Statistica Sinica Preprint No: SS-2024-0159	
Title	Dimension Reduction for Extreme Regression via Contour Projection
Manuscript ID	SS-2024-0159
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0159
Complete List of Authors	Liujun Chen and Jing Zeng
Corresponding Authors	Jing Zeng
E-mails	zengjxl@ustc.edu.cn

Statistica Sinica

Dimension Reduction for Extreme Regression via Contour Projection

Liujun Chen and Jing Zeng

University of Science and Technology of China

Abstract: In extreme regression problems, a primary objective is to infer extreme values of the response given a set of predictors. The high dimensionality and heavy-tailedness of the predictors limit the applicability of classical tools for inferring conditional extremes. In this paper, we focus on the central extreme subspace (CES), whose existence and uniqueness are guaranteed under fairly mild conditions. By projecting the data onto the CES, the dimension of the predictors is reduced while all the information for inferring conditional extremes is retained, which effectively addresses the high dimensionality issue. We propose the novel COPES method to estimate the CES by utilizing contour projection. Notably, COPES is robust against heavy-tailed predictors. The theoretical justification for the consistency of COPES is established. Overall, our proposal not only extends the toolkit for extreme regression but also broadens the scope of dimension reduction techniques. The effectiveness of our proposal is demonstrated through extensive simulation studies and an application to Chinese stock market data.

Keywords and phrases: conditional extremes, sufficient dimension reduction, elliptically contoured distribution, heavy-tailedness.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0159

1. Introduction

In regression problems, predicting the response becomes particularly challenging when the dimension of predictors gets large. This issue is more pronounced when we focus on the tail behaviour of the response, such as extreme quantile (Chernozhukov 2005), extreme expectile (Girard et al. 2021) and extreme probability (Hall & Weissman 1997), which is referred to as *extreme regression*. The primary reason for this challenge is that when we study the tail of a distribution, the effective sample size is only a small fraction of the total sample size. Therefore, even moderate-dimensional predictors can easily deteriorate the state-of-the-art methods. To mitigate the impact of high dimensionality in extreme regression, some progress has been made in reducing the dimension of the predictors (e.g., Aghbalou et al. 2024, Gardes 2018, Bousebata et al. 2023). Most existing methods for dimension reduction in extreme regression are grounded in the concept of sufficient dimension reduction (SDR; Li 1991, Cook 1998), which seeks a low-dimensional subspace that retains all the relevant information about the conditional distribution of the response given the predictors.

However, in extreme regressions, none of the existing dimension reduction methods offer theoretical guarantees when the predictors have heavy tails. In an extensive scope of areas, heavy-tailedness has been one of

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0159

the most common characteristics of the collected data. For instance, financial asset returns are widely recognized as heavy-tailed (Zhao et al. 2018), and the distribution of advertiser values in online advertising exhibits heavy-tailed characteristics (Arnosti et al. 2016). Moreover, investigating the heavy-tailedness is a key problem in extreme value statistics (Resnick 2007). Researchers have identified the adverse impact of heavy-tailedness on numerous state-of-the-art methods, prompting the development of robust techniques that enhance the statistical inference accuracy. These works include linear regression (Loh 2017, Fan et al. 2021), generalized linear regression (Zhu & Zhou 2021), and classification (Hall et al. 2009). This motivates our investigation on the dimension reduction techniques designed specifically for extreme regressions with the presence of heavy-tailed predictors.

In this paper, we introduce the COPES method, short for **CO**ntour **P**rojected Estimation for Central **E**xtreme **S**ubspace. We first define the *extreme dimension reduction* (EDR) subspace and the *central extreme subspace* (CES). By projecting the predictors onto an EDR subspace, we retain all the information for the conditional extremes of the response given the predictors. The CES is the minimal EDR subspace, defined as the intersection of all EDR subspaces, and it exists under fairly mild conditions.

A careful examination of the distinctions between the EDR subspace and dimension reduction subspaces in related works is also conducted. Subsequently, under the elliptically-contoured distribution assumption for predictors, we project the predictors onto an elliptical contour, and we establish three working subspaces for the response and projected predictors. These working subspaces are shown to be associated with the CES under certain method-specific assumptions. We proceed to estimate the CES through estimating the working subspaces from samples.

Our contributions are multi-fold. First, our method demonstrates robustness to heavy-tailed predictors and expands the applicability of dimension reduction in extreme regression. Second, we develop both first-order and second-order methods for estimating the CES, offering a comprehensive methodology for dimension reduction in extreme regression. Our numerical studies underscore the complementary effects of these methods. Third, recognizing that the contour-projected predictors deviate from the constant variance condition typically assumed in SDR, the direct adaptation of second-order SDR methods to extreme regression is inapplicable. Thus, we carefully design two second-order methods, which strategically circumvent the constant variance condition by leveraging the properties of the elliptically-contoured distribution.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0159

Our work is closely related to Aghbalou et al. (2024), which introduced the concept of the extreme SDR space. However, it is not well understood whether the minimal extreme SDR space is unique, which brings ambiguities in parameter estimation. In contrast, the existence and uniqueness of our newly defined minimal EDR subspace, i.e., the CES, are ensured, provides us with a uniquely defined target parameter. Moreover, the proposal in Aghbalou et al. (2024) is vulnerable to certain classes of predictors, particularly those from heavy-tailed distributions. In comparison, our COPES enjoys favorable properties, even in the presence of heavy-tailed predictors. The numerical findings also demonstrate that our proposal exhibits superior performance over the proposal in Aghbalou et al. (2024) when the predictors have heavy tails. Another related work is Gardes (2018), which introduced the notion of the tail dimension reduction subspace. This subspace subtly differs from the extreme SDR space in Aghbalou et al. (2024); see Aghbalou et al. (2024) for discussions on such distinctions. Although Gardes (2018) established the uniqueness of the minimal tail dimension reduction subspace under a certain condition, this condition is stronger than ours. Additionally, while Gardes (2018) offered an estimation of the tail dimension reduction subspace, the theoretical justification of the consistency of the estimator is not provided. Hence, the performance of their proposal under heavy-

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0159

tailed predictors remains unclear. Another line of related works assumes semi-parametric models, including the single- or multiple-index model. For instance, Xu et al. (2022) studied the estimation of conditional extreme quantiles under the tail single-index model. Meanwhile, Bousebata et al. (2023) proposed the partial least-square approach for dimension reduction in conditional extremes under single- and multiple-index models.

The rest of the paper is organized as follows. After the review of some background knowledge in Section 2, we study the properties of EDR subspace and CES in Section 3. In Section 4, we detail three specific COPES methods. The estimations and asymptotic theories for three COPES methods are studied in Section 5. In Section 6, we evaluate the finite sample performance of the COPES methods through synthetic data. In Section 7, we demonstrate the effectiveness and efficiency of the COPES methods through a Chinese stock market example. All technical proofs are gathered in the Supplementary Material.

2. Background

2.1 Notations

For vector $\mathbf{u} = (u_1, \ldots, u_p)^\top \in \mathbb{R}^p$, let $\|\mathbf{u}\| = (\sum_{j=1}^p u_j^2)^{1/2}$ denote the Euclidean norm. For a matrix $\mathbf{A} \in \mathbb{R}^{p\times q}$, we interchangeably use $\mathrm{span}(\mathbf{A})$ and $\mathcal{S}_{\mathbf{A}}$ to denote the subspace spanned by the columns of $\mathbf{A}$. Let $\sigma_i(\mathbf{A})$

denote the i-th largest singular value of $\mathbf{A}$, for $i = 1, \ldots, \min\{p, q\}$ and $\boldsymbol{\sigma}(\mathbf{A}) = \{\sigma_1(\mathbf{A}), \ldots, \sigma_{\min\{p,q\}}(\mathbf{A})\}$. Define the Frobenius norm as $\|\mathbf{A}\|_F = \{\sum_{i=1}^{\min\{p,q\}} \sigma_i^2(\mathbf{A})\}^{1/2}$ and the spectral norm as $\|\mathbf{A}\| = \sigma_1(A)$. Let $\mathrm{SVD}_d(\mathbf{A}) \in \mathbb{R}^{p\times d}$ denote the matrix composed of left singular vectors of $\mathbf{A}$ corresponding to its largest d singular values. For a subspace $\mathcal{S} \subseteq \mathbb{R}^p$ spanned by some matrix $\mathbf{A} \in \mathbb{R}^{p\times q}$, i.e., $\mathcal{S} = \mathrm{span}(\mathbf{A})$, let $\mathbf{P}_{\mathcal{S}} \equiv \mathbf{P}_{\mathbf{A}} = \mathbf{A}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top$ denote the projection onto the subspace $\mathcal{S}$. Moreover, let $\mathbf{Q}_{\mathcal{S}} \equiv \mathbf{Q}_{\mathbf{A}} = \mathbf{I} - \mathbf{P}_{\mathbf{A}}$. For a positive definite matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, define $\mathbf{P}_{\mathbf{A}(\boldsymbol{\Sigma})} = \mathbf{A}(\mathbf{A}^\top\boldsymbol{\Sigma}\mathbf{A})^{-1}\mathbf{A}^\top\boldsymbol{\Sigma}$ and $\mathbf{Q}_{\mathbf{A}(\boldsymbol{\Sigma})} = \mathbf{I} - \mathbf{P}_{\mathbf{A}(\boldsymbol{\Sigma})}$.

2.2 Sufficient dimension reduction

Our proposal is intricately related to the sufficient dimension reduction (SDR; Li 1991, Cook 1998). For a response variable $Y \in \mathbb{R}$ and predictors $\mathbf{X} \in \mathbb{R}^p$, SDR seeks a lower-dimensional subspace $\mathcal{S} \subseteq \mathbb{R}^p$ such that

$$Y \mid \mathbf{X} \sim Y \mid \mathbf{P}_{\mathcal{S}}\mathbf{X}, \tag{2.1}$$

where "$\sim$" denotes "the same distribution as" and $\mathbf{P}_{\mathcal{S}}$ denotes the projection matrix onto the subspace $\mathcal{S}$. The subspace satisfying (2.1) is referred to as the SDR subspace, and the intersection of all SDR subspaces, if itself is an SDR subspace, is known as the central subspace, denoted by $\mathcal{S}_{Y|\mathbf{X}}$ (Cook 1998). The existence of central subspace is ensured under mild conditions (Cook 1998). In the spirit of (2.1), by replacing the original data $\mathbf{X}$ with the reduced data $\mathbf{P}_{\mathcal{S}_{Y|\mathbf{X}}}\mathbf{X}$, all the information about the conditional distribution of $Y \mid \mathbf{X}$ is preserved. Various SDR methods have been proposed to estimate the central subspace, including inverse regression methods (Li 1991, Cook & Weisberg 1991, Li & Wang 2007, Zhu et al. 2010, Cook & Li 2002, Chen et al. 2010, Yu et al. 2016), likelihood-based methods (Cook & Forzani 2008, Bura et al. 2016), and semi-parametric or non-parametric methods (Xia et al. 2002, Ma & Zhu 2012, Fukumizu & Leng 2014). For an overview of SDR methods, interested readers can refer to the monograph Li (2018).

