

Article

Instrumental Variable Method for Regularized Estimation in Generalized Linear Measurement Error Models

Lin Xue and Liqun Wang * 

Department of Statistics, University of Manitoba, Winnipeg, MB R3T 2N2, Canada; linx.stats@gmail.com

* Correspondence: liqun.wang@umanitoba.ca

Abstract: Regularized regression methods have attracted much attention in the literature, mainly due to its application in high-dimensional variable selection problems. Most existing regularization methods assume that the predictors are directly observed and precisely measured. It is well known that in a low-dimensional regression model if some covariates are measured with error, then the naive estimators that ignore the measurement error are biased and inconsistent. However, the impact of measurement error in regularized estimation procedures is not clear. For example, it is known that the ordinary least squares estimate of the regression coefficient in a linear model is attenuated towards zero and, on the other hand, the variance of the observed surrogate predictor is inflated. Therefore, it is unclear how the interaction of these two factors affects the selection outcome. To correct for the measurement error effects, some researchers assume that the measurement error covariance matrix is known or can be estimated using external data. In this paper, we propose the regularized instrumental variable method for generalized linear measurement error models. We show that the proposed approach yields a consistent variable selection procedure and root-n consistent parameter estimators. Extensive finite sample simulation studies show that the proposed method performs satisfactorily in both linear and generalized linear models. A real data example is provided to further demonstrate the usage of the method.

Keywords: regularization method; penalized estimation; measurement error; instrumental variable; generalized linear model; variable selection; consistency; oracle property



Citation: Xue, Lin, and Liqun Wang. 2024. Instrumental Variable Method for Regularized Estimation in Generalized Linear Measurement Error Models. *Econometrics* 12: 21. <https://doi.org/10.3390/econometrics12030021>

Academic Editor: Xinzheng Shi

Received: 26 April 2024

Revised: 16 June 2024

Accepted: 5 July 2024

Published: 12 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Regularization is an important approach for estimation in regression models with a relatively large number of parameters because it provides a stable numerical procedure and better prediction while avoiding the overfitting problem. This approach has attracted much attention in the recent literature, mainly due to its applications in variable selection problems in high-dimensional models where the conventional statistical methods are infeasible theoretically and computationally. To address the variable selection problem in sparse regression, various regularization methods have been proposed, e.g., the bridge regression (Frank and Friedman 1993), Lasso (Tibshirani 1996), SCAD (Fan and Li 2001), adaptive Lasso (Zou 2006), MCP (Zhang et al. 2010), elastic net (Zou and Hastie 2005) and Dantzig selector (Candes et al. 2007). A more detailed review of regularization methods can be found in Fan and Lv (2010) and Negahban et al. (2012).

In real data analysis, it is common that some predictors cannot be observed directly or measured precisely. For example, the long-term average systolic blood pressure and cholesterol level are important factors of cardiovascular disease, which are usually measured with error. In a lung cancer risk study, the inhaled dose of air pollutants cannot be measured precisely and are approximated by the average level of pollutants within a certain area. In regression models, it is well known that if some predictors are measured with error, ordinary estimation procedures ignoring the ME are biased and inconsistent. However, the impact of ME on regularized estimation procedures is not clear. For example,

in a linear model, the naive least squares estimate for the regression coefficient of the mismeasured predictor is attenuated towards zero and, on the other hand, the variance of the corresponding observed surrogate predictor is inflated. Therefore, the combined effect of these two factors may cause false-positive or -negative results in the selection outcome, as illustrated in Example 1.

Research on regularized estimation in ME models is sparse. Some authors considered the penalized version of the usual error correction methods, assuming the ME covariance matrix is known or can be estimated using replicate data. For example, Liang and Li (2009) applied the penalized least squares method with attenuation correction and quantile estimation with orthogonal regression adjustment in a partial linear model. Ma and Li (2010) studied general parametric and semiparametric models using the method of penalized estimating equations. Further, Huang and Zhang (2013) used the penalized score functions, while Zhang et al. (2017) used a prediction criterion for variable selection in linear ME models.

Another major approach to estimation in ME models is the instrumental variable (IV) method. This method has been used to treat the endogeneity problem in high-dimensional regression models by Fan and Liao (2014) who proposed the focused generalized method of moments estimator. Lin et al. (2015) studied a two-stage regularization method for selecting relevant instruments and predictors in linear models under the assumption that the random errors are jointly normally distributed. Zhong et al. (2020) proposed a two-stage estimation procedure with instrumental variables for a dummy endogenous variable.

All these works mainly focus on a general endogeneity problem in linear models or binary response variables. So far, there is very few, if any, published studies focusing specifically on the IV approach to measurement error problems. In this paper, we try to fill in this gap. Specifically, we extend the IV method to study the variable selection and estimation problem in generalized linear models with ME. This method does not require the distribution or covariance matrix of the ME to be known. It is an extension of the method of conditional moments by Wang and Hsiao (2011). The proposed selection procedure and estimator are consistent and enjoy the oracle property under general conditions.

The rest of the paper is organized as follows: In Section 2, we introduce the regularized instrumental variable method and study its asymptotic properties. Section 3 contains the special case of linear model. Numerical examples are given in Section 4 followed by a real example in Section 5. Technical details are relegated to Appendix A.

2. The Model and Estimation Method

Suppose the response variable Y has the conditional mean function

$$E(Y|X, Z) = g(\alpha + \beta_x^T X + \beta_z^T Z), \quad (1)$$

where $X \in \mathbb{R}^p$ is a vector of error-prone predictors in low dimension, $Z \in \mathbb{R}^q$ is a vector of error-free predictors and $g(\cdot)$ is a link function. Equation (1) includes the generalized linear models as well as the so-called single index models as special cases. We assume that the observed surrogate predictors are

$$X^* = X + \delta, \quad (2)$$

where δ is a random ME. Further, we assume that there are instrumental variables (IV) $W \in \mathbb{R}^l$ besides the main sample (Y, X^*, Z) . The usual requirement for an IV is that it is correlated with the unobserved predictor X but independent of the ME δ and is conditionally independent of Y given (X, Z) . Following the literature (Wang 2021; Wang and Hsiao 2011), we assume that the IV W is related with X through

$$X = \Gamma W + U, \quad (3)$$

where Γ is the $p \times l$ matrix of unknown parameters which is assumed to have full rank p , \mathbf{U} is independent of \mathbf{W}, \mathbf{Z} , has mean $E(\mathbf{U}) = \mathbf{0}$ and density $f_U(\mathbf{u}; \boldsymbol{\phi})$ with unknown parameters $\boldsymbol{\phi} \in \mathbb{R}^k$. It is further assumed that the ME δ in (2) satisfies $E(\delta|\mathbf{X}, \mathbf{Z}, \mathbf{W}) = \mathbf{0}$. Throughout this paper, we assumed that \mathbf{Z} is exogenous and all expectations on it are taken conditionally; however, \mathbf{Z} is suppressed to simplify notations. We also adopt the common assumption in the ME literature that the ME δ is nondifferential, which implies that $E(Y|\mathbf{X}, \mathbf{X}^*, \mathbf{W}) = E(Y|\mathbf{X})$.

