Review

Overview of High-Dimensional Measurement Error Regression Models

Jingxuan Luo [1], Lili Yue [2] and Gaorong Li [1,*] (ID)

[1]   School of Statistics, Beijing Normal University, Beijing 100875, China; luojingxuan@mail.bnu.edu.cn
[2]   School of Statistics and Data Science, Nanjing Audit University, Nanjing 211815, China; yuell91@nau.edu.cn
[*]   Correspondence: ligaorong@bnu.edu.cn

Abstract: High-dimensional measurement error data are becoming more prevalent across various fields. Research on measurement error regression models has gained increasing attention due to the risk of drawing inaccurate conclusions if measurement errors are ignored. When the dimension p is larger than the sample size n, it is challenging to develop statistical inference methods for high-dimensional measurement error regression models due to the existence of bias, nonconvexity of objective function, high computational cost and many other difficulties. Over the past few years, some works have overcome the aforementioned difficulties and proposed several statistical inference methods. This paper mainly reviews the current development on estimation, hypothesis testing and variable screening methods for high-dimensional measurement error regression models and shows the theoretical results of these methods with some directions worthy of exploring for future research.

Keywords: convex optimization; estimation; high-dimensional data; hypothesis testing; measurement error; variable selection

---

1. Introduction

Measurement error data inevitably exists in applications and has raised significant concerns in various fields including biology, medicine, epidemiology, economics, finance and remote sensing. So far, there have been a wealth of research achievements on classical low-dimensional measurement error regression models under various assumptions. Numerous studies focus on parameter estimation for low-dimensional measurement error regression models, with primary techniques listed below: (1) corrected regression estimation methods [1]; (2) Simulation-Extrapolation (SIMEX) estimation methods [2,3]; (3) deconvolution methods [4]; (4) corrected empirical likelihood methods [5,6]. For more detailed discussions on other estimation and hypothesis testing methods for classical low-dimensional measurement error models, please refer to the literature [7–29], as well as the monographs [30–35].

As one of the most popular research fields in statistics, high-dimensional regression has been widely used in various fields including genetics, economics, medical imaging, meteorology and sensor networks. Over the past two decades, various high-dimensional regression methods have been widely proposed such as Lasso [36], smoothly clipped absolute deviation (SCAD) [37], Elastic Net [38], Adaptive Lasso [39], Dantzig Selector [40], smooth integration of counting and absolute deviation (SICA) [41], minimax concave penalty (MCP) [42], and among many others. These methods have been widely applied to estimate regression coefficients while also achieving the goal of variable selection by adding penalties to objective functions, please refer to the literature review [43–45], as well as the monographs [46–48].

For the variable screening methods of ultrahigh-dimensional regression models where dimension p and sample size n satisfying $\log p = O(n^{\kappa}), \kappa > 0$, Fan and Lv [49] proposed the sure independence screening (SIS) method, which is a pioneering method in this field. For the estimation and variable selection of ultrahigh-dimensional regression models, it is suggested applying SIS method for variable screening first. Then, based on the variables screened in the first step, we can utilize regularization methods with penalties to estimate the regression coefficients and identify the significant variables

simultaneously. Due to the operability and effectiveness of SIS method in applications, numerous works have been done to extend the method, see [50–59].

However, most of the aforementioned theories and applications for high-dimensional regression models focused on clean data. In the era of big data, researchers frequently collect high-dimensional data with measurement errors. Typical instances include gene expression data [61] and sensor network data [60]. The imprecise measurements are the result of poorly managed and defective data collection processes, as well as the imprecise measuring instruments. It is well known that ignoring the influence of measurement errors will result in biased estimators and erroneous conclusions. Therefore, developing statistical inference methods for high-dimensional measurement error regression models have drawn a lot of interest.

Based on the types of measurement errors, research on high-dimensional measurement error regression models can be divided into the following three categories: covariates containing measurement errors; response variables containing measurement errors; both covariates and response variables containing measurement errors. In this paper, we mainly focus on the category that covariates contain measurement errors. When the dimension p is larger than the sample size n, parameter estimation can be challenging due to the nonconvexity of penalized objective function caused by correction for the bias. This further makes it impossible to obtain the optimal solution of optimization problem. We utilize the following linear regression model to illustrate this problem

$$y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $y = (y_1, \ldots, y_n)^{\mathrm{T}} \in \mathbb{R}^n$ is the $n \times 1$ response vector, $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)^{\mathrm{T}} \in \mathbb{R}^{n \times p}$ is the $n \times p$ fixed design matrix with $\mathbf{X}_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}} \in \mathbb{R}^p$ is the sparse regression coefficient vector with only s nonzero components, and assume that model error vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}} \in \mathbb{R}^n$ is independent of $\mathbf{X}$. In order to obtain a sparse estimator of the true regression coefficient vector $\boldsymbol{\beta}_0 = (\beta_{01}, \cdots, \beta_{0p})^{\mathrm{T}} \in \mathbb{R}^p$, we can minimize the following penalized least square objective function

$$\frac{1}{2n}\|y - \mathbf{X}\boldsymbol{\beta}\|_2^2 + \|p_\lambda(\boldsymbol{\beta})\|_1, \tag{2}$$

which is equivalent to minimizing

$$\frac{1}{2}\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\beta} - \boldsymbol{\rho}^{\mathrm{T}}\boldsymbol{\beta} + \|p_\lambda(\boldsymbol{\beta})\|_1, \tag{3}$$

where $\boldsymbol{\Sigma} = n^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{X}$, $\boldsymbol{\rho} = n^{-1}\mathbf{X}^{\mathrm{T}}y$, $p_\lambda(\cdot)$ is a penalty function with regularization parameter $\lambda \geq 0$. If the covariates matrix $\mathbf{X}$ can be precisely measured, the penalized objective functions (2) and (3) are convex. Thus, we can obtain a sparse estimator of $\boldsymbol{\beta}_0$ by minimizing the penalized objective function (2) or (3).

However, it is common that the covariates matrix $\mathbf{X}$ cannot be accurately observed in practice. Let $\mathbf{W} = (\mathbf{W}_1, \ldots, \mathbf{W}_n)^{\mathrm{T}} = (w_{ij})_{n \times p}$ be the observed covariates matrix with additive measurement errors satisfying $\mathbf{W} = \mathbf{X} + \mathbf{U}$, where $\mathbf{U} = (\mathbf{U}_1, \ldots, \mathbf{U}_n)^{\mathrm{T}}$ is the matrix of measurement errors, $\mathbf{U}_i = (u_{i1}, \cdots, u_{ip})^{\mathrm{T}}$ follows a sub-Gaussian distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}_u$, and it is assumed to be independent of $(\mathbf{X}, y)$. To reduce the influence of measurement errors, Loh and Wainwright [62] proposed to replace $\boldsymbol{\Sigma}$ and $\boldsymbol{\rho}$ in the penalized objective function (3) by their consistent estimators $\widehat{\boldsymbol{\Sigma}} = n^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{W} - \boldsymbol{\Sigma}_u$ and $\widetilde{\boldsymbol{\rho}} = n^{-1}\mathbf{W}^{\mathrm{T}}y$, respectively. Then we can obtain the sparse estimator of $\boldsymbol{\beta}_0$ by minimizing the following penalized objective function

$$\frac{1}{2}\boldsymbol{\beta}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}\boldsymbol{\beta} - \widetilde{\boldsymbol{\rho}}^{\mathrm{T}}\boldsymbol{\beta} + \|p_\lambda(\boldsymbol{\beta})\|_1. \tag{4}$$

Note that when the dimension p is fixed or smaller than the sample size n, it can be guaranteed that $\widehat{\boldsymbol{\Sigma}}$ is a positive definite or semi positive-definite matrix. It further ensures that the penalized objective

function (4) remains convex. Thus, the global optimal solution of β can be obtained by minimizing the penalized objective function (4).

However, for high-dimensional or ultrahigh-dimensional regression models, i.e., $p > n$ or $p \gg n$, there are two key problems: (i) the penalized objective function (4) is no longer convex and unbounded from below because the corrected estimator $\widehat{\Sigma}$ of Σ is no longer a semi-positive definite matrix. This further makes it impossible to obtain the estimator of β_0 by minimizing the penalty objective function (4); (ii) In order to construct an objective function similar to that of standard Lasso and solve the corresponding optimization problem using R package "glmnet" or "lars", it is necessary to decompose $\widehat{\Sigma}$ by Cholesky decomposition method and obtain the substitution of response vector and covariates matrix. However, this process results in an error accumulation and makes it challenging to guarantee valid theoretical results, and please see the detailed discussions in [63,64].

For problem (i), Loh and Wainwright [62] changed the unconstrained optimization problem into a constrained optimization problem by adding restrictions to β. They suggested applying the projected gradient descent algorithm to solve the restricted optimization problem and acquire the global optimal solution of true regression coefficient vector β_0. Nevertheless, the penalized objective function of the optimization problem is still nonconvex. To address this issue, Datta and Zou [63] suggested substituting $\widehat{\Sigma}$ by its semi-positive definite projection matrix $\widetilde{\Sigma}$, and they proposed convex conditioned Lasso (CoCoLasso). Further, Zheng et al. [64] introduced a balanced estimation that prevented overfitting while maintaining the estimation accuracy by combining l_1 and concave penalty. Tao et al. [65] constructed a modified least-squares loss function using a semi-positive definite projection matrix for estimated covariance matrix and proposed calibrated zero-norm regularized least squares (CaZnRLS) estimation of regression coefficients. Rosenbaum and Tsybakov [66,67] proposed a matrix uncertainty (MU) selector and its improved version compensated MU selector for high-dimensional linear models with additive measurement errors in covariates. Sørensen et al. [68] extended MU selector to generalized linear models and developed the generalized matrix uncertainty (GMU) selector. Sørensen et al. [69] showed the theoretical results of relevant variable selection methods. Based on MU selector, Belloni et al. [70] introduced an estimator that can achieve the minimax efficiency bound. They proved that the corresponding optimization problem can be converted into a second-order cone programming problem, which can be solved in polynomial time. Romeo and Thoresen [71] evaluated the performance of MU selector in [66], nonconvex Lasso in [62], and CoCoLasso in [63] using simulation studies. Brown et al. [72] proposed a path-following iterative algorithm called Measurement Error Boosting (MEBoost), which is a computationally effective method for variable selection in high-dimensional measurement error regression models. Nghiem and Potgieter [73] introduced a new estimation method called simulation-selection-extrapolation (SIMSELEX), which used Lasso in the simulation step and group Lasso in the selection step. Jiang and Ma [74] drew on the idea of nonconvex Lasso in [62] and proposed an estimator of the regression coefficients for high-dimensional Poisson models with measurement errors. Byrd and McGee [75] developed an iterative estimation method for high-dimensional generalized linear models with additive measurement errors based on the imputation-regularized optimization (IRO) algorithm in [76]. However, the error accumulation issue mentioned in problem (ii) has not been addressed in the literature.

The aforementioned works place more emphasis on estimation and variable selection problems rather than hypothesis testing. For high-dimensional regression models with clean data, research on hypothesis testing problems has made significant progress under various settings in [77–84]. For high-dimensional measurement error models, the hypothesis testing methods are equally crucial. However, the bias and instability caused by measurement errors make hypothesis testing extremely difficult. Recently, some progress has been achieved in statistical inference methods. Based on multiplier bootstrap, Belloni [85] constructed simultaneous confidence intervals for the target parameters in high-dimensional linear measurement error models. Focused on the case where a fixed number of covariates contain measurement errors, Li et al. [86] proposed a corrected decorrelated score test for parameters corresponding to the error-prone covariates and created asymptotic confidence

intervals for them. Huang et al. [87] proposed a new variable selection method based on debiased CoCoLasso and proved that it can achieve false discovery rate (FDR) control. Jiang et al. [88] developed Wald and score tests for high-dimensional Poisson measurement error models.

Compared to the above estimation and hypothesis testing methods, the screening techniques for ultrahigh-dimensional measurement error models is relatively few. Nghiem et al. [89] introduced two screening methods named corrected penalized marginal screening (PMSc) and corrected sure independence screening (SISc) for ultrahigh-dimensional linear measurement error models.

This paper gives an overview of the estimation and hypothesis testing methods for high-dimensional measurement error regression models, as well as the variable screening methods for ultrahigh-dimensional measurement error models. The rest of this paper is organized as follows. In Section 2, we review some estimation methods for linear models. We survey the estimation methods for generalized linear models in Section 3. Section 4 presents the recent advances in hypothesis testing methods for high-dimensional measurement error models. Section 5 introduces the variable screening techniques for ultrahigh-dimensional linear measurement error models. We conclude the paper with some discussions in Section 6.

