

Instrumental Variables Regression with Measurement Errors and Multicollinearity in Instruments

Ayyub Sheikhi ¹, Mohsen Rezapour ², and Hamid Hoseinkhani ³

¹ Department of Statistics, Faculty of Mathematics and Computer, Shahid Bahonar University of Kerman, Kerman, Iran.

² Department of Biostatistics and Data Science, School of Public Health, University of Texas Health Science Center at Houston (UTHealth), Houston, Texas, USA.

³ Department of Statistics, University of Birjand, Birjand, Iran.

Received: 17/08/2019, Revision received: 07/03/2021, Published online: 17/04/2021

Abstract. In this paper we obtain a consistent estimator when there exist some measurement errors and multicollinearity in the instrumental variables in a two stage least square estimation of parameters. We investigate the asymptotic distribution of the proposed estimator and discuss its properties using some theoretical proofs and a simulation study. A real numerical application is also provided for more illustration.

Keywords. Endogeneity, Shrinkage Estimators, Measurement Error Models, Instrumental Variable, Two Stage Least Square.

MSC: 62J05, 62J07.

1 Introduction

The ordinary least-squares estimator (OLS) provides a convenient solution of the linear regression analysis. One of the most important assumptions in the linear regression analysis is that the value of the error terms is independent of (some) predictor variables. When this assumption is violated the ordinary least square method arises a biased estimator and referred as the endogeneity problem, see e.g., Ebbes (2004) and Bowden and Turkington (1984). There are several methods to overcome endogeneity in which instrumental variable (IV) method has long been one of the most popular procedures among epidemiologists, econometricists, social scientists, etc. See for example, Wooldridge (2016), Burgess et al. (2015) and Lawlor (2008). Instrumental variables are independent of the error term and help to obtain unbiased and consistent estimator of the regression coefficients. When the instrumental variables implemented in the model, a statistical technique, named Two Stage Least Squares (2SLS) method may be used to estimate parameters. The 2SLS regression analysis which introduced by Bassmann (1957) assumes that there is a secondary predictor that is correlated to the predictor but not with the error term. This secondary predictor is named instrumental variable. Also, The 2SLS estimator is a generalization of the instrumental variable estimator and it reduces to the IV estimator if the equation is exactly identified.

A valid IV must have at least two conditions: First, the *relevance* assumption, which means that the IV has a causal effect on endogenous variable and second, the *exclusion restriction*; which indicates that the IV affects the outcome only through the endogenous variables. In some situations it is quite difficult to find valid instrumental and in fact, may only exist some ill-conditioned IVs. Some researchers caution that the challenge of identifying a credible instrument is not trivial (see for example, Martens et al. (2006) and Lousdal (2018)).

Another most important assumptions in regression analysis is that all the observations are correctly observed. However, in many applications this assumption may not hold and the data are contaminated by measurement errors. In these models, estimation based on the standard assumption leads to inconsistent estimates, meaning that the parameter estimates do not tend to the true values even in very large samples. In fact, measurement error is known to cause biased parameter estimates (Carroll et al. (2006)) and lack of power is a direct consequence of this misestimates. Error ignoring measurement error often leads to incorrect inferences about parameters (see, for example, Cock and Campbel (1979)). Measurement error estimation has extensively been employed in theoretical and applied statistics; Rasekh and Fieller (2001) discussed the Influence functions in functional measurement error models with replicated data,

while Devanarayana and stefanski (2002) and Nawarathna and Choudhary (2015) considered a heteroscedastic measurement error models with replicate measurements. Singh et al. (2014) considered a measurement error model under exact linear restrictions. Wimmer and Witkovsky (2007) considered a measurement error model with replicated data in comparative calibration problem, see also, Bolfarine and Lachos (2016) which considered Berkson measurement errors. Dalen et. al. (2009) used replication to correct misclassification of a categorized exposure in binary regression. Chan and Mak (1979) and Isogawa (1985) studied the structural form of the replicated measurement error model under the condition of normally distributed measurement errors, while Cao et al. (2015) dealt with multivariate measurement error models for replicated data under heavy-tailed distributions. Recently, Shalabh et al. (2016) immaculated the inconsistent estimator of parameter in ultrastructural measurement error model with replicated data, see also Ullah et al. (2001) and Shalabh et al. (2009) for non normal measurement errors in such a case.

Besides, another important assumption in the classical regression analysis is that predictor variables are linearly independent. When this assumption is violated, the predictor variables are nearly dependent which refers as multicollinearity problem and yields poor estimators of interest parameters. In order to resolve this problem, several approaches have been considered, among them, the ridge regression was introduced by Horel and Kennard (1970) and considers a shrinkage method to overcome the problem of multi- collinearity in the estimation of regression parameters. For more information about this type of estimators and their properties and extensions, one may refer to Saleh et al. (2019) which is a most significant book in this field, along with Arashi et al. (2014) and Fallah et al. (2017).