Now, we consider the estimation of unknown parameters in (1)–(3) given an iid random sample $Y_i, \mathbf{X}_i^*, \mathbf{W}_i, i = 1, 2, \dots, n$. First, substituting (3) into (2) results in a usual linear regression equation

$$\mathbf{X}^* = \Gamma\mathbf{W} + \mathbf{U} + \delta$$

and, therefore, Γ can be consistently estimated by the least squares estimator

$$\hat{\Gamma} = \left(\sum_{i=1}^n \mathbf{X}_i^* \mathbf{W}_i^T\right) \left(\sum_{i=1}^n \mathbf{W}_i \mathbf{W}_i^T\right)^{-1}.$$

In the following, we focus on the estimation of other parameters of main interest $\boldsymbol{\psi} = (\alpha, \boldsymbol{\beta}_x^T, \boldsymbol{\beta}_z^T, \boldsymbol{\phi}^T)^T$ in model (1)–(3). Specifically, we propose an estimator based on the first two conditional moments $E(Y|\mathbf{W})$ and $E(Y\mathbf{X}^*|\mathbf{W})$. To simplify notation, we denote $\tilde{\mathbf{X}}^* = (1, \mathbf{X}^{*T})^T$ and $\tilde{\mathbf{X}} = (1, \mathbf{X}^T)^T$. Then, the two conditional moments can be written together as

$$\begin{aligned} E(\tilde{\mathbf{X}}^* Y|\mathbf{W}) &= \int \tilde{\mathbf{X}}^* g(\alpha + \boldsymbol{\beta}_x^T \Gamma \mathbf{W} + \boldsymbol{\beta}_z^T \mathbf{Z} + \boldsymbol{\beta}_x^T \mathbf{u}) f_U(\mathbf{u}; \boldsymbol{\phi}) d\mathbf{u} \\ &= \int \tilde{\mathbf{x}} g(\alpha + \boldsymbol{\beta}_x^T \mathbf{x} + \boldsymbol{\beta}_z^T \mathbf{Z}) f_U(\mathbf{x} - \Gamma \mathbf{W}; \boldsymbol{\phi}) d\mathbf{x} \\ &:= m(\Gamma \mathbf{W}; \boldsymbol{\psi}), \end{aligned} \tag{4}$$

where

$$m(\mathbf{v}; \boldsymbol{\psi}) = \int \tilde{\mathbf{x}} g(\alpha + \boldsymbol{\beta}_x^T \mathbf{x} + \boldsymbol{\beta}_z^T \mathbf{Z}) f_U(\mathbf{x} - \mathbf{v}; \boldsymbol{\phi}) d\mathbf{x}.$$

Then, the loss function for estimating $\boldsymbol{\psi}$ is defined as

$$L_n(\boldsymbol{\psi}) = \frac{1}{2} \sum_{i=1}^n \hat{\rho}_i(\boldsymbol{\psi})^T A_i \hat{\rho}_i(\boldsymbol{\psi}), \tag{5}$$

where $\hat{\rho}_i(\boldsymbol{\psi}) = Y_i \tilde{\mathbf{X}}_i^* - m(\hat{\Gamma} \mathbf{W}_i; \boldsymbol{\psi})$ and $A_i = A(\mathbf{W}_i)$ are a semipositive definite matrix which may depend on \mathbf{W}_i .

One of the main features in the high-dimensional variable selection framework is the sparsity of the model, where many regression parameters in $\boldsymbol{\beta} = (\boldsymbol{\beta}_x^T, \boldsymbol{\beta}_z^T)^T$ have a true value zero. In the following, we denote the true parameter values of $\boldsymbol{\beta}$ as $\boldsymbol{\beta}_0$, the index set of non-zero coefficients as $J = \{j : \beta_{0j} \neq 0\}$ and its complement set as $J^c = \{j : \beta_{0j} = 0\}$. We further denote $\boldsymbol{\beta}_J = \{\beta_j, j \in J\}$, $\boldsymbol{\beta}_{J^c} = \{\beta_j, j \in J^c\}$ and $\boldsymbol{\psi}_J = (\alpha, \boldsymbol{\beta}_J^T, \boldsymbol{\phi}^T)^T$. Similarly, let Γ_J be the matrix consisting of rows of Γ corresponding to the index set J , and $\boldsymbol{\gamma} = \text{vec}(\Gamma^T)$ as the vector consisting of the columns of Γ^T . Finally, the proposed regularized IV estimator is defined as the minimizer of the objective function

$$Q_n(\boldsymbol{\psi}) = L_n(\boldsymbol{\psi}) + n \sum_{j=1}^d p_{\lambda_n}(|\beta_j|), \tag{6}$$

where $p_{\lambda_n}(|b|)$ is a penalty function. Let $\boldsymbol{\psi}_0 = (\alpha_0, \boldsymbol{\beta}_0^T, \boldsymbol{\phi}_0^T)^T$ be the true value of model parameters.

Theorem 1. Under Assumptions A1–A5 in Appendix A, suppose the penalty function satisfies

$$a_n = \max\{p'_{\lambda_n}(|\beta_{0j}|) : \beta_{0j} \neq 0\} = O(n^{-1/2}),$$

and

$$b_n = \max\{p''_{\lambda_n}(|\beta_{0j}|) : \beta_{0j} \neq 0\} = o(1).$$

Then there exists a local minimizer $\hat{\psi}$ of the objective function (6) such that $\|\hat{\psi} - \psi_0\| = O_p(n^{-1/2})$.

Further, let $\mathbf{b} = (0, p'_{\lambda_n}(|\beta_{0j}^T|), 0)^T \circ \text{sign}(\psi_{0j})$ and $\Sigma = \text{diag}(0, p''_{\lambda_n}(|\beta_{0j}^T|), 0)$. We have the following results.

Theorem 2. If $\lambda_n \rightarrow 0$, $\sqrt{n}\lambda_n \rightarrow \infty$ and $\liminf_{n \rightarrow \infty} \liminf_{\xi \rightarrow 0^+} p'_{\lambda_n}(\xi) / \lambda_n > 0$, then with probability approaching 1, the root n consistent estimator $\hat{\psi}$ in Theorem 1, satisfies

- (a) $\hat{\beta}_{jc} = \mathbf{0}$,
- (b) $\hat{\psi}_j$ has asymptotic distribution

$$\sqrt{n}(H + \Sigma)(\hat{\psi}_j - \psi_{0j}) + \sqrt{n}\mathbf{b} \xrightarrow{d} N(0, DCD^T),$$

where

$$H = E \left[\frac{\partial \rho^T(\psi_{0j})}{\partial \psi_j} A(\mathbf{W}) \frac{\partial \rho(\psi_{0j})}{\partial \psi_j^T} \right],$$

$$D = \left[I_{s+2}, E \left(\frac{\partial \rho^T(\psi_{0j})}{\partial \psi_j} A(\mathbf{W}) \frac{\partial \rho(\psi_{0j})}{\partial \gamma^T} \right) (I_p \otimes E(\mathbf{W}\mathbf{W}^T)^{-1}) \right],$$

$$C = E(KK^T)$$

and

$$K = \begin{pmatrix} \partial \rho^T(\psi_{0j}) / \partial \psi_j \cdot A(\mathbf{W}) \rho(\psi_{0j}) \\ (\mathbf{X}_j^* - \Gamma_{0j}\mathbf{W}) \otimes \mathbf{W} \end{pmatrix}.$$

From the proof of the above theorem in the Appendix, it can be seen that the covariance matrix DCD^T can be estimated by

$$\frac{1}{n} \frac{\partial L_n(\hat{\psi}_j)}{\partial \psi_j} \frac{\partial L_n(\hat{\psi}_j)}{\partial \psi_j^T} \xrightarrow{p} DCD^T,$$

where

$$\frac{\partial L_n(\psi_j)}{\partial \psi_j} = \sum_{i=1}^n \frac{\partial \rho_i^T(\psi_j)}{\partial \psi_j} A_i \rho_i(\psi_j).$$

Though the estimator is consistent regardless of the choice of $A(\mathbf{W})$, there exists an optimal weight $A(\mathbf{W})$ matrix theoretically for the most efficient estimator. Following Wang and Hsiao (2011), the optimal weight matrix is given by

$$A(\mathbf{W}) = E[\rho(\psi_{0j})\rho^T(\psi_{0j})|\mathbf{W}].$$

Since the optimal weight matrix involves unknown parameters, $A(\mathbf{W})$ can be calculated via a two-stage estimation procedure. First, the objective function is minimized using the identity matrix as a weight matrix. In the second stage, the estimators are obtained with the optimal weight matrix, which is calculated with the estimates from the first stage.