Notations. Let $\mathbb{S}^p$ be the set of all $p \times p$ real symmetric matrices and $\mathbb{S}^p_+$ be the subset of $\mathbb{S}^p$ containing all positive semi-definite matrix in $\mathbb{S}^p$. We use $|\mathcal{A}|$ to denote the cardinality of set $\mathcal{A}$. Let $\mathcal{S} = \{j : \beta_{0j} \neq 0, j = 1, \ldots, p\}$ be the index set of nonzero parameters. For a vector $\boldsymbol{a} = (a_1, \ldots, a_m) \in \mathbb{R}^m$, let $\|\boldsymbol{a}\|_q = (\sum_{\ell=1}^m |a_\ell|^q)^{1/q}, 1 \leq q < \infty$ denote its l_q norm, and write $\|\boldsymbol{a}\|_\infty = \max_{1 \leq \ell \leq m} |a_\ell|$. Denote $\boldsymbol{a}_\mathcal{A} \in \mathbb{R}^{|\mathcal{A}|}$ the subvector of $\boldsymbol{a}$ with index set $\mathcal{A} \subset \{1, \ldots, m\}$. Denote by $\boldsymbol{e}$ the vector of all ones. For a matrix $\mathbf{B} = (b_{ij})$, let $\|\mathbf{B}\|_1 = \max_j \sum_i |b_{ij}|, \|\mathbf{B}\|_{\max} = \max_{i,j} |b_{ij}|$ and $\|\mathbf{B}\|_\infty = \max_i \sum_j |b_{ij}|$. For constants a and b, define $a \vee b = \max\{a, b\}$. We use c and C to denote positive constants that may vary throughout the paper. Finally, let $\xrightarrow{d}$ denote convergence in distribution.

2. Estimation Methods for Linear Models

This section mainly focuses on the linear model (1) with high-dimensional settings where the dimension p is larger than the sample size n. When the data can be observed precisely, we can estimate the true regression coefficient vector $\boldsymbol{\beta}_0$ by minimizing the penalized objective function (2) or (3). However, we frequently come across cases where the measured covariates contain measurement errors. There are various types of measurement error data, and we primarily focus on the two categories below.

(1) *Covariates with additive errors*. The observed error-prone of covariate $\boldsymbol{W}_i = \boldsymbol{X}_i + \boldsymbol{U}_i$, where the measurement error $\boldsymbol{U}_i$ is independent of $\boldsymbol{X}_i$ and independently generated from a distribution with mean zero and known covariance matrix $\boldsymbol{\Sigma}_u$.

(2) *Covariates with multiplicative errors*. The observed error-prone of covariates $\boldsymbol{W}_i = \boldsymbol{X}_i \odot \boldsymbol{M}_i$, where $\odot$ denotes the Hadamard product, the measurement error $\boldsymbol{M}_i$ is independent of $\boldsymbol{X}_i$ and follows from a distribution with mean $\boldsymbol{\mu}_M$ and known covariance matrix $\boldsymbol{\Sigma}_M$.

Our main goal is to obtain the sparse estimator $\widehat{\boldsymbol{\beta}}$ of true regression coefficient vector $\boldsymbol{\beta}_0$ in the presence of measurement errors. As we introduced in Section 1, we will run into the issue of the penalized objective function being nonconvex and unbounded from below after correcting the bias caused by measurement errors. This prevents us solving the optimization problem. Several works focused on this issue and proposed some estimation methods.

2.1. Nonconvex Lasso

In order to resolve the issue of objective function being unbounded from below and unsolvable in the presence of measurement errors, Loh and Wainwright [62] added restrictions to regression

coefficients $\boldsymbol{\beta}$ and adopted l_1 penalty. Then the estimator of $\boldsymbol{\beta}_0$ can be obtained by the following l_1-constrained quadratic program

$$\widehat{\boldsymbol{\beta}}_{\mathrm{NCL}} \in \arg \min_{\|\boldsymbol{\beta}\|_1 \le c_0 \sqrt{s}} \left\{ \frac{1}{2} \boldsymbol{\beta}^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}} \boldsymbol{\beta} - \widetilde{\boldsymbol{\rho}}^{\mathrm{T}} \boldsymbol{\beta} + \lambda \|\boldsymbol{\beta}\|_1 \right\} =: \arg \min_{\|\boldsymbol{\beta}\|_1 \le c_0 \sqrt{s}} \left\{ \mathcal{L}(\boldsymbol{\beta}) + \lambda \|\boldsymbol{\beta}\|_1 \right\}, \tag{5}$$

where $c_0 > 0$ is a constant, $s = |\mathcal{S}|$ denotes the number of nonzero components of $\boldsymbol{\beta}_0$, $\mathcal{L}(\boldsymbol{\beta}) = 2^{-1} \boldsymbol{\beta}^{\mathrm{T}} \widehat{\boldsymbol{\Sigma}} \boldsymbol{\beta} - \widetilde{\boldsymbol{\rho}}^{\mathrm{T}} \boldsymbol{\beta}$ is the loss function, $\widehat{\boldsymbol{\Sigma}}$ and $\widetilde{\boldsymbol{\rho}}$ are the consistent estimators of covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{X}_i$ and marginal correlation coefficient vector $\boldsymbol{\rho}$ of $(\boldsymbol{X}_i, \boldsymbol{y}_i)$, and they may differ in terms of various kinds of measurement error data. Under the additive error setting,

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{add}} = n^{-1} \mathbf{W}^{\mathrm{T}} \mathbf{W} - \boldsymbol{\Sigma}_u, \quad \widetilde{\boldsymbol{\rho}}_{\mathrm{add}} = n^{-1} \mathbf{W}^{\mathrm{T}} \boldsymbol{y}. \tag{6}$$

Under the multiplicative error setting,

$$\widehat{\boldsymbol{\Sigma}}_{\mathrm{mul}} = n^{-1} \mathbf{W}^{\mathrm{T}} \mathbf{W} \oslash (\boldsymbol{\Sigma}_m + \boldsymbol{\mu}_m \boldsymbol{\mu}_m^{\mathrm{T}}), \quad \widetilde{\boldsymbol{\rho}}_{\mathrm{mul}} = n^{-1} \mathbf{W}^{\mathrm{T}} \boldsymbol{y} \oslash \boldsymbol{\mu}_m, \tag{7}$$

where $\oslash$ denotes elementwise division operator, and let $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Sigma}}_{\mathrm{add}}$ or $\widehat{\boldsymbol{\Sigma}}_{\mathrm{mul}}$ throughout the sequel. The reason for using "$\in$" rather than "$=$" in (5) is that several local minima might exist in the objective function. Note that this method still relies on a nonconvex objective function to obtain the estimator of $\boldsymbol{\beta}_0$. Thus, we refer to it as "nonconvex Lasso".

The nonconvexity of the penalized objective function makes it challenging to obtain the global minimum of the optimization problem (5). To solve the optimization problem (5), Loh and Wainwright [62] applied the projected gradient descent algorithm and demonstrated that even if the penalized objective function is nonconvex, the solution produced by this algorithm can reach the global minimum with high probability. The algorithm finds the global minimum in an iterative way as follows. At $(k+1)$th iteration,

$$\boldsymbol{\beta}_{\mathrm{NCL}}^{(k+1)} = \arg \min_{\|\boldsymbol{\beta}\|_1 \le c_0 \sqrt{s}} \left\{ \mathcal{L}(\boldsymbol{\beta}_{\mathrm{NCL}}^{(k)}) + \nabla \mathcal{L}(\boldsymbol{\beta}_{\mathrm{NCL}}^{(k)})^{\mathrm{T}} (\boldsymbol{\beta} - \boldsymbol{\beta}_{\mathrm{NCL}}^{(k)}) + \frac{\eta}{2} \|\boldsymbol{\beta} - \boldsymbol{\beta}_{\mathrm{NCL}}^{(k)}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \right\}, \tag{8}$$

where $\nabla \mathcal{L}(\boldsymbol{\beta}) = \widehat{\boldsymbol{\Sigma}} \boldsymbol{\beta} - \widetilde{\boldsymbol{\rho}}$ is the gradient of loss function $\mathcal{L}(\boldsymbol{\beta})$, $\eta > 0$ denotes the stepsize parameter. For details of this algorithm, please see [62,90–92]. Loh and Wainwright [62] proved that the solution obtained by iteration (8) is quite near to the global minimum in both l_1-norm and l_2-norm under some conditions. Specifically, for all $t \ge 0$,

$$\|\boldsymbol{\beta}_{\mathrm{NCL}}^{(k)} - \widehat{\boldsymbol{\beta}}_{\mathrm{NCL}}\|_2^2 \le \gamma^k \|\boldsymbol{\beta}_{\mathrm{NCL}}^{(0)} - \widehat{\boldsymbol{\beta}}_{\mathrm{NCL}}\|_2^2 + C_1 \frac{\log p}{n} \|\widehat{\boldsymbol{\beta}}_{\mathrm{NCL}} - \boldsymbol{\beta}_0\|_1^2 + C_2 \|\widehat{\boldsymbol{\beta}}_{\mathrm{NCL}} - \boldsymbol{\beta}_0\|_2^2,$$

$$\|\boldsymbol{\beta}_{\mathrm{NCL}}^{(k)} - \widehat{\boldsymbol{\beta}}_{\mathrm{NCL}}\|_1 \le 2\sqrt{k} \|\boldsymbol{\beta}_{\mathrm{NCL}}^{(k)} - \widehat{\boldsymbol{\beta}}_{\mathrm{NCL}}\|_2 + 2\sqrt{k} \|\widehat{\boldsymbol{\beta}}_{\mathrm{NCL}} - \boldsymbol{\beta}_0\|_2 + 2\|\widehat{\boldsymbol{\beta}}_{\mathrm{NCL}} - \boldsymbol{\beta}_0\|_1,$$

where C_1 and C_2 are positive constants, $\gamma \in (0, 1)$ is a contraction coefficient independent of (n, p, k). For the estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{NCL}}$ of the true regression coefficient vector $\boldsymbol{\beta}_0$, Loh and Wainwright [62] showed that, with any $c_0 \ge \|\boldsymbol{\beta}_0\|_2$ and $\lambda = O(\sqrt{\log p / n})$, the l_q-estimation error of $\widehat{\boldsymbol{\beta}}_{\mathrm{NCL}}$ satisfies the bounds

$$\|\widehat{\boldsymbol{\beta}}_{\mathrm{NCL}} - \boldsymbol{\beta}_0\|_q = O \left(s^{1/q} \sqrt{\frac{\log p}{n}} \right), \quad q = 1, 2.$$

When $q = 1$, l_1-estimation error can reach the convergence rate $s\sqrt{\log p / n}$; when $q = 2$, l_2-estimation error can reach the convergence rate $\sqrt{s \log p / n}$. However, Loh and Wainwright [62] did not establish the variable selection consistency and oracle inequality for prediction error of nonconvex Lasso estimator.

2.2. Convex Conditioned Lasso

Nonconvex Lasso [62] overcomes the problem of unsolvability caused by nonconvex objective function in the presence of measurement errors. However, there are some drawbacks to this method. First, the nonconvex Lasso solves the problem by adding constraints to β, but the penalized objective function remains nonconvex. It is well recognized that the convexity of the penalized objective function will be incredibly useful for theoretical analysis and computation. Second, two important unknown parameters c_0 and s are included in the optimization problem (5). These two parameters have a direct impact on the estimation results, but we are not sure about their magnitudes in applications. Third, Loh and Wainwright [62] have not established the variable selection results of nonconvex Lasso estimator. To remedy these issues, Datta and Zou [63] proposed Convex Conditioned Lasso (CoCoLasso) based on a convex objective function, which possesses computational and theoretical superiority brought by convexity.

