimension-reduction context, MAVE aims to solve

$$\min_{\boldsymbol{B}} \left\{ \mathrm{E}[(Y - \mathrm{E}[Y \mid \boldsymbol{X}^\top \boldsymbol{B}])^2] \right\}. \tag{2.2}$$

Let $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)$ be an orthonormal matrix of size $p \times d$ ($\boldsymbol{B}^\top \boldsymbol{B} = \boldsymbol{I}_d$). For each $\boldsymbol{B}$, define the conditional variance function stated as

$$\sigma^2_{\boldsymbol{B}}(\boldsymbol{X}^\top \boldsymbol{B}) = \mathrm{E}[(Y - \mathrm{E}[Y \mid \boldsymbol{X}^\top \boldsymbol{B}])^2 \Big| \boldsymbol{X}^\top \boldsymbol{B}], \tag{2.3}$$

so that

$$\mathrm{E}\Big[\big(Y - \mathrm{E}[Y \mid \boldsymbol{X}^\top \boldsymbol{B}]\big)^2\Big] = \mathrm{E}\Big[\sigma^2_{\boldsymbol{B}}(\boldsymbol{X}^\top \boldsymbol{B})\Big]. \tag{2.4}$$

Since $\boldsymbol{B}^\top \boldsymbol{B} = \boldsymbol{I}_d$, minimizing the left-hand side of Equation (2.4) is equivalent to minimizing

$$\mathrm{E}\Big[\sigma^2_{\boldsymbol{B}}(\boldsymbol{X}^\top \boldsymbol{B})\Big]. \tag{2.5}$$

Hence, we have that

$$\min_{\boldsymbol{B}} \Big\{\mathrm{E}\Big[\big(Y - \mathrm{E}[Y \mid \boldsymbol{X}^\top \boldsymbol{B}]\big)^2\Big]\Big\} = \min_{\boldsymbol{B}} \Big\{\mathrm{E}\Big[\sigma^2_{\boldsymbol{B}}(\boldsymbol{X}^\top \boldsymbol{B})\Big]\Big\}. \tag{2.6}$$

In other words, finding $\boldsymbol{B}$ that minimizes $\mathrm{E}[\sigma^2_{\boldsymbol{B}}(\cdot)]$ yields the optimal dimension reduction in the sense of preserving $\mathrm{E}[Y \mid \boldsymbol{X}]$.

3. Implementation and practical considerations

This section focuses on the practical aspects of applying the MAVE method within the APLM framework. We discuss implementation, including local smoothing, bandwidth selection, and the iterative procedure for estimating the projection matrix.

3.1 *Local smoothing and kernel weights*

Suppose we have an independent and identically distributed sample $\{(\boldsymbol{X}_i, Y_i)\}_{i=1}^n$ from the joint distribution of $(\boldsymbol{X}, Y)$. Define $g_{\boldsymbol{B}}(v_1, \ldots, v_d) = \mathrm{E}[Y \mid \boldsymbol{b}_1^\top \boldsymbol{X} = v_1, \ldots, \boldsymbol{b}_d^\top \boldsymbol{X} = v_d]$, where $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)$ is the projection matrix of size $p \times d$. For any fixed point $\boldsymbol{X}_0$, we can approximate the function $\mathrm{E}[Y \mid \boldsymbol{B}^\top \boldsymbol{X} = \boldsymbol{B}^\top \boldsymbol{X}_0]$ using a local linear expansion.

In addition to the partially linear structure, suppose the model includes a nonparametric term that can be expressed as

$$g(\boldsymbol{Z}) = \gamma^\top b(\boldsymbol{Z}), \tag{3.7}$$

where b is a vector of known basis functions (for example, spline basis) and γ is a vector of unknown coefficients. Then, for the i-th observation, $g(\boldsymbol{Z}_i) = \gamma^\top b(\boldsymbol{Z}_i)$.

Hence, the difference $Y_i - g_{\boldsymbol{B}}(\boldsymbol{B}^\top \boldsymbol{X}_i)$ can be approximated by $Y_i - a - b^\top \boldsymbol{B}^\top (\boldsymbol{X}_i - \boldsymbol{X}_0) - \gamma^\top B(\boldsymbol{Z}_i)$. Using a local linear smoothing framework, we fit (a, b) by minimizing the weighted sum of squared residuals: $\sum_{i=1}^n (Y_i - (a + b^\top \boldsymbol{B}^\top (\boldsymbol{X}_i - \boldsymbol{X}_0) + \gamma^\top B(\boldsymbol{Z}_i)))^2 W_{i0}$, where $W_{i0} \geq 0$ is a kernel weight centered around $\boldsymbol{B}^\top \boldsymbol{X}_0$ such that $\sum_{i=1}^n W_{i0} = 1$. Specifically, we get

$$W_{i0} = \frac{K_h(\boldsymbol{B}^\top (\boldsymbol{X}_i - \boldsymbol{X}_0))}{\sum_{\ell=1}^n K_h(\boldsymbol{B}^\top (\boldsymbol{X}_\ell - \boldsymbol{X}_0))}, \tag{3.8}$$

with $K_h(u) = h^{-d} K(u/h)$ for a chosen kernel function K and bandwidth $h > 0$. Determining h appropriately (for example, via cross-validation —CV—) is crucial to balancing bias and variance in local linear estimation.

To estimate (a, b) for each local neighborhood, we use the criterion implied by Equation (2.4). Specifically, for a given $\boldsymbol{B}$ and a point $\boldsymbol{B}^\top \boldsymbol{X}_0$, we approximate $\sigma^2_{\boldsymbol{B}}(\boldsymbol{B}^\top \boldsymbol{X}_0)$ —from Equation (2.3)— via the formulation presented as

$$\widehat{\sigma}^2_{\boldsymbol{B}}(\boldsymbol{B}^\top \boldsymbol{X}_0) = \min_{a,b} \left\{ \sum_{i=1}^{n} \Big(Y_i - a - b^\top \boldsymbol{B}^\top (\boldsymbol{X}_i - \boldsymbol{X}_0) - \gamma^\top B(\boldsymbol{Z}_i) \Big) 2W_{i0} \right\}.$$

Under regular conditions, it holds that $\widehat{\sigma}^2_{\boldsymbol{B}}(\boldsymbol{B}^\top \boldsymbol{X}_0) \approx \sigma^2_{\boldsymbol{B}}(\boldsymbol{B}^\top \boldsymbol{X}_0) = O_p(1)$, where $O_p(1)$ denotes a sequence of random variables that is bounded in probability, and p refers to the dimension of the original covariate space.