Many authors described the connection between endogeneity and multicollinearity in the literature. Recently, Philips and Evans (2016) discussed approximating and reducing bias in 2SLS estimation while Hansen and Kuzbur (2014) proposed a jackknife instrumental variables estimator with regularization at each jackknife iteration that helps alleviate the bias. Carrasco (2012) proposed an original approach based on regularized 2SLS. Carrasco and Tchuente (2015) regularized versions of the limited information maximum likelihood (LIML). Also, Rasekh (2006), Ghapani and Babadi (2016) and Saleh, E., Shalabh (2014) presented the connection between measurement error models and ridge estimators. In this paper we employ the ridge regression method to combat multicollinearity in the estimation of IV models in the presence of some measurement errors.

The organization of this paper is as follows. In Section 2, we obtain a ridge estimator

in a two stage least square estimation where there are some errors in measurement of the instrumental variables. Some large sample properties and theoretical comparison will be presented in Section 3 and finally, in Section 4, we present a simulation study and a real numerical application of our results.

2 Model Specification and Estimation

Consider relationship between the $n \times 1$ vector of outcome variable \mathbf{y} and the explanatory matrix \mathbf{X} and the relation between the exploratory variables and their corresponding instruments \mathbf{Z} are respectively as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} , \quad (2.1a)$$

$$\mathbf{X} = \mathbf{Z}\boldsymbol{\eta} + \mathbf{U} , \quad (2.1b)$$

where \mathbf{X} is an $n \times p$ matrix of the endogenous regressors, potentially correlated with $\boldsymbol{\epsilon}$, whose i th row is \mathbf{x}_i^T and $\boldsymbol{\beta}$ is the unknown regression coefficients. Also, assume that \mathbf{Z} is an $n \times p$ matrix of the instrumental variables or the excluded exogenous variables which is uncorrelated with $\boldsymbol{\epsilon}$ and $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_p)$, where η_i , $i = 1, 2, \dots, p$, is the vector of coefficients associated with these instruments. The $n \times p$ matrix of instrumental model error $\mathbf{U} = (u_1, u_2, \dots, u_n)^T$ is assumed to have normally distributed rows, with mean zero and covariance matrix $\sigma_u^2 \mathbf{I}_p$ and hence $\mathbf{U} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I}_{np})$, where \mathbf{I} is an identity matrix. Moreover, assume that $E(\mathbf{Z}\mathbf{U}) = E(\mathbf{Z}\boldsymbol{\epsilon}) = \mathbf{0}$, and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. In addition, suppose that the observations of the latent (unobservable) IVs \mathbf{Z} are contaminated with measurement errors, as

$$\mathbf{V} = \mathbf{Z} + \boldsymbol{\Delta} , \quad (2.2)$$

where \mathbf{V} is the matrix of observed variables and $\boldsymbol{\Delta} = (\delta_1, \delta_2, \dots, \delta_n)^T$ is associated with \mathbf{Z} and normally distributed rows, with mean zero and covariance $\boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda}$ is a diagonal $p \times p$ matrix of known values with non-negative diagonal elements. Thus, $\boldsymbol{\Delta} \sim N(\mathbf{0}, \mathbf{I}_n \otimes \boldsymbol{\Lambda})$ where \otimes is the Kronecker product.

With these assumptions, each column of the matrix \mathbf{X} , i.e., \mathbf{x}_i , $i = 1, 2, \dots, p$, in (2.1b) follow normal distribution with mean $\boldsymbol{\mu}_i = \mathbf{Z}\boldsymbol{\eta}_i$ and variance-covariance matrix $\boldsymbol{\Sigma} = \sigma \mathbf{I}$, so, the log-likelihood functions can be written as,

$$\begin{aligned} \ell(\boldsymbol{\eta}_i, \mathbf{Z}, \mathbf{x}_i) &= -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln(\boldsymbol{\Sigma}) - \frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_i) \\ &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{x}_i - \mathbf{Z}\boldsymbol{\eta}_i)^T (\mathbf{x}_i - \mathbf{Z}\boldsymbol{\eta}_i), \quad i = 1, 2, \dots, p. \end{aligned}$$

Nakamura (1990) has proposed an approach in measurement error models to find a corrected score function whose expectation with respect to the measurement error distribution coincides with the usual score function based on the unknown true independent variables. See, e.g, Zare and Rasekh (2011) and Fung (2003) for more details. According to Fung et al. (2003), the appropriate corrected log-likelihood function is defined as

$$\ell^*(\eta_i, \mathbf{V}, \mathbf{x}_i) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{x}_i - \mathbf{V}\eta_i)^T(\mathbf{x}_i - \mathbf{V}\eta_i) - n\eta_i^T \Lambda \eta_i, i = 1, 2, \dots, p,$$

where, $E^*(\ell^*(\eta_i, \mathbf{V}, \mathbf{x}_i) = \ell(\eta_i, \mathbf{Z}, \mathbf{x}_i), i = 1, 2, \dots, p$, and E^* denotes the conditional mean with respect to \mathbf{Z} given \mathbf{x} .