2.3 Elliptically-contoured distribution

For a random vector $\mathbf{X}$ following an elliptically-contoured (EC) distribution with a mean vector $\boldsymbol{\mu} \in \mathbb{R}^p$ and a scatter matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$, its probability density function is given by $f(\boldsymbol{x}) = k_p|\boldsymbol{\Sigma}|^{-1/2} g\{(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}$ for $\boldsymbol{x} \in \mathbb{R}^p$, where the one-dimensional function g is independent of p, and k_p is a normalizing constant (Johnson 1987). To ensure the identifiability of $\boldsymbol{\Sigma}$, we assume $\mathrm{tr}(\boldsymbol{\Sigma}) = p$.

We focus on the EC distribution for several reasons. First, it covers a broad spectrum of distributions, including light-tailed and heavy-tailed distributions. Commonly used distributions, such as multivariate normal, multivariate t, multivariate Laplacian, multivariate slash, and multivariate contaminated normal distributions fall into the range of the EC distribution. This flexibility enables us to accommodate various types of data and handle more complicated situations. Second, the EC distribution assumption has been widely employed in SDR, especially for modeling the heavy-tailed data, see e.g., Wang et al. (2008), Luo et al. (2009), and Chen et al. (2022). One possible reason is that the linearity condition (Eaton 1986), a widely used assumption in inverse-regression based SDR methods, is automatically satisfied by the EC-distributed data. Moreover, SDR in the context of extreme regression also requires the linearity condition, see e.g., Aghbalou et al. (2024). Since our proposal relies on the linearity condition and accommodates the heavy-tailedness, the EC distribution assumption is a natural choice. Third, the linear combination of EC random vectors remains an EC random vector, making the reduced predictors more interpretable.

3. Dimension reduction subspace for extreme regression

In extreme regression, a central focus lies in inferring the extreme values of $Y \mid \mathbf{X}$. Without loss of generality, we work on the right tail of Y, i.e., $Y \mid (Y > y)$, where y represents a large thresholding value. Note that studying the left tail of Y is equivalent to studying the right tail of $-Y$. Therefore, once the dimension reduction theory of the right tail of Y is well understood, the results can be easily generalized to the left tail of Y. Let y^+

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0159

denote the supremum of the support of Y, we assume that $\Pr(Y > y) \to 0$ as $y \to y^+$, excluding the special case where a point mass is at the right endpoint of Y. The EDR subspace is introduced in the following.

Definition 1 (EDR subspace)**.** The subspace $\mathcal{S}_{\boldsymbol{\beta}} \subseteq \mathbb{R}^p$, spanned by some basis matrix $\boldsymbol{\beta} \in \mathbb{R}^{p\times d}$, is called an EDR subspace of Y given $\mathbf{X}$ if and only if, for any $\varepsilon > 0$, there exists some constant y_0 such that for all $y \geq y_0$,

$$\left| \frac{\Pr(Y > y \mid \mathbf{X} = \boldsymbol{x}) - \Pr(Y > y \mid \boldsymbol{\beta}^\top \mathbf{X} = \boldsymbol{\beta}^\top \boldsymbol{x})}{\Pr(Y > y)} \right| \leq \varepsilon, \quad \text{for all } \boldsymbol{x} \in \Omega_{\mathbf{X}},$$

where $\Omega_{\mathbf{X}}$ is the support of $\mathbf{X}$.

It is evident that any SDR subspace is also an EDR subspace. A similar relationship between the SDR subspace and the extreme SDR subspace defined in Aghbalou et al. (2024) has been discussed in Aghbalou et al. (2024). Aghbalou et al. (2024) has further clarified through concrete examples that the SDR subspace contains redundant information immaterial to the inference for the conditional extremes in some cases.

Moreover, any EDR subspace is also an extreme SDR subspace, which is defined in Aghbalou et al. (2024). However, the question remains whether the intersection of the extreme SDR subspaces is still an extreme SDR subspace, thus the minimal extreme SDR subspace may not be uniquely defined. In comparison, as we will clarify later, the intersection of all EDR subspaces is still an EDR subspace under very mild condition, guaranteeing

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0159

the uniqueness of the minimal EDR subspace, which we refer to as the central extreme subspace (CES). The CES is formally defined as follows.

Definition 2 (CES)**.** The intersection of all EDR subspaces of Y given $\mathbf{X}$ is called the CES if it is itself an EDR subspace, denoted by $\mathcal{S}_{Y_\infty|\mathbf{X}}$.

If the CES exists, it is uniquely defined. Since the CES is the minimal EDR subspace in the sense that it belongs to any EDR subspace, it stands as the most economical and efficient dimension reduction subspace, preserving all the information for inferring the extreme values of Y given $\mathbf{X}$. A concrete guarantee of the existence of the CES is provided in the following theorem, which only requires a mild condition on the distribution of $\mathbf{X}$.

Theorem 1. *Assume that* $\mathbf{X}$ *is supported on a convex set* $\Omega_{\mathbf{X}} \subseteq \mathbb{R}^p$*, then* $\mathcal{S}_{Y_\infty|\mathbf{X}}$ *exists.*

The convexity condition for $\Omega_{\mathbf{X}}$ is considered mild and is satisfied by many commonly adopted distributions. It is also a requirement for the existence of the central subspace (Cook 1998, Proposition 6.4). Under this weak assumption, the minimal EDR subspace is uniquely defined. In contrast, the uniqueness of the minimal extreme SDR space in Aghbalou et al. (2024) is uncertain. Gardes (2018) established the uniqueness of the minimal tail dimension reduction subspace under a stronger condition, requiring

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0159

$\mathbf{X}$ to be supported on $\mathbb{R}^p$. In the remainder of the paper, we assume that $\mathbf{X}$ is supported on a convex set $\Omega_{\mathbf{X}} \subseteq \mathbb{R}^p$ so that $\mathcal{S}_{Y_\infty|\mathbf{X}}$ exists. Denote the dimension $d^* = \dim(\mathcal{S}_{Y_\infty|\mathbf{X}})$.

4. COPES

Heavy-tailed data poses challenges to dimension reduction in extreme regression, as confirmed by our empirical findings. In this section, we assume that $\mathbf{X}$ follows an EC distribution and develop COPES methods, which efficiently estimate $\mathcal{S}_{Y_\infty|\mathbf{X}}$ even when the predictors have heavy tails.

We first project the predictors onto an elliptical contour, constructing the contour-projected predictors $\overrightarrow{\mathbf{X}}$. Then, we develop specific COPES methods based on $\overrightarrow{\mathbf{X}}$ from three classical SDR methods, namely, sliced inverse regression (SIR; Li 1991), sliced average variance estimate (SAVE; Cook & Weisberg 1991), and directional regression (DR; Li & Wang 2007). We choose SIR, SAVE, and DR for several reasons. First, they are important representative first-order and second-order SDR methods, which have gained comprehensive studies in the literature. Second, their implementation is easy, facilitating the computation of our proposal, as we will elaborate in Section 5.1 and 5.2. Third, all these inverse moment methods demand the linearity condition, which is automatically satisfied by the EC distribution. For each specific COPES method, we define the working subspace and provide their non-asymptotic approximations.

4.1 Contour projection

We project the centered EC-distributed predictors $\mathbf{X}-\boldsymbol{\mu}$ onto the elliptical contour $\mathcal{C}=\{\mathbf{u}:\|\mathbf{u}\|_{\boldsymbol{\Sigma}}=1\}$, where $\|\mathbf{u}\|_{\boldsymbol{\Sigma}}=(\mathbf{u}^\top\boldsymbol{\Sigma}^{-1}\mathbf{u})^{1/2}$, resulting the contour-projected predictors $\overrightarrow{\mathbf{X}}=(\mathbf{X}-\boldsymbol{\mu})/\|\mathbf{X}-\boldsymbol{\mu}\|_{\boldsymbol{\Sigma}}$. Due to its bounded nature, $\overrightarrow{\mathbf{X}}$ possess finite moments of all orders. Such a favorable finite-moment property brings ease to the parameter estimation. For instance, the contour projection method also proves effective in SDR under the heavy-tailed predictors (Wang et al. 2008, Luo et al. 2009). If we are able to establish some connection between an EDR subspace of Y given $\mathbf{X}$ and that of Y given $\overrightarrow{\mathbf{X}}$, the challenges posed by heavy-tailed predictors can be mitigated. The following result implies such a connection.

Lemma 1. *Assume that* $\mathbf{X}$ *follows an EC distribution. Then, any EDR subspace of* Y *given* $\mathbf{X}$ *is also an EDR subspace of* Y *given* $\overrightarrow{\mathbf{X}}$. *Moreover, the CES* $\mathcal{S}_{Y_\infty|\mathbf{X}}$ *is an EDR subspace of* Y *given* $\overrightarrow{\mathbf{X}}$.

Remark 1. Lemma 1 assumes that $\mathbf{X}$ follows an EC distribution. This assumption can be relaxed to that $\overrightarrow{\mathbf{X}}$ and $\|\mathbf{X}-\boldsymbol{\mu}\|_{\boldsymbol{\Sigma}}$ are independent.

The direct implication of Lemma 1 is that the intersection of all EDR subspaces of Y given $\overrightarrow{\mathbf{X}}$ belongs to $\mathcal{S}_{Y_\infty|\mathbf{X}}$. One can further define the CES

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0159

of Y given $\overrightarrow{\mathbf{X}}$ as the intersection of all EDR subspaces of Y given $\overrightarrow{\mathbf{X}}$ if it is itself an EDR subspace of Y given $\overrightarrow{\mathbf{X}}$. The existence of the CES of Y given $\overrightarrow{\mathbf{X}}$ is under investigation and left for future research.

In fact, the existence of the CES of Y given $\overrightarrow{\mathbf{X}}$ is not quite crucial since the target parameter in the current paper is $\mathcal{S}_{Y_\infty|\mathbf{X}}$. In the following, we propose three working subspaces contained in an EDR subspace of Y given $\overrightarrow{\mathbf{X}}$. More importantly, we establish the connections between the working subspaces and $\mathcal{S}_{Y_\infty|\mathbf{X}}$.