As noted in Abarin and Wang (2012), for some models like gamma log-linear and Poisson log-linear model, the analytical form of the expectation (4) can be obtained for some

error distribution $f_U(u)$. For example, when the random error u follows an univariate normal distribution $u \sim N(0, \phi)$, the integral in (4) has the following closed-form expression

$$E(\tilde{X}^*Y|W) = \tilde{a}\zeta,$$

where $\tilde{a} = (1, \Gamma W + \beta_x \phi)^T$ and $\zeta = \exp(\alpha + \beta_x \Gamma W + \beta_z^T Z + \beta_x^2 \phi / 2)$. With the closed-form expression, the burden of computation is eased a lot. On the other hand, in situations where the integral in (4) does not have analytical form, Monte Carlo methods (e.g., importance sampling) can be used to approximate the integral. Specifically, we follow the suggestions in Wang and Hsiao (2011) to calculate (4) as follows.

- (1) Choose a candidate distribution whose density function $h(x)$ is known;
- (2) Generate i.i.d. random sample $\{x_{is}, s = 1, 2, \dots, S, S + 1, \dots, 2S; i = 1, 2, \dots, n\}$ from density function $h(x)$;
- (3) Calculate the Monte Carlo approximation of $m(\Gamma w_i; \psi)$ as

$$m_{S1}(\Gamma w_i; \psi) = \frac{1}{S} \sum_{s=1}^S \frac{\tilde{x}_{is} g(\alpha + \beta_x^T x_{is} + \beta_z^T z_i) f_U(x_{is} - \Gamma w_i; \phi)}{h(x_{is})},$$

and

$$m_{S2}(\Gamma w_i; \psi) = \frac{1}{S} \sum_{s=S+1}^{2S} \frac{\tilde{x}_{is} g(\alpha + \beta_x^T x_{is} + \beta_z^T z_i) f_U(x_{is} - \Gamma w_i; \phi)}{h(x_{is})};$$

- (4) Apply the gradient descent method on approximated loss function

$$L_n(\psi) = \frac{1}{2} \sum_{i=1}^n \hat{\rho}_{i,S1}^T(\psi) A_i \hat{\rho}_{i,S2}(\psi) + n \sum_{j=1}^d p_{\lambda_n}(|\beta_j|),$$

where $\hat{\rho}_{i,S1}(\psi) = y_i \tilde{x}_i^* - m_{S1}(\hat{\Gamma} w_i; \psi)$ and $\hat{\rho}_{i,S2}(\psi) = y_i \tilde{x}_i^* - m_{S2}(\hat{\Gamma} w_i; \psi)$.

For some penalty functions like SCAD and MCP, \mathbf{b} and Σ are both zero when the tuning parameter λ_n is sufficiently small. Hence the resulting estimator has the oracle performance such that $\hat{\beta}_{jc} = \mathbf{0}$ and the asymptotic distribution of $\hat{\psi}_j$ is given by

$$\sqrt{n}(\hat{\psi}_j - \psi_{0j}) \xrightarrow{d} N(0, H^{-1} D C D^T H^{-1}).$$

3. Linear ME Model

For the linear regression model, the proposed regularized IV method simplifies to a regularized two-stage least squares method when the weight matrix $A = I$. Specifically, consider a linear model

$$Y = \alpha + \beta_x^T X + \beta_z^T Z + \epsilon, \tag{7}$$

where $\epsilon \sim N(0, \sigma^2)$, $E(\mathbf{U}|W, Z) = \mathbf{0}$ and $E(\mathbf{U}\mathbf{U}^T|W, Z) = \Sigma_u$. Without loss of generality, assume the intercept α is zero. The regularized instrumental variable estimator is defined as the minimizer of the following objective function

$$\frac{1}{2} \sum_{i=1}^n (y_i - \beta_x^T \hat{x}_i - \beta_z^T z_i)^2 + n \sum_{j=1}^d p_{\lambda_n}(|\beta_j|). \tag{8}$$

where $\hat{x}_i = \hat{\Gamma} w_i$. Since the naive estimator is inconsistent in estimation and selection in general, the observed covariates are replaced by its corrected version \hat{x} based on instrumental variables. Furthermore, since the objective function in (8) involves the non-independence of a random sample (y_i, \hat{x}_i, z_i) due to the involvement of $\hat{\Gamma}$, the standard results for regularized

linear regression cannot be applied directly. For the linear regression model, we have the following results.

Corollary 1. If $a_n = O(n^{-1/2})$, $b_n = o(1)$ and $E(\tilde{W}\tilde{W}^T)$ is positive definite, where $\tilde{W} = (\mathbf{W}^T, \mathbf{Z}^T)^T$, then there exists a local minimizer $\hat{\beta}$ of $Q(\beta)$ such that $\|\hat{\beta} - \beta_0\| = O_p(n^{-1/2})$.

Corollary 2. If $\lambda_n \rightarrow 0$, $\sqrt{n}\lambda_n \rightarrow \infty$ and $\liminf_{n \rightarrow \infty} \liminf_{\xi \rightarrow 0^+} p'_{\lambda_n}(\xi)/\lambda_n > 0$, then with probability approaching 1, the root n consistent estimator $\hat{\beta}$ in (8) satisfies

- (a) $\hat{\beta}_{J^c} = 0$,
- (b) $\hat{\beta}_J$ has the following asymptotic normal distribution

$$\sqrt{n}(H + \Sigma)(\hat{\beta}_J - \beta_{0J} + (H + \Sigma)^{-1}\mathbf{b}) \rightarrow_d N(0, E[(Y - \beta_{0J}^T \tilde{X}^*)^2 \tilde{\Gamma}_{0J} \tilde{W} \tilde{W}^T \tilde{\Gamma}_{0J}^T]),$$

where $H = \tilde{\Gamma}_{0J} E(\tilde{W}\tilde{W}^T) \tilde{\Gamma}_{0J}^T$, $\tilde{\Gamma} = \text{diag}(\hat{\Gamma}, I_q)$ and $\tilde{\Gamma}_J$ is the matrix consisting of rows of $\tilde{\Gamma}$ corresponding to the index set J .

4. Numerical Examples

In this section, we conduct simulations to assess the finite sample performance of the proposed instrumental variable estimator (IVE) on variable selection as well as parameter estimation. For comparison purposes, we also calculate the regularized estimator (TRE) using the true data $(y_i, \mathbf{x}_i^T, \mathbf{z}_i^T)$, and the naive estimator (NAE) using the observed sample $(y_i, \mathbf{x}_i^{*T}, \mathbf{z}_i^T)$. The proposed method is implemented with SCAD penalty function. The tuning parameter is selected by BIC that has the property of recovering the true model consistently for SCAD penalty (Wang et al. 2007). To assess the selection performance, we calculate the false-positive (FP) rate that is the average number of zero coefficients incorrectly estimated as non-zero, and the false negative (FN) rate that is the average number of non-zero coefficients incorrectly estimated as zero. We also calculate the Matthews correlation coefficient (MCC) that is a general measure of describing the confusion matrix of true/false positives/negatives and is defined as $(TP \times TN - FP \times FN) / [(TP + FP)(TP + FN)(TN + FP)(TN + FN)]^{1/2}$. The MCC ranges from -1 to 1 , where the large value indicates good prediction. Finally, we calculate the mean squared error (MSE) $\|\hat{\beta} - \beta_0\|^2$ to assess estimation accuracy.