In order to construct the convex objective function, Datta and Zou [63] introduced a nearest positive semi-definite matrix projection operator for the square matrix, which is defined as

$$(\mathbf{A})_+ = \arg \min_{\mathbf{A}_1 \geq 0} \|\mathbf{A} - \mathbf{A}_1\|_{\max}, \tag{9}$$

where $\mathbf{A}$ is a square matrix. Let $\widetilde{\boldsymbol{\Sigma}} = (\widehat{\boldsymbol{\Sigma}})_+$, and the alternating direction method of multipliers (ADMM) algorithm [93] can be utilized to derive $\widetilde{\boldsymbol{\Sigma}}$ from $\widehat{\boldsymbol{\Sigma}}$. Based on $\widetilde{\boldsymbol{\Sigma}}$, the following convex objective function can be constructed, and it yields CoCoLasso estimator

$$\widehat{\boldsymbol{\beta}}_{\text{coco}} = \arg \min_{\boldsymbol{\beta}} \left\{ \frac{1}{2} \boldsymbol{\beta}^{\mathrm{T}} \widetilde{\boldsymbol{\Sigma}} \boldsymbol{\beta} - \widetilde{\boldsymbol{\rho}}^{\mathrm{T}} \boldsymbol{\beta} + \lambda \|\boldsymbol{\beta}\|_1 \right\}. \tag{10}$$

When the covariates contain additive measurement errors,

$$\widetilde{\boldsymbol{\Sigma}}_{\text{add}} = (\widehat{\boldsymbol{\Sigma}}_{\text{add}})_+, \quad \widetilde{\boldsymbol{\rho}}_{\text{add}} = n^{-1}\mathbf{W}^{\mathrm{T}}\boldsymbol{y}, \quad \widehat{\boldsymbol{\Sigma}}_{\text{add}} = n^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{W} - \boldsymbol{\Sigma}_u. \tag{11}$$

When the covariates contain multiplicative measurement errors,

$$\widetilde{\boldsymbol{\Sigma}}_{\text{mul}} = (\widehat{\boldsymbol{\Sigma}}_{\text{mul}})_+, \quad \widetilde{\boldsymbol{\rho}}_{\text{mul}} = n^{-1}\mathbf{W}^{\mathrm{T}}\boldsymbol{y} \oslash \boldsymbol{\mu}_m, \quad \widehat{\boldsymbol{\Sigma}}_{\text{mul}} = n^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{W} \oslash (\boldsymbol{\Sigma}_m + \boldsymbol{\mu}_m\boldsymbol{\mu}_m^{\mathrm{T}}). \tag{12}$$

Note that $\widetilde{\boldsymbol{\Sigma}}$ not only contributes to the construction of the convex objective function but also possesses the same level of estimation accuracy as $\widehat{\boldsymbol{\Sigma}}$ in [62]. It can be guaranteed by the following equation

$$\|\widetilde{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\max} \leq \|\widetilde{\boldsymbol{\Sigma}} - \widehat{\boldsymbol{\Sigma}}\|_{\max} + \|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\max} \leq 2\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\max}.$$

Since $\widetilde{\boldsymbol{\Sigma}}$ is semi-positive definite, we can perform Cholesky decomposition on $\widetilde{\boldsymbol{\Sigma}}$. Then, the cholesky factor of $\widetilde{\boldsymbol{\Sigma}}$ can be used to simplify computations by rewriting (10) as

$$\widehat{\boldsymbol{\beta}}_{\text{coco}} = \arg \min_{\boldsymbol{\beta}} \frac{1}{2n} \|\widetilde{\boldsymbol{y}} - \widetilde{\mathbf{W}}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1, \tag{13}$$

where $\widetilde{\mathbf{W}}$ denotes Cholesky factor of $\widetilde{\boldsymbol{\Sigma}}$ satisfying $n^{-1}\widetilde{\mathbf{W}}^{\mathrm{T}}\widetilde{\mathbf{W}} = \widetilde{\boldsymbol{\Sigma}}$, and $\widetilde{\boldsymbol{y}}$ is the vector satisfying $n^{-1}\widetilde{\mathbf{W}}^{\mathrm{T}}\widetilde{\boldsymbol{y}} = \widetilde{\boldsymbol{\rho}}$. The penalized objective function in (13) is similar to that of standard Lasso. Thus, we can utilize the coordinate descent algorithm to obtain CoCoLasso estimator, please see the details in [63,94,95]. Theoretically, Datta and Zou [63] established the l_q-estimation ($q = 1, 2$) and prediction error bounds of CoCoLasso estimator. Suppose that

$$\psi = \min_{\delta \neq 0, \|\delta_{\mathcal{S}^c}\|_1 \leq 3\|\delta_{\mathcal{S}}\|_1} \frac{\delta^{\mathrm{T}}\boldsymbol{\Sigma}\delta}{\|\delta\|_2^2} > 0.$$

For $s\sqrt{\zeta \log p/n} < \lambda \le \min\{\epsilon_0, 12\epsilon_0\|\beta_{0\mathcal{S}}\|_\infty\}$, where $\zeta = \max\{\sigma_\varepsilon^4, \sigma_U^4, 1\}$, $\epsilon_0 = \sigma_U^2$, σ_ε^2 and σ_U^2 are sub-Gaussian parameters of model error and measurement error, respectively, CoCoLasso estimator $\widehat{\beta}_{\text{coco}}$ satisfies that with probability at least $1 - C\exp(-c\log p)$,

$$\|\widehat{\beta}_{\text{coco}} - \beta_0\|_q = O\left(\frac{\lambda s^{1/q}}{\psi}\right), \quad q = 1, 2, \tag{14}$$

$$n^{-1/2}\|\mathbf{X}(\widehat{\beta}_{\text{coco}} - \beta_0)\|_2 = O\left(\lambda\sqrt{\frac{s}{\psi}}\right). \tag{15}$$

The fomulas (14) and (15) show the oracle inequalities for l_q-estimation error with $q = 1, 2$ and prediction error. Further, Datta and Zou [63] established the sign consistency of CoCoLasso estimator under additional irrepresentable condition and minimum signal strength condition. While there was no variable selection result provided for the nonconvex Lasso estimator $\widehat{\beta}_{\text{NCL}}$ in [62]. Thus, CoCoLasso estimation method not only enjoys the computational convenience of convexity, but also possesses excellent theoretical results. However, when the dimension of covariates p is large, the computation of $\widetilde{\Sigma}$ is expensive. To improve the computational efficiency, Escribe et al. [96] applied a two-step block descent algorithm and proposed a block coordinate descent convex conditioned Lasso (BDCoCoLasso), which is designed for the case that the covariate matrix is only partially corrupted.

2.3. Balanced Estimation

CoCoLasso is effective in parameter estimation of high-dimensional measurement error models, but it suffers from overfitting. To overcome this drawback, Zheng et al. [64] replaced Lasso penalty in CoCoLasso with the combined l_1 and concave penalty and developed the balanced estimator, which can be obtained by

$$\widehat{\beta}_{\text{bal}} = \arg\min_{\beta}\left\{\frac{1}{2}\beta^{\mathrm{T}}\widetilde{\Sigma}\beta - \widetilde{\rho}^{\mathrm{T}}\beta + \lambda_0\|\beta\|_1 + \|p_\lambda(\beta)\|_1\right\}, \tag{16}$$

where $\lambda_0 = c_1\sqrt{\log p/n}$ is the regularization parameter for the l_1 penalty with c_1 being a positive constant, $p_\lambda(\beta) = [p_\lambda(|\beta_1|), \ldots, p_\lambda(|\beta_p|)]^{\mathrm{T}}$, and $p_\lambda(u), u \in [0, +\infty)$ is a concave penalty function with the tuning parameter $\lambda \ge 0$. The definitions of $\widetilde{\Sigma}$ and $\widetilde{\rho}$ are the same as those in (11) and (12) with the two kinds of measurement error data. In contrast to CoCoLasso estimator, balanced estimator strikes a perfect balance between prediction and variable selection. Surprisingly, excellent variable selection results promote the estimation and prediction accuracy of balanced estimator. The simulation studies in [64] demonstrate the estimation and prediction accuracy, as well as the better variable selection results of balanced estimator. As for asymptotic properties of $\widehat{\beta}_{\text{bal}}$, Zheng et al. [64] established the oracle inequalities for l_q-estimation and prediction error,

$$\|\widehat{\beta}_{\text{bal}} - \beta_0\|_q = O_p\left(\frac{\lambda_0 s^{1/q}}{\phi^2}\right), \quad q = 1, 2, \tag{17}$$

$$n^{-1/2}\|\mathbf{X}(\widehat{\beta}_{\text{bal}} - \beta_0)\|_2 = O_p\left(\frac{\lambda_0\sqrt{s}}{\phi}\right), \tag{18}$$

where

$$\phi = \min_{\delta \ne 0, \|\delta_{\mathcal{S}^c}\|_1 \le 7\|\delta_{\mathcal{S}}\|_1} \frac{n^{-1/2}\|\mathbf{X}\delta\|_2}{\|\delta_{\mathcal{S}}\|_2 \vee \|\delta_{\mathcal{S}^c}^*\|_2} > 0,$$

and $\delta_{\mathcal{S}^c}^* \in \mathbb{R}^s$ contains the s largest absolute vaules of $\delta_{\mathcal{S}^c}$. It can be seen from (17) and (18) that the bounds of l_q-estimation ($q = 1, 2$) and prediction error are free of regularization parameter λ for

the concave penalty. Also, the upper bound of falsely discovered signs is provided in [64]. Denote $\text{FS}(\widehat{\boldsymbol{\beta}}) = |\{1 \leq j \leq p : \text{sgn}(\widehat{\beta}_j) \neq \text{sgn}(\beta_{0,j})\}|$, then

$$\text{FS}(\widehat{\boldsymbol{\beta}}) = O_p\left(\frac{\lambda_0^2 s}{\lambda^2 \phi^4}\right). \tag{19}$$

From (19), we can see that if $\min_{j \in \mathcal{S}} |\beta_{0j}| \gg \sqrt{s \log p / n}$ such that $\lambda^2 \gg \lambda_0^2 s$, balanced estimator can achieve sign consistency, which is stronger than the variable selection consistency. Compared with balanced estimator, CoCoLasso estimator requires additional irrepresentable condition to achieve this property.

2.4. Calibrated Zero-norm Regularized Least Square Estimation

The nearest positive semi-definite matrix projection operator defined in [63] solves the problem that the penalized objective function is nonconvex in high-dimensional measurement error models. However, with the constraint of the positive semi-definite matrix, the computation cost of $\widetilde{\boldsymbol{\Sigma}}$ is high. Tao et al. [65] demonstrated that as the dimension p increases, the time required to calculate $\widetilde{\boldsymbol{\Sigma}}$ using the ADMM algorithm will increase significantly. Thus, Tao et al. [65] suggested replacing $\widetilde{\boldsymbol{\Sigma}}$ with an approximation of $\widehat{\boldsymbol{\Sigma}}$ that is easy to obtain but less precise. To achieve this purpose, consider the eigendecomposition of $\widehat{\boldsymbol{\Sigma}}$ as follows

$$\widehat{\boldsymbol{\Sigma}} = \mathbf{V}\text{diag}(\theta_1, \ldots, \theta_p)\mathbf{V}^{\mathrm{T}},$$

where $\text{diag}(\theta_1, \ldots, \theta_p)$ is a diagonal matrix containing the eigenvalues of $\widehat{\boldsymbol{\Sigma}}$ with $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_p$, $\mathbf{V} \in \mathbb{R}^{p \times p}$ is an orthonormal matrix consisting of the corresponding eigenvectors. Then, Tao et al. [65] substituted Frobenius norm for elementwise maximum norm in (9) and obtained a positive definite approximation of $\widehat{\boldsymbol{\Sigma}}$ as follows

$$\widetilde{\boldsymbol{\Sigma}}_F = \arg \min_{\mathbf{W} \geq \xi \mathbf{I}} \|\widehat{\boldsymbol{\Sigma}} - \mathbf{W}\|_F \quad \text{for some } \xi > 0. \tag{20}$$

Note that the optimal solution of (20) is the same as that of the problem

$$\min_{\mathbf{W} \geq \xi \mathbf{I}} \|\widehat{\boldsymbol{\Sigma}} - \mathbf{W}\|_F^2. \tag{21}$$

Thus, we have

$$\widetilde{\boldsymbol{\Sigma}}_F = \xi \mathbf{I} + \Pi_{\mathbb{S}_+^p}(\widehat{\boldsymbol{\Sigma}} - \xi \mathbf{I}) = \mathbf{V}\text{diag}[\max(\theta_1, \xi), \ldots, \max(\theta_p, \xi)]\mathbf{V}^{\mathrm{T}}, \tag{22}$$

where $\Pi_{\mathbb{S}_+^p}(\cdot)$ denotes the projection of a matrix on $\mathbb{S}_+^p$. Similar to $\widetilde{\boldsymbol{\Sigma}}$, we have $\widetilde{\boldsymbol{\Sigma}}_F = n^{-1}\widetilde{\mathbf{W}}_F^{\mathrm{T}}\widetilde{\mathbf{W}}_F$, where $n^{-1/2}\widetilde{\mathbf{W}}_F$ is Cholesky factor of $\widetilde{\boldsymbol{\Sigma}}_F$. Let $\widetilde{\boldsymbol{y}}_F$ be the vector satisfying $n^{-1}\widetilde{\mathbf{W}}_F^{\mathrm{T}}\widetilde{\boldsymbol{y}}_F = \widetilde{\boldsymbol{\rho}}$. By some simple calculation, we can obtain that

$$\begin{cases} \widetilde{\mathbf{W}}_F = \sqrt{n}\mathbf{V}\,\text{diag}\left(\sqrt{\max(\theta_1, \xi)}, \ldots, \sqrt{\max(\theta_p, \xi)}\right)\mathbf{V}^{\mathrm{T}}, \\[4mm] \widetilde{\boldsymbol{y}}_F = \sqrt{n}\mathbf{V}\,\text{diag}\left(\dfrac{1}{\sqrt{\max(\theta_1, \xi)}}, \ldots, \dfrac{1}{\sqrt{\max(\theta_p, \xi)}}\right)\mathbf{V}^{\mathrm{T}}\widetilde{\boldsymbol{\rho}}. \end{cases} \tag{23}$$

Based on equation (22), $\widetilde{\boldsymbol{\Sigma}}_F$ can be obtained easily. This implies that computing $\widetilde{\boldsymbol{\Sigma}}_F$ requires substantially less time than computing $\widetilde{\boldsymbol{\Sigma}}$. However, the approximation accuracy of $\widetilde{\boldsymbol{\Sigma}}_F$ to $\widehat{\boldsymbol{\Sigma}}$ is not as good as that of $\widetilde{\boldsymbol{\Sigma}}$ because minimizing Frobenius norm may yield larger components compared with the elementwise maximum norm. To get an excellent estimator of $\boldsymbol{\beta}_0$, it is reasonable to find a more effective regression

method to replace Lasso. Tao et al. [65] considered the zero norm penalty and defined the following calibrated zero-norm regularized least squares (CaZnRLS) estimator

$$\widehat{\boldsymbol{\beta}}_{\text{zn}} \in \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{ \frac{1}{2n\lambda} \|\widetilde{\mathbf{W}}_F\boldsymbol{\beta} - \widetilde{\boldsymbol{y}}_F\|_2^2 + \|\boldsymbol{\beta}\|_0 \right\}. \tag{24}$$