Using Equations (2.2), (2.5), and (2.6), we can aggregate these local estimates to obtain the objective given by

$$\min_{\boldsymbol{B}:\boldsymbol{B}^\top \boldsymbol{B}=\boldsymbol{I}} \left\{ \sum_{j=1}^{n} \widehat{\sigma}^2_{\boldsymbol{B}}(\boldsymbol{B}^\top \boldsymbol{X}_j) \right\} = \min_{\substack{\boldsymbol{B}:\boldsymbol{B}^\top \boldsymbol{B}=\boldsymbol{I} \\ a_j, b_j;\, j \in \{1,\ldots,n\}}} \left\{ \sum_{j=1}^{n} \sum_{i=1}^{n} \Big(Y_i - a_j - b_j^\top \boldsymbol{B}^\top (\boldsymbol{X}_i - \boldsymbol{X}_j) - \gamma^\top B(\boldsymbol{Z}_i) \Big)^2 W_{ij} \right\},$$

where $b_j^\top = (b_{j1}, \ldots, b_{jd})$ and $K(\boldsymbol{z}) = (1/\sqrt{2\pi}) \exp(-\boldsymbol{z}^\top \boldsymbol{z}/2)$ denotes the multivariate Gaussian kernel used in the definition of W_{ij} stated in Equation (3.8). The optimal bandwidth $h_{\text{opt}} = A_{(d)} n^{-1/(4+d)}$, where $A_{(d)} = (4/(d+2))^{1/(4+d)}$, is a commonly used rule-of-thumb for local linear smoothing in d-dimensional settings. Correct bandwidth selection is crucial, as it balances bias and variance in the local linear estimates. Several error metrics, such as the average squared error or the integrated squared error, can guide the selection of h (Hmood and Stadtmuller, 2013).

A common practical approach is CV, which balances squared bias and variance by leaving out one observation at a time (the leave-one-out method) and choosing the bandwidth h that minimizes a CV function, defined as

$$\text{CV}(h) = \frac{1}{n} \sum_{i=1}^{n} \Big(Y_i - \widehat{m}_{h,-i}(\boldsymbol{X}_i) \Big)^2, \tag{3.9}$$

where

$$\widehat{m}_{h,-i}(\boldsymbol{X}_i) = \sum_{\ell \neq i} W_{h,-i}(\boldsymbol{X}_\ell) Y_\ell \Big/ \sum_{\ell \neq i} W_{h,-i}(\boldsymbol{X}_\ell), \tag{3.10}$$

and $W_{h,-i}$ denotes kernel weights computed without using observation i. Algorithm 3.1 outlines the leave-one-out CV procedure for selecting $\widehat{h}$.

Algorithm 3.1: Cross-validation for bandwidth selection.

Input: A set of candidate bandwidths $\{h_1, h_2, \ldots, h_k\}$ and a dataset $\{(\boldsymbol{X}_i, Y_i)\}_{i=1}^n$.

1: Iterate over each candidate bandwidth h:

1.1 Compute $W_{h,-i}$ for observation $i \in \{1, \ldots, n\}$, excluding case i, and calculate $\widehat{m}_{h,-i}(\boldsymbol{X}_i)$ using Equation (3.10).

1.2 Evaluate the CV(h) criterion utilizing Equation (3.9).

2: Select the optimal bandwidth $\widehat{h} = \arg\min_h \{\text{CV}(h)\}$.

Output: The optimal bandwidth $\widehat{h}$.

3.2 *MAVE algorithm for APLMs*

We now present the general MAVE procedure when combining a partially linear structure with a nonparametric component of the form $g(Z) = \gamma^\top b(z)$. In what follows, let $\boldsymbol{B} \in \mathrm{R}^{p\times d}$ denote the projection matrix we seek to estimate (subject to $\boldsymbol{B}^\top\boldsymbol{B} = I_d$), and let $\{(a_j, b_j)\}_{j=1}^n$ be local linear approximation parameters. The general MAVE procedure has the following steps:

Step 1 —State a local MAVE criterion

Given $\boldsymbol{B}$ (and a provisional estimate $\widehat{\gamma}$ for the nonparametric term), we estimate (a_j, b_j) for $j \in \{1, \ldots, n\}$ by solving the problem formulated as

$$\min_{\substack{\boldsymbol{B}:\boldsymbol{B}^\top\boldsymbol{B}=\boldsymbol{I}\\ a_j,b_j;j\in\{1,\ldots,n\}}} \left\{ \sum_{j=1}^n \sum_{i=1}^n \Big(Y_i - a_j - b_j^\top \boldsymbol{B}^\top(\boldsymbol{X}_i - \boldsymbol{X}_j) - \widehat{\gamma}^\top b(\boldsymbol{Z}_i)\Big)^2 W_{ij} \right\}. \tag{3.11}$$

Here, W_{ij} are kernel weights as defined in Equation (3.8) —or the relevant local weighting scheme.

Step 2 —Update each direction

Once $\{(\widehat{a}_j, \widehat{b}_j)\}_{j=1}^n$ are obtained, we update each column $\boldsymbol{b}_m$ of $\boldsymbol{B}$ (for $m \in \{1, \ldots, d\}$) by fixing all previously estimated directions $\widehat{\boldsymbol{b}}_1, \ldots, \widehat{\boldsymbol{b}}_{m-1}$ and solving the problem stablished as

$$\min_{\boldsymbol{b}_m;a_j,b_j;j\in\{1,\ldots,n\}} \left\{ \sum_{j=1}^n \sum_{i=1}^n \Big(Y_i - \widehat{a}_j - \widehat{b}_j^\top \big(\widehat{\boldsymbol{b}}_1, \ldots, \widehat{\boldsymbol{b}}_{m-1}, \boldsymbol{b}_m\big)^\top (\boldsymbol{X}_i - \boldsymbol{X}_j) - \widehat{\gamma}^\top b(\boldsymbol{Z}_i)\Big)^2 W_{ij} \right\}. \tag{3.12}$$

The iterative procedure for updating each column of the projection matrix $\boldsymbol{B}$ is outlined in Algorithm 3.2. At each iteration, previously estimated directions are fixed, and the next column is updated by minimizing the objective function defined in Equation (3.12).

Algorithm 3.2: Procedure for updating each column of the projection matrix $\boldsymbol{B}$.

Input: A dataset $\{(\boldsymbol{X}_i, \boldsymbol{Z}_i, Y_i)\}_{i=1}^n$, an initial projection matrix $\boldsymbol{B}^{(0)} \in \mathrm{R}^{p\times d}$, tolerance for convergence, and the dimension d.

1: Compute the nonparametric component $g(\boldsymbol{Z})$ using Equation (3.7), and obtain an initial estimate $\widehat{g}(\boldsymbol{Z}) = \widehat{\gamma}^\top b(\boldsymbol{Z})$.

2: Initialize $m = 1$. Choose an arbitrary initial projection matrix $\boldsymbol{B}^{(0)} \in \mathrm{R}^{p\times d}$ with orthonormal columns (for example, random or via SVD).

3: Perform local fitting. For each candidate $\boldsymbol{B}$, estimate (a_j, b_j) for $j \in \{1, \ldots, n\}$ by solving the minimization problem stated in Equation (3.11), subject to $\boldsymbol{B}^\top\boldsymbol{B} = \boldsymbol{I}$. Store the estimates $(\widehat{a}_j, \widehat{b}_j)$.

4: Update the m-th column of $\boldsymbol{B}$ by minimizing Equation (3.12) utilizing $(\widehat{a}_j, \widehat{b}_j)$. Denote the solution by $\widehat{\boldsymbol{b}}_m$.

5: Replace the m-th column of $\boldsymbol{B}$ with $\widehat{\boldsymbol{b}}_m$. Repeat the local fitting and column update (Steps 3–5) until convergence for this particular m.

6: Increment $m \leftarrow m+1$. If $m \le d$, return to Step 3 to update the next column. Otherwise, stop when all d directions are estimated, or use a suitable information criterion to determine d.

Output: The optimal projection matrix $\widehat{\boldsymbol{B}} \in \mathrm{R}^{p\times d}$.