4.2 COPES-SIR

The working subspace for COPES-SIR is called the extreme SIR subspace, denoted by $\mathcal{S}_{\mathrm{eSIR}}$. Its existence depends on the following convergence assumption related to the tail inverse moment of $\overrightarrow{\mathbf{X}}$:

(A1) There exists a non-zero vector $\boldsymbol{\nu} \in \mathbb{R}^p$ such that $\boldsymbol{\nu} = \lim_{y \to y^+} \mathbb{E}(\overrightarrow{\mathbf{X}} \mid Y > y)$.

In Aghbalou et al. (2024), a convergence assumption similar to Assumption (A1) was also required. They assumed the tail inverse moment convergence for Y and original predictors $\mathbf{X}$, i.e., $\lim_{y \to y^+} \mathbb{E}(\mathbf{X}|Y > y)$ exists. However, $\mathbb{E}(\mathbf{X}|Y > y)$ may diverge to infinity for many common distributions, including the multivariate normal and t distributions. To illustrate how the convergence assumption of their tail inverse moment is violated and how the contour projection alleviates such an issue, we provide a toy example in Section S2 of the Supplementary Material.

Under Assumption (A1), we define the extreme SIR subspace as $\mathcal{S}_{\text{eSIR}} = \boldsymbol{\Sigma}^{-1}\text{span}(\boldsymbol{\nu})$. Note that the non-zero assumption on $\boldsymbol{\nu}$ eliminates the degenerate case where $\mathcal{S}_{\text{eSIR}}$ is null. In classical SIR, the linearity condition is necessarily assumed, ensuring that SIR recovers at least a portion of the central subspace (Li 1991). The linearity condition states that $\mathbb{E}(\mathbf{X} \mid \mathbf{P}_{\mathcal{S}_{Y|\mathbf{X}}}\mathbf{X})$ is a linear function of $\mathbf{P}_{\mathcal{S}_{Y|\mathbf{X}}}\mathbf{X}$, a condition satisfied when $\mathbf{X}$ follows the EC distribution. Since $\overrightarrow{\mathbf{X}}$ is elliptically contoured, it automatically satisfies the linearity condition. This property facilitates the connection between $\mathcal{S}_{\text{eSIR}}$ and an EDR subspace of $Y \mid \overrightarrow{\mathbf{X}}$, formally stated in the following theorem.

Theorem 2. *Let* $\text{span}(\boldsymbol{\beta}) \subseteq \mathbb{R}^p$ *denote an EDR subspace of* Y *given* $\overrightarrow{\mathbf{X}}$. *Under the EC distribution assumption of* $\mathbf{X}$ *and Assumption (A1), we have* $\mathcal{S}_{\text{eSIR}} \subseteq \text{span}(\boldsymbol{\beta})$. *Moreover, we have* $\mathcal{S}_{\text{eSIR}} \subseteq \mathcal{S}_{Y_\infty|\mathbf{X}}$.

The direct estimation of $\mathcal{S}_{\text{eSIR}}$ is tricky since $\mathcal{S}_{\text{eSIR}}$ is defined in a limiting manner. We overcome this issue by approximating $\mathcal{S}_{\text{eSIR}}$ with a non-asymptotic surrogate. Let $\widetilde{Y} = -Y$ denote the opposite of the response and F denote the cumulative distribution function of $\widetilde{Y}$. For $h \in (0, 1]$, we introduce the kernel matrix $\mathbf{M}^h_{\text{eSIR}} = \int_0^1 \mathbf{C}_h(u)\mathbf{C}_h^\top(u)\text{d}u$, where $\mathbf{C}_h(u) \in \mathbb{R}^p$ is a vector-valued function of u, specified by $\mathbf{C}_h(u) = (1/h)\mathbb{E}\{\overrightarrow{\mathbf{X}}I(\widetilde{Y} <$

$F^{-1}(uh))\}$, for $0 \leq u \leq 1$. Such a cumulative slicing estimation is in the same spirit of Aghbalou et al. (2024) and Zhu et al. (2010). The following lemma affirms that $\boldsymbol{\Sigma}^{-1}\text{span}(\mathbf{M}^h_{\text{eSIR}})$ is a valid approximation of $\mathcal{S}_{\text{eSIR}}$.

Lemma 2. *Under Assumption (A1), there exists a matrix* $\mathbf{M}_{\text{eSIR}} \in \mathbb{R}^{p\times p}$ *such that* $\mathbf{M}^h_{\text{eSIR}} \to \mathbf{M}_{\text{eSIR}}$ *as* $h \to 0$*. Moreover,* $\boldsymbol{\Sigma}^{-1}\text{span}(\mathbf{M}_{\text{eSIR}}) = \mathcal{S}_{\text{eSIR}}$.

According to Lemma 2, as h converges to zero, the matrix $\boldsymbol{\Sigma}^{-1}\text{span}(\mathbf{M}^h_{\text{eSIR}})$ converges to $\mathcal{S}_{\text{eSIR}}$. Since $\mathbf{M}^h_{\text{eSIR}}$ is defined in a non-asymptotic manner, it can be readily estimated from samples to approximate $\mathcal{S}_{\text{eSIR}}$. By consistently estimating $\boldsymbol{\Sigma}$ and $\mathbf{M}^h_{\text{eSIR}}$ with a sufficiently small h, we can obtain the accurate estimation of $\mathcal{S}_{\text{eSIR}}$.

Since the limit $\boldsymbol{\nu} = \lim_{y\to y^+} \mathbb{E}(\overrightarrow{\mathbf{X}} \mid Y > y)$ is unique, $\mathcal{S}_{\text{eSIR}}$ captures at most one direction, which is different from the classical SIR method based on $\mathbb{E}(\overrightarrow{\mathbf{X}}|Y)$. By varying Y over its range, classical SIR may detect multiple directions. Therefore, for scenarios where multiple directions are on demand, higher-order moments become imperative for a more exhaustive search of dimension reduction directions.

4.3 COPES-SAVE

In this section, we introduce the COPES-SAVE method, using higher-order moments to discover additional dimension reduction directions. The construction of the extreme SAVE subspace $\mathcal{S}_{\text{eSAVE}}$, the working subspace for

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0159

COPES-SAVE, relies on the following convergence assumption:

(A2) There exists a positive semi-definite matrix $\mathbf{T} \in \mathbb{R}^{p\times p}$ such that $\mathbf{T} = \lim_{y\to y^+} \mathbb{E}(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top \mid Y > y)$.

While it might be tempting to directly follow the SAVE method and define a naive extreme SAVE subspace as $\boldsymbol{\Sigma}^{-1}\mathrm{span}(\mathbf{T}-\boldsymbol{\Sigma})$, its connection to an EDR subspace of Y given $\overrightarrow{\mathbf{X}}$, $\mathrm{span}(\boldsymbol{\beta})$, requires the linearity condition and the constant covariance condition. The constant covariance condition assumes that $\mathrm{Cov}(\overrightarrow{\mathbf{X}} \mid \mathbf{P}_{\boldsymbol{\beta}}\overrightarrow{\mathbf{X}})$ is a deterministic matrix. However, these two conditions essentially imply the normality for $\overrightarrow{\mathbf{X}}$ (Ma & Zhu 2013), which is inherently impossible for $\overrightarrow{\mathbf{X}}$ distributed on the sphere.

In fact, the constant covariance condition is not structurally necessary for constructing the extreme SAVE subspace; instead, the explicit form of the covariance $\mathrm{Cov}(\overrightarrow{\mathbf{X}} \mid \mathbf{P}_{\boldsymbol{\beta}}\overrightarrow{\mathbf{X}})$ is crucial. Leveraging the properties of the EC distribution and contour projection, Luo et al. (2009) established, in the proof of their Lemma 3, that the conditional covariance $\mathrm{Cov}(\overrightarrow{\mathbf{X}} \mid \mathbf{P}_{\boldsymbol{\beta}}\overrightarrow{\mathbf{X}})$ has an explicit form. The following lemma is a variant of this conclusion.

Lemma 3. *Assume that* $\mathbf{X}$ *follows the EC distribution, then for any matrix* $\boldsymbol{\eta} \in \mathbb{R}^{p\times d}$, *we have* $\mathrm{Cov}(\overrightarrow{\mathbf{X}} \mid \boldsymbol{\eta}^\top\overrightarrow{\mathbf{X}}) = \zeta(\boldsymbol{\eta}^\top\overrightarrow{\mathbf{X}})\boldsymbol{\Sigma}\mathbf{Q}_{\boldsymbol{\eta}(\boldsymbol{\Sigma})}$, *where* $\zeta(\boldsymbol{\eta}^\top\overrightarrow{\mathbf{X}}) = \|\mathbf{Q}_{\boldsymbol{\eta}(\boldsymbol{\Sigma})}^\top\overrightarrow{\mathbf{X}}\|_{\boldsymbol{\Sigma}}^2/(p-d)$.

Lemma 3 differs from the corresponding result in Luo et al. (2009)

in that $\mathbf{X}$ is not necessarily standardized in our setting. Define $\tau_{\boldsymbol{\beta}} = \lim_{y\to y^+} \mathbb{E}\{\zeta(\boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}) \mid Y > y\}$. By rewriting $\zeta(\boldsymbol{\beta}^\top \overrightarrow{\mathbf{X}}) = \mathrm{tr}\{\mathbf{Q}_{\boldsymbol{\beta}(\boldsymbol{\Sigma})}\boldsymbol{\Sigma}^{-1}\mathbf{Q}_{\boldsymbol{\beta}(\boldsymbol{\Sigma})}^\top \overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top\}/(p-d)$, it is easy to verify that under Assumption (A2), the limit $\tau_{\boldsymbol{\beta}}$ exists. We define the extreme SAVE subspace in terms of $\mathrm{span}(\boldsymbol{\beta})$ as

$$\mathcal{S}_{\mathrm{eSAVE}} = \boldsymbol{\Sigma}^{-1}\mathrm{span}(\mathbf{T} - \tau_{\boldsymbol{\beta}}\boldsymbol{\Sigma}). \tag{4.2}$$

The following theorem shows that $\mathcal{S}_{\mathrm{eSAVE}}$ is a subset of $\mathrm{span}(\boldsymbol{\beta})$.

Theorem 3. *Let* $\mathrm{span}(\boldsymbol{\beta}) \subseteq \mathbb{R}^p$ *denote an EDR subspace of* Y *given* $\overrightarrow{\mathbf{X}}$. *Under the EC distribution assumption of* $\mathbf{X}$ *and Assumption (A2), we have* $\mathcal{S}_{\mathrm{eSAVE}} \subseteq \mathrm{span}(\boldsymbol{\beta})$. *In special, for the extreme SAVE subspace defined in terms of* $\mathcal{S}_{Y_\infty|\mathbf{X}}$*, we have* $\mathcal{S}_{\mathrm{eSAVE}} \subseteq \mathcal{S}_{Y_\infty|\mathbf{X}}$.