However, it is difficult to solve (24) directly. Thus, to give an equivalent form for (24) that can be solved, Tao et al. [65] defined

$$\phi(u) := \frac{a-1}{a+1}u^2 + \frac{2}{a+1}u \quad (a>1), \quad u\in\mathbb{R}.$$

It is easy to verify that for any $\boldsymbol{\beta}\in\mathbb{R}^p$,

$$\|\boldsymbol{\beta}\|_0 = \min_{\boldsymbol{w}\in\mathbb{R}^p} \left\{ \sum_{i=1}^p \phi(w_i) : (\boldsymbol{e}-\boldsymbol{w})^{\mathrm{T}}|\boldsymbol{\beta}| = 0, 0\le\boldsymbol{w}\le\boldsymbol{e} \right\}, \tag{25}$$

where $|\boldsymbol{\beta}| = (|\beta_1|,\ldots,|\beta_p|)^{\mathrm{T}}$. The formula (25) implies that the optimization problem (24) can be rewritten as the following mathematical program with equilibrium constraints (MPEC)

$$\min_{\boldsymbol{\beta},\boldsymbol{w}\in\mathbb{R}^p} \left\{ \frac{1}{2n\lambda} \|\widetilde{\mathbf{W}}_F\boldsymbol{\beta} - \widetilde{\boldsymbol{y}}_F\|_2^2 + \sum_{i=1}^p \phi(w_i) : (\boldsymbol{e}-\boldsymbol{w})^{\mathrm{T}}|\boldsymbol{\beta}| = 0, 0\le\boldsymbol{w}\le\boldsymbol{e} \right\}. \tag{26}$$

Note that if the optimal solution of optimization problem (24) is $\widehat{\boldsymbol{\beta}}^*$, then the corresponding optimal solution of optimization problem (26) is $(\widehat{\boldsymbol{\beta}}^*, \text{sign}(|\widehat{\boldsymbol{\beta}}^*|))$.

However, it can be seen that the annoying nonconvexity is introduced by the restriction $(\boldsymbol{e}-\boldsymbol{w})^{\mathrm{T}}|\boldsymbol{\beta}| = 0$ in (26), and it is the cause of the difficulty in obtaining the estimator $\widehat{\boldsymbol{\beta}}_{\text{zn}}$. Accordingly, Tao et al. [65] considered the following penalized version of optimization problem (26)

$$\min_{\boldsymbol{\beta},\boldsymbol{w}\in\mathbb{R}^p} \left\{ \frac{1}{2n\lambda} \|\widetilde{\mathbf{W}}_F\boldsymbol{\beta} - \widetilde{\boldsymbol{y}}_F\|_2^2 + \sum_{i=1}^p \phi(w_i) + \rho(\boldsymbol{e}-\boldsymbol{w})^{\mathrm{T}}|\boldsymbol{\beta}|, 0\le\boldsymbol{w}\le\boldsymbol{e} \right\}, \tag{27}$$

where $\rho>0$ is the penalty parameter. Tao et al. [65] proved that the global optimal solution of optimization problem (27) with $\rho \ge \overline{\rho} := (4aL_f)[(a+1)\lambda]^{-1}$ is the same as that of optimization problem (26), where L_f is Lipschitz constant of the function $f(\boldsymbol{\beta}) := (2n)^{-1}\|\widetilde{\mathbf{W}}_F\boldsymbol{\beta} - \widetilde{\boldsymbol{y}}_F\|_2^2$ on the ball $\{\boldsymbol{\beta}\in\mathbb{R}^p : \|\boldsymbol{\beta}\|_2 \le R\}$, and R is a constant. Thus, $\widehat{\boldsymbol{\beta}}_{\text{zn}}$ can be obtained by solving the following optimization problem with $\rho \ge \overline{\rho}$

$$\widehat{\boldsymbol{\beta}}_{\text{zn}} \in \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p,\boldsymbol{w}\in[\boldsymbol{0},\boldsymbol{e}]} \left\{ \frac{1}{2n} \|\widetilde{\mathbf{W}}_F\boldsymbol{\beta} - \widetilde{\boldsymbol{y}}_F\|_2^2 + \sum_{i=1}^p \lambda\big[\phi(w_i) + \rho(1-w_i)|\beta_i|\big] \right\}. \tag{28}$$

Tao et al. [65] recommended using the multi-stage convex relaxation approach (GEP–MSCRA) to obtain $\widehat{\boldsymbol{\beta}}_{\text{zn}}$. This approach solves (28) in an iterative way with the main steps summarized as follows.

Step 1. Initialize the algorithm with $\boldsymbol{w}^{(0)}\in[\boldsymbol{0},2^{-1}\boldsymbol{e}]$, $\rho^{(0)}=1$, $\lambda>0$, $k=1$.

Step 2. Solve the following optimization problem and get $\widehat{\boldsymbol{\beta}}_{\text{zn}}^{(k)}$

$$\widehat{\boldsymbol{\beta}}_{\text{zn}}^{(k)} = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{ \frac{1}{2n} \|\widetilde{\mathbf{W}}_F\boldsymbol{\beta} - \widetilde{\boldsymbol{y}}_F\|_2^2 + \lambda\sum_{i=1}^p (1-w_i^{(k-1)})|\beta_i| \right\}.$$

Step 3. If $k=1$, choose an appropriate $\rho^{(1)} > \rho^{(0)}$ using the information from $\|\widehat{\boldsymbol{\beta}}_{\text{zn}}^{(1)}\|_\infty$; if $1<k\le 3$, choose $\rho^{(k)}$ satisfying $\rho^{(k)} > \rho^{(k-1)}$; if $k>3$, let $\rho^{(k)} = \rho^{(k-1)}$.

Step 4. Obtain $w_i^{(k)}(i = 1, \ldots, p)$ through the following optimization problem

$$w_i^{(k)} = \arg \min_{0 \le w_i \le 1} \left\{ \phi(w_i) - \rho^{(k)} w_i |\widehat{\beta}_{zn,i}^{(k)}| \right\}.$$

Step 5. Let $k \leftarrow k + 1$ and repeat Steps 2–4 until the stopping conditions are satisfied.

Note that the initial $w^{(0)}$ in Step 1 is an arbitrary vector from the interval $[\mathbf{0}, 2^{-1}e]$ rather than the feasible set $[\mathbf{0}, e]$ in (28). The reason is to obtain a better initial estimator $\widehat{\beta}_{zn}^{(1)}$. In addition, $w_i^{(k)}$ in Step 4 has the following closed form based on the convexity of ϕ

$$w_i^{(k)} = \min \left[1, \max \left(\frac{(a+1)\rho^{(k)}|\beta_i^{(k)}| - 2}{2(a+1)}, 0 \right) \right], \quad i = 1, \ldots, p.$$

Consequently, the primary calculation in each iteration is to solve a weighted l_1-norm regularized least square problem. Under some regularity conditions, $\widehat{\beta}_{zn}^{(k)}$ satisfies

$$\|\widehat{\beta}_{zn}^{(k)} - \beta_0\|_2 = O_p(\lambda\sqrt{s}) \quad \forall k \in \mathbb{N}^+. \tag{29}$$

It can be seen from (29) that the l_2-estimation error bound of CaZnRLS estimator possesses the same order as those of nonconvex Lasso and CoCoLasso estimators. Tao et al. [65] further showed that the error bound of $\widehat{\beta}_{zn}^{(k+1)}$ is better than that of $\widehat{\beta}_{zn}^{(k)}$ for all $k \in \mathbb{N}^+$. Furthermore, Tao et al. [65] demonstrated that GEP-MSCRA will produce a $\widehat{\beta}_{zn}^{(k)}$ such that $\text{supp}(\widehat{\beta}_{zn}^{(k)}) = \text{supp}(\beta_0)$ in a finite number of iterations if the minimum nonzero value of smallest nonzero entries of β_0 is not too small.

2.5. Linear and Conic Programming Estimation

In addition to the approaches mentioned above, another class of methods is based on the idea of Dantzig selector to acquire estimator of true regression coefficients β_0. Rosenbaum and Tsybakov [66] proposed the following matrix uncertainty (MU) selector

$$\widehat{\beta}_{\text{MU}} = \arg \min_{\beta} \left\{ \|\beta\|_1 : \|n^{-1}\mathbf{W}^{\mathsf{T}}(y - \mathbf{W}\beta)\|_\infty \le \delta\|\beta\|_1 + \lambda \right\}, \tag{30}$$

where $\delta \ge 0$ and $\lambda \ge 0$ are tuning parameters depending on the level of measurement error U and model error ε, respectively.

However, the $n^{-1}\mathbf{W}^{\mathsf{T}}\mathbf{W}$ is included in (30) rather than $n^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{X}$ due to the unobservability of $\mathbf{X}$. Obviously, the matrix $n^{-1}\mathbf{W}^{\mathsf{T}}\mathbf{W}$ contains bias caused by measurement errors. To address this issue, Rosenbaum and Tsybakov [67] proposed an improved version of MU selector called compensated MU selector. It is applicable to the case that the entries of measurement error U_i is independent such that $\sigma_{U,j}^2 = n^{-1}\sum_{i=1}^n \mathbb{E}(U_{ij}^2)$ is finite for $j = 1, \ldots, p$. The compensated MU selector is defined as

$$\widehat{\beta}_{\text{CMU}} = \arg \min_{\beta} \left\{ \|\beta\|_1 : \|n^{-1}\mathbf{W}^{\mathsf{T}}(y - \mathbf{W}\beta) + \widehat{\mathbf{D}}\beta\|_\infty \le \delta\|\beta\|_1 + \lambda \right\}, \tag{31}$$

where $\widehat{\mathbf{D}}$ is a diagonal matrix consisting of $\widehat{\sigma}_{U,j}^2, j = 1, \ldots, p$, and constants δ and λ are the same as those in (30). Rosenbaum and Tsybakov [67] showed that the l_q-estimation error of the estimator $\widehat{\beta}_{\text{CMU}}$ satisfies

$$\|\widehat{\beta}_{\text{CMU}} - \beta_0\|_q = O_p \left(s^{1/q}(\|\beta_0\|_1 + 1)\sqrt{\frac{\log p}{n}} \right), \quad 1 \le q \le \infty.$$

MU selector and compensated MU selector provide two alternative estimation methods for high-dimensional measurement error models, but there remains an issue. The optimization problem in (31) may be nonconvex, and Rosenbaum and Tsybakov [67] did not offer a suitable algorithm to the

general case. To remedy this issue, Belloni et al. [70] proposed the conic-programming based estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{cp}}$. Consider the following optimization problem

$$\min_{\boldsymbol{\beta},t}\{\|\boldsymbol{\beta}\|_1 + \kappa t\},$$

$$\text{s.t. } \|n^{-1}\mathbf{W}^{\mathrm{T}}(\boldsymbol{y} - \mathbf{W}\boldsymbol{\beta}) + \widehat{\mathbf{D}}\boldsymbol{\beta}\|_\infty \le \delta t + \lambda, \quad \|\boldsymbol{\beta}\|_2 \le t, \quad t \in \mathbb{R}^+, \tag{32}$$

where κ, δ and λ are positive tuning parameters. Suppose that the solution of (32) is $(\widehat{\boldsymbol{\beta}}_{\mathrm{cp}}, \widehat{t})$, then $\widehat{\boldsymbol{\beta}}_{\mathrm{cp}}$ is defined as the conic-programming based estimator of true regression coefficients $\boldsymbol{\beta}_0$. It is easy to verify that the optimization problem (32) can be solved efficiently in a polynomial time as it is a second-order cone programming problem. To analyze the asymptotic properties of $\widehat{\boldsymbol{\beta}}_{\mathrm{cp}}$, assume that $\kappa \in [2^{-1}, 2]$, $\delta = O(\sqrt{\log p/n})$, and $\lambda = O(\sqrt{\log p/n})$. Then, Belloni et al. [70] showed that l_q-estimation ($1 \le q \le \infty$) and prediction error of $\widehat{\boldsymbol{\beta}}_{\mathrm{cp}}$ satisfy

$$\|\widehat{\boldsymbol{\beta}}_{\mathrm{cp}} - \boldsymbol{\beta}_0\|_q = O_p\left(s^{1/q}(\|\boldsymbol{\beta}_0\|_2 + 1)\sqrt{\frac{\log p}{n}}\right), \quad 1 \le q \le \infty, \tag{33}$$

$$n^{1/2}\|\mathbf{X}(\widehat{\boldsymbol{\beta}}_{\mathrm{cp}} - \boldsymbol{\beta}_0)\|_2 = O_p\left(s^{1/2}(\|\boldsymbol{\beta}_0\|_2 + 1)\sqrt{\frac{\log p}{n}}\right). \tag{34}$$

In contrast to nonconvex Lasso in [62], the conic-programming based estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{cp}}$ can achieve the convergence rate in (33) and (34) without any information of the parameters $\|\boldsymbol{\beta}_0\|_1$, $\|\boldsymbol{\beta}_0\|_2$ or s. Compared with compensated MU selector in [67], the conic-programming based estimator $\widehat{\boldsymbol{\beta}}_{\mathrm{cp}}$ can be obtained in the general case without the computational difficulty of nonconvexity.