Example 1. First, we consider a linear model $Y = \beta_x X + \beta_z^T Z + \epsilon$, where $(\beta_x, \beta_z^T) = (3, 1.5, 0, 0, 2, 0, 0, 0)$ and $(Z_1, W, Z_2, \dots, Z_7)^T$ are jointly generated from $N(0, \Sigma)$ with $\Sigma_{ij} = 0.7^{|i-j|}$. In addition, the true covariate X is generated as $X = -1.5W + U$, where ϵ and U are the standard normal. The observed surrogate X^* is generated as $X^* = X + \delta$, where δ follows a normal distribution with mean zero and variance σ_δ^2 .

Figure 1 shows the estimated coefficients, FP and FN, for various values of σ_δ^2 with a sample size $n = 200$. The results of the naive method (NAE) are on the left-hand side, while the results of the IVE method are on the right-hand side. Both the FP and FN increase with σ_δ^2 for the naive method, as seen from the bottom left graph. In contrast, the IV estimator is robust against the magnitude of σ_δ^2 . The simulation results with $\sigma_\delta^2 = 2$ are reported in Table 1. It can be seen that the naive method has both high FP and FN in the selection results. The increase in FN is due to the fact that covariate z_1 is dropped from the model incorrectly, as shown in Table 1. On the other hand, the TR and IV methods perform well in recovering the true model.

The selection results of three methods ($\sigma_\delta^2 = 1$) with sample sizes $n = 50, 100, 200$ are reported in Table 2. As the sample size increases, it can be seen that both FP and FN decrease for TR and IV methods, whereas the FP increases for the naive method. In addition, the performance of MCC and MSE is better for TR and IV methods than that of the naive method. The selection is biased for the naive method regardless of the sample size.

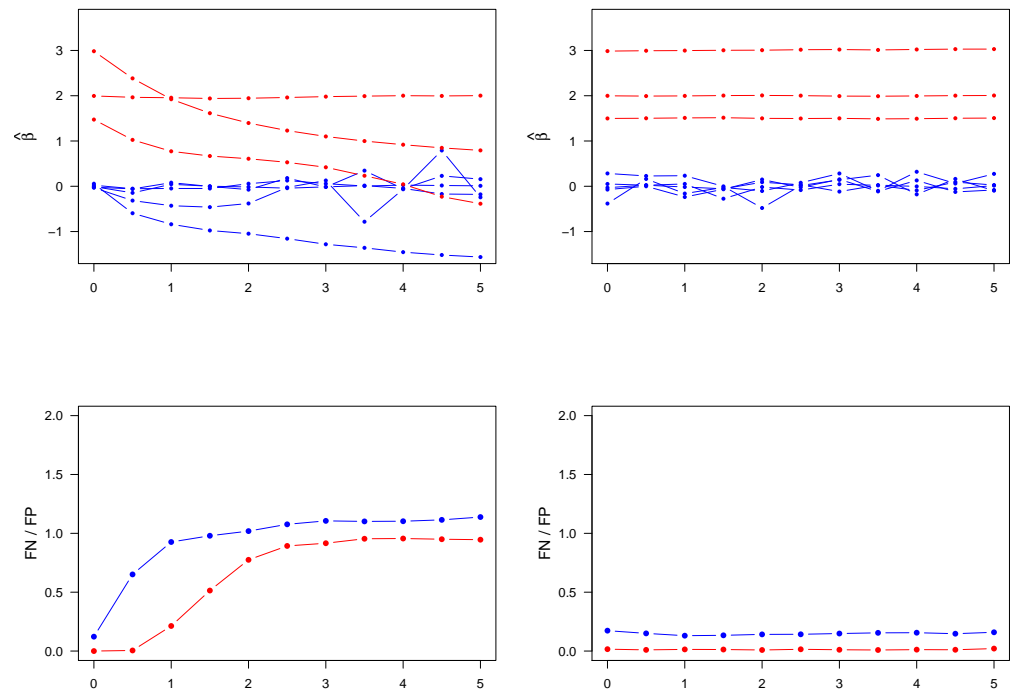


Figure 1. Estimation and selection results for Example 2 with $n = 200$. In the upper panel, the estimated non-zero parameters are marked red, while the estimated zero parameters are marked blue. In the lower panel, FN is marked red and FP is marked blue.

Table 1. Simulation results of Example 1 with $n = 200$, $\sigma_\delta^2 = 2$.

	FP	FN	MCC	MSE		z_1	x	z_2	z_3	z_4	z_5	z_6	z_7
TR	0.1	0.0	0.98	0.02	TR	100	100	99	98	100	98	98	98
IV	0.2	0.0	0.96	0.34	IV	100	100	98	98	100	97	97	95
NA	1.0	0.8	0.54	5.69	NA	23	100	17	96	100	96	95	96

Table 2. Simulation results of Example 1 with different sample sizes and $\sigma_\delta^2 = 1$.

	$n = 50$				$n = 100$				$n = 200$			
	FP	FN	MCC	MSE	FP	FN	MCC	MSE	FP	FN	MCC	MSE
TR	0.3	0	0.92	0.08	0.1	0	0.96	0.03	0.1	0	0.98	0.01
IV	0.3	0	0.92	0.52	0.2	0	0.96	0.19	0.1	0	0.98	0.07
NA	0.5	0	0.87	0.87	0.6	0	0.85	0.71	0.8	0	0.82	0.64

Example 2. In this example, we consider a logistic model where $Y|(X, \mathbf{Z})$ follows the Bernoulli distribution with mean function $g(\alpha + \beta_x X + \beta_z^T \mathbf{Z})$, where $g(\eta) = \exp(\eta) / (1 + \exp(\eta))$ and $(\alpha, \beta_x, \beta_z^T) = (1, 3, 1.5, 0, 0, 2, 0, 0, 0)$. The covariates $(Z_1, W, Z_2, \dots, Z_7)^T$ are jointly generated from $N(0, \Sigma)$ with $\Sigma_{ij} = 0.7^{|i-j|}$. Further, $X = 1.5W + U$ and the rest of the model setting is the same as in Example 1. The simulation results are shown on the left-hand side of Table 3. The results show similar patterns as in Example 1, where values of FP and FN are both low for TR and IV methods, compared with the NA method.

Example 3. In this example, we consider the Poisson model for $Y|(X, \mathbf{Z})$ with mean function $\exp(\alpha + \beta_x X + \beta_z^T \mathbf{Z})$, and the rest of model setting is the same as in Example 2. The simulation results are shown on the right-hand side of Table 3. It can be seen that in the Poisson log-linear model, the naive method performs the worst among all three methods, where FP and FN remain at a

high level. In contrast, the results from the IV method is similar to that of the TR method, where values of FP, FN and MSE are close to zero and MCC is close to one.

Table 3. Simulation results of Examples 2 and 3 with $n = 200, \sigma_\delta^2 = 5$.