In practice, Xia et al. (2014) suggested a multidimensional Nadaraya–Watson approach for calculating the weights W_{ij}. Namely, if $\widehat{\boldsymbol{B}}$ is the current estimate of the projection matrix (of size $p \times d$), then we have

$$W_{ij} = \frac{K_h\big(\widehat{\boldsymbol{B}}^\top(\boldsymbol{X}_j - \boldsymbol{X}_i)\big)}{\sum_{\ell=1}^{n} K_h\big(\widehat{\boldsymbol{B}}^\top(\boldsymbol{X}_\ell - \boldsymbol{X}_i)\big)},$$

where K_h is a smoothing kernel with bandwidth h. Although MAVE efficiently reduces dimensionality, the resulting predictors remain linear combinations of the original covariates. To enhance model parsimony and achieve simultaneous variable selection, one can incorporate penalization techniques such as ALASSO or SCAD.

4. MAVE with adaptive penalization: ALASSO and SCAD

In this section, we introduce the use of adaptive penalization within the MAVE framework to achieve both dimension reduction and variable selection. Specifically, we discuss the ALASSO and SCAD, highlighting their integration with MAVE and their advantages in constructing sparse, interpretable models.

4.1 *MAVE-ALASSO method*

ALASSO was originally introduced by Zou (2006) to mitigate certain drawbacks of the standard LASSO, notably its bias toward larger coefficients and occasional inconsistency in variable selection. In the MAVE framework, Hua et al. (2012) further developed an ALASSO variant by adding adaptive weights to the penalization term, thereby improving variable selection while preserving the efficiency of MAVE. Specifically, ALASSO assigns distinct weights to each coefficient, overcoming the limitations of ordinary LASSO, which uses a uniform penalty. As a result, ALASSO estimators can satisfy the so-called oracle property, yielding consistent and unbiased estimates (Leng et al., 2009; Hua et al., 2012).

Formally, consider the penalized objective function given by

$$\min_{\boldsymbol{B}}\left\{\sum_{j=1}^{n}\sum_{i=1}^{n}\Big(Y_i - a_j + b_j^\top \boldsymbol{B}^\top(\boldsymbol{X}_i - \boldsymbol{X}_j) + \gamma^\top b(\boldsymbol{Z}_i)\Big)^2 W_{ij} + \lambda \sum_{m=1}^{d}\sum_{t=1}^{p} w_t |\boldsymbol{B}_{t,m}|\right\}, \quad (4.13)$$

subject to $\boldsymbol{B}^\top\boldsymbol{B} = \boldsymbol{I}_d$, where:

- $\lambda \geq 0$ is the penalty parameter controlling the shrinkage,
- $w_t = 1/|\widehat{\beta}_{\mathrm{OLS},t}|^\theta$ are adaptive weights computed from an initial estimator —often ordinary least squares (OLS)— for some $\theta > 0$,
- p is the number of linear covariates,
- $\{(a_j, b_j)\}_{j=1}^n$ are the local linear approximation parameters for MAVE,
- $\gamma^\top b(\boldsymbol{Z}_i)$ represents additional nonparametric spline components if needed,
- n is the sample size, d is the dimension of the reduction space,
- $\boldsymbol{B} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d)$ is the projection matrix, and $\boldsymbol{B}_{t,m}$ is the t-th entry of the m-th column $\boldsymbol{b}_m$.

The first term in Equation (4.13) is the MAVE least-squares criterion, while the second term is the ALASSO penalty that simultaneously shrinks and selects coefficients. Under suitable regularity conditions, this adaptive penalty can yield consistent variable selection with minimal bias for large coefficients (Hmood and Saleh, 2016).

Following the general MAVE framework, we now incorporate an adaptive LASSO penalty. Recall that $\boldsymbol{B} \in \mathrm{R}^{p\times d}$ is our projection matrix (subject to $\boldsymbol{B}^\top\boldsymbol{B} = \boldsymbol{I}_d$), and let $\{(a_j, b_j)\}_{j=1}^n$ be local linear parameters as before. Suppose the nonparametric component of the APLM stated in Equation (3.7). For a given $\boldsymbol{B}$ (and a provisional estimate $\widehat{\gamma}$), the local MAVE criterion stated in Equation (3.11) can be written as

$$\min_{\substack{\boldsymbol{B}:\boldsymbol{B}^\top\boldsymbol{B}=\boldsymbol{I}_d \\ a_j,b_j;\, j\in\{1,\ldots,n\}}} \left\{ \sum_{j=1}^{n}\sum_{i=1}^{n} \Big(Y_i - a_j - b_j^\top \boldsymbol{B}^\top(\boldsymbol{X}_i - \boldsymbol{X}_j) - \widehat{\gamma}^\top b(\boldsymbol{Z}_i)\Big)^2 W_{ij} \right\}. \tag{4.14}$$

Once $(\widehat{a}_j, \widehat{b}_j)$ are obtained, we penalize the projection coefficients in $\boldsymbol{B}$ via an adaptive LASSO term stated in Equation (4.13). Specifically, for the m-th column $\boldsymbol{b}_m$ of $\boldsymbol{B}$, we have

$$\min_{\substack{\boldsymbol{B}:\boldsymbol{B}^\top\boldsymbol{B}=\boldsymbol{I}_d \\ a_j,b_j;\, j\in\{1,\ldots,n\}}} \left\{ \sum_{j=1}^{n}\sum_{i=1}^{n} \Big(Y_i - \widehat{a}_j - \widehat{b}_j^\top (\widehat{\boldsymbol{b}}_1, \ldots, \widehat{\boldsymbol{b}}_{m-1}, \boldsymbol{b}_m)^\top (\boldsymbol{X}_i - \boldsymbol{X}_j) - \widehat{\gamma}^\top b(\boldsymbol{Z}_i)\Big)^2 W_{ij} + \lambda \sum_{t=1}^{p} w_t |\boldsymbol{b}_{m,t}| \right\}, \tag{4.15}$$

where $\boldsymbol{b}_{m,t}$ is the t-th component of the m-th column of $\boldsymbol{B}$, $\lambda \geq 0$ is the penalty parameter, and w_t are the adaptive weights.

Algorithm 4.1 details the implementation of MAVE-ALASSO. This procedure incorporates adaptive penalization into the MAVE framework, iteratively refining the projection matrix $\boldsymbol{B}$ while balancing dimension reduction and variable selection through the use of adaptive weights.

Algorithm 4.1: Implementation of the MAVE-ALASSO method.

Input: A dataset $\{(\boldsymbol{X}_i, \boldsymbol{Z}_i, Y_i)\}_{i=1}^n$, an initial projection matrix $\boldsymbol{B}^{(0)} \in \mathrm{R}^{p\times d}$, tolerance for convergence, and the dimension d.

1: Compute the nonparametric component $g(\boldsymbol{Z})$ as in Equation (3.7), and obtain an initial estimate $\widehat{g}(\boldsymbol{Z}) = \widehat{\gamma}^\top b(\boldsymbol{Z})$.

2: Initialize $m = 1$. Choose an arbitrary initial projection matrix $\boldsymbol{B}^{(0)} \in \mathrm{R}^{p\times d}$ with orthonormal columns (for example, random or via SVD).

3: Perform the local MAVE step (unpenalized). For the current $\boldsymbol{B}$, estimate (a_j, b_j) for $j \in \{1, \ldots, n\}$ by solving the unpenalized MAVE problem stated in Equation (4.14). Store the estimates $(\widehat{a}_j, \widehat{b}_j)$.