3. Estimation Methods for Generalized Linear Models

The above methods are mainly for linear models. This section introduces the estimation methods for high-dimensional generalized linear models with measurement errors.

3.1. Estimation Method for Poisson Models

Count data is commonly encountered in various fields including finance, economics and social sciences. In order to analyze count data, Poisson regression models are a popular choice in practice. Jiang and Ma [74] studied the high-dimensional Poisson regression models with additive measurement errors and proposed a novel optimization algorithm to obtain the estimator of true regression coefficient vector $\boldsymbol{\beta}_0$. Suppose that Y_i is the response variable following a Poisson distribution satisfying $\mathbb{E}(Y_i|X_i) = \exp(X_i^{\mathrm{T}}\boldsymbol{\beta})$, where $X_i \in \mathbb{R}^p$ is an unobservable covariate. Its error-prone surrogate $W_i = X_i + U_i$, and the measurement error U_i follows from a sub-Gaussian distribution with known covariance matrix $\boldsymbol{\Sigma}_u$. It is easy to verify that

$$\mathbb{E}\left\{Y_i W_i^{\mathrm{T}}\boldsymbol{\beta} - \exp\left(\boldsymbol{\beta}^{\mathrm{T}}W_i - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_u\boldsymbol{\beta}/2\right) \mid X_i, Y_i\right\} = Y_i X_i^{\mathrm{T}}\boldsymbol{\beta} - \exp\left(\boldsymbol{\beta}^{\mathrm{T}}X_i\right). \tag{35}$$

From (35), Jiang and Ma [74] imposed restriction on $\boldsymbol{\beta}$ similar to it in [62] and estimated $\boldsymbol{\beta}$ by solving the following optimization problem

$$\widehat{\boldsymbol{\beta}}_{\mathrm{P}} = \arg \min_{\|\boldsymbol{\beta}\|_1 \le c_p\sqrt{s}, \|\boldsymbol{\beta}\|_2 \le c_p} \{\mathcal{L}(\boldsymbol{\beta}) + \lambda\|\boldsymbol{\beta}\|_1\}, \tag{36}$$

where

$$\mathcal{L}(\boldsymbol{\beta}) = -\frac{1}{n}\sum_{i=1}^n\left\{Y_i W_i^{\mathrm{T}}\boldsymbol{\beta} - \exp\left(\boldsymbol{\beta}^{\mathrm{T}}W_i - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_u\boldsymbol{\beta}/2\right)\right\}. \tag{37}$$

The estimator $\widehat{\beta}_{\mathrm{P}}$ can be obtained by the composite gradient descent algorithm. Specifically, at $(k+1)$th iteration, first solve the following optimization problem without any restrictions on β

$$\widetilde{\beta}_{\mathrm{P}}^{(k+1)} = \arg\min_{\beta} \left\{ \partial\mathcal{L}(\beta_{\mathrm{P}}^{(k)})/\partial\beta^{\mathrm{T}}(\beta - \beta^{(k)}) + \eta/2\|\beta - \beta^{(k)}\|_2^2 + \lambda\|\beta\|_1 \right\},$$

where $\eta > 0$ is a stepsize parameter. Next, apply the projection method in [90] to project $\widetilde{\beta}_{\mathrm{P}}^{(k+1)}$ onto the l_1 ball with radius $c_p\sqrt{s}$ and produce $\breve{\beta}_{\mathrm{P}}^{(k+1)}$. If $\|\breve{\beta}_{\mathrm{P}}^{(k+1)}\|_2 > c_p$, let $\widehat{\beta}_{\mathrm{P}}^{(k+1)} = \breve{\beta}_{\mathrm{P}}^{(k+1)} c_p/\|\breve{\beta}_{\mathrm{P}}^{(k+1)}\|_2$, otherwise let $\widehat{\beta}_{\mathrm{P}}^{(k+1)} = \breve{\beta}_{\mathrm{P}}^{(k+1)}$. Repeat the above steps until the stopping condition is satisfied. Jiang and Ma [74] proved the convergence of this algorithm. Under some regularity conditions, they further showed that the global minimum $\widehat{\beta}_{\mathrm{P}}$ of (36) satisfies

$$\|\widehat{\beta}_{\mathrm{P}} - \beta_0\|_q = O(s^{1/q}\lambda). \tag{38}$$

There is an usual requirement that $\lambda \geqslant 2\|\partial\mathcal{L}(\beta)/\partial\beta\|_\infty$ in Poisson models, and the term $\|\partial\mathcal{L}(\beta)/\partial\beta\|_\infty = O(\sqrt{n/\log p})$. Thus, the convergence rate of $\widehat{\beta}_{\mathrm{P}}$ is slower than those of nonconvex Lasso, CoCoLasso and balanced estimators in linear models.

3.2. Generalized Matrix Uncertainty Selector

The method in [74] is only designed for high-dimensional Poisson models with measurement errors. To develop a method that is applicable to generalized linear models, Sørensen et al. [68] drew on the idea of MU selector and proposed the generalized matrix uncertainty (GMU) selector for high-dimensional generalized linear models with additive measurement errors.

Consider a generalized linear model with response variable Y distributed according to

$$f_Y(y;\theta,\phi) = \exp\left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi) \right\},$$

where $\theta = X^{\mathrm{T}}\beta_0$, $X \in \mathbb{R}^p$ are the covariates. The expected response is given by the mean function $\mu(\theta) = b'(\theta)$, and Taylor expansion of the mean function $\mu(X_i^{\mathrm{T}}\beta_0)$ at point $W_i^{\mathrm{T}}\beta_0$ is

$$\mu(X_i^{\mathrm{T}}\beta_0) = \sum_{\ell=0}^{\infty} \frac{\mu^{(\ell)}(W_i^{\mathrm{T}}\beta_0)}{\ell!}(-U_i^{\mathrm{T}}\beta_0)^\ell, \tag{39}$$

where $\mu^{(\ell)}(\cdot)$ is the ℓth derivative of function $\mu(\cdot)$. With Taylor expansion (39) of the mean function, the generalized matrix uncertainty selector can be defined as

$$\widehat{\beta}_{\mathrm{GMU}}^L = \arg\min_{\beta} \left\{ \|\beta\|_1 : \beta \in \Theta^L \right\},$$

$$\Theta^L = \left\{ \beta \in \mathbb{R}^p : \max_{1 \leq j \leq p} \left| \frac{1}{n}w_{ij}[Y_i - \mu(W_i^{\mathrm{T}}\beta)] \right| \leq \lambda + \sum_{\ell=1}^{L} \frac{\delta^\ell}{\ell!\sqrt{n}}\|\beta\|_1^\ell\|\mu^{(\ell)}(W\beta)\|_2 \right\}, \tag{40}$$

where δ is the positive parameter satisfying

$$\|\mathbf{U}\|_\infty \leq \delta, \quad \mu^{(\ell)}(W\beta) = [\mu^{(\ell)}(W_1^{\mathrm{T}}\beta),\ldots,\mu^{(\ell)}(W_n^{\mathrm{T}}\beta)]^{\mathrm{T}}.$$

In practice, Sørensen et al. [68] recommended using $L = 1$ for computational convenience and demonstrated that the first-order approximation produces satisfactory results.

To solve the optimation problem (40) and obtain the estimator $\widehat{\beta}^{L}_{\text{GMU}}$, we can utilize the iterative reweighing algorithm. The main iteration step of the algorithm is stated as follows

$$\widehat{\beta}^{(k+1)}_{\text{GMU}} = \arg\min_{\beta} \left\{ \|\beta\|_1 : \frac{1}{n} \left\| \widetilde{\mathbf{W}}^{(k)\mathrm{T}}_{\text{g}} (\widetilde{z}^{(k)} - \widetilde{\mathbf{W}}^{(k)}_{\text{g}} \beta) \right\|_\infty \leq \lambda + \sum_{\ell=1}^{L} \frac{\delta^\ell}{\ell!\sqrt{n}} \|\beta\|^\ell_1 \|V^{(\ell,k)}\|_2 \right\}, \qquad (41)$$

where $\widetilde{\mathbf{W}}_{\text{g}} \in \mathbb{R}^{n\times p}$ is a matrix of weighted error-prone surrogate of covariates with elements $\widetilde{w}^{(k)}_{\text{g},ij} = w_{ij}\sqrt{V^{(1,k)}_i}$, $\widetilde{z}^{(k)} \in \mathbb{R}^n$ is a vector with the elements $\widetilde{z}^{(k)}_i = z^{(k)}_i \sqrt{V^{(1,k)}_i}$,

$$z^{(k)}_i = W^{\mathrm{T}}_i \widehat{\beta}^{(k)}_{\text{GMU}} + \left[Y_i - \mu\left\{ W^{\mathrm{T}}_i \widehat{\beta}^{(k)}_{\text{GMU}} \right\} \right] \mu'\left\{ W^{\mathrm{T}}_i \widehat{\beta}^{(k)}_{\text{GMU}} \right\}^{-1}, \quad i = 1,\ldots,n,$$

and

$$V^{(\ell,k)} = \left[\mu^{(\ell)}\left\{ W^{\mathrm{T}}_1 \widehat{\beta}^{(k)}_{\text{GMU}} \right\}, \ldots, \mu^{(\ell)}\left\{ W^{\mathrm{T}}_n \widehat{\beta}^{(k)}_{\text{GMU}} \right\} \right]^{\mathrm{T}} = \left(V^{(\ell,k)}_1, \ldots, V^{(\ell,k)}_n \right)^{\mathrm{T}}, \quad \ell = 1,\ldots,L$$

is the weight vector in Taylor expansion with L terms. When $L = 1$ is applied, it is easy to verify that (41) is a linear program. For more details about the algorithm, please see [68,97]. However, Sørensen et al. [68] did not establish any asymptotic properties of GMU selector.

4. Hypothesis Testing Methods

The aforementioned works on high-dimensional measurement error models mainly investigate estimation problems and numerical algorithms of optimization problems, as well as the theoretical properties of estimators. Recently, some works have studied the hypothesis testing problems for high-dimensional measurement error regression models, which will be introduced in this section.

4.1. Corrected Decorrelated Score Test

The above methods are proposed under the setting that all covariates are corrupted. In practice, it is common that not all covariates are measured with errors. Thus, Li et al. [86] investigated high-dimensional measurement error models where a fixed number of covariates contain measurement errors and proposed statistical inference methods for the regression coefficients corresponding to these covariates.

Consider the following high-dimensional linear model with one of the covariates containing additive errors

$$\begin{cases} y_i = \beta_0 X_i + \gamma^{\mathrm{T}}_0 Z_i + \varepsilon_i, \\ W_i = X_i + U_i, \qquad i = 1,\ldots,n, \end{cases}$$

where $X_i \in \mathbb{R}$ is an unobservable covariate, and W_i is its error-prone surrogate, $Z_i \in \mathbb{R}^{p-1}$ is an observed covariate precisely. The measurement error U_i follows from sub-Gaussian distribution with mean zero and variance σ^2_U, and U_i is independent of $(X_i, Z_i, \varepsilon_i)$. Denote $y = (y_1, \ldots, y_n)^{\mathrm{T}}, X = (X_1, \ldots, X_n)^{\mathrm{T}}, W = (W_1, \ldots, W_n)^{\mathrm{T}}$ and $Z = (Z_1, \ldots, Z_n)^{\mathrm{T}}$. This subsection aims to test the hypothesis:

$$H_0 : \beta_0 = \beta^* \longleftrightarrow H_1 : \beta_0 \neq \beta^* (\beta^* \in \mathbb{R}),$$

and construct a confidence interval for β_0 under high-dimensional settings.