	FP	FN	MCC	MSE		FP	FN	MCC	MSE
TR	0.4	0.0	0.90	1.51	TR	0	0	0.99	0.03
IV	0.6	0.0	0.87	1.62	IV	0.3	0	0.94	0.07
NA	1.1	0.1	0.75	3.20	NA	2.6	0.7	0.32	25.68

Example 4. In this example, we consider the linear model of Example 1 with relatively high dimension p . In particular, we simulate the data with sample size $n = 50$ and $p = 100$. The true parameter values are $(\beta_x, \beta_z) = (3, 1.5, 0, 0, 2, 0, 0, \dots, 0)$. The simulation results in Table 4 show that the proposed IVE method performs similarly to the small p scenarios, and, in particular, it outperforms the naive method clearly.

Table 4. Simulation results of Example 4 with $n = 50, p = 100$.

	$\sigma_\delta^2 = 1$				$\sigma_\delta^2 = 2$				$\sigma_\delta^2 = 5$			
	FP	FN	MCC	MSE	FP	FN	MCC	MSE	FP	FN	MCC	MSE
TR	0.70	0.00	0.90	0.19	0.60	0.00	0.91	0.12	0.60	0.00	0.91	0.17
IV	5.80	0.10	0.56	4.70	4.60	0.00	0.60	4.19	5.40	0.00	0.57	4.32
NA	8.10	0.20	0.46	11.40	6.30	0.20	0.51	10.90	7.70	0.20	0.47	11.45

Example 5. In this example, we consider the linear model of Example 5 with different parameter settings where $(\beta_x, \beta_z) = (0, 1.5, 2, 0, 1, 0, 0, 0)$. In this case, the corresponding coefficient is 0 for the error prone covariate x . The simulation results are presented in Table 5. It can be observed that the ME has virtually no effects on the FN. Also, as the sample size increases, the IV estimation performs nearly the same as the TR model.

Table 5. Simulation results of Example 5.

	$n = 100$				$n = 200$				$n = 500$			
	FP	FN	MCC	MSE	FP	FN	MCC	MSE	FP	FN	MCC	MSE
TR	0.18	0.00	14.46	0.02	0.06	0.00	14.82	0.12	0.04	0.00	14.88	0.00
IV	0.32	0.00	14.04	0.22	0.20	0.00	14.40	4.19	0.02	0.00	14.94	0.03
NA	0.32	0.00	14.04	0.68	0.36	0.00	13.92	10.90	0.82	0.00	12.54	0.63

Example 6. In this example, we consider the linear model. First, we consider a linear model $Y = \beta_x X + \beta_z^T Z + \epsilon$, where $(\beta_x, \beta_z^T) = (3, 1.5, 0, 0, 2, 0, 0, 0)$ and $(Z_1, X, W, Z_2, \dots, Z_7)^T$ are jointly generated from $N(0, \Sigma)$ with $\Sigma_{ij} = 0.7^{|i-j|}$. Note that the covariate X and W are jointly generated together with all other covariates in this case. The rest of the model setting remains the same as in Example 1. The results are shown in Table 6, which are similar as those in Example 1, regardless of the data-generating mechanism. The IV estimator performs better compared with NA estimator as the sample size increases.

Table 6. Simulation results of Example 6.

	$n = 100$				$n = 200$				$n = 500$			
	FP	FN	MCC	MSE	FP	FN	MCC	MSE	FP	FN	MCC	MSE
TR	0.26	0.00	14.22	0.04	0.08	0.00	14.76	0.01	0.00	0.00	15.00	0.00
IV	0.36	0.00	13.92	0.37	0.36	0.00	13.92	0.12	0.20	0.00	14.40	0.05
NA	0.70	0.00	12.90	0.89	0.84	0.00	12.48	0.86	1.06	0.00	11.82	0.92

5. Real Data Example

We applied the proposed method on a real dataset in this section. The Mobility Program Clinical Research Unit of St. Michael's Hospital conducted research studying the prognostic factors of work productivity after a limb injury. The dataset was collected through the Work Limitations Questionnaire (WLQ) from a group of injured workers attending Shoulder & Elbow Specialty clinic, which is managed by Workplace Safety & Insurance Board of Ontario, Canada. The WLQ developed by Lerner et al. (2001) and Lerner et al. (2012) offers a way of measuring how the health problems affect the job performance and the productivity loss at work. The WLQ has shown its good criterion validity and is adopted by several research institutes such as Ida et al. (2012) and Tang et al. (2011). There were 168 recruited participants who were worker compensation claimants and may or may not be working at the time of initial clinic attendance. Typically, injured workers were referred to these clinics if they have a chronic work-related upper limb injury greater than 6 months in duration without sufficient recovery.

In this paper, we are interested in exploring the prognostic factors of the WLQ index. The response variable, the work limitations questionnaire index, evaluates the proportion of time where difficulty is experienced in the following four different domains: time management, physical demands, and mental–interpersonal and output demands. This index quantifies the productivity loss at work as a result of health disorders. The predictors (prognostic factors) are supervisor support (x^*); lower quick disabilities of the arm, shoulder and hand (DASH) score (z_1); better mental health factor score (z_2); better physical health factor score (z_3); age (z_4); lower von Korff pain intensity score (z_5); lower von Korff pain intensity score (z_6); and lower shoulder pain and disability index (z_7). The instrumental variables are organization support and decision authority. The work disability is an important issue in public health, caused by whether the productivity loss at work can exceed the direct medical cost. In the literature, supervisor support is associated with the productivity and health outcomes of workers. Physical and mental disorders are also significantly related to work loss. For example, positive support from a supervisor is associated with low degree of stress and low sickness absence of the employees (Nielsen et al. 2006; Stansfeld et al. 1997). Physical–mental comorbidity is also found to have an additive increase effect in work loss (Buist-Bouwman et al. 2005). The estimation results are presented in Table 7. It can be observed that, besides the covariates lower quick DASH score and better mental health factor score that are retained in the model for the naive method, the IV method keeps the supervisor support and lower shoulder pain and disability index.

Table 7. Estimation results of WLQ data.

	IV		Naive		Full	
	coef	se	coef	se	coef	se
int	8.19	0.090	8.19	0.404	8.19	0.408
x^*	−0.09	0.090	−	−	0.13	0.445
z_1	−1.83	0.164	−1.69	0.457	−1.49	0.706
z_2	−1.56	0.119	−1.60	0.457	−1.74	0.501
z_3	−	−	−	−	−0.52	0.555
z_4	−	−	−	−	−0.33	0.415
z_5	−	−	−	−	−0.25	0.583
z_6	−	−	−	−	0.26	0.451
z_7	0.23	0.128	−	−	0.05	0.629

6. Conclusions and Discussion

Although the regularized regression methods have been widely investigated in the literature, most of the published works assume the data are precisely measured. Some researchers study the high-dimensional measurement error models, assuming the ME covariance matrix is known or can be estimated using replicate data. However, the replicate data are not always available in real applications. Instead, instrumental data are more

flexible and relatively easy to obtain. Technically, the assumption of instrumental variable is weaker than the replicate measurements. Although the IV approach is used by some authors to study general endogeneity problem in linear models, very few studies focus specifically on ME problems. Developing methodologies in this particular context allow us to obtain more insights into ME issues such as its impact on variable selection and parameter estimation in high-dimensional models.

In this paper, we extended the instrument variable method to the regularization estimation setup to correct for ME effects in both linear and generalized linear ME models. Besides the attenuation effect, the ME also affects the selection results in various settings. The proposed estimator is shown to have the oracle property, which is consistent in both variable selection and parameter estimation. The asymptotic distribution is derived for the proposed estimator in both linear and generalized linear ME models. Extensive simulation studies for linear, logistic and Poisson log-linear models are conducted examining the performance of the proposed estimator, as well as the naive estimator. Simulation results show that the proposed estimator performs well in various model settings with a finite sample size. The extension of the proposed method to nonlinear models is of interest for future research.