It is worth noting that the constant covariance condition is not required for Theorem 3. While the ultimate goal of proposing $\mathcal{S}_{\mathrm{eSAVE}}$ is to approach $\mathrm{span}(\boldsymbol{\beta})$, the parameter $\tau_{\boldsymbol{\beta}}$ in (4.2) itself, however, contains the target parameter $\boldsymbol{\beta}$, making $\mathcal{S}_{\mathrm{eSAVE}}$ more like a recursive definition of $\mathrm{span}(\boldsymbol{\beta})$. We realize that when $\dim(\mathcal{S}_{\mathrm{eSAVE}}) < p$, i.e., the matrix $\mathbf{T} - \tau_{\boldsymbol{\beta}}\boldsymbol{\Sigma}$ is rank-deficient, $\tau_{\boldsymbol{\beta}}$ must coincide with some eigenvalues of $\boldsymbol{\Sigma}^{-1/2}\mathbf{T}\boldsymbol{\Sigma}^{-1/2}$. If we further assume that $\dim(\mathcal{S}_{\mathrm{eSAVE}}) < p/2$, then more than half of eigenvalues of $\boldsymbol{\Sigma}^{-1/2}\mathbf{T}\boldsymbol{\Sigma}^{-1/2}$ equal $\tau_{\boldsymbol{\beta}}$. Consequently, the scalar parameter $\tau_{\boldsymbol{\beta}}$ can be substituted by the median of the eigenvalues of $\boldsymbol{\Sigma}^{-1/2}\mathbf{T}\boldsymbol{\Sigma}^{-1/2}$, providing a computationally feasible parameterization of $\mathcal{S}_{\mathrm{eSAVE}}$ as follows.

Lemma 4. *Assume that* $\dim(\mathcal{S}_{\text{eSAVE}}) < p/2$*, then* $\mathcal{S}_{\text{eSAVE}} = \mathbf{\Sigma}^{-1}\text{span}(\mathbf{T} - \tau\mathbf{\Sigma})$ *with* $\tau = \text{median}\{\boldsymbol{\sigma}(\mathbf{\Sigma}^{-1/2}\mathbf{T}\mathbf{\Sigma}^{-1/2})\}$.

For the extreme SAVE subspace defined in terms of $\mathcal{S}_{Y_\infty|\mathbf{X}}$, the assumption that $\dim(\mathcal{S}_{\text{eSAVE}}) < p/2$ is essentially mild, implied by requiring that $\dim(\mathcal{S}_{Y_\infty|\mathbf{X}}) < p/2$. In the context of dimension reduction, we expect a low dimension of the CES $\mathcal{S}_{Y_\infty|\mathbf{X}}$ such that the dimensionality of predictors is reduced to a low level for more efficient subsequent analysis.

Similar to COPES-SIR, we provide a non-asymptotic approximation of $\mathcal{S}_{\text{eSAVE}}$ for an easier estimation. For $h \in (0, 1]$, define the kernel matrix $\mathbf{M}^h_{\text{eSAVE}} = \int_0^1 \mathbf{D}_h(u)\mathbf{D}_h^\top(u)\text{d}u$, where the matrix-valued function $\mathbf{D}_h(u) \in \mathbb{R}^{p\times p}$ is defined as $\mathbf{D}_h(u) = \mathbf{T}_h(u) - \text{median}[\boldsymbol{\sigma}\{\mathbf{\Sigma}^{-1/2}\mathbf{T}_h(u)\mathbf{\Sigma}^{-1/2}\}]\mathbf{\Sigma}$, for $0 \le u \le 1$, and $\mathbf{T}_h(u) = (1/h)\mathbb{E}[\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top I\{\widetilde{Y} < F^{-1}(uh)\}]$. When $\dim(\mathcal{S}_{\text{eSAVE}}) < p/2$, $\mathbf{\Sigma}^{-1}\text{span}(\mathbf{M}^h_{\text{eSAVE}})$ provides a proper approximation of $\mathcal{S}_{\text{eSAVE}}$.

Lemma 5. *Under Assumption (A2), there exists a matrix* $\mathbf{M}_{\text{eSAVE}} \in \mathbb{R}^{p\times p}$ *such that* $\mathbf{M}^h_{\text{eSAVE}} \to \mathbf{M}_{\text{eSAVE}}$ *as* $h \to 0$*. Further assume that* $\dim(\mathcal{S}_{\text{eSAVE}}) < p/2$*, then we have* $\mathbf{\Sigma}^{-1}\text{span}(\mathbf{M}_{\text{eSAVE}}) = \mathcal{S}_{\text{eSAVE}}$.

COPES-SIR and COPES-SAVE bear similarities with TIREX1 and TIREX2 methods which are proposed by Aghbalou et al. (2024), as both are inspired by SIR and SAVE methods and introduce kernel matrices to facilitate the estimation of working subspaces. However, their proposal

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0159

assumes the existence of tail inverse moments $\lim_{y\to y^+}\mathbb{E}(\mathbf{X}|Y>y)$ and $\lim_{y\to y^+}\mathbb{E}(\mathbf{X}\mathbf{X}^\top|Y>y)$ and requires uniform integrability conditions concerning the tail inverse moments of $\|\mathbf{X}\|$ and $\|\mathbf{X}\|^2$. As clarified earlier, these conditions are quite stringent even for common distributions. In contrast, our approach relaxes these conditions by exploiting the contour-projected predictors. While the contour projection approach avoids the stringent conditions required by Aghbalou et al. (2024), it faces the new challenge not encountered in their work, that is, the constant covariance condition is not met for the contour-projected predictors. To address this challenge, we exploit the distributional characteristics of $\overrightarrow{\mathbf{X}}$ and explicitly incorporate the conditional variance $\mathrm{Cov}(\overrightarrow{\mathbf{X}}\mid\mathbf{P}_{\boldsymbol{\beta}}\overrightarrow{\mathbf{X}})$ in the construction of $\mathcal{S}_{\mathrm{eSAVE}}$. These aspects distinguish our methodology from that of Aghbalou et al. (2024).

4.4 COPES-DR

It is known in the literature that SIR and SAVE have acknowledged limitations: SIR may fall short in identifying SDR directions when the model exhibits a symmetric structure, and SAVE could be inefficient in capturing the monotone trend, especially when the sample size is limited. In response to these limitations, directional regression (DR) was introduced to integrate the advantages of both SIR and SAVE. Li & Wang (2007) and subsequent works have demonstrated the high effectiveness of DR across a broader

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0159

range of models. Motivated by these findings, we extend the DR methodology to an extreme version, namely, COPES-DR, making our study more comprehensive.

Let $(\overrightarrow{\mathbf{X}}^*, Y^*)$ be an independent copy of $(\overrightarrow{\mathbf{X}}, Y)$. Define the matrix $\mathbf{A} = \lim_{y,y^* \to y^+} \mathbb{E}\{(\overrightarrow{\mathbf{X}} - \overrightarrow{\mathbf{X}}^*)(\overrightarrow{\mathbf{X}} - \overrightarrow{\mathbf{X}}^*)^\top \mid Y > y, Y^* > y^*\}$, which is a crucial component in the extreme DR subspace $\mathcal{S}_{\text{eDR}}$, the working subspace of COPES-DR. The next lemma claims the existence of $\mathbf{A}$.

Lemma 6. *Under Assumptions (A1) & (A2), the matrix* $\mathbf{A}$ *exists.*

Motivated by DR, a naive extreme DR subspace can be defined as $\mathbf{\Sigma}^{-1}\text{span}(\mathbf{A} - 2\mathbf{\Sigma})$. However, the constant covariance condition is also required by DR, which is absent when we consider the contour-projected predictors $\overrightarrow{\mathbf{X}}$. Hence, analogous to COPES-SAVE, we define the extreme DR subspace in terms of an EDR subspace of Y given $\overrightarrow{\mathbf{X}}$, $\text{span}(\boldsymbol{\beta})$, as follows, $\mathcal{S}_{\text{eDR}} = \mathbf{\Sigma}^{-1}\text{span}(\mathbf{A} - 2\tau_{\boldsymbol{\beta}}\mathbf{\Sigma})$. We connect $\mathcal{S}_{\text{eDR}}$ and $\text{span}(\boldsymbol{\beta})$ as follows.

Theorem 4. *Let* $\text{span}(\boldsymbol{\beta}) \subseteq \mathbb{R}^p$ *denote an EDR subspace of* Y *given* $\overrightarrow{\mathbf{X}}$. *Under the EC distribution assumption of* $\mathbf{X}$ *and Assumptions (A1) & (A2), we have* $\mathcal{S}_{\text{eDR}} \subseteq \text{span}(\boldsymbol{\beta})$. *In special, for the extreme DR subspace defined in terms of* $\mathcal{S}_{Y_\infty|\mathbf{X}}$, *we have* $\mathcal{S}_{\text{eDR}} \subseteq \mathcal{S}_{Y_\infty|\mathbf{X}}$.

In defining $\mathcal{S}_{\text{eDR}}$, an issue similar to $\mathcal{S}_{\text{eSAVE}}$ arises as the parameter $\tau_{\boldsymbol{\beta}}$

contains the target parameter $\boldsymbol{\beta}$. To address this, we restrict the dimension of $\mathcal{S}_{\mathrm{eDR}}$, allowing for an estimable parameterization.