Since that we are only concerned with the inference of the parameter β, then the parameter γ is regarded as a nuisance. Following the idea in [81], Li et al. [86] defined the corrected score function as

$$S_\theta(\theta) = \widehat{\Sigma}\theta - \widehat{\rho} = \frac{1}{n}\sum_{i=1}^{n} S_{i\theta}(\theta) = \begin{pmatrix} S_\beta(\beta,\gamma) \\ S_\gamma(\beta,\gamma) \end{pmatrix} = \begin{pmatrix} \widehat{\Sigma}_{11}\beta + \widehat{\Sigma}_{12}\gamma - \widehat{\rho}_1 \\ \widehat{\Sigma}_{21}\beta + \widehat{\Sigma}_{22}\gamma - \widehat{\rho}_2 \end{pmatrix},$$

where $\boldsymbol{\theta} = (\beta, \boldsymbol{\gamma}^{\mathrm{T}})^{\mathrm{T}}$,

$$
\widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} \widehat{\boldsymbol{\Sigma}}_{11} & \widehat{\boldsymbol{\Sigma}}_{12} \\ \widehat{\boldsymbol{\Sigma}}_{21} & \widehat{\boldsymbol{\Sigma}}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}/n - \sigma_U^2 & \boldsymbol{W}^{\mathrm{T}}\boldsymbol{Z}/n \\ \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{W}/n & \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}/n \end{pmatrix} \text{ and } \widehat{\boldsymbol{\rho}} = \begin{pmatrix} \widehat{\boldsymbol{\rho}}_1 \\ \widehat{\boldsymbol{\rho}}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{W}^{\mathrm{T}}\boldsymbol{y}/n \\ \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{y}/n \end{pmatrix}
$$

are consistent estimators of $\boldsymbol{\Sigma} = (\boldsymbol{X}, \boldsymbol{Z})^{\mathrm{T}}(\boldsymbol{X}, \boldsymbol{Z})/n$ and $\boldsymbol{\rho} = (\boldsymbol{X}, \boldsymbol{Z})^{\mathrm{T}}\boldsymbol{y}/n$, respectively. The corrected score covariance matrix is defined as

$$
\mathbf{I}(\boldsymbol{\theta}) = \mathbb{E}\left\{ \boldsymbol{S}_{i\boldsymbol{\theta}}(\boldsymbol{\theta})\boldsymbol{S}_{i\boldsymbol{\theta}}(\boldsymbol{\theta})^{\mathrm{T}} \right\} = \begin{pmatrix} I_{\beta\beta} & \boldsymbol{I}_{\beta\gamma} \\ \boldsymbol{I}_{\gamma\beta} & \mathbf{I}_{\gamma\gamma} \end{pmatrix}.
$$

To conduct statistical inference on the target parameter β, it is crucial to eliminate the influence of nuisance parameter $\boldsymbol{\gamma}$. Thus, Li et al. [86] developed the corrected decorrelated score function for target parameter β as

$$
S(\beta, \boldsymbol{\gamma}) = S_\beta(\beta, \boldsymbol{\gamma}) - \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{S}_\gamma(\beta, \boldsymbol{\gamma}),
$$

where $\boldsymbol{\omega}^{\mathrm{T}} = \boldsymbol{I}_{\beta\gamma}\mathbf{I}_{\gamma\gamma}^{-1} = \mathbb{E}(X_i\boldsymbol{Z}_i^{\mathrm{T}})\mathbb{E}(\boldsymbol{Z}_i\boldsymbol{Z}_i^{\mathrm{T}})^{-1}$. It easy to verify that $\mathbb{E}[S(\beta_0, \boldsymbol{\gamma}_0)\boldsymbol{S}_\gamma(\beta_0, \boldsymbol{\gamma}_0)] = 0$, which indicates that $S(\beta, \boldsymbol{\gamma})$ and nuisance score function $\boldsymbol{S}_\gamma(\beta, \boldsymbol{\gamma})$ are uncorrelated. Obviously, we can obtain that $\mathrm{Var}[S(\beta, \boldsymbol{\gamma})] = I_{\beta\beta} - \boldsymbol{I}_{\beta\gamma}\mathbf{I}_{\gamma\gamma}^{-1}\boldsymbol{I}_{\gamma\beta} =: \sigma_{\beta|\gamma}^2$. Then, Li et al. [86] constructed the test statistic and the confidence interval for β_0 based on the estimated decorrelated score function. This statistical inference procedure is summarized as follows.

Step 1. Apply CoCoLasso estimation method in [63] to calculate initial estimator $\widetilde{\boldsymbol{\theta}} = (\widetilde{\beta}, \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}})^{\mathrm{T}}$, and utilize following Dantzig type estimator to estimate $\boldsymbol{\omega}$

$$
\widehat{\boldsymbol{\omega}} = \arg\min_{\boldsymbol{\omega}}\|\boldsymbol{\omega}\|_1, \quad \text{s.t.} \quad \|\widehat{\boldsymbol{\Sigma}}_{12} - \boldsymbol{\omega}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{22}\|_\infty \leq \lambda',
$$

where $\lambda' = O(\sqrt{\log p/n})$.

Step 2. Estimate the decorrelated score function by

$$
\widehat{S}(\beta, \widetilde{\boldsymbol{\gamma}}) = S_\beta(\beta, \widetilde{\boldsymbol{\gamma}}) - \widehat{\boldsymbol{\omega}}^{\mathrm{T}}\boldsymbol{S}_\gamma(\beta, \widetilde{\boldsymbol{\gamma}}),
$$

and calculate the test statistic $\widehat{T} = \sqrt{n}\widehat{S}(\beta^*, \widetilde{\boldsymbol{\gamma}})(\widehat{\sigma}_{\beta|\gamma, H_0}^2)^{-1/2}$, where

$$
\begin{aligned}
\widehat{\sigma}_{\beta|\gamma, H_0}^2 &= \left\{ \widehat{I}_{\beta\beta} - \widehat{\boldsymbol{\omega}}^{\mathrm{T}}\widehat{\boldsymbol{I}}_{\gamma\beta} \right\}\Big|_{\beta=\beta^*} \\
&= (\widehat{\sigma}_{\varepsilon, H_0}^2 + \beta^{*2}\sigma_U^2)(1 - \widehat{\boldsymbol{\omega}}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{21}) + \beta^{*2}\mathbb{E}(U_i^4) + \widehat{\sigma}_{\varepsilon, H_0}^2\sigma_U^2 - \beta^{*2}\sigma_U^4.
\end{aligned}
$$

Step 3. Estimate β as

$$
\widehat{\beta} = \widetilde{\beta} - \widehat{S}(\widetilde{\boldsymbol{\theta}})/(\widehat{\boldsymbol{\Sigma}}_{11} - \widehat{\boldsymbol{\omega}}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{21}),
$$

and construct $(1-\alpha)100\%$ confidence interval for β_0 as

$$
\left[\widehat{\beta} - u_{1-\alpha/2}\sqrt{\widehat{\sigma}_\beta^2/n}, \widehat{\beta} + u_{1-\alpha/2}\sqrt{\widehat{\sigma}_\beta^2/n} \right],
$$

where $u_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of standard normal distribution,

$$
\widehat{\sigma}_\beta^2 = (1 - \widehat{\boldsymbol{\omega}}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{21})^{-2}\left\{ (\widehat{\sigma}_\varepsilon^2 + \widehat{\beta}^2\sigma_U^2)(1 - \widehat{\boldsymbol{\omega}}^{\mathrm{T}}\widehat{\boldsymbol{\Sigma}}_{21}) + \widehat{\beta}^2\mathbb{E}(U_i^4) + \widehat{\sigma}_\varepsilon^2\sigma_U^2 - \widehat{\beta}^2\sigma_U^4 \right\}
$$

is the estimator of the asymptotic variance σ_β^2 of $\widehat{\beta}$, and $\widehat{\sigma}_\varepsilon^2 = n^{-1}\sum_{i=1}^n (y_i - \widehat{\beta}W_i - \widetilde{\boldsymbol{\gamma}}^{\mathrm{T}}\boldsymbol{Z}_i)^2 - \widehat{\beta}^2\sigma_U^2$ is the estimator of the variance σ_ε^2 of ε_i.

Note that the methods used to estimate $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ in Step 1 can be varying, as long as the corresponding estimators are consistent, please see more discussions in [86]. Li et al. [86] showed that, under some regularity conditions,

$$\sqrt{n}\widehat{S}\left(\beta^*, \widetilde{\gamma}\right)\left(\widehat{\sigma}_{\beta|\gamma, H_0}^2\right)^{-1/2} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as} \quad n \to \infty.$$

Further, the asymptotic normality of the test statistic $\widehat{T}_n$ at local alternatives was also established in [86] without any additional condition. Li et al. [86] also constructed the asymptotic confidence interval for target parameter β in Step 3 based on the asymptotic normality of $\widehat{\beta}$, which is given as follows

$$\sqrt{n}(\widehat{\beta} - \beta_0) = -\left[\mathbb{E}\left\{\left.\frac{\partial S\left(\beta, \gamma_0\right)}{\partial \beta}\right|_{\beta=\beta_0}\right\}\right]^{-1}\sqrt{n}S\left(\beta_0, \gamma_0\right) + o_P(1) \xrightarrow{d} \mathcal{N}(0, \sigma_\beta^2) \quad \text{as} \quad n \to \infty,$$

where $\sigma_\beta^2 = \left\{\mathbb{E}\left(X_i^2\right) - \boldsymbol{\omega}^{\mathrm{T}}\mathbb{E}\left(X_i\mathbf{Z}_i\right)\right\}^{-2}\sigma_{\beta|\gamma,0}^2$, and

$$\sigma_{\beta|\gamma,0}^2 = (\sigma_\varepsilon^2 + \beta_0^2\sigma_U^2)\left\{1 - \boldsymbol{\omega}^{\mathrm{T}}\mathbb{E}\left(X_i\mathbf{Z}_i\right)\right\} + \beta_0^2\mathbb{E}(U_i^4) + \sigma_\varepsilon^2\sigma_U^2 - \beta_0^2\sigma_U^4.$$

4.2. Wald and Score Tests for Poisson Models

In addition to linear models, researchers have made some progress on hypothesis testing problems for Poisson models. Jiang et al. [88] studied hypothesis testing problems for high-dimensional Poisson measurement error models, and they proposed Wald and score tests for the linear function of regression coefficients.

Consider the following hypothesis test

$$H_0: \mathbf{C}\boldsymbol{\beta}_{0\mathcal{M}} = \boldsymbol{b} \longleftrightarrow H_1: \mathbf{C}\boldsymbol{\beta}_{0\mathcal{M}} = \boldsymbol{b} + \boldsymbol{h}_n \text{ for some } \boldsymbol{h}_n \in \mathbb{R}^r,$$

where $\mathbf{C} \in \mathbb{R}^{r\times m}$ is a matrix with $r \leq m$, $\boldsymbol{\beta}_{0\mathcal{M}} \in \mathbb{R}^{m\times 1}$ is a subvector of the true regression coefficient vector $\boldsymbol{\beta}_0 = (\beta_{01}, \ldots, \beta_{0p})^{\mathrm{T}}$ formed by $\beta_{0j}(j \in \mathcal{M})$. To construct a valid test statistic, Jiang et al. [88] drew on the idea of estimation method in [74] and suggested estimating regression coefficients under the null hypothesis by

$$\widehat{\boldsymbol{\beta}}_{\mathrm{pn}} = \arg\min_{\|\boldsymbol{\beta}\|_1 \leqslant R_1, \|\boldsymbol{\beta}\|_2 \leqslant R_2}\left\{\mathcal{L}(\boldsymbol{\beta}) + p_\lambda(\boldsymbol{\beta}_{\mathcal{M}^c})\right\}, \quad \text{s.t. } \mathbf{C}\boldsymbol{\beta}_{\mathcal{M}} = \boldsymbol{b}, \tag{42}$$

where $p_\lambda(\cdot)$ is a penalty function, and $\mathcal{L}(\boldsymbol{\beta})$ is defined in (37). Similarly, the following estimator of $\boldsymbol{\beta}_0$ can be considered without assuming the null hypothesis

$$\widehat{\boldsymbol{\beta}}_{\mathrm{pw}} = \arg\min_{\|\boldsymbol{\beta}\|_1 \leqslant R_1, \|\boldsymbol{\beta}\|_2 \leqslant R_2}\left\{\mathcal{L}(\boldsymbol{\beta}) + p_\lambda(\boldsymbol{\beta}_{\mathcal{M}^c})\right\}. \tag{43}$$

The estimators $\widehat{\boldsymbol{\beta}}_{\mathrm{pn}}$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{pw}}$ can be obtained by ADMM algorithm, please see more details in [88]. It can be seen that optimization problems (42) and (43) can be distinguished from the method in (36) because we do not impose penalties on the components of the target parameter $\boldsymbol{\beta}_{\mathcal{M}}$ to avoid forcing them to be zeros. Then, based on the above estimators of $\boldsymbol{\beta}_0$, Jiang et al. [88] proposed the following score statistic and Wald statistic to test whether $\mathbf{C}\boldsymbol{\beta}_{0\mathcal{M}} = \boldsymbol{b}$ or not

$$T_S = n\left\{\frac{\partial\mathcal{L}(\widehat{\boldsymbol{\beta}})}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\right\}_{\mathcal{M}\cup\mathcal{S}}\mathbf{A}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}(\widehat{\boldsymbol{\Sigma}}^r, \widehat{\mathbf{Q}}, \widehat{\boldsymbol{\beta}})\mathbf{A}\left\{\frac{\partial\mathcal{L}(\widehat{\boldsymbol{\beta}})}{\partial\boldsymbol{\beta}}\right\}_{\mathcal{M}\cup\mathcal{S}},$$

$$T_W = n(\mathbf{C}\widehat{\boldsymbol{\beta}}_{\mathrm{pw},\mathcal{M}} - \boldsymbol{b})^{\mathrm{T}}\boldsymbol{\Psi}(\widehat{\boldsymbol{\Sigma}}^r, \widehat{\mathbf{Q}}, \widehat{\boldsymbol{\beta}}_{\mathrm{pw}})^{-1}(\mathbf{C}\widehat{\boldsymbol{\beta}}_{\mathrm{pw},\mathcal{M}} - \boldsymbol{b}),$$

where $\mathbf{A} = \mathbf{C}\left[\mathbf{I}_{m\times m}, \mathbf{0}_{m\times k}\right]\widehat{\mathbf{Q}}_{\mathcal{M}\cup\mathcal{S},\mathcal{M}\cup\mathcal{S}}^{-1}(\widehat{\boldsymbol{\beta}})$,

$$\boldsymbol{\Psi}(\boldsymbol{\Sigma}, \mathbf{Q}, \boldsymbol{\beta}) \equiv \mathbf{C}\left[\mathbf{I}_{m\times m}, \mathbf{0}_{m\times k}\right]\mathbf{Q}_{\mathcal{M}\cup\mathcal{S},\mathcal{M}\cup\mathcal{S}}^{-1}(\boldsymbol{\beta})\boldsymbol{\Sigma}_{\mathcal{M}\cup\mathcal{S},\mathcal{M}\cup\mathcal{S}}(\boldsymbol{\beta})\mathbf{Q}_{\mathcal{M}\cup\mathcal{S},\mathcal{M}\cup\mathcal{S}}^{-1}(\boldsymbol{\beta})\left[\mathbf{I}_{m\times m}, \mathbf{0}_{m\times k}\right]^{\mathrm{T}}\mathbf{C}^{\mathrm{T}},$$