In this paper, we assume that the possibly mismeasured covariates are of low dimension. In future, it is important to study the case where a large number of covariates are measured with errors.

Author Contributions: Methodology, L.X. and L.W.; software programming, L.X.; validation, L.X. and L.W.; investigation, L.X. and L.W.; data curation, L.X.; writing—original draft preparation, L.X.; writing—review and editing, L.X. and L.W.; visualization, L.X.; supervision, L.W.; project administration, L.W.; funding acquisition, L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the Natural Sciences and Engineering Research Council of Canada (RGPIN 2023-04924).

Data Availability Statement: Data is unavailable due to privacy restrictions.

Acknowledgments: The authors thank the editor and three anonymous referees for their helpful comments and suggestions.

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

Appendix A

We assume the following regularity conditions.

Assumption A1. $E((\rho(\psi) - \rho(\psi_0))^T A(\mathbf{W})(\rho(\psi) - \rho(\psi_0))) = 0$ if and only if $\psi = \psi_0$, where $\rho(\psi) = \mathbf{Y}\tilde{\mathbf{X}}^* - m(\mathbf{\Gamma}\mathbf{W}; \psi)$.

Assumption A2. The parameter spaces $\Psi \subset \mathbb{R}^{d+k+1}$ and $\Gamma \subset \mathbb{R}^{p \times 1}$ are compact.

Assumption A3. $f_U(u; \phi)$ is continuously differentiable with respect to u and ϕ respectively. Furthermore, $E\|A(\mathbf{W})\|(|\mathbf{Y}|^2 + \|\mathbf{Y}\tilde{\mathbf{X}}^*\|^2) < \infty$, functions $G(\mathbf{X}, \mathbf{Z}; \theta)f_U(\mathbf{X} - \mathbf{\Gamma}\mathbf{W}; \phi)$ and $G(\mathbf{X}, \mathbf{Z}; \theta)\partial f_U(\mathbf{X} - \mathbf{\Gamma}\mathbf{W}; \phi)/\partial u^T$ and their first-order partial derivatives with respect to θ and ϕ respectively are dominated by a function $\eta(\mathbf{X}, \mathbf{Z}, \mathbf{W})$ that satisfies

$$E\|A(\mathbf{W})\| \left(\int \int \eta(x, z, \mathbf{W})(\|x\| + \|z\| + 1) dx dz \right)^2 < \infty.$$

Assumption A4. $f_U(u; \phi)$ is twice continuously differentiable with respect to ϕ in an open neighbourhood of ϕ_0 and the first two partial derivatives of $G(\mathbf{X}, \mathbf{Z}; \theta)f_U(\mathbf{X} - \mathbf{\Gamma}\mathbf{W}; \phi)$ with respect to ψ satisfies the similar dominating condition as in Assumption A3.

Assumption A5. The matrix $E \left[\frac{\partial \rho^T(\boldsymbol{\psi}_{0j})}{\partial \boldsymbol{\psi}_j} A(\mathbf{W}) \frac{\partial \rho(\boldsymbol{\psi}_{0j})}{\partial \boldsymbol{\psi}_j^T} \right]$ is non-singular.

Proof of Theorem 1. The score and hessian function of $L_n(\boldsymbol{\psi})$ are written as $G_n(\boldsymbol{\psi})$ and $H_n(\boldsymbol{\psi})$, respectively. By Assumption A4, using the Taylor expansion, the score of the objective function $Q_n(\boldsymbol{\psi})$ can be written as

$$S_n(\boldsymbol{\psi}) = G_n(\boldsymbol{\psi}_0) + \tilde{p}'_{\lambda_n}(|\boldsymbol{\beta}_0|) \circ \text{sign}(\boldsymbol{\psi}_0) + H_n(\boldsymbol{\psi}^*)(\boldsymbol{\psi} - \boldsymbol{\psi}_0) + \tilde{p}''_{\lambda_n}(|\boldsymbol{\beta}_0|)(\boldsymbol{\psi} - \boldsymbol{\psi}_0)(1 + o_p(1)),$$

where $\boldsymbol{\psi}^*$ is in between $\boldsymbol{\psi}$ and $\boldsymbol{\psi}_0$. It is sufficient to show that $S_n(\boldsymbol{\psi}) = 0$ has a solution $\boldsymbol{\psi}$ satisfying $\|\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0\| = O_p(n^{-1/2})$. To this end, we show for any $\boldsymbol{\psi}$ such that $\|\boldsymbol{\psi} - \boldsymbol{\psi}_0\| = n^{-1/2}C$ the inequality $(\boldsymbol{\psi} - \boldsymbol{\psi}_0)^T S_n(\boldsymbol{\psi}) > 0$ holds with probability approaching 1. By Assumption A5 and $b_n = 0(1)$, it follows that

$$(\boldsymbol{\psi} - \boldsymbol{\psi}_0)^T S_n(\boldsymbol{\psi}) = (\boldsymbol{\psi} - \boldsymbol{\psi}_0)^T (G_n(\boldsymbol{\psi}_0) + \tilde{p}'_{\lambda_n}(|\boldsymbol{\beta}_0|) \circ \text{sign}(\boldsymbol{\psi}_0)) + n\|\boldsymbol{\psi} - \boldsymbol{\psi}_0\|^2(1 + o_p(1)).$$

It can be seen that by Assumptions A1–A3, the first term is of order $O_p(C)$, and the second term is of order $O_p(C^2)$. Hence, for a sufficiently large C , the second term dominates the others. $(\boldsymbol{\psi} - \boldsymbol{\psi}_0)^T S_n(\boldsymbol{\psi})$ is shown to be positive with probability tending to 1, which completes the proof. \square

Proof of Theorem 2a. By Assumption A4, the Taylor expansion of $S_n(\boldsymbol{\psi})$ around $\boldsymbol{\psi}_0$ is given by

$$\begin{aligned} S_n(\boldsymbol{\psi}) &= G_n(\boldsymbol{\psi}_0) + H_n(\boldsymbol{\psi}^*)(\boldsymbol{\psi} - \boldsymbol{\psi}_0) + n\tilde{p}'_{\lambda_n}(|\boldsymbol{\beta}|) \circ \text{sign}(\boldsymbol{\psi}) \\ &= n\lambda_n \left\{ \frac{1}{n\lambda_n} G_n(\boldsymbol{\psi}_0) + \frac{1}{n\lambda_n} H_n(\boldsymbol{\psi}^*)(\boldsymbol{\psi} - \boldsymbol{\psi}_0) + \frac{1}{\lambda_n} \tilde{p}'_{\lambda_n}(|\boldsymbol{\beta}|) \circ \text{sign}(\boldsymbol{\psi}) \right\}. \end{aligned}$$

For $j \in J^c$, $\epsilon_n = Cn^{-1/2}$, by Assumptions A1–A5, it can be shown that

$$S_n(\beta_j) = n\lambda_n \left\{ O_p\left(\frac{1}{\sqrt{n}\lambda_n}\right) + \frac{p'_{\lambda_n}(|\beta_j|)}{\lambda_n} \text{sign}(\beta_j) \right\}.$$