4: Update the m-th column of $\boldsymbol{B}$ by minimizing the adaptive LASSO objective presented in Equation (4.15) using $(\widehat{a}_j, \widehat{b}_j)$. Denote the solution by $\widehat{\boldsymbol{b}}_{m,\mathrm{ALASSO}}$.

5: Replace the m-th column of $\boldsymbol{B}$ with $\widehat{\boldsymbol{b}}_{m,\mathrm{ALASSO}}$. Repeat the local MAVE step and the ALASSO update (Steps 3–5) until convergence for this particular m.

6: Increment $m \leftarrow m+1$. If $m \leq d$, return to Step 3 to update the next column. Otherwise, stop when all d directions are estimated, or use a suitable information criterion to determine d.

Output: The optimal projection matrix $\widehat{\boldsymbol{B}} \in \mathrm{R}^{p\times d}$.

Remark 1

- In Step 3 of Algorithm 4.1, we solve Equation (4.14) without the penalty term to obtain local linear fits $(\widehat{a}_j, \widehat{b}_j)$.
- In Step 4 of Algorithm 4.1, we impose the ALASSO penalty formulated in Equation (4.15) for the m-th direction. We typically cycle through $m \in \{1, \ldots, d\}$ repeatedly until convergence.
- The weight w_t can be derived from a pilot estimator such as OLS (with exponent $\theta > 0$), as in Zou (2006).

4.2 *MAVE-SCAD method*

The MAVE framework is combined with the SCAD penalty to achieve both accurate estimation and sparse variable selection (Fan and Li, 2001; Hmood and Saleh, 2016; Ahmed and Hmood, 2021; Huda and Hmood, 2021). Specifically, let $\boldsymbol{B} \in \mathrm{R}^{p\times d}$ be the projection matrix (with $\boldsymbol{B}^\top\boldsymbol{B} = \boldsymbol{I}_d$), and let $\{(a_j, b_j)\}_{j=1}^n$ be the local MAVE parameters as before. Then, the penalized objective function becomes given by

$$\min_{\boldsymbol{B}:\boldsymbol{B}^\top\boldsymbol{B}=\boldsymbol{I}_d}\left\{\sum_{j=1}^{n}\sum_{i=1}^{n}\Big(Y_i-a_j-b_j^\top\boldsymbol{B}^\top(\boldsymbol{X}_i-\boldsymbol{X}_j)-\gamma^\top b(\boldsymbol{Z}_i)\Big)^2 W_{ij}+n\sum_{m=1}^{d}\sum_{t=1}^{p}p_{\mathrm{SCAD},\lambda,a}\Big(|\boldsymbol{B}_{t,m}|\Big)\right\}, \tag{4.16}$$

where:

- $\lambda \geq 0$ is the regularization parameter,
- $a > 1$ is the SCAD tuning constant (sometimes denoted by γ in the literature),
- $p_{\mathrm{SCAD},\lambda,a}$ is the SCAD penalty function given in Equation (4.17),
- p is the total number of linear covariates, and d is the dimension of the reduction space,
- $\boldsymbol{B}_{t,m}$ denotes the t-th entry in the m-th column of $\boldsymbol{B}$,
- (a_j, b_j) are the local linear parameters in MAVE, and $\gamma^\top b(\boldsymbol{Z}_i)$ is a possible spline-based nonparametric component.

The first term in Equation (4.16) is the usual MAVE least-squares part, while the second term introduces the SCAD penalty scaled by n. This scaling helps to control model complexity while preserving larger coefficients, thanks to the reduced shrinkage imposed by the SCAD form for values exceeding $a\lambda$.

Fan and Li (2001) proposed the SCAD penalty stated as

$$p_{\mathrm{SCAD},\lambda,a}(|x|) = \begin{cases} \lambda|x|, & \text{if } 0 \leq |x| < \lambda; \\ \dfrac{(a^2-1)\lambda^2 - (|x|-a\lambda)^2}{2(a-1)}, & \text{if } \lambda \leq |x| < a\lambda; \\ \dfrac{(a+1)\lambda^2}{2}, & \text{if } |x| \geq a\lambda. \end{cases} \tag{4.17}$$

SCAD achieves oracle-like properties by penalizing small coefficients heavily while leaving large coefficients almost unshrunk once $|x| > a\lambda$. This facilitates both consistent variable selection and reduced bias for relevant covariates (Huang et al., 2008; Huda and Hmood, 2021).

Following the same rationale as MAVE-ALASSO, we now incorporate the SCAD penalty into the MAVE framework to enable sparse estimation of the projection matrix $\boldsymbol{B} \in \mathrm{R}^{p\times d}$. Let $\{(a_j, b_j)\}_{j=1}^n$ be local linear parameters. Suppose again that a nonparametric component is given in Equation (3.7).

As in previously stated, for a given $\boldsymbol{B}$ (and estimate $\widehat{\gamma}$ of the nonparametric term), we solve the unpenalized local MAVE objective, given by

$$\min_{\substack{\boldsymbol{B}:\boldsymbol{B}^\top\boldsymbol{B}=\boldsymbol{I}_d \\ a_j,b_j;j\in\{1,\ldots,n\}}}\sum_{j=1}^{n}\sum_{i=1}^{n}\Big(Y_i - a_j - b_j^\top\boldsymbol{B}^\top(\boldsymbol{X}_i-\boldsymbol{X}_j) - \widehat{\gamma}^\top b(\boldsymbol{Z}_i)\Big)^2 W_{ij}. \tag{4.18}$$

Once $(\widehat{a}_j, \widehat{b}_j)$ are obtained, we introduce the SCAD penalty term presented in Equation (4.16) for the m-th column $\boldsymbol{b}_m$ of $\boldsymbol{B}$. Specifically, we have

$$\min_{\substack{\boldsymbol{B}:\boldsymbol{B}^\top\boldsymbol{B}=\boldsymbol{I}_d \\ a_j,b_j;j\in\{1,\ldots,n\}}} \left\{ \sum_{j=1}^{n}\sum_{i=1}^{n}\Big(Y_i - \widehat{a}_j - \widehat{b}_j^\top(\widehat{\boldsymbol{b}}_1,\ldots,\widehat{\boldsymbol{b}}_{m-1},\boldsymbol{b}_m)^\top(\boldsymbol{X}_i-\boldsymbol{X}_j) - \widehat{\gamma}^\top b(\boldsymbol{Z}_i)\Big)^2 W_{ij} \right. \tag{4.19}$$

$$\left. + n\sum_{t=1}^{p} p_{\mathrm{SCAD},\lambda,a}(|\boldsymbol{b}_{m,t}|) \right\},$$

where $\boldsymbol{b}_{m,t}$ is the t-th entry of the m-th column $\boldsymbol{b}_m$ and $p_{\mathrm{SCAD},\lambda,a}$ is the SCAD penalty defined in Equation (4.17).

The steps for integrating the SCAD penalty into the MAVE framework are provided in Algorithm 4.2. This integration achieves sparse estimation of the projection matrix $\boldsymbol{B}$ by iteratively applying penalized regression techniques as stated in Equation (4.17).

Algorithm 4.2: Integration of SCAD penalty into the MAVE method.