Lemma 7. *Assume that* $\dim(\mathcal{S}_{\mathrm{eDR}}) < p/2 - 1$*, then* $\mathcal{S}_{\mathrm{eDR}} = \boldsymbol{\Sigma}^{-1}\mathrm{span}(\mathbf{A} - 2\tau\boldsymbol{\Sigma})$ *with* $\tau = \mathrm{median}\{\boldsymbol{\sigma}(\boldsymbol{\Sigma}^{-1/2}\mathbf{T}\boldsymbol{\Sigma}^{-1/2})\}$*, where* $\mathbf{T} = \lim_{y \to y^+} \mathbb{E}(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top \mid Y > y)$*.*

The non-asymptotic approximation of $\mathcal{S}_{\mathrm{eDR}}$ is introduced in the following. Let $\widetilde{Y}^* = -Y^*$, for $h \in (0, 1]$, define the kernel matrix $\mathbf{M}_{\mathrm{eDR}}^h = \int_0^1 \int_0^1 \mathbf{G}_h(u, u^*)\mathbf{G}_h^\top(u, u^*)\mathrm{d}u\mathrm{d}u^*$, where $\mathbf{G}_h(u, u^*) \in \mathbb{R}^{p\times p}$ is the matrix-valued function of the pair (u, u^*) and is defined as

$$\mathbf{G}_h(u, u^*) = \frac{1}{h^2}\mathbb{E}\left[(\overrightarrow{\mathbf{X}} - \overrightarrow{\mathbf{X}}^*)(\overrightarrow{\mathbf{X}} - \overrightarrow{\mathbf{X}}^*)^\top I\{\widetilde{Y} < F^{-1}(uh)\}I\{\widetilde{Y}^\star < F^{-1}(u^*h)\}\right]$$
$$- u^* \cdot \mathrm{median}\{\boldsymbol{\sigma}(\boldsymbol{\Sigma}^{-1/2}\mathbf{T}_h(u)\boldsymbol{\Sigma}^{-1/2})\}\boldsymbol{\Sigma} - u \cdot \mathrm{median}[\boldsymbol{\sigma}\{\boldsymbol{\Sigma}^{-1/2}\mathbf{T}_h(u^*)\boldsymbol{\Sigma}^{-1/2}\}]\boldsymbol{\Sigma},$$

and $\mathbf{T}_h(u) = (1/h)\mathbb{E}[\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top I\{\widetilde{Y} < F^{-1}(uh)\}]$. The following lemma confirms the integrality of $\boldsymbol{\Sigma}^{-1}\mathrm{span}(\mathbf{M}_{\mathrm{eDR}}^h)$ as an approximation of $\mathcal{S}_{\mathrm{eDR}}$.

Lemma 8. *Under Assumptions (A1) & (A2), there exists a matrix* $\mathbf{M}_{\mathrm{eDR}} \in \mathbb{R}^{p\times p}$ *such that* $\mathbf{M}_{\mathrm{eDR}}^h \to \mathbf{M}_{\mathrm{eDR}}$ *as* $h \to 0$*. Further assume that* $\dim(\mathcal{S}_{\mathrm{eDR}}) < p/2 - 1$*, then we have* $\boldsymbol{\Sigma}^{-1}\mathrm{span}(\mathbf{M}_{\mathrm{eDR}}) = \mathcal{S}_{\mathrm{eDR}}$*.*

The kernel matrix $\mathbf{M}_{\mathrm{eDR}}^h$ has a complicated form, and its estimation demands intense computations. We alleviate the computational burden by re-expressing $\mathbf{M}_{\mathrm{eDR}}^h$ as $\mathbf{M}_{\mathrm{eDR}}^h = 2\sum_{i=1}^8 \mathbf{K}_i$, where the expressions of $\mathbf{K}_i$,

$i = 1, \ldots, 8$, are given in Section S1 of the Supplementary Material. It can be seen that $\mathbf{M}^h_{\text{eDR}}$ is nothing special but the combinations of integrals of $\mathbf{C}_h(u)$ and $\mathbf{D}_h(u)$, which are key components of $\mathbf{M}^h_{\text{eSIR}}$ and $\mathbf{M}^h_{\text{eSAVE}}$, respectively. Thus, the estimation of $\mathbf{M}^h_{\text{eDR}}$ is no harder than that of $\mathbf{M}^h_{\text{eSIR}}$ and $\mathbf{M}^h_{\text{eSAVE}}$. Once we obtain the empirical estimates of $\mathbf{C}_h(u)$ and $\mathbf{D}_h(u)$, the estimation of $\mathbf{M}^h_{\text{eDR}}$ can be easily obtained.

We make some remarks on the exhaustiveness of the three working subspaces, i.e., $\mathcal{S}_{\text{eSIR}}$, $\mathcal{S}_{\text{eSAVE}}$, and $\mathcal{S}_{\text{eDR}}$, where a working subspace $\mathcal{S}$ is called *exhaustive* if $\mathcal{S} = \mathcal{S}_{Y_\infty|\mathbf{X}}$, following the terminology in Li & Wang (2007). According to Lemma 1, the CES $\mathcal{S}_{Y_\infty|\mathbf{X}}$ is also an EDR subspace of Y and $\overrightarrow{\mathbf{X}}$. In Theorems 2–4, we have established that each working subspace covers at least a portion of $\mathcal{S}_{Y_\infty|\mathbf{X}}$. When the dimension $d^* = \dim(\mathcal{S}_{Y_\infty|\mathbf{X}}) = 1$, the three working subspaces are exhaustive if they are non-degenerate. However, when $d^* > 1$, the dimension-deficient $\mathcal{S}_{\text{eSIR}}$ is definitely not exhaustive since it has at most one direction. For $\mathcal{S}_{\text{eSAVE}}$ and $\mathcal{S}_{\text{eDR}}$, they still reflect important information in the extreme regression if they are only subsets of $\mathcal{S}_{Y_\infty|\mathbf{X}}$. In practice, the exhaustiveness assumption for $\mathcal{S}_{\text{eSAVE}}$ and $\mathcal{S}_{\text{eDR}}$ is often mild. In real-world applications, it is often assumed that d^* is small, allowing $\mathcal{S}_{\text{eSAVE}}$ and $\mathcal{S}_{\text{eDR}}$ to potentially recover all directions. The exhaustiveness is often assumed in SDR literature (Cook & Ni 2005, 2006, Li &

Wang 2007) and commonly regarded as mild. In addition, the exhaustiveness of $\mathcal{S}_{\text{eSAVE}}$ and $\mathcal{S}_{\text{eDR}}$ has also been supported by our empirical findings in Section S4.3 of the Supplementary Material, where COPES-SAVE and COPES-DR provide the accurate dimension determination for CES. To fix our focus, we assume throughout the paper that $\mathcal{S}_{\text{eSAVE}} = \mathcal{S}_{Y_\infty|\mathbf{X}}$ and $\mathcal{S}_{\text{eDR}} = \mathcal{S}_{Y_\infty|\mathbf{X}}$ so that the working subspaces in COPES-SAVE and COPES-DR reflect the complete information in $\mathcal{S}_{Y_\infty|\mathbf{X}}$.

Moreover, when the EC distribution assumption is violated, we are still able to estimate working subspaces. In fact, the existence of the CES and three working subspaces is independent of the EC distribution assumption. Without the EC distribution assumption, we can still construct the contour-projected predictors $\overrightarrow{\mathbf{X}} = \{\mathbf{X} - \mathbb{E}(\mathbf{X})\}/\|\mathbf{X} - \mathbb{E}(\mathbf{X})\|_{\text{Cov}(\mathbf{X})}$, given that the first two moments of $\mathbf{X}$ exist, and the approximation results in Lemmas 2, 5, and 8 remain valid. The EC distribution is required only when one connects working subspaces to the CES (cf. Lemma 1, Theorems 2, 3, and 4). In Section S4.5 of the Supplementary Material, we also demonstrate through empirical results that COPES methods remain high estimation accuracy even when the EC distribution assumption is violated.

It is also possible to develop new methods based on other linearity-condition-free SDR methods, if the linearity condition is a concern. For

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0159

instance, non-parametric methods (e.g., Xia et al. 2002), semi-parametric methods (e.g., Ma & Zhu 2012, Huang & Chiang 2017), and distance-based methods (e.g., Sheng & Yin 2016) are free from the linearity condition. We leave the development of these methods as a future direction.

5. Estimation and asymptotic theory

In this section, we present the estimators for three specific COPES methods, i.e., COPES-SIR, COPES-SAVE, and COPES-DR, along with a study of their asymptotic properties. For the exhaustive COPES-SAVE and COPES-DR, we develop their estimation and theoretical results given the true d^*. When d^* is unknown, we also provide an approach to determine it, elaborated in Section 5.4.

Let $(\boldsymbol{x}_i, y_i)$, $i = 1, \ldots, n$, denote i.i.d. samples. The first step in COPES is to construct the contour-projected predictors, for which we need estimators of the unknown $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Once the corresponding estimators $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ are obtained, we construct the sample contour-projected predictors as $\overrightarrow{\boldsymbol{x}}_i = (\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}})/\|\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}\|_{\widehat{\boldsymbol{\Sigma}}}$, for $i = 1, \ldots, n$. We adopt the $\sqrt{n}$-consistent estimators for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ from Luo et al. (2009). The estimator $\widehat{\boldsymbol{\mu}}$ is the element-wise median of samples $\boldsymbol{x}_i$, $i = 1, \ldots, n$, and is $\sqrt{n}$-consistent. The estimator $\widehat{\boldsymbol{\Sigma}}$ is obtained from an iterative algorithm. It has been demonstrated in Luo et al. (2009) that when $\boldsymbol{x}_i$ follows an EC distribution, the iterative algorithm is guaranteed to converge with probability one, leading to a $\sqrt{n}$-consistent estimator $\widehat{\boldsymbol{\Sigma}}$.

5.1 Estimation of extreme SIR subspace

In COPES-SIR, the kernel matrix $\mathbf{M}^{h}_{\text{eSIR}}$ is indexed by h. As we have shown in Lemma 2, h is expected to converge to zero for the recovery of $\mathcal{S}_{\text{eSIR}}$. In this section, in the context of n i.i.d. samples, we replace h with k/n, where k denotes the actual number of samples used in the estimation. We choose $k = k(n)$ such that $k \to \infty$ and $k/n \to 0$ as $n \to \infty$.

Let $\widetilde{y}_i = -y_i$, $i = 1, \ldots, n$. For the vector-valued function $\mathbf{C}_{k/n}(u)$, we estimate it by $\widehat{\mathbf{C}}_{k/n}(u) = (1/k)\sum_{i=1}^{n} \overrightarrow{\boldsymbol{x}}_i I\{\widetilde{y}_i \le \widehat{F}^{-1}(uk/n)\}$, where $\widehat{F}(t)$ is the empirical cumulative distribution function computed from $\widetilde{y}_i$, defined as $\widehat{F}(t) = n^{-1}\sum_{i=1}^{n} I(\widetilde{y}_i \le t)$. The inverse of $\widehat{F}(t)$ is $\widehat{F}^{-1}(x) = \inf\{t : \widehat{F}(t) \ge x\}$ for $0 \le x \le 1$. For simplicity, we assume that the responses $\widetilde{y}_i$, $i = 1, \ldots, n$, are all distinct. Let $\widetilde{y}_{(i)}$ denote the sorted responses for $\widetilde{y}_i$, $i = 1, \ldots, n$, such that $\widetilde{y}_{(1)} < \widetilde{y}_{(2)} < \cdots < \widetilde{y}_{(n)}$, and let $\overrightarrow{\boldsymbol{x}}_{(i)}$ denote the correspondingly sorted contour-projected predictors. Then, when u is in $((m-1)/k, m/k]$, the estimator $\widehat{\mathbf{C}}_{k/n}(u) = \widehat{\mathbf{C}}_{k/n}(m/k) = k^{-1}\sum_{i=1}^{m} \overrightarrow{\boldsymbol{x}}_{(i)}$, for $m = 1, \ldots, k$. Accordingly, the estimator for the kernel matrix $\mathbf{M}^{k/n}_{\text{eSIR}}$ is

$$\widehat{\mathbf{M}}^{k/n}_{\text{eSIR}} = \int_0^1 \widehat{\mathbf{C}}_{k/n}(u)\widehat{\mathbf{C}}^{\top}_{k/n}(u)du = \frac{1}{k}\sum_{m=1}^{k} \widehat{\mathbf{C}}_{k/n}(m/k)\,\widehat{\mathbf{C}}^{\top}_{k/n}(m/k)\,.$$

Since $\dim(\mathcal{S}_{\text{eSIR}}) = 1$, we estimate the basis of $\mathcal{S}_{\text{eSIR}}$ by taking the leading left singular vector of $\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{M}}_{\text{eSIR}}^{k/n}$, that is, $\widehat{\boldsymbol{\beta}}_{\text{eSIR}}^{k/n} = \text{SVD}_1(\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{M}}_{\text{eSIR}}^{k/n})$.