$\widehat{\boldsymbol{\Sigma}}^r(\boldsymbol{\beta})$ and $\widehat{\mathbf{Q}}(\boldsymbol{\beta})$ are estimators of $\boldsymbol{\Sigma}^r(\boldsymbol{\beta})$ and $\mathbf{Q}(\boldsymbol{\beta}) = \mathbb{E}\left\{\exp\left(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}\right)\mathbf{X}\mathbf{X}^{\mathrm{T}}\right\}$ respectively, and

$$\boldsymbol{\Sigma}^r(\boldsymbol{\beta}) = \mathbb{E}\left[\left\{Y_i\mathbf{W}_i - \exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{W}_i - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_u\boldsymbol{\beta}/2)\left(\mathbf{W}_i - \boldsymbol{\Sigma}_u\boldsymbol{\beta}\right)\right\}^{\otimes 2}\right]$$

is the covariance of the residuals.

Jiang et al. [88] established the consistency of $\widehat{\boldsymbol{\beta}}_{\mathrm{pn}}$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{pw}}$ with λ larger than $O(\{\log p/n\}^{1/4})$, $m = o(\{\log p/n\}^{1/2})$ and $s = o(\{\log p/n\}^{1/2})$. Further, the asymptotic distributions of the two test statistics are established, specifically, as $n \to \infty$, we have

$$T_S \xrightarrow{d} \chi^2\left(r, n\mathbf{h}_n^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\left(\boldsymbol{\Sigma}, \mathbf{Q}, \boldsymbol{\beta}_t\right)\mathbf{h}_n\right), \qquad T_W \xrightarrow{d} \chi^2\left(r, n\mathbf{h}_n^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\left(\boldsymbol{\Sigma}, \mathbf{Q}, \boldsymbol{\beta}_t\right)\mathbf{h}_n\right).$$

Thus, we reject the null hypothesis if $T_S > \chi_{1-\alpha}^2$ for score test with the nominal significance level $\alpha > 0$, and reject the null hypothesis if $T_W > \chi_{1-\alpha}^2$ for Wald test, where $\chi_{1-\alpha}^2$ is the $(1-\alpha)$ quantile of the chi-square distribution $\chi^2(r)$.

5. Screening Methods

As the dimension of data becomes higher and higher, we often encounter ultrahigh-dimensional data. For the ultrahigh-dimensional models, we frequently reduce dimension using variable screening techniques and then apply other estimation or hypothesis testing methods. The variable screening technique SIS [49] designed for ultrahigh-dimensional clean data has achieved great success and has been extended to various settings. SIS screens the variables according to the magnitudes of their marginal correlations with the response variable. Nghiem et al. [89] drew inspiration from the ideas of SIS in [49] and marginal bridge estimation in [98], and proposed the corrected sure independence screening (SISc) method and corrected penalized marginal screening method (PMSc). Consider the following optimization problem

$$\begin{aligned}\widetilde{\boldsymbol{\beta}}_{\mathrm{sc}} &= \arg\min_{\boldsymbol{\beta}} L(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}}\left\{\sum_{j=1}^{p} L_j(\beta_j)\right\} \\ &= \arg\min_{\boldsymbol{\beta}}\left\{\frac{1}{n}\sum_{j=1}^{p}\left[\sum_{i=1}^{n}\left(y_i - w_{ij}\beta_j\right)^2 - \sigma_j^2\beta_j^2 + p_\lambda\left(|\beta_j|\right)\right]\right\},\end{aligned} \tag{44}$$

where $p_\lambda(\cdot)$ is a penalty function, and the bridge penalty is adopted in [89]. Based on (44), Nghiem et al. [89] proposed PMSc and SISc methods. For PMSc method, it suggested taking the selected submodel as

$$\widehat{\mathcal{S}}_{\mathrm{PMSc}} = \left\{j : \widetilde{\beta}_{\mathrm{sc},j} \neq 0\right\}.$$

Under some regularity conditions, Nghiem et al. [89] showed that $\mathbb{P}(\mathcal{S} \subset \widehat{\mathcal{S}}_{\mathrm{PMSc}}) \to 1$. Furthermore, when $\lambda = 0$, we can obtain that

$$\widetilde{\beta}_{\mathrm{sc},j} = \frac{\sum_{i=1}^{n} w_{ij}y_i}{\sum_{i=1}^{n} w_{ij}^2 - n\sigma_{u,j}^2}, \quad j = 1, \ldots, p,$$

which measures the marginal correlation between the jth variable and the response variable. The SISc selects the variable according to the magnitude of $\widetilde{\beta}_{\mathrm{sc},j}$. The corresponding selected set is

$$\widehat{\mathcal{S}}_{\mathrm{SISc}} = \left\{1 \leq j \leq p : |\widetilde{\beta}_j| \text{ is among the } d \text{ largest of all}\right\}.$$

Nghiem et al. [89] proved that $\mathbb{P}(\mathcal{S} \subset \widehat{\mathcal{S}}_{\text{SISc}}) = 1 - O\{p\exp(-Cn)\}$ for some constant $C > 0$ under some regularity conditions.

6. Conclusions

With the advent of big data era, high-dimensional measurement error data have proliferated in various fields. Over the past few years, many statistical inference methods for high-dimensional measurement error regression models have been developed to overcome the difficulties in scientific research and provide effective approaches for tackling problems in applications. This paper reviews the research advances in estimation and hypothesis testing methods for high-dimensional measurement error models, as well as variable screening methods for ultrahigh-dimensional measurement error models. Due to the prevalence of high-dimensional measurement error data in daily life and the growing demand for the statistical inference methods of measurement error regression models in applications, the related research is still one of the crucial aspects in statistical research. At present, the statistical inference methods and the theoretical system of high-dimensional measurement error models are far from complete. Further research in this area includes the following aspects.

1. Existing estimation methods for high-dimensional measurement error regression models are mainly for linear or generalized linear models. Therefore, it is urgent to develop estimation methods for nonlinear models with high-dimensional measurement error data such as nonparametric and semiparametric models.
2. Existing works mainly focus on independent and identically distributed data. It is worthwhile to extend the estimation and hypothesis testing methods to measurement error models with complex data such as panel data and functional data.
3. In most studies of high-dimensional measurement error models, it is assumed that the covariance structure of the measurement errors is specific or that the covariance matrix of measurement errors is known. Thus, it is a challenging problem to develop estimation and hypothesis testing methods in the case that the covariance matrix of measurement errors is completely unknown.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Abbreviations

The following abbreviations are used in this manuscript:

SIMEX	Simulation-extrapolation
SCAD	Smoothly clipped absolute deviation
SICA	Smooth integration of counting and absolute deviation
MCP	Minimax concave penalty
SIS	Sure independence screening
CoCoLasso	Convex conditioned Lasso
CaZnRLS	Calibrated zero-norm regularized least squares
MU	Matrix uncertainty
MEBoost	Measurement error boosting

SIMSELEX	Simulation-selection-extrapolation
IRO	Imputation-regularized optimization
FDR	False discovery rate
PMSc	Corrected penalized marginal screening
SISc	Corrected sure independence screening
ADMM	Alternating direction method of multipliers
BDCoCoLasso	Block coordinate descent convex conditioned Lasso
MPEC	Mathematical program with equilibrium constraints
GEP–MSCRA	Multi-stage convex relaxation approach
GMU	Generalized matrix uncertainty

References

1. Liang, H.; Härdle, W.; Carroll, R.J. Estimation in a semiparametric partially linear errors-in-variables model. *The Annals of Statistics* **1999**, *27(5)*, 1519–1535.
2. Cook, J.; Stefanski, L.A. Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* **1994**, *89(428)*, 1314–1328.
3. Carroll, R.J.; Lombard, F.; Kuchenhoff, H.; Stefanski, L.A. Asymptotics for the SIMEX estimator in structural measurement error models. *Journal of the American Statistical Association* **1996**, *91(433)*, 242–250.
4. Fan, J.Q.; Truong, Y.K. Nonparametric regression with errors in variables. *The Annals of Statistics* **1993**, *21(4)*, 1900–1925.
5. Cui, H.J.; Chen, S.X. Empirical likelihood confidence region for parameter in the errors-in-variables models. *Journal of Multivariate Analysis* **2003**, *84(1)*, 101–115.
6. Cui, H.J.; Kong, E.F. Empirical likelihood confidence region for parameters in semi-linear errors-in-variables models. *Scandinavian Journal of statistics* **2006**, *33(1)*, 153–168.
7. Cheng, C.L.; Tsai, J.R.; Schneeweiss H. Polynomial regression with heteroscedastic measurement errors in both axes: estimation and hypothesis testing. *Statistical Methods in Medical Research* **2019**, *28(9)*, 2681–2696.
8. He, X.M.; Liang, H. Quantile regression estimates for a class of linear and partially linear errors-in-variables models. *Statistica Sinica* **2000**, *10*, 129–140.
9. Carroll, R.J.; Delaigle, A.; Hall, P. Nonparametric prediction in measurement error models. *Journal of the American Statistical Association* **2009**, *104(487)*, 993–1003.
10. Jeon, J.M.; Park, B.U.; Keilegom, I.V. Nonparametric regression on lie groups with measurement errors. *The Annals of Statistics* **2022**, *50(5)*, 2973–3008.
11. Chen, L.P.; Yi, G.Y. Model selection and model averaging for analysis of truncated and censored data with measurement error. *Electronic Journal of Statistics* **2020**, *14(2)*, 4054–4109.
12. Shi, P.X.; Zhou, Y.C.; Zhang, A.R. High-dimensional log-error-in-variable regression with applications to microbial compositional data analysis. *Biometrika* **2022**, *109(2)*, 405–420.
13. Li, B.; Yin, X.R. On surrogate dimension reduction for measurement error regression: an invariance law. *The Annals of Statistics* **2007**, *35(5)*, 2143–2172.
14. Staudenmayer, J.; Buonaccorsi, J.P. Measurement error in linear autoregressive models. *Journal of the American Statistical Association* **2005**, *100(471)*, 841–852.
15. Wei, Y.; Carroll, R.J. Quantile regression with measurement error. *Journal of the American Statistical Association* **2009**, *104(487)*, 1129–1143.
16. Liang, H.; Li, R.Z. Variable selection for partially linear models with measurement errors. *Journal of the American Statistical Association* **2009**, *104(485)*, 234–248.
17. Hall, P.; Ma, Y.Y. Estimation in a semiparametric partially linear errors-in-variables model. *The Annals of Statistics* **2007**, *35(6)*, 2620–2638.
18. Hall, P.; Ma, Y.Y. Semiparametric estimators of functional measurement error models with unknown error. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **2007**, *69(3)*, 429–446.
19. Ma, Y.Y.; Carroll, R.J. Locally efficient estimators for semiparametric models with measurement error. *Journal of the American Statistical Association* **2006**, *101(476)*, 1465–1474.
20. Ma, Y.Y.; Li, R.Z. Variable selection in measurement error models. *Bernoulli* **2010**, *16(1)*, 274–300.

21. Ma, Y.Y.; Hart, J.D.; Janicki, R.; Carroll, R.J. Local and omnibus goodness-of-fit tests in classical measurement error models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **2011**, *73(1)*, 81–98.

22. Wang, L.Q. Estimation of nonlinear models with Berkson measurement errors. *The Annals of Statistics* **2004**, *32(6)*, 2559–2579.

23. Nghiem, L.H.; Byrd, M.C.; Potgieter, C.J. Estimation in linear errors-in-variables models with unknown error distribution. *Biometrika* **2020**, *107(4)*, 841–856.

24. Pan, W.Q.; Zeng, D.L.; Lin, X.H. Estimation in semiparametric transition measurement error models for longitudinal data. *Biometrics* **2009**, *65(3)*, 728–736.