Together with the condition $\liminf_{n \rightarrow \infty} \liminf_{\xi \rightarrow 0^+} p'_{\lambda_n}(\xi)/\lambda_n > 0$, we have $G_n(\beta_j) > 0$ if $0 < \beta_j < \epsilon_n$; $G_n(\beta_j) < 0$ if $-\epsilon_n < \beta_j < 0$. Hence, $P(\beta_j = 0) \rightarrow 1$ for $j \in J^c$. \square

Proof of Theorem 2b. By Assumption A4, the Taylor expansion of $S_n(\boldsymbol{\psi}_J)$ around $\boldsymbol{\psi}_{0j}$ is given by

$$S_n(\boldsymbol{\psi}_J) = G_n(\boldsymbol{\psi}_{0j}) + H_n(\boldsymbol{\psi}_J^*)(\boldsymbol{\psi}_J - \boldsymbol{\psi}_{0j}) + n\mathbf{b} + n\tilde{p}''_{\lambda_n}(|\boldsymbol{\beta}_J^*|)(\boldsymbol{\psi}_J - \boldsymbol{\psi}_{0j}),$$

where

$$\tilde{p}''_{\lambda_n}(|\boldsymbol{\beta}_J^*|) = \text{diag}(0, p''_{\lambda_n}(|\boldsymbol{\beta}_J^{*T}|), 0),$$

$$G_n(\boldsymbol{\psi}_{0j}) = \sum_{i=1}^n \frac{\partial \hat{\rho}_i^T(\boldsymbol{\psi}_{0j})}{\partial \boldsymbol{\psi}_j} A_i \hat{\rho}_i(\boldsymbol{\psi}_{0j})$$

and

$$H_n(\boldsymbol{\psi}_J^*) = \sum_{i=1}^n \left[\frac{\partial \hat{\rho}_i^T(\boldsymbol{\psi}_J^*)}{\partial \boldsymbol{\psi}_j} A_i \frac{\partial \hat{\rho}_i(\boldsymbol{\psi}_J^*)}{\partial \boldsymbol{\psi}_j^T} + \left(\hat{\rho}_i^T(\boldsymbol{\psi}_J^*) A_i \otimes I_{s+2} \right) \frac{\partial \text{vec}(\partial \hat{\rho}_i^T(\boldsymbol{\psi}_J^*) / \partial \boldsymbol{\psi}_j)}{\partial \boldsymbol{\psi}_j^T} \right].$$

By rearranging the terms, we obtain

$$-\frac{1}{\sqrt{n}}G_n(\boldsymbol{\psi}_{0J}) = \sqrt{n}\left(\frac{1}{n}H(\boldsymbol{\psi}_J^*) + \check{p}''_{\lambda_n}(|\boldsymbol{\beta}_J^*|)\right)(\boldsymbol{\psi}_J - \boldsymbol{\psi}_{0J}) + \sqrt{n}\check{p}'_{\lambda_n}(|\boldsymbol{\beta}_{0J}|) \circ \text{sign}(\boldsymbol{\psi}_{0J}).$$

Note that

$$\begin{aligned} \frac{1}{n}H(\boldsymbol{\psi}_J^*) &\rightarrow_p E\left[\frac{\partial\rho^T(\boldsymbol{\psi}_{0J})}{\partial\boldsymbol{\psi}_J}A(\mathbf{W})\frac{\partial\rho(\boldsymbol{\psi}_{0J})}{\partial\boldsymbol{\psi}_J^T} + (\rho^T(\boldsymbol{\psi}_{0J})A \otimes I_{s+2})\frac{\partial\text{vec}(\partial\rho^T(\boldsymbol{\psi}_{0J})/\partial\boldsymbol{\psi}_J)}{\partial\boldsymbol{\psi}_J^T}\right] \\ &= H, \end{aligned} \tag{A1}$$

since the expectation of the second term is

$$\begin{aligned} &E\left[\rho^T(\boldsymbol{\psi}_J^*)A \otimes I_{s+2}\frac{\partial\text{vec}(\partial\rho^T(\boldsymbol{\psi}_J^*)/\partial\boldsymbol{\psi}_J)}{\partial\boldsymbol{\psi}_J^T}\right] \\ &= E\left[E(\rho^T(\boldsymbol{\psi}_J^*)|\tilde{\mathbf{W}})A \otimes I_{s+2}\frac{\partial\text{vec}(\partial\rho^T(\boldsymbol{\psi}_J^*)/\partial\boldsymbol{\psi}_J)}{\partial\boldsymbol{\psi}_J^T}\right] \\ &= 0. \end{aligned}$$

Now, consider the first-order Taylor expansion of $G_n(\boldsymbol{\psi}_{0J})$ around $\boldsymbol{\gamma}_{0J}$. Again by Assumption A4,

$$G_n(\boldsymbol{\psi}_{0J}) = \sum_{i=1}^n \frac{\partial\rho_i^T(\boldsymbol{\psi}_{0J})}{\partial\boldsymbol{\psi}_J}A_i\rho_i(\boldsymbol{\psi}_{0J}) + \frac{\partial^2\tilde{L}_n(\boldsymbol{\psi}_{0J})}{\partial\boldsymbol{\psi}_J\partial\boldsymbol{\gamma}_J^T}(\hat{\boldsymbol{\gamma}}_J - \boldsymbol{\gamma}_{0J}), \tag{A2}$$

where

$$\begin{aligned} &\frac{\partial^2\tilde{L}_n(\boldsymbol{\psi}_{0J})}{\partial\boldsymbol{\psi}_J\partial\boldsymbol{\gamma}_J^T} \\ &= \sum_{i=1}^n \left[\frac{\partial\rho_i^T(\boldsymbol{\psi}_{0J}, \boldsymbol{\gamma}_J^*)}{\partial\boldsymbol{\psi}_J}A_i\frac{\partial\rho_i(\boldsymbol{\psi}_{0J}, \boldsymbol{\gamma}_J^*)}{\partial\boldsymbol{\gamma}_J^T} + (\rho_i^T(\boldsymbol{\psi}_{0J}, \boldsymbol{\gamma}_J^*)A_i \otimes I_{s+2})\frac{\partial\text{vec}(\partial\rho_i^T(\boldsymbol{\psi}_{0J}, \boldsymbol{\gamma}_J^*)/\partial\boldsymbol{\psi}_J)}{\partial\boldsymbol{\gamma}_J^T}\right]. \end{aligned}$$

Using a similar argument of Equation (A1), it can be shown that

$$\frac{1}{n}\frac{\partial^2\tilde{L}_n(\boldsymbol{\psi}_{0J})}{\partial\boldsymbol{\psi}_J\partial\boldsymbol{\gamma}_J^T} \rightarrow_p E\left(\frac{\partial\rho^T(\boldsymbol{\psi}_{0J})}{\partial\boldsymbol{\psi}_J}A\frac{\partial\rho(\boldsymbol{\psi}_{0J})}{\partial\boldsymbol{\gamma}_J^T}\right).$$

In addition, the term $\hat{\boldsymbol{\gamma}}_J - \boldsymbol{\gamma}_{0J}$ in Equation (A2) can be written as

$$\hat{\boldsymbol{\gamma}}_J - \boldsymbol{\gamma}_{0J} = (\sum I_p \otimes \mathbf{W}_i\mathbf{W}_i^T)^{-1}(\sum(\mathbf{X}_{ji}^* - \boldsymbol{\Gamma}_{0J}\mathbf{W}_i) \otimes \mathbf{W}_i).$$