5.2 Estimations of extreme SAVE and DR subspaces

The estimations for $\mathcal{S}_{\text{eSAVE}}$ and $\mathcal{S}_{\text{eDR}}$ are conducted in a similar manner as that for $\mathcal{S}_{\text{eSIR}}$. We begin by estimating the matrix-valued functions $\mathbf{T}_{k/n}(u)$ and $\mathbf{D}_{k/n}(u)$ by $\widehat{\mathbf{T}}_{k/n}(u) = (1/k)\sum_{i=1}^{n} \overrightarrow{\boldsymbol{x}}_i \overrightarrow{\boldsymbol{x}}_i^{\top} I\{\widetilde{y}_i \leq \widehat{F}^{-1}(uk/n)\}$ and $\widehat{\mathbf{D}}_{k/n}(u) = \widehat{\mathbf{T}}_{k/n}(u) - \text{median}\{\boldsymbol{\sigma}(\widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\mathbf{T}}_{k/n}(u)\widehat{\mathbf{\Sigma}}^{-1/2})\}\widehat{\mathbf{\Sigma}}$. The function $\widehat{\mathbf{T}}_{k/n}(u)$ is constant with respect to u on $((m-1)/k, m/k]$, for $m = 1, \ldots, k$, given by $\widehat{\mathbf{T}}_{k/n}(u) = \widehat{\mathbf{T}}_{k/n}(m/k) = k^{-1}\sum_{i=1}^{m} \overrightarrow{\boldsymbol{x}}_{(i)} \overrightarrow{\boldsymbol{x}}_{(i)}^{\top}$. The function $\widehat{\mathbf{D}}_{k/n}(u)$ is also constant within $((m-1)/k, m/k]$, for $m = 1, \ldots, k$. Subsequently, we define the estimator for the kernel matrix $\mathbf{M}_{\text{eSAVE}}^{k/n}$ as

$$\widehat{\mathbf{M}}_{\text{eSAVE}}^{k/n} = \int_0^1 \widehat{\mathbf{D}}_{k/n}(u)\widehat{\mathbf{D}}_{k/n}^{\top}(u)du = \frac{1}{k}\sum_{m=1}^{k} \widehat{\mathbf{D}}_{k/n}\left(\frac{m}{k}\right)\widehat{\mathbf{D}}_{k/n}^{\top}\left(\frac{m}{k}\right).$$

Given the oracle knowledge of the dimension d^*, we estimate the basis of $\mathcal{S}_{\text{eSAVE}}$ by taking the leading d^* left singular vectors of $\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{M}}_{\text{eSAVE}}^{k/n}$ as $\widehat{\boldsymbol{\beta}}_{\text{eSAVE}}^{k/n} = \text{SVD}_{d^*}(\widehat{\mathbf{\Sigma}}^{-1}\widehat{\mathbf{M}}_{\text{eSAVE}}^{k/n})$.

Next, we estimate the kernel matrix $\mathbf{M}_{\text{eDR}}^{k/n}$. As outlined in Section 4.4, $\mathbf{M}_{\text{eDR}}^{k/n}$ can be re-expressed as the sum of different integrals of $\mathbf{C}_{k/n}(u)$ and $\mathbf{D}_{k/n}(u)$. Then we estimate $\mathbf{M}_{\text{eDR}}^{k/n}$ by $\widehat{\mathbf{M}}_{\text{eDR}}^{k/n} = 2\sum_{i=1}^{8} \widehat{\mathbf{K}}_i^{k/n}$, where each component $\widehat{\mathbf{K}}_i^{k/n}$, $i = 1, \ldots, 8$, is obtained by substituting $\mathbf{C}_{k/n}(u)$ and $\mathbf{D}_{k/n}(u)$ with the corresponding sample estimators. The specific formula for $\widehat{\mathbf{K}}_i^{k/n}$ can be found in Section S1 of the Supplementary Material. With the known d^*, the basis of $\mathcal{S}_{\mathrm{eDR}}$ is estimated by $\widehat{\boldsymbol{\beta}}^{k/n}_{\mathrm{eDR}} = \mathrm{SVD}_{d^*}(\widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\mathbf{M}}^{k/n}_{\mathrm{eDR}})$.

5.3 Asymptotic theory

In this section, we investigate the asymptotic property of the estimated kernel matrix $\widehat{\mathbf{M}}^{k/n}_{\mathrm{eDR}}$ and the estimated extreme DR subspace $\mathrm{span}(\widehat{\boldsymbol{\beta}}^{k/n}_{\mathrm{eDR}})$. The asymptotic theory for COPES-SIR and COPES-SAVE is similar to that for COPES-DR and is gathered in Section S3 of the Supplementary Material. Before presenting our asymptotic result, we introduce the following technical assumption to ensure the $\sqrt{n}$-consistency of $\widehat{\boldsymbol{\Sigma}}$.

(A3) Assume that $\|\mathbf{X}\|^2$ has a continuous distribution with probability density $f(\cdot)$. Further assume that there exist some constants $\alpha > 1$ and $C_\alpha > 0$ such that $t^{-\alpha} f(t) \to C_\alpha$ as $t \to 0$.

This assumption is quite mild and satisfied by many EC distributions, including multivariate normal and multivariate t distributions. We now derive the asymptotic property for the estimated kernel matrix and the corresponding estimated working subspace.

Theorem 5. *Assume that* $\mathbf{X}$ *follows an EC distribution and Assumptions (A1)–(A3) hold. Moreover, assume that* $\mathrm{Cov}(\overrightarrow{\mathbf{X}}|Y > y)$ *and* $\mathrm{Cov}\{\mathrm{vec}(\overrightarrow{\mathbf{X}}\overrightarrow{\mathbf{X}}^\top)|Y > y\}$ *converge as* $y \to y^+$*, where* $\mathrm{vec}(\cdot)$ *is the vectorization operator concatenating all columns of a matrix. As* $n \to \infty$*, we have (i)* $\|\widehat{\mathbf{M}}^{k/n}_{\mathrm{eDR}} - \mathbf{M}^{k/n}_{\mathrm{eDR}}\|_F = O_P(k^{-1/2})$*; (ii)* $\|\widehat{\mathbf{M}}^{k/n}_{\mathrm{eDR}} - \mathbf{M}_{\mathrm{eDR}}\|_F = o_P(1)$*. Moreover, under the additional* *assumption that* $\dim(\mathcal{S}_{\text{eDR}}) < p/2 - 1$, *we have (iii)* $\|\mathbf{P}_{\widehat{\boldsymbol{\beta}}^{k/n}_{\text{eDR}}} - \mathbf{P}_{\mathcal{S}_{Y_\infty|\mathbf{X}}}\|_F = o_P(1)$ *as* $n \to \infty$.

The statement (i) in Theorems 5 suggests that the difference between the estimated kernel matrix $\widehat{\mathbf{M}}^{k/n}_{\text{eDR}}$ and its population counterpart $\mathbf{M}^{k/n}_{\text{eDR}}$ is $\sqrt{k}$-consistent. By statement (ii), the estimated kernel matrix is a consistent estimator for the limit matrix $\mathbf{M}_{\text{eDR}}$. We have also shown in statement (iii) that the estimated subspace $\text{span}(\widehat{\boldsymbol{\beta}}^{k/n}_{\text{eDR}})$ is consistent estimator for $\mathcal{S}_{\text{eDR}}$. However, claiming specific convergence rates for $\|\widehat{\mathbf{M}}^{k/n}_{\text{eDR}} - \mathbf{M}_{\text{eDR}}\|_F$ and $\|\mathbf{P}_{\widehat{\boldsymbol{\beta}}^{k/n}_{\text{eDR}}} - \mathbf{P}_{\mathcal{S}_{Y_\infty|\mathbf{X}}}\|_F$ requires additional assumptions on the convergence of $\mathbf{M}^{k/n}_{\text{eDR}}$. To understand this more clearly, note that $\|\widehat{\mathbf{M}}^{k/n}_{\text{eDR}} - \mathbf{M}_{\text{eDR}}\|_F \leq \|\widehat{\mathbf{M}}^{k/n}_{\text{eDR}} - \mathbf{M}^{k/n}_{\text{eDR}}\|_F + \|\mathbf{M}^{k/n}_{\text{eDR}} - \mathbf{M}_{\text{eDR}}\|_F$. Thus, the convergence rate of $\|\widehat{\mathbf{M}}^{k/n}_{\text{eDR}} - \mathbf{M}_{\text{eDR}}\|_F$ is determined by $\|\widehat{\mathbf{M}}^{k/n}_{\text{eDR}} - \mathbf{M}^{k/n}_{\text{eDR}}\|_F$ and $\|\mathbf{M}^{k/n}_{\text{eDR}} - \mathbf{M}_{\text{eDR}}\|_F$. In Section S3 of the Supplementary Material, we show that $\widehat{\mathbf{M}}^{k/n}_{\text{eDR}}$ and $\text{span}(\widehat{\boldsymbol{\beta}}^{k/n}_{\text{eDR}})$ exhibit $\sqrt{k}$-consistency in estimating $\mathbf{M}_{\text{eDR}}$ and $\mathcal{S}_{Y_\infty|\mathbf{X}}$ under additional convergence assumptions.