Input: A dataset $\{(\boldsymbol{X}_i, \boldsymbol{Z}_i, Y_i)\}_{i=1}^n$, an initial projection matrix $\boldsymbol{B}^{(0)} \in \mathrm{R}^{p\times d}$, tolerance for convergence, and the dimension d.

1: Compute the nonparametric component $g(\boldsymbol{Z})$ defined in Equation (3.7), and obtain an initial estimate $\widehat{g}(\boldsymbol{Z}) = \widehat{\gamma}^\top b(\boldsymbol{Z})$.

2: Initialize $m = 1$. Choose an arbitrary initial projection matrix $\boldsymbol{B}^{(0)} \in \mathrm{R}^{p\times d}$ with orthonormal columns (for example, random or via SVD).

3: Perform the local MAVE step (unpenalized). For the current $\boldsymbol{B}$, estimate (a_j, b_j) for $j \in \{1,\ldots,n\}$ by solving the unpenalized MAVE problem stated in Equation (4.18). Store the estimates $(\widehat{a}_j, \widehat{b}_j)$.

4: Update the m-th column of $\boldsymbol{B}$ by minimizing the SCAD-penalized objective presented in Equation (4.19) using $(\widehat{a}_j, \widehat{b}_j)$. Denote the solution by $\widehat{\boldsymbol{b}}_{m,\mathrm{SCAD}}$.

5: Replace the m-th column of $\boldsymbol{B}$ with $\widehat{\boldsymbol{b}}_{m,\mathrm{SCAD}}$. Repeat the local MAVE step and the SCAD update (Steps 3–5) until convergence for this particular m.

6: Increment $m \leftarrow m+1$. If $m < d$, return to Step 3 to update the next column. Otherwise, stop when all d directions are estimated, or use a criterion to select d.

Output: The optimal projection matrix $\widehat{\boldsymbol{B}} \in \mathrm{R}^{p\times d}$.

Remark 2

- The choice of (λ, a) strongly influences sparsity and shrinkage; these can be selected by CV or a model selection criterion.
- The factor n multiplying the SCAD penalty stated in Equation (4.19) is a scaling convention ensuring that the penalty remains comparable to the sum of squared residuals.
- SCAD exerts less shrinkage on large coefficients, mitigating the bias inherent in standard LASSO methods while still performing effective variable selection.

4.3 *Performance evaluation criteria*

Several criteria can be used to assess the performance of regression function estimators in APLMs. Following Li et al. (2018), we adopt the following three measures:

(i) The mean average squared error (MASE) is defined for an estimator $\widehat{Y}_i$ of Y_i as $\mathrm{MASE} = (1/n)\mathrm{E}(\sum_{i=1}^n (Y_i - \widehat{Y}_i)^2)$. In practice, this is typically approximated by the sample mean of the squared residuals.

(ii) Criterion C represents the number of truly zero coefficients that are correctly identified as zero. A higher value of C indicates better detection of irrelevant variables, leading to more accurate sparse estimation (Horowitz, 2015).
(iii) Criterion I quantifies the number of truly nonzero coefficients that are incorrectly estimated as zero (false zeros). A smaller value of I reflects fewer omitted relevant variables, reducing the rate of type-II errors in variable selection (Horowitz, 2015; Cook and Forzani, 2018).

5. Simulation study

In this section, we evaluate the finite-sample performance of the proposed methods, MAVE-ALASSO and MAVE-SCAD, through an extensive Monte Carlo simulation study. The results highlight the trade-offs between sparsity and prediction accuracy under varying conditions.

5.1 *Simulation setup*

We investigate the finite-sample performance of our proposed methods via Monte Carlo simulations. All computations were carried out in the R environment using the functions `mave`, `bs`, `cv.glmnet`, and `ncvreg` (R Core Team, 2024). The simulations were performed on an Intel(R) Core(TM) i5-6200U CPU, 64-bit operating system.

We consider the APLM from Equation (2.1) with $k = 2$ nonlinear components, where the two nonparametric functions are given by $g_1(\boldsymbol{Z}_1) = 8\cos(7\pi \boldsymbol{Z}_1)$, $g_2(\boldsymbol{Z}_2) = 30(\exp(-4.75\boldsymbol{Z}_2) - 2\exp(-8.1\boldsymbol{Z}_2) + 5\exp(-3.25\boldsymbol{Z}_2))$. We assume there are $p = 60$ linear covariates, plus two nonlinear variables $(\boldsymbol{Z}_1, \boldsymbol{Z}_2)$. The unknown parameter vector is $\boldsymbol{\beta} = (2, 1.2, 0.8, -0.9, 1.7, 3, 5.1, -6.1, 4.2, -2.1, 0, \ldots, 0, 1.3, -1.7, 1.9, 3.2, -3.1)$, where the remaining coefficients (beyond those explicitly listed) are set to zero to reflect true sparsity. The error term ε is independently drawn from a standard normal distribution with variance σ^2, where $\sigma \in \{1, 2, 6\}$. The covariates $\boldsymbol{X}$ are sampled from a multivariate normal distribution with correlation $\rho^{|i-j|}$, for $\rho \in \{0.3, 0.5, 0.8\}$. Meanwhile, $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are drawn independently from a uniform distribution on $[0, 1]$. Cubic B-splines are used to approximate g_1 and g_2, varying the number of internal knots from 2 to 10.

5.2 *Design factors and summary of findings*

We replicate each scenario 1000 times, considering sample sizes $n \in \{20, 50, 100, 150\}$. We then estimate the model using both MAVE-ALASSO and MAVE-SCAD, evaluating their performance via MASE, C, and I. Tables 1–9 report these results across different values of $\rho \in \{0.3, 0.5, 0.8\}$ and $\sigma \in \{1, 2, 6\}$.

Next, we summarize our simulation findings:

- Sample size effect —As n increases, both MAVE-ALASSO and MAVE-SCAD exhibit improvements in MASE (lower errors), higher C (better identification of zero coefficients), and lower I (fewer falsely omitted variables). For small $n = 20$, the results can be more variable.
- Comparison of MAVE-ALASSO versus MAVE-SCAD.
  - Overall, MAVE-SCAD tends to yield smaller MASE and lower I (fewer false zeros) for larger sample sizes and higher dimensionality/nonlinearity.
  - Conversely, MAVE-ALASSO often attains slightly larger C, indicating stronger detection of truly zero coefficients, which can be advantageous with smaller sample sizes or when variable selection is paramount.

- Influence of ρ and σ —Higher correlation ($\rho = 0.8$) and larger noise levels ($\sigma = 6$) naturally increase estimation challenges. Nonetheless, both methods remain reasonably robust, with MAVE-SCAD often preserving accuracy due to its less-aggressive shrinkage on large coefficients.

Hence, MAVE-SCAD appears preferable for large-scale problems or highly complex models, thanks to its balance between sparsity and reduced shrinkage bias. In addition, MAVE-ALASSO may be more appealing for small or moderate sample sizes, as it can more reliably retain zero coefficients at the expense of a slight increase in estimation bias for nonzero coefficients.