By differentiating from $\ell^*(\eta_i, \mathbf{V}, \mathbf{x}_i)$ with respect to η_i , we obtain the corrected log-likelihood estimator of η_i as

$$\hat{\eta}_{iME} = (\mathbf{V}^T \mathbf{V} - n\Lambda)^{-1} \mathbf{V}^T \mathbf{x}_i, \quad i = 1, 2, \dots, p, \tag{2.3}$$

and so,

$$\hat{\eta}_{ME} = (\mathbf{V}^T \mathbf{V} - n\Lambda)^{-1} \mathbf{V}^T \mathbf{X}. \tag{2.4}$$

Then, We use the predicted matrix $\hat{\mathbf{X}}$ as an estimation for \mathbf{X} and apply the IV estimation using $\hat{\mathbf{X}}_{ME} = \mathbf{Z}\hat{\eta}_{ME}$ (see for instance, Wooldridge (2016)). Then, we regress \mathbf{y} on the predicted values from the first stage, so, the second stage regression coefficient using OLS method will be obtained as

$$\hat{\beta}_{2SLS-ME} = (\hat{\mathbf{X}}_{ME}^T \hat{\mathbf{X}}_{ME})^{-1} \hat{\mathbf{X}}_{ME}^T \mathbf{y}, \tag{2.5}$$

where the subscript 2SLS – ME stands for the two stage least squares estimator with measurement error in instrumental variables. Now, consider in model (2.1) there are some multicollinearity within columns of the matrix \mathbf{Z} . In what follows, we use the ridge penalty to combat this multicollinearity. Define

$$\mathbf{x}_i^* = \begin{pmatrix} \mathbf{x}_i \\ \mathbf{0} \end{pmatrix}, \mathbf{Z}^* = \begin{pmatrix} \mathbf{Z} \\ \sqrt{\lambda} \mathbf{I}_k \end{pmatrix} \text{ and } \mathbf{U}_i^* = \begin{pmatrix} \mathbf{U}_i - \Lambda \eta_i \\ \boldsymbol{\varphi} \end{pmatrix},$$

in which λ is the tuning parameter, and $\boldsymbol{\varphi}$ is an appropriate random vector with $E(\boldsymbol{\varphi}) = \mathbf{0}$ and $Var(\boldsymbol{\varphi}) = \sigma^2 \mathbf{I}_k$. Then the appropriate corrected log-likelihood in the presence of measurement errors and the multicollinearity, which satisfies

$$E^*(\ell^*(\eta_i, \mathbf{V}, \mathbf{x}_i) = \ell(\eta_i, \mathbf{Z}, \mathbf{x}_i), i = 1, 2, \dots, p,$$

is defined as

$$\ell^*(\boldsymbol{\eta}_i, \mathbf{V}, \mathbf{x}_i) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \{(\mathbf{x}_i - \mathbf{V}\boldsymbol{\eta}_i)^T (\mathbf{x}_i - \mathbf{V}\boldsymbol{\eta}_i) - n\boldsymbol{\eta}_i^T \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \lambda\boldsymbol{\eta}_i^T \boldsymbol{\eta}_i\},$$

for $i = 1, 2, \dots, p$ and $\lambda > 0$. By differentiating from $\ell^*(\boldsymbol{\eta}_i, \mathbf{V}, \mathbf{x}_i)$ with respect to $\boldsymbol{\eta}_i$ we obtain the corrected log-likelihood estimator of $\boldsymbol{\eta}_i$, as

$$\hat{\boldsymbol{\eta}}_i = (\mathbf{V}^T \mathbf{V} - n\boldsymbol{\Lambda} + \lambda \mathbf{I})^{-1} \mathbf{V}^T \mathbf{x}_i, \quad i = 1, 2, \dots, p, \quad (2.6)$$

and hence the ridge estimator when the instrumental variables have multicollinearity and measurement error is as

$$\hat{\boldsymbol{\eta}}_{RME} = (\mathbf{V}^T \mathbf{V} - n\boldsymbol{\Lambda} + \lambda \mathbf{I})^{-1} \mathbf{V}^T \mathbf{X}. \quad (2.7)$$

Now, similar to equation (2.3) by plug-in the predicted matrix $\hat{\mathbf{X}}_{RME} = \mathbf{Z}\hat{\boldsymbol{\eta}}_{RME}$ to equation (2.1a), the regularized two stage estimator of $\boldsymbol{\beta}$, in the presence of measurement errors, can be obtained as

$$\hat{\boldsymbol{\beta}}_{2SLS-RME} = (\hat{\mathbf{X}}_{RME}^T \hat{\mathbf{X}}_{RME})^{-1} \hat{\mathbf{X}}_{RME}^T \mathbf{y}, \quad (2.8)$$

in which the subscript *2SLS – RME* indicates the two stage least squares estimator when the instrumental variables have multicollinearity and measurement error.

It should be noted that the estimator in (2.8) will reduce to (2