5.4 Determination of structural dimension

We provide a dimension determination approach by exploiting the spectral structures of the kernel matrices in $\mathcal{S}_{\text{eSAVE}}$ and $\mathcal{S}_{\text{eDR}}$, which recover $\mathcal{S}_{Y_\infty|\mathbf{X}}$ exhaustively. We take $\mathcal{S}_{\text{eDR}}$ as an example. Recall that under the method-specific assumptions, By Lemma 8, $\mathcal{S}_{\text{eDR}} = \boldsymbol{\Sigma}^{-1}\text{span}(\mathbf{M}_{\text{eDR}})$, then $d^* =$

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0159

$\max\{i \mid \sigma_i(\mathbf{M}_{\text{eDR}}) > 0\}$.

For some constant $\varepsilon > 0$, we define the adjusted ratio of the consecutive singular values of $\mathbf{M}_{\text{eDR}}$ as

$$r_i = \{\sigma_i(\mathbf{M}_{\text{eDR}}) + \varepsilon\}/\{\sigma_{i+1}(\mathbf{M}_{\text{eDR}}) + \varepsilon\}, \quad i = 1, \ldots, \lfloor p/2 \rfloor - 1,$$

where $\lfloor a \rfloor$ denotes the largest integer less or equal to a. By taking sufficiently small ε, we have $d^* = \operatorname{argmax}_i \{r_i\}$. Intuitively, we estimate d^* by $\widehat{d} = \operatorname{argmax}_i \{\widehat{r}_i\}$, where

$$\widehat{r}_i := \{\sigma_i(\widehat{\mathbf{M}}^{k/n}_{\text{eDR}}) + \varepsilon\}/\{\sigma_{i+1}(\widehat{\mathbf{M}}^{k/n}_{\text{eDR}}) + \varepsilon\}, \quad i = 1, \ldots, \lfloor p/2 \rfloor - 1.$$

The following result shows that $\widehat{d}$ is a consistent estimator for d^*.

Lemma 9. *Assume the same assumptions as in Theorem 5 and* $d^* < p/2 - 1$. *By taking the constant* $\varepsilon > 0$ *such that* $\varepsilon\{\sigma_1(\mathbf{M}_{\text{eDR}}) - 2\sigma_{d^*}(\mathbf{M}_{\text{eDR}})\} < \sigma^2_{d^*}(\mathbf{M}_{\text{eDR}})$, *we have* $\widehat{d} \to d^*$ *in probability as* $n \to \infty$.

The dimension determination based on $\mathcal{S}_{\text{eSAVE}}$ is similar to that based on $\mathcal{S}_{\text{eDR}}$. As supported by our empirical findings in Section S4.3 of the Supplementary Material, our proposed dimension determination procedure consistently estimates the structural dimension, especially when sample size n is large.

6. Simulation studies

We generate the predictors $\mathbf{X} = (X_1, \ldots, X_p)^\top$ from the multivariate t distribution and set $p = 10$. Specifically, we construct $\mathbf{X} = \mathbf{W}/\sqrt{u/\nu}$, where $\mathbf{W} \sim N(\mathbf{0}, \mathbf{I}_p)$, $u \sim \chi^2_\nu$, and u is independent of $\mathbf{W}$. We consider degrees of freedom $\nu = 2, 3, 5$, and ∞, representing different magnitudes of heavy-tailedness. Specifically, $\mathbf{X}$ reduces to the multivariate normal random vector when $\nu = \infty$. Let ξ_1 and ξ_2 be two random variables following the Pareto distribution such that $P(\xi_1 > t) = (1+t)^{-1}$ and $P(\xi_2 > t) = (1+t)^{-2}$, for $t \geq 0$. Assume that the random elements $\{\mathbf{X}, \xi_1, \xi_2\}$ are mutually independent.

Let B denote a Bernoulli random variable with a success probability of 0.5, independent of $\{\mathbf{X}, \xi_1, \xi_2\}$. Let U denote a uniform random variable supported on $[0, 1]$, independent of $\mathbf{X}$. Additionally, let F_{ξ_i} denote the cumulative distribution function for the variable ξ_i, $i = 1, 2$. We consider four different models, described in the following.

(Model A) $Y = B\xi_1 \sin(X_1/2) + (1 - B)\xi_2 \sin(X_2/2)$.
(Model B) $Y = B\xi_1\{4X_1 + (X_2 + 0.25)^2\} + (1 - B)\xi_2/\{(X_3 + 0.25)^2 + 0.25\}$.
(Model C) $Y = F^{-1}_{\xi_1}(U)\{\cos(X_1 + \pi/4) + I(U \leq 0.95)\cos(X_2 + \pi/4)\}$.
(Model D) $Y = F^{-1}_{\xi_1}(U)\sin(X_1)/\{(X_2 + 0.25)^2 + 0.25\} + I(U \leq 0.95)\, F^{-1}_{\xi_2}(U)\sin(X_3)$.

In Models A and C, we have $\mathcal{S}_{Y_\infty|\mathbf{X}} = \text{span}(\mathbf{e}_1)$ and $d^* = 1$, and in Models B and D, we have $\mathcal{S}_{Y_\infty|\mathbf{X}} = \text{span}(\mathbf{e}_1, \mathbf{e}_2)$ and $d^* = 2$. In each model, we generate i.i.d. samples $(y_i, \boldsymbol{x}_i)$, $i = 1, \ldots, n$, where $n = 5000$. We compare our three COPES methods, COPES-SIR, COPES-SAVE, and COPES-DR, with two other extreme dimension reduction approaches, namely, TIREX1

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0159

and TIREX2. In this section, we assume that the dimension d^* is given. Note that in Models B and D, TIREX1 and COPES-SIR can only estimate one direction. Despite this limitation, we include these two methods in the comparison under Models B and D for demonstration purposes. Additionally, we estimate the rank-d^* subspace for TIREX2.

We measure the subspace estimation error via the subspace distance. Given two basis matrices $\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta} \in \mathbb{R}^{p\times d}$, the subspace distance between $\mathrm{span}(\widehat{\boldsymbol{\beta}})$ and $\mathrm{span}(\boldsymbol{\beta})$ is defined as $D(\widehat{\boldsymbol{\beta}}, \boldsymbol{\beta}) = \|\mathbf{P}_{\widehat{\boldsymbol{\beta}}} - \mathbf{P}_{\boldsymbol{\beta}}\|_F^2/(2d)$. Similar definition of subspace distance is adopted in SDR literature (Tan et al. 2020, Zeng et al. 2024). We report the mean squared error (MSE) of the subspace estimation for each method over 200 data replicates.

In Figure 1, we present the MSE results versus k for Model B with different ν's. The results under other models are displayed in Section S4.2 of the Supplementary Material. In general, the curve of COPES decreases as k increases up to some point, and then it starts to increase. This phenomenon is commonly observed in extreme value analysis. When k is small, the used samples reflect the extreme nature more accurately, resulting in low bias but high variance in the estimation. With larger k, since more samples are exploited, the estimation variance is then reduced. However, this potentially introduces bias, which gradually becomes the dominant term in the

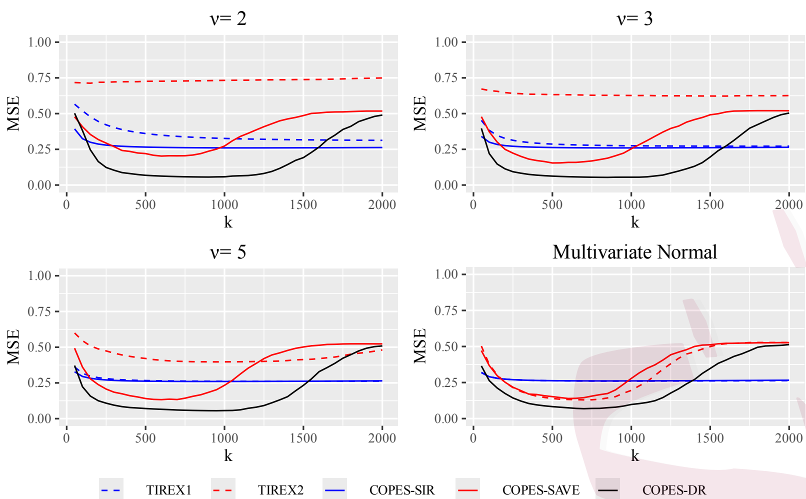

Figure 1: MSEs for different competitors under various ν's under Model B.

estimation error. Across all models, when $\mathbf{X}$ exhibits heavy-tailedness, i.e., when ν is small, TIREX1 and TIREX2 are inferior to our proposal. As ν increases from 2 to ∞, the estimation of COPES is relatively stable. In comparison, TIREX1 and TIREX2 are sensitive to the tails. When $\nu = \infty$, TIREX1 and TIRES2 perform similarly to COPES-SIR and COPES-SAVE, respectively. In this sense, our COPES methods are more robust against heavy-tailedness. In Models B and D, where $d^* = 2$, since COPES-SIR and TIREX1 only estimate one direction, perform the worst in all situations.

We observe that our COPES methods partly inherit the characteristics of their counterparts in SDR; more detailed discussion is provided in Section S4.1 of the Supplementary Material. In addition, the accuracy of our

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0159

dimension determination procedure and the robustness of COPES against the violation of the EC distribution assumption are supported by the empirical results in Sections S4.3 and S4.5.

7. Application

In this section, we demonstrate the effectiveness of our methodology using an example from the Chinese stock market. The dataset contains accounting information for Chinese stock market firms from 1997 to 2000, previously studied by Luo et al. (2009) and Wang & Tsai (2009). The data set comprises 2951 observations, where the original response variable Y_0 represents the next year's return on equity (ROEt) of the firm. To be consistent with notations in our theory, let $Y = -Y_0$, meaning the loss of the firm in the next year. The predictors $\mathbf{X} \in \mathbb{R}^6$ are current year accounting variables, including return on equity (ROE), log-transformed total assets (ASSET), profit margin ratio (PM), sales growth rate (GROWTH), leverage level (LEV), and asset turnover ratio (ATO). Our focus lies in the right tail of the response, denoted by $Y|(Y > y_0)$, where y_0 is selected as the 90% quantile of Y.

In this study, we compare our proposed methods, COPES-SIR, COPES-SAVE, and COPES-DR, with TIREX1 and TIREX2. Additionally, we include robust SDR methods, CP-SIR, CP-SAVE, and CP-DR, proposed by

Luo et al. (2009), into comparison. First, we reduce the dimension of $\mathbf{X}$ by projecting it onto the estimated subspace from each competitor. Subsequently, we employ the K-nearest neighbor (KNN) algorithm with $K = 10$ to predict the event $\{Y > y_0\}$ using the reduced predictors, and we measure the performance using the AUC score as the evaluation metric. Additionally, to verify the effectiveness of dimension reduction, we implement the classification without dimension reduction (WODR).