Table 1. MASE, C and I with $n = 20, 50, 100, 150$, $\rho = 0.3$, $\sigma = 1$ and $p = 60$.

n	Methods	C	I	MASE	Best method
20	MAVE-ALASSO	40.883	6.265	3.6888	MAVE-SCAD
	MAVE-SCAD	42.768	5.171	3.3863	
50	MAVE-ALASSO	41.292	5.695	3.3535	MAVE-SCAD
	MAVE-SCAD	43.196	4.701	3.0785	
100	MAVE-ALASSO	41.705	5.178	3.0486	MAVE-SCAD
	MAVE-SCAD	43.628	4.274	2.7986	
150	MAVE-ALASSO	42.122	4.707	2.7715	MAVE-SCAD
	MAVE-SCAD	44.064	3.885	2.5442	

Table 2. MASE, C and I with $n = 20, 50, 100, 150$, $\rho = 0.5$, $\sigma = 1$ and $p = 60$.

n	Methods	C	I	MASE	Best method
20	MAVE-ALASSO	41.688	5.472	3.2705	MAVE-ALASSO
	MAVE-SCAD	41.077	5.968	3.3521	
50	MAVE-ALASSO	42.240	3.862	2.2117	MAVE-ALASSO
	MAVE-SCAD	41.636	4.424	2.2668	
100	MAVE-ALASSO	42.049	3.128	2.1150	MAVE-SCAD
	MAVE-SCAD	42.660	2.731	2.0635	
150	MAVE-ALASSO	42.171	2.043	1.2815	MAVE-SCAD
	MAVE-SCAD	44.064	3.885	2.5442	

Table 3. MASE, C and I with $n = 20, 50, 100, 150$, $\rho = 0.8$, $\sigma = 1$ and $p = 60$.

n	Methods	C	I	MASE	Best method
20	MAVE-ALASSO	37.356	6.511	3.6893	MAVE-ALASSO
	MAVE-SCAD	37.216	6.659	3.7402	
50	MAVE-ALASSO	38.119	5.861	3.3203	MAVE-ALASSO
	MAVE-SCAD	37.975	5.993	3.3662	
100	MAVE-ALASSO	38.504	4.689	2.6563	MAVE-SCAD
	MAVE-SCAD	38.358	4.494	2.5246	
150	MAVE-ALASSO	38.542	3.516	1.9921	MAVE-ALASSO
	MAVE-SCAD	38.746	3.371	1.8935	

Table 4. MASE, C and I with $n = 20, 50, 100, 150$, $\rho = 0.3$, $\sigma = 2$ and $p = 60$.

n	Methods	C	I	MASE	Best method
20	MAVE-ALASSO	40.731	5.223	3.4202	MAVE-ALASSO
	MAVE-SCAD	38.936	6.328	3.7257	
50	MAVE-ALASSO	41.139	4.748	3.1092	MAVE-ALASSO
	MAVE-SCAD	39.326	5.752	3.3870	
100	MAVE-ALASSO	41.550	4.316	2.8266	MAVE-ALASSO
	MAVE-SCAD	39.719	5.229	3.0791	
150	MAVE-ALASSO	41.966	3.924	2.5696	MAVE-ALASSO
	MAVE-SCAD	40.116	4.754	2.7992	

Table 5. MASE, C and I with $n = 20, 50, 100, 150$, $\rho = 0.5$, $\sigma = 2$ and $p = 60$.

n	Methods	C	I	MASE	Best method
20	MAVE-ALASSO	41.146	5.544	3.3136	MAVE-ALASSO
	MAVE-SCAD	40.543	6.047	3.3963	
50	MAVE-ALASSO	41.691	3.912	2.2407	MAVE-ALASSO
	MAVE-SCAD	41.095	4.482	2.2967	
100	MAVE-ALASSO	41.503	3.169	2.1429	MAVE-SCAD
	MAVE-SCAD	42.106	2.766	2.0907	
150	MAVE-ALASSO	41.623	2.070	1.2984	MAVE-SCAD
	MAVE-SCAD	42.213	1.807	1.2666	

Table 6. MASE, C and I with $n = 20, 50, 100, 150$, $\rho = 0.8$, $\sigma = 2$ and $p = 60$.

n	Methods	C	I	MASE	Best method
20	MAVE-ALASSO	36.609	6.645	4.0541	MAVE-ALASSO
	MAVE-SCAD	36.471	6.794	4.1101	
50	MAVE-ALASSO	37.356	5.980	3.6487	MAVE-ALASSO
	MAVE-SCAD	37.216	6.115	3.6991	
100	MAVE-ALASSO	37.734	4.784	2.9190	MAVE-SCAD
	MAVE-SCAD	37.591	4.587	2.7743	
150	MAVE-ALASSO	37.771	3.588	2.1893	MAVE-SCAD
	MAVE-SCAD	37.971	3.440	2.0807	

Table 7. MASE, C and I with $n = 20, 50, 100, 150$, $\rho = 0.3$, $\sigma = 6$ and $p = 60$.

n	Methods	C	I	MASE	Best method
20	MAVE-ALASSO	39.970	5.430	3.5556	MAVE-ALASSO
	MAVE-SCAD	38.208	6.578	3.8732	
50	MAVE-ALASSO	40.370	4.936	3.2324	MAVE-ALASSO
	MAVE-SCAD	38.590	5.980	3.5211	
100	MAVE-ALASSO	40.773	4.487	2.9385	MAVE-ALASSO
	MAVE-SCAD	38.976	5.437	3.2010	
150	MAVE-ALASSO	41.181	4.079	2.6714	MAVE-ALASSO
	MAVE-SCAD	39.366	4.942	2.9100	

Table 8. MASE, C and I with $n = 20, 50, 100, 150$, $\rho = 0.5$, $\sigma = 6$ and $p = 60$.

n	Methods	C	I	MASE	Best method
20	MAVE-ALASSO	40.611	5.991	3.3572	MAVE-ALASSO
	MAVE-SCAD	40.017	6.126	3.4409	
50	MAVE-ALASSO	41.149	3.964	2.2703	MAVE-ALASSO
	MAVE-SCAD	40.560	4.541	2.3269	
100	MAVE-ALASSO	40.963	3.210	2.1711	MAVE-SCAD
	MAVE-SCAD	41.558	2.803	2.1181	
150	MAVE-ALASSO	41.082	2.097	1.3155	MAVE-SCAD
	MAVE-SCAD	41.664	1.830	1.2834	

Table 9. MASE, C and I with $n = 20, 50, 100, 150$, $\rho = 0.8$, $\sigma = 6$ and $p = 60$.

n	Methods	C	I	MASE	Best method
20	MAVE-ALASSO	35.877	6.780	4.4550	MAVE-ALASSO
	MAVE-SCAD	35.742	6.934	4.5165	
50	MAVE-ALASSO	36.609	6.102	4.0095	MAVE-ALASSO
	MAVE-SCAD	36.471	6.240	4.0649	
100	MAVE-ALASSO	36.979	4.881	3.2077	MAVE-SCAD
	MAVE-SCAD	36.840	4.680	3.0487	
150	MAVE-ALASSO	37.016	3.662	2.4057	MAVE-SCAD
	MAVE-SCAD	37.212	3.510	2.2865	

6. Application with real-world data

This section provides an application with real-world data.

6.1 *Description of the data*

The dataset used for this study corresponds to financial data from the Baghdad Soft Drinks Company, a commercial enterprise listed on the Iraqi Stock Exchange. It spans the period from 2015 to 2022, with four observations (quarters) per year, totaling 32 data points. This dataset provides a unique opportunity to evaluate the performance of MAVE-ALASSO and MAVE-SCAD in a real-world context characterized by high-dimensionality and potential nonlinear relationships.