Hence, Equation (A2) can be written as

$$G_n(\boldsymbol{\psi}_{0J}) = D_n \sum_{i=1}^n K_i,$$

where

$$\begin{aligned}
 D_n &= \left(I_{s+2}, \frac{\partial^2 \tilde{L}_n(\boldsymbol{\psi}_{0J})}{\partial \boldsymbol{\psi}_J \partial \boldsymbol{\gamma}_J^T} (I_p \otimes (\sum_{i=1}^n \mathbf{W}_i \mathbf{W}_i^T)^{-1}) \right) \\
 &\rightarrow \left(I_{s+2}, E \left(\frac{\partial \rho^T(\boldsymbol{\psi}_{0J})}{\partial \boldsymbol{\psi}_J} A \frac{\partial \rho(\boldsymbol{\psi}_{0J})}{\partial \boldsymbol{\gamma}_J^T} \right) (I_p \otimes E(\mathbf{W} \mathbf{W}^T)^{-1}) \right) \\
 &= D, \\
 K_i &= \begin{pmatrix} \partial \rho_i^T(\boldsymbol{\psi}_{0J}) / \partial \boldsymbol{\psi}_J \cdot A_i \rho_i(\boldsymbol{\psi}_{0J}) \\ (\mathbf{X}_{ji}^* - \boldsymbol{\Gamma}_{0J} \mathbf{W}_i) \otimes \mathbf{W}_i \end{pmatrix}.
 \end{aligned}$$

Then, together with Assumption A5,

$$\sqrt{n}(H + \Sigma)(\hat{\boldsymbol{\psi}}_J - \boldsymbol{\psi}_{0J}) + \sqrt{n}\mathbf{b} \rightarrow_d N(0, DCD^T).$$

□

References

- Abarin, Taraneh, and Liqun Wang. 2012. Instrumental variable approach to covariate measurement error in generalized linear models. *Annals of the Institute of Statistical Mathematics* 64: 475–93. [\[CrossRef\]](#)
- Buist-Bouwman, M. A., Ron de Graaf, W. A. M. Vollebergh, and Johan Ormel. 2005. Comorbidity of physical and mental disorders and the effect on work-loss days. *Acta Psychiatrica Scandinavica* 111: 436–43. [\[CrossRef\]](#) [\[PubMed\]](#)
- Candes, Emmanuel, and Terence Tao. 2007. The dantzig selector: Statistical estimation when p is much larger than n. *The Annals of Statistics* 35: 2313–51.
- Fan, Jianqing, and Jinchi Lv. 2010. A selective overview of variable selection in high dimensional feature space. *Statistica Sinica* 20: 101.
- Fan, Jianqing, and Runze Li. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96: 1348–60. [\[CrossRef\]](#)
- Fan, Jianqing, and Yuan Liao. 2014. Endogeneity in high dimensions. *The Annals of Statistics* 42: 872. [\[CrossRef\]](#)
- Frank, Ldiko E., and Jerome H. Friedman. 1993. A statistical view of some chemometrics regression tools. *Technometrics* 35: 109–35. [\[CrossRef\]](#)
- Huang, Xianzheng, and Hongmei Zhang. 2013. Variable selection in linear measurement error models via penalized score functions. *Journal of Statistical Planning and Inference* 143: 2101–11. [\[CrossRef\]](#)
- Ida, Hiromasa, Kazumi Nakagawa, Masako Miura, Kyoko Ishikawa, and Naonori Yakura. 2012. Development of the work limitations questionnaire japanese version (wlq-j): Fundamental examination of the reliability and validity of the wlq-j. *Sangyo eiseigaku zasshi = Journal of Occupational Health* 54: 101–7. [\[CrossRef\]](#)
- Lerner, Debra, Benjamin C. Amick, III, William H. Rogers, Susan Malspeis, Kathleen Bungay, and Diane Cynn. 2001. The work limitations questionnaire. *Medical Care* 39: 72–85. [\[CrossRef\]](#)
- Lerner, Debra, David Adler, Richard C. Hermann, Hong Chang, Evette J. Ludman, Annabel Greenhill, Katherine Perch, William C. McPeck, and William H. Rogers. 2012. Impact of a work-focused intervention on the productivity and symptoms of employees with depression. *Journal of Occupational and Environmental Medicine* 54: 128. [\[CrossRef\]](#) [\[PubMed\]](#)
- Liang, Hua, and Runze Li. 2009. Variable selection for partially linear models with measurement errors. *Journal of the American Statistical Association* 104: 234–48. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lin, Wei, Rui Feng, and Hongzhe Li. 2015. Regularization methods for high-dimensional instrumental variables regression with an application to genetical genomics. *Journal of the American Statistical Association* 110: 270–88. [\[CrossRef\]](#)
- Ma, Yanyuan, and Runze Li. 2010. Variable selection in measurement error models. *Bernoulli: Official Journal of the Bernoulli Society for Mathematical Statistics and Probability* 16: 274. [\[CrossRef\]](#) [\[PubMed\]](#)
- Negahban, Sahand N., Pradeep K. Ravikumar, Martin J. Wainwright, and Bin Yu. 2012. A unified framework for high-dimensional analysis of m -estimators with decomposable regularizers. *Statistical Science* 27: 538–57. [\[CrossRef\]](#)
- Nielsen, Martin L., Reiner Rugulies, Karl B. Christensen, Lars Smith-Hansen, and Tage S. Kristensen. 2006. Psychosocial work environment predictors of short and long spells of registered sickness absence during a 2-year follow up. *Journal of Occupational and Environmental Medicine* 48: 591–98. [\[CrossRef\]](#) [\[PubMed\]](#)
- Stansfeld S. A., G. S. Rael, J. Head, M. Shipley, and M. Marmot. 1997. Social support and psychiatric sickness absence: A prospective study of british civil servants. *Psychological Medicine* 27: 35–48. [\[CrossRef\]](#) [\[PubMed\]](#)

- Tang, Kenneth, Dorcas E. Beaton, Annelies Boonen, Monique A. Gignac, and Claire Bombardier. 2011. Measures of work disability and productivity: Rheumatoid arthritis specific work productivity survey (wps-ra), workplace activity limitations scale (wals), work instability scale for rheumatoid arthritis (ra-wis), work limitations questionnaire (wlq), and work productivity and activity impairment questionnaire (wpai). *Arthritis Care & Research* 63: S337.
- Tibshirani, Robert. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 58: 267–88. [[CrossRef](#)]
- Wang, Hansheng, Runze Li, and Chih-Ling Tsai. 2007. Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika* 94: 553–68. [[CrossRef](#)]
- Wang, Liqun. 2021. Identifiability in measurement error models. In *Handbook of Measurement Error Models*. Boca Raton: Chapman and Hall/CRC, pp. 55–70.
- Wang, Liqun, and Cheng Hsiao. 2011. Method of moments estimation and identifiability of semiparametric nonlinear errors-in-variables models. *Journal of Econometrics* 165: 30–44. [[CrossRef](#)]
- Zhang, Cun-Hui. 2010. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics* 38: 894–942. [[CrossRef](#)] [[PubMed](#)]
- Zhang, Xinyu, Haiying Wang, Yanyuan Ma, and Raymond J. Carroll. 2017. Linear model selection when covariates contain errors. *Journal of the American Statistical Association* 112: 1553–61. [[CrossRef](#)] [[PubMed](#)]
- Zhong, Wei, Wei Zhou, Qingliang Fan, and Yang Gao. 2020. Dummy endogenous treatment effect estimation using high-dimensional instrumental variables. *Canadian Journal of Statistics* 50: 795–819. [[CrossRef](#)]
- Zou, Hui. 2006. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* 101: 1418–29. [[CrossRef](#)]
- Zou, Hui, and Trevor Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 67: 301–20. [[CrossRef](#)]

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