The dataset is randomly split into a training set and a testing set at a ratio of 75/25 for 100 replicates. For COPES methods and TIREX methods, the parameter k is selected via five-fold cross-validation on the training set. The optimal k is determined as the one that yields the highest averaged AUC score across all five folds. Using the selected optimal k, each competitor is evaluated on the independent test set through the AUC score. Notably, when $d^* = 1$, all methods exhibit poor performance. Thus, we set $d^* = 2$ for all competitors. Since COPES-SIR and TIREX1 can at most recover one direction, we omit their results. The averaged AUC scores and standard deviations for the competitors are presented in Table 1.

From Table 1, it is evident that all the dimension reduction methods exhibit higher AUC scores than WODR, indicating that proper dimension reduction improves classification. Moreover, COPES methods outperform

Table 1: Averaged AUC scores (and standard errors) on Chinese stock market data by using the KNN and AdaBoost algorithms.

Classifier	COPES-SAVE	COPES-DR	TIREX2	WODR	CP-SIR	CP-SAVE	CP-DR
KNN	**0.737 (0.003)**	0.721 (0.003)	0.637 (0.007)	0.687 (0.004)	0.727 (0.003)	0.716(0.003)	0.716 (0.003)
AdaBoost	**0.774 (0.002)**	0.757 (0.003)	0.639 (0.008)	0.757 (0.004)	0.762 (0.003)	0.757(0.003)	0.756 (0.004)

Table 2: The estimated directions from COPES-SAVE on Chinese stock market data.

	ROE	ASSET	PM	GROWTH	LEV	ATO
$\widehat{\boldsymbol{\beta}}_1$	-0.980	0.001	0.130	-0.029	0.110	0.093
$\widehat{\boldsymbol{\beta}}_2$	-0.088	0.130	0.250	0.137	-0.288	0.901

SDR methods, with COPES-SAVE achieving the highest AUC score. This suggests that SDR is less efficient on this dataset since the central subspace might contain redundant information irrelevant to the inference of the conditional extremes of $Y \mid \mathbf{X}$. In addition, by replacing the KNN with the AdaBoost in classification, our proposal still performs the best as indicated by the second row of Table 1.

In Table 2, we present the two directions, $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$, estimated by COPES-SAVE on the original dataset. It can be seen that ROE and ATO have relatively larger absolute coefficients and contribute the most to the association with the firm's tail risk. The results in Luo et al. (2009) suggest that ROE, PM, and ATO are the most important variables in SDR. This interesting finding implies that PM is possibly a redundant variable for inferring the extremely high values of the response. Moreover, we observed the heavy-tailedness of ROE and ATO, which partly explains the advantage of our robust COPES methods over the non-robust TIREX methods.

Visualization is another intriguing feature of COPES. We display the scatter plot of the reduced predictors $\widehat{\boldsymbol{\beta}}_1^\top \mathbf{X}$ and $\widehat{\boldsymbol{\beta}}_2^\top \mathbf{X}$ from COPES-SAVE in Figure S11 of the Supplementary Material. The distribution pattern of the two classes is well observed. It can be seen that most of the tail observations are located in the right-upper part of the plots.

Supplementary Material

The Supplementary Material includes additional discussions, theories, numerical results, and technical proofs.

Acknowledgments

The authors are grateful to the Editor, Associate Editor, and two anonymous referees, whose suggestions led to great improvement of this work. All authors contributed equally and are listed in alphabetical order. Liujun Chen's research was partially supported by Grants 12301387 and 12471279 from National Natural Science Foundation of China (NNSFC). Jing Zeng's research was partially supported by Grant 12301365 from NNSFC and Grant WK2040000075 from Fundamental Research Funds for the Central

Universities.

References

Aghbalou, A., Portier, F., Sabourin, A. & Zhou, C. (2024), 'Tail inverse regression: dimension reduction for prediction of extremes', *Bernoulli* **30**(1), 503–533.

Arnosti, N., Beck, M. & Milgrom, P. (2016), 'Adverse selection and auction design for internet display advertising', *American Economic Review* **106**(10), 2852–2866.

Bousebata, M., Enjolras, G. & Girard, S. (2023), 'Extreme partial least-squares', *Journal of Multivariate Analysis* **194**, 105101.

Bura, E., Duarte, S. & Forzani, L. (2016), 'Sufficient reductions in regressions with exponential family inverse predictors', *Journal of the American Statistical Association* **111**(515), 1313–1329.

Chen, X., Zhang, J. & Zhou, W. (2022), 'High-dimensional elliptical sliced inverse regression in non-gaussian distributions', *Journal of Business & Economic Statistics* **40**(3), 1204–1215.

Chen, X., Zou, C. & Cook, R. D. (2010), 'Coordinate-independent sparse sufficient dimension reduction and variable selection', *The Annals of Statistics* **38**(6), 3696–3723.

Chernozhukov, V. (2005), 'Extramal quantile regression', *The Annals of Statistics* **33**(2), 806–839.

Cook, R. D. (1998), *Regression Graphics: Ideas for Studying Regressions Through Graphics*, Vol. 318, John Wiley & Sons.

Cook, R. D. & Forzani, L. (2008), 'Principal Fitted Components for Dimension Reduction in Regression', *Statistical Science* **23**(4), 485–501.

Cook, R. D. & Li, B. (2002), 'Dimension reduction for conditional mean in regression', *The Annals of Statistics* **30**(2), 455–474.

Cook, R. D. & Ni, L. (2005), 'Sufficient dimension reduction via inverse regression: A minimum discrepancy approach', *Journal of the American Statistical Association* **100**(470), 410–428.

Cook, R. D. & Ni, L. (2006), 'Using intraslice covariances for improved estimation of the central subspace in regression', *Biometrika* **93**(1), 65–74.

Cook, R. D. & Weisberg, S. (1991), 'Discussion of sliced inverse regression for dimension reduction', *Journal of the American Statistical Association* **86**(414), 328–332.

Eaton, M. L. (1986), 'A characterization of spherical distributions', *Journal of Multivariate Analysis* **20**(2), 272–276.

Fan, J., Wang, W. & Zhu, Z. (2021), 'A shrinkage principle for heavy-tailed data: High-dimensional robust low-rank matrix recovery', *The Annals of Statistics* **49**(3), 1239–1266.

Fukumizu, K. & Leng, C. (2014), 'Gradient-based kernel dimension reduction for regression', *Journal of the American Statistical Association* **109**(505), 359–370.

Gardes, L. (2018), 'Tail dimension reduction for extreme quantile estimation', *Extremes* **21**, 57–95.

Girard, S., Stupfler, G. & Usseglio-Carleve, A. (2021), 'Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models', *The Annals of Statistics* **49**(6), 3358–3382.

Hall, P., Titterington, D. & Xue, J.-H. (2009), 'Median-based classifiers for high-dimensional data', *Journal of the American Statistical Association* **104**(488), 1597–1608.

Hall, P. & Weissman, I. (1997), 'On the estimation of extreme tail probabilities', *The Annals of Statistics* **25**(3), 1311–1326.

Huang, M.-Y. & Chiang, C.-T. (2017), 'An effective semiparametric estimation approach for the sufficient dimension reduction model', *Journal of the American Statistical Association* **112**(519), 1296–1310.

Johnson, M. E. (1987), *Multivariate statistical simulation: A guide to selecting and generating continuous multivariate distributions*, Vol. 192, John Wiley & Sons.

Li, B. (2018), *Sufficient dimension reduction: Methods and applications with R*, Chapman and Hall/CRC.

Li, B. & Wang, S. (2007), 'On directional regression for dimension reduction', *Journal of the American Statistical Association* **102**(479), 997–1008.

Li, K.-C. (1991), 'Sliced inverse regression for dimension reduction', *Journal of the American Statistical Association* **86**(414), 316–327.

Loh, P.-L. (2017), 'Statistical consistency and asymptotic normality for high-dimensional robust M-estimators', *The Annals of Statistics* **45**(2), 866–896.

Luo, R., Wang, H. & Tsai, C.-L. (2009), 'Contour projected dimension reduction', *The Annals of Statistics* **37**(6), 3743–3778.

Ma, Y. & Zhu, L. (2012), 'A semiparametric approach to dimension reduction', *Journal of the American Statistical Association* **107**(497), 168–179.

Ma, Y. & Zhu, L. (2013), 'A review on dimension reduction', *International Statistical Review* **81**(1), 134–150.

Resnick, S. I. (2007), *Heavy-tail phenomena: probabilistic and statistical modeling*, Springer Science & Business Media.

Sheng, W. & Yin, X. (2016), 'Sufficient dimension reduction via distance covariance', *Journal of Computational and Graphical Statistics* **25**(1), 91–104.

Tan, K., Shi, L. & Yu, Z. (2020), 'Sparse SIR: Optimal rates and adaptive estimation', *The Annals of Statistics* **48**(1), 64–85.

Wang, H., Ni, L. & Tsai, C.-L. (2008), 'Improving dimension reduction via contour-projection', *Statistica Sinica* **18**(1), 299–311.

Wang, H. & Tsai, C.-L. (2009), 'Tail index regression', *Journal of the American Statistical Association* **104**(487), 1233–1240.

Xia, Y., Tong, H., Li, W. K. & Zhu, L.-X. (2002), 'An adaptive estimation of dimension reduction space', *Journal of the Royal Statistical Society Series B: Statistical Methodology* **64**(3), 363–410.

Xu, W., Wang, H. J. & Li, D. (2022), 'Extreme quantile estimation based on the tail single-index model', *Statistica Sinica* **32**(2), 893–914.

Yu, Z., Dong, Y. & Zhu, L.-X. (2016), 'Trace pursuit: A general framework for model-free variable selection', *Journal of the American Statistical Association* **111**(514), 813–821.

Zeng, J., Mai, Q. & Zhang, X. (2024), 'Subspace estimation with automatic dimension and variable selection in sufficient dimension reduction', *Journal of the American Statistical Association* **119**(545), 343–355.

Zhao, Z., Zhang, Z. & Chen, R. (2018), 'Modeling maxima with autoregressive conditional fréchet model', *Journal of Econometrics* **207**(2), 325–351.

Zhu, L.-P., Zhu, L.-X. & Feng, Z.-H. (2010), 'Dimension reduction in regressions through cumulative slicing estimation', *Journal of the American Statistical Association* **105**(492), 1455–1466.

Zhu, Z. & Zhou, W. (2021), Taming heavy-tailed features by shrinkage, *in* 'Proceedings of The 24th International Conference on Artificial Intelligence and Statistics', Vol. 130 of *Proceedings of Machine Learning Research*, PMLR, pp. 3268–3276.

International Institute of Finance, School of Management, University of Science and Technology of China, Hefei, Anhui 230026, China.

E-mails: ljchen22@ustc.edu.cn; zengjxl@ustc.edu.cn