The covariates include 45 linear predictors and 2 nonlinear predictors, capturing diverse aspects of financial performance and operational characteristics. These variables are detailed next, which provides their descriptions and roles in the analysis. The response variable, Y, represents the company stock market value.

Both MAVE-ALASSO and MAVE-SCAD were implemented in the statistical software R using the procedures and algorithms outlined in Sections 3 and 4. Specifically, the iterative algorithms described in Algorithm 4.1 (for MAVE-ALASSO) and Algorithm 4.2 (for MAVE-SCAD) were employed to estimate the projection matrix $\boldsymbol{B}$ and identify the most relevant covariates. By integrating adaptive penalization with SDR, these methods balance interpretability and accuracy, making them well-suited for high-dimensional data.

Next, we describe the variables included in the real-data application:

- Y — stock market value (response)
- Z_1 — dividend distribution
- Z_2 — operating cash flow
- X_1 — financial leverage
- X_2 — return on assets
- X_3 — sales growth rate
- X_4 — annual growth rate
- X_5 — company size
- X_6 — liquidity ratios
- X_7 — investment spending
- X_8 — issuance of new stocks
- X_9 — net working capital
- X_{10} — accounts receivable turnover
- X_{11} — average collection period
- X_{12} — inventory turnover
- X_{13} — average age of inventory
- X_{14} — operating cycle
- X_{15} — payable turnover
- X_{16} — payable turnover in days
- X_{17} — fixed asset turnover
- X_{18} — total asset turnover
- X_{19} — debt ratio
- X_{20} — leverage ratio
- X_{21} — debt/equity ratio
- X_{22} — times interest earned
- X_{23} — return on common equity
- X_{24} — return on total assets
- X_{25} — return on investment
- X_{26} — gross profit margin
- X_{27} — profit margin
- X_{28} — earnings per stock
- X_{29} — price/earnings ratio
- X_{30} — dividend yield
- X_{31} — dividend payout ratio
- X_{32} — book value of stock
- X_{33} — indebtedness ratio
- X_{34} — ratio of total liabilities to total assets
- X_{35} — ratio of total assets to equity
- X_{36} — ratio of liabilities to equity
- X_{37} — rate of interest coverage
- X_{38} — return on equity
- X_{39} — return on invested money
- X_{40} — total profit to sales
- X_{41} — earnings per share
- X_{42} — profitable repeater
- X_{43} — cash earnings per share
- X_{44} — cash distribution ratio
- X_{45} — book value per share

6.2 *Results of variable selection*

Table 10 summarizes the variable selection process using MAVE-ALASSO and MAVE-SCAD methods. This table includes estimates for all variables analyzed, offering additional numeric details compared to the results presented in the main text.

Table 10. Results of variable selection for MAVE-ALASSO and MAVE-SCAD methods with real-world data.

j	MAVE-ALASSO	MAVE-SCAD	j	MAVE-ALASSO	MAVE-SCAD
	β_j	β_j		β_j	β_j
1	1.0242	1.4106	2	0.1392	0.5341
3	Removed	Removed	4	Removed	0.5026
5	0.4344	0.3495	6	0.4787	0.8699
7	Removed	Removed	8	0.1728	0.2756
9	Removed	Removed	10	-0.6551	-0.8308
11	0.3801	0.2966	12	Removed	Removed
13	0.5753	0.7112	14	-0.4036	-0.1944
15	Removed	Removed	16	0.876	0.6237
17	Removed	-1.2874	18	-0.4936	-0.6638
19	Removed	Removed	20	Removed	0.5516
21	0.2965	0.4072	22	0.7665	0.9136
23	-0.2759	-0.4491	24	Removed	Removed
25	-0.305	-0.3576	26	Removed	Removed
27	Removed	Removed	28	3.5736	2.9314
29	Removed	Removed	30	Removed	Removed
31	Removed	Removed	32	Removed	Removed
33	Removed	Removed	34	Removed	Removed
35	Removed	Removed	36	Removed	Removed
37	Removed	Removed	38	Removed	Removed
39	Removed	1.1302	40	Removed	Removed
41	Removed	-1.6994	42	Removed	Removed
43	Removed	Removed	44	Removed	Removed
45	Removed	Removed	46	-0.5728	Removed
47	0.649	0.7817			

The results of the variable selection process are summarized in Table 11, which lists the variables retained ($\beta_j \neq 0$) and removed by each method. These results are based on the APLM framework described in Section 2. MAVE-ALASSO and MAVE-SCAD were both effective in identifying a subset of predictors that explain the variability in Y, with some differences in the number and type of variables selected.

Remark 3 "Retained" indicates that the corresponding coefficient was nonzero; "Removed" indicates a zero (or negligible) estimate. Variables with indices j beyond the original range may represent nonlinear components or interactions derived from the APLM framework. For a comprehensive list of all variables and their estimates from MAVE-ALASSO and MAVE-SCAD, refer to Table 10.

Table 11. Variables selected by MAVE-ALASSO and MAVE-SCAD for modeling Y with real-world data.

j	MAVE-ALASSO		j	MAVE-SCAD	
	β_j	Status		β_j	Status
1	1.0242	Retained	2	0.5341	Retained
3	–	Removed	4	0.5026	Retained
5	0.4344	Retained	6	0.8699	Retained
7	–	Removed	8	0.2756	Retained
10	-0.6551	Retained	11	0.2966	Retained
13	0.5753	Retained	14	-0.1944	Retained
16	0.876	Retained	17	-1.2874	Retained
18	-0.4936	Retained	20	0.5516	Retained
21	0.2965	Retained	22	0.9136	Retained
23	-0.2759	Retained	25	-0.3576	Retained
28	3.5736	Retained	39	1.1302	Retained
41	–	Removed	46	-0.5728	Retained
47	0.649	Retained	41	-1.6994	Retained

6.3 *Comparison of methods*

Next, we examine the performance of MAVE-ALASSO and MAVE-SCAD through metrics such as MSE, percent error (PE), and the coefficient of determination (R^2), highlighting the trade-offs between sparsity and predictive accuracy.

Table 12 summarizes the results for the real-data application. While both methods achieve high R^2 values (indicating that the selected variables explain a substantial portion of the variability in Y), MAVE-ALASSO retains fewer variables, resulting in a more parsimonious model. In contrast, MAVE-SCAD selects a slightly larger set of predictors but offers similar predictive performance.

Table 12. Comparison of MAVE-ALASSO and MAVE-SCAD with real-world data.

Method	MSE	PE (%)	R^2
MAVE-ALASSO	0.0808	2.5854	91.66%
MAVE-SCAD	0.0825	2.6392	91.49%

6.4 *Discussion*

Next, we discuss some key aspects identified in our real-data application:

- Overall fit —Both methods demonstrate high predictive accuracy, with R^2 values exceeding 91%. This suggests that the APLM framework, combined with adaptive penalization, is effective for capturing key patterns in the data.
- Sparsity and interpretability —MAVE-ALASSO achieves greater sparsity by selecting 18 variables, which may be preferable when model interpretability and parsimony are critical.
- Comprehensive selection —MAVE-SCAD retains 22 variables, offering a comprehensive set of predictors. This may be advantageous when the goal is to ensure that potentially important variables are not omitted, albeit at the cost of a slightly larger model.