25. Zhang, J.; Zhou, Y. Calibration procedures for linear regression models with multiplicative distortion measurement errors. *Brazilian Journal of Probability and Statistics* **2020**, *34(3)*, 519–536.

26. Zhang, J. Estimation and variable selection for partial linear single-index distortion measurement errors models. *Statistical Papers* **2021**, *62*, 887–913.

27. Wang, L.Q.; Hsiao, C. Method of moments estimation and identifiability of semiparametric nonlinear errors-in-variables models. *Journal of Econometrics* **2011**, *165*, 30–44.

28. Schennach, S.M.; Hu, Y.Y. Nonparametric identification and semiparametric estimation of classical measurement error models without side information. *Journal of the American Statistical Association* **2013**, *108(501)*, 177–186.

29. Zhang, X.Y.; Ma, Y.Y.; Carroll, R.J. MALMEM: model averaging in linear measurement error models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **2019**, *81(4)*, 763–779.

30. Carroll, R.J.; Ruppert, D.; Stefanski, L.A.; Crainiceanu, C.M. *Measurement Error in Nonlinear Models*, 2nd ed.; Chapman and Hall: New York, America, 2006.

31. Cheng, C.L.; Van Ness, J.W. *Statistical Regression With Measurement Error*; Oxford University Press: New York, America, 1999.

32. Fuller, W.A. *Measurement Error Models*; John Wiley & Sons: New York, America, 1987.

33. Li, G.R.; Zhang, J.; Feng, S.Y. *Modern Measurement Error Models*; Science Press: Beijing, China, 2016.

34. Yi, G.Y. *Statistical Analysis with Measurement Error or Misclassification*; Springer: New York, America, 2017.

35. Yi, G.Y.; Delaigle, A.; Gustafson, P. *Handbook of Measurement Error Models*; Chapman and Hall: New York, America, 2021.

36. Tibshirani, R. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **1996**, *58*, 267–288.

37. Fan, J.Q.; Li, R.Z. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **2001**, *96(456)*, 1348–1360.

38. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **2005**, *67*, 301–320.

39. Zou, H. The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association* **2006**, *101(476)*, 1418–1429.

40. Candès, E.J.; Tao, T. The Dantzig selector: statistical estimation when p is much larger than n. *The Annals of Statistics* **2007**, *35(6)*, 2313–2351.

41. Lv, J.C.; Fan, Y.Y. A unified approach to model selection and sparse recovery using regularized least squares. *The Annals of Statistics* **2009**, *37(6A)*, 3498–3528.

42. Zhang, C.-H. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics* **2010**, *38(2)*, 894-942.

43. Fan, J.Q.; Lv, J.C. A selective overview of variable selection in high dimensional feature space. *Statistica Sinica* **2010**, *20*, 101–148.

44. Wu, Y.N.; Wang, L. A survey of tuning parameter selection for high-dimensional regression. *Annual Review of Statistics and Its Application* **2020**, *7*, 209–226.

45. Kuchibhotla, A.K.; Kolassa, J.E.; Kuffner, T.A. Post-selection inference. *Annual Review of Statistics and Its Application* **2022**, *9*, 1–23.

46. Bühlmann, P.; van de Geer, S. *Statistics for High-Dimensional Data: Methods, Theory and Applications*; Springer-Verlag: Heidelberg, Germany, 2011.

47. Hastie, T.; Tibshirani, R.; Wainwright, M. *Statistical Learning with Sparsity: The Lasso and Generalizations*; Taylor & Francis Group, CRC: Boca Raton, America, 2015.

48.   Fan, J.Q.; Li, R.Z.; Zhang, C.-H.; Zou, H. *Statistical Foundations of Data Science*; Chapman and Hall: Boca Raton, America, 2020.

49.   Fan, J.Q.; Lv, J.C. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **2008**, *70(5)*, 849–911.

50.   Barut, E.; Fan, J.Q.; Verhasselt, A. Conditional sure independence screening. *Journal of the American Statistical Association* **2016**, *111(515)*, 1266–1277.

51.   Fan, J.Q.; Song, R. Sure independence screening in generalized linear models with NP-dimensionality. *The Annals of Statistics* **2010**, *38(6)*, 3567–3604.

52.   Fan, J.Q.; Feng, Y.; Song, R. Nonparametric independence screening in sparse ultrahigh-dimensional additive models. *Journal of the American Statistical Association* **2011**, *106(494)*, 544–557.

53.   Li, G.R.; Peng, H.; Zhang, J.; Zhu, L.X. Robust rank correlation based screening. *The Annals of Statistics* **2012**, *40(3)*, 1846–1877.

54.   Ma, S.J.; Li, R.Z.; Tsai, C.L. Variable screening via quantile partial correlation. *Journal of the American Statistical Association* **2017**, *112(518)*, 650–663.

55.   Pan, W.L.; Wang, X.Q.; Xiao, W.N.; Zhu, H.T. A generic sure independence screening procedure. *Journal of the American Statistical Association* **2019**, *114(526)*, 928–937.

56.   Tong, Z.X.; Cai, Z.R.; Yang, S.S.; Li, R.Z. Model-free conditional feature screening with FDR control. *Journal of the American Statistical Association* **2022**, *in press*.

57.   Wen, C.H.; Pan, W.L.; Huang, M.; Wang, X.Q. Sure independence screening adjusted for confounding covariates with ultrahigh dimensional data. *Statistica Sinica* **2018**, *28(1)*, 293–317.

58.   Wang, L.M.; Li, X.X.; Wang, X.Q.; Lai, P. Unified mean-variance feature screening for ultrahigh-dimensional regression. *Computational Statistics* **2022**, *37*, 1887–1918.

59.   Zhao, S.F.; Fu, G.F. Distribution-free and model-free multivariate feature screening via multivariate rank distance correlation. *Journal of Multivariate Analysis* **2022**, *192*, article 105081.

60.   Slijepcevic, S.; Megerian, S; Potkonjak, M. Location errors in wireless embedded sensor networks: sources, models, and effects on applications. *Mobile Computing and Communications Review* **2002**, *6(3)*, 67–78.

61.   Purdom, E.; Holmes, S.P. Error distribution for gene expression data. *Statistical Applications in Genetics and Molecular Biology* **2005**, *4(1)*, 16–16.

62.   Loh, P.-L.; Wainwright, M.J. High-dimensional regression with noisy and missing data: provable guarantees with nonconvexity. *The Annals of Statistics* **2012**, *40(3)*, 1637–1664.

63.   Datta, A.; Zou, H. CoCoLasso for high-dimensional error-in-variables regression. *The Annals of Statistics* **2017**, *45(6)*, 2400–2426.

64.   Zheng, Z.M.; Li, Y.; Yu, C.X.; Li, G.R. Balanced estimation for high-dimensional measurement error models. *Computational Statistics & Data Analysis* **2018**, *126*, 78–91.

65.   Tao, T.; Pan, S.H.; Bi, S.J. Calibrated zero-norm regularized LS estimator for high-dimensional error-in-variables regression. *Statistica Sinica* **2018**, *31(2)*, 909–933.

66.   Rosenbaum, M.; Tsybakov, A. Sparse recovery under matrix uncertainty. *The Annals of Statistics* **2010**, *38(5)*, 2620–2651.

67.   Rosenbaum, M.; Tsybakov, A. Improved matrix uncertainty selector. *From Probability to Statistics and Back: High-Dimensional Models and Processes* **2013**, *9*, 276–290.

68.   Sørensen, Ø.; Hellton, K.H.; Frigessi, A.; Thoresen, M. Covariate selection in high-dimensional generalized linear models with measurement error. *Journal of Computational and Graphical Statistics* **2018**, *27(4)*, 739–749.

69.   Sørensen, Ø.; Frigessi, A.; Thoresen, M. Measurement error in Lasso: impact and likelihood bias correction. *Statistics Sinica* **2015**, *25(2)*, 809–829.

70.   Belloni, A.; Rosenbaum, M.; Tsybakov, A.B. Linear and conic programming estimators in high dimensional errors-in-variables models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **2017**, *79(3)*, 939–956.

71.   Romeo, G.; Thoresen, M. Model selection in high-dimensional noisy data: a simulation study. *Journal of Statistical Computation and Simulation* **2019**, *89(11)*, 2031–2050.

72.   Brown, B.; Weaver, T.; Wolfson, J. Meboost: variable selection in the presence of measurement error. *Statistics in Medicine* **2019**, *38*, 2705–2718.

73.   Nghiem, L.H.; Potgieter, C.J. Simulation-selection-extrapolation: estimation in high-dimensional errors-in-variables models. *Biometrics* **2019**, *75*, 1133–1144.

74. Jiang, F.; Ma, Y.Y. Poisson regression with error corrupted high dimensional features. *Statistica Sinica* **2022**, *32*, 2023–2046.

75. Byrd, M.; McGee, M. A simple correction procedure for high-dimensional generalized linear models with measurement error. *arXiv preprint* **2019**, arXiv:1912.11740.

76. Liang, F.M.; Jia, B.C.; Xue, J.N.; Li, Q.Z.; Luo, Y. An imputation–regularized optimization algorithm for high dimensional missing data problems and beyond. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **2018**, *80(5)*, 899–926.

77. van de Geer, S.; Bühlmann, P.; Ritov, Y.; Dezeure, R. On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics* **2014**, *42(3)*, 1166–1202.

78. Zhang, C.-H.; Zhang, S.S. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **2014**, *76*, 217–242.

79. Ma, S.J.; Carroll, R.J.; Liang, H.; Xu, S.Z. Estimation and inference in generalized additive coefficient models for nonlinear interactions with high-dimensional covariates. *The Annals of Statistics* **2015**, 43(5), 2102–2131.

80. Dezeure, R.; Bühlmann, P.; Meier, L.; Meinshausen, N. High-dimensional inference: confidence intervals, p-values and R-software hdi. *Statistical Science* **2015**, 30(4), 533–558.

81. Ning, Y.; Liu, H. A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *The Annals of Statistics* **2017**, *45(1)*, 158–195.

82. Zhang, X.Y.; Cheng, G. Simultaneous inference for high-dimensional linear models. *Journal of the American Statistical Association* **2017**, *112(518)*, 757–768.

83. Vandekar, S.N.; Reiss, P.T.; Shinohara, R.T. Interpretable high-dimensional inference via score projection with an application in neuroimaging. *Journal of the American Statistical Association* **2019**, *114(526)*, 820–830.

84. Ghosh, S.; Tan, Z.Q. Doubly robust semiparametric inference using regularized calibrated estimation with high-dimensional data. *Bernoulli* **2022**, *28(3)*, 1675–1703.

85. Belloni, A.; Chernozhukov, V.; Kaul, A. Confidence bands for coefficients in high dimensional linear models with error-in-variables. *arXiv preprint* **2017**, arXiv:1703.00469.

86. Li, M.Y.; Li, R.Z.; Ma, Y.Y. Inference in high dimensional linear measurement error models. *Journal of Multivariate Analysis* **2021**, *184*, article 104759.

87. Huang, X.D.; Bao, N.N.; Xu, K.; Wang, G.P. Variable selection in high-dimensional error-in-variables models via controlling the false discovery proportion. *Communications in Mathematics and Statistics* **2022**, *10*, 123–151.

88. Jiang, F.; Zhou, Y.Q.; Liu, J.X.; Ma, Y.Y. On high dimensional Poisson models with measurement error: hypothesis testing for nonlinear nonconvex optimization. *The Annals of Statistics* **2023**, *51(1)*, 233–259.

89. Nghiem, L.H.; Hui, F.K.C.; Müller, S.; Welsh, A.H. Screening methods for linear errors-in-variables models in high dimensions. *Biometrics* **2022**, *in press*.

90. Duchi, J.; Shalev-Shwartz,S.; Singer, Y.; Chandra, T. Efficient projections onto the l_1-ball for learning in high dimensions. In Proceedings of International Conference on Machine Learning, New York, America, July 2008.

91. Agarwal, A.; Negahban, S.; Wainwright, M.J. Fast global convergence of gradient methods for high-dimensional statistical recovery. *The Annals of Statistics* **2012**, *40(5)*, 2452–2482.

92. Chen, Y.D.; Caramanis, C. Noisy and missing data regression: distribution-oblivious support recovery. *Journal of Machine Learning Research* **2013**, *28*, 383–391.

93. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning* **2011**, *3(1)*, 1–122.

94. Efron, B.; Hastie, T.; Johnstone, I.; Tibshirani, R. Least angle regression. *The Annals of Statistics* **2004**, *32(2)*, 407–499.

95. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* **2010**, *33(1)*, 1–22.

96. Escribe, C.; Lu, T.Y.; Keller-Baruch, J.; Forgetta, V.; Xiao, B.W.; Richards, J.B.; Bhatnagar, S.; Oualkacha, K.; Greenwood, C.M.T. Block coordinate descent algorithm improves variable selection and estimation in error-in-variables regression. *Genetic Epidemiology* **2021**, *45*, 874–890.

97. James, G.M.; Radchenko, P. A generalized Dantzig selector with shrinkage tuning. *Biometrika* **2009**, *96(2)*, 323–337.
98. Huang, J.; Horowitz, J.L.; Ma, S.G. Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *The Annals of Statistics* **2008**, *36(2)*, 587–613.