In summary, MAVE-ALASSO is advantageous for applications requiring a more concise and interpretable model, while MAVE-SCAD provides a broader perspective on potential predictors with comparable accuracy. The choice between the two methods ultimately depends on the specific priorities of the analysis, such as parsimony versus coverage of relevant variables.

7. Conclusions

The findings from our simulation study demonstrate that the minimum average variance estimation with adaptive least absolute shrinkage and selection operator (MAVE-ALASSO) exhibits a clear advantage for smaller or moderate sample sizes, particularly under conditions of low to moderate correlation and variance. Conversely, for larger sample sizes and scenarios where correlations are intermediate to high (such as $\rho = 0.5$ or 0.8), the minimum average variance estimation with smoothly clipped absolute deviation (MAVE-SCAD) generally outperforms MAVE-ALASSO. These results highlight the complementary strengths of the two methods and their applicability under different settings.

The real-data application, based on financial data from the Baghdad Soft Drinks Company, further illustrates the utility of these methods in a practical context. In this case, MAVE-ALASSO provided a marginally better fit than MAVE-SCAD, achieving slightly lower mean squared error and percent error, as well as a higher coefficient of determination. Several key predictors of market value were identified, including variables related to dividend distribution, operating cash flow, sales and annual growth rates, liquidity ratios, net working capital, and several financial leverage and profitability measures. These findings underscore the potential of MAVE-ALASSO to yield parsimonious and interpretable models that can inform managerial decision-making and financial forecasting.

Despite these promising results, the study is not without limitations. The conclusions drawn from the simulations are based on specific settings, including sample sizes, correlation structures, and error variances. While these scenarios were designed to represent a range of practical situations, caution is needed when generalizing to other contexts.

Similarly, the real-data application is limited to a single company operating in a specific financial market, and different data characteristics in other contexts may influence model performance. Additionally, the results may be sensitive to methodological choices such as bandwidth selection, spline basis construction, and tuning parameter values for ALASSO and SCAD. Future work could explore more refined or adaptive strategies for selecting these parameters to further enhance the robustness of the methods.

Building on the insights gained from this work, several directions for future research are apparent. Extending the use of MAVE-ALASSO and MAVE-SCAD to other semiparametric or high-dimensional regression models, such as partially linear single-index models or generalized additive models, could provide additional applications.

Developing robust variants of MAVE that account for outliers or heavy-tailed error distributions would further broaden its utility in practical settings. Moreover, incorporating advanced regularization techniques, such as MCP or elastic net, into the MAVE framework could enhance its capacity for variable selection in challenging datasets. Applying these methods to more complex data structures, such as multiresponse or longitudinal data, represents a promising avenue for extending the theoretical and practical contributions of MAVE-ALASSO and MAVE-SCAD.

In conclusion, this study demonstrates the flexibility and effectiveness of integrating minimum average variance estimation with adaptive penalization techniques for dimension reduction and variable selection in semiparametric modeling. The complementary strengths of MAVE-ALASSO and MAVE-SCAD, coupled with their strong theoretical foundations and practical performance, pave the way for further methodological advancements and applications in high-dimensional data analysis.

Statements

Acknowledgement

The authors would like to thank the editors and the anonymous reviewers for their valuable comments and suggestions which improved substantially the quality of this article.

Author contributions

Conceptualization, M.Y.H, H.R.T.; data curation, M.Y.H, H.R.T.; formal analysis, M.Y.H, H.R.T.; investigation, M.Y.H, H.R.T.; methodology, M.Y.H, H.R.T.; software,M.Y.H, H.R.T.; supervision, M.Y.H, H.R.T.; validation, M.Y.H, H.R.T.; visualization, M.Y.H, H.R.T.; writing-original draft, M.Y.H, H.R.T.; writing-review and editing, M.Y.H, H.R.T. All authors have read and agreed to the published version of the article.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Funding

The authors have not received funding from any institutions or organizations.

References

Ahmed, H.Y. and Hmood, M.Y., 2021. Comparison of some semi-parametric methods in partial linear single-index model. Journal of Economics and Administrative Sciences, 27, 170–184.

Alkenani, A. and Yu, K., 2013. Sparse MAVE with oracle penalties. Advances and Applications in Statistics, 34, 85–105.

Leng, C., Xia, Y., and Xu, J., 2009. An adaptive estimation method for semiparametric models and dimension reduction. In Chan, K.S. (Ed.). Exploration of a Nonlinear World: An Appreciation of Howell Tong's Contributions to Statistics. World Scientific, London, UK, pp. 347–360.

Cook, R.D. and Bing, L., 2002. Dimension reduction for conditional mean in regression. Annals of Statistics, 30, 455–474.

Cook, R.D. and Forzani, L., 2009. Likelihood-based sufficient dimension reduction. Journal of the American Statistical Association, 104, 197–208.

Cook, R.D. and Forzani, L., 2018. Big data and partial least-squares prediction. The Canadian Journal of Statistics, 46, 62–78.

Da Silva, R.D., Schwartz, W.R., and Pedrini, H. 2011. Image segmentation based on wavelet feature descriptor and dimensionality reduction applied to remote sensing. Chilean Journal of Statistics, 2(2), 51–60.

Fan, J. and Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association, 96, 1348–1360.

Hmood, M.Y. and Saleh, T.A., 2016. Comparison of some semi-parametric methods in the analysis of the single index model. Journal of Economics and Administrative Sciences, 91, 367–388.

Hmood, M.Y. and Stadtmuller, U., 2013. A new version of local linear estimators. Chilean Journal of Statistics, 4(2), 61–74.

Hua L., Hansheng W., and Tsai C., 2012. Profiled forward regression for ultrahigh dimensional of variable screening in semiparametric partially linear models. Statistica Sinica, 22, 531–554.

Huang, J., Horowitz, J.L., and Ma, S., 2008. Asymptotic properties of bridge estimators in sparse high-dimension regression model. Annals of Statistics, 36, 587–613.

Huda, Y.A. and Hmood, M.Y., 2021. Comparison of some semiparametric methods in partial linear single-index model. Journal of Economics and Administrative Sciences, 27, 170–184.

Huda, Y.A. and Hmood, M.Y., 2022. Estimating partial linear single index model by MAVE and two stage estimation procedures. The International Journal of Nonlinear Analysis and Applications, 13, 2609–2619.

Horowitz J.L., 2015. Variable selection and estimation in high-dimensional models. The Canadian Journal of Economics, 48, 389–407.

Li W., Lan X., Annie Q., and Hua L., 2018. Estimation of errors-in-variables partially linear additive models. Statistica Sinica, 28, 2353–2373.

Lian, H., Hua, L., and Ruppert, D., 2014. Separation of covariates into nonparametric and parametric parts in high dimensional partially linear additive models. Statistica Sinica, 25, 591–607.

R Core Team, 2024. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.r-project.org/.

Wang, L., Lan X., Annie Q., and Hua L., 2014. Estimation and model selection in generalized additive partial linear models for correlated date with diverging number of covariates. Annals of Statistics, 42, 592–624.

Xia, Y., Tong, H., Li, W.K., and Zhu, L.X. 2002. An adaptive estimation of dimension reduction space. Journal of the Royal Statistical Society B, 64, 363–410.

Xinyu, Z. and Wendun W., 2019. Optimal model averaging estimation for partially linear models. Statistica Sinica, 29, 693–718.

Zou, H., 2006. The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 101, 1418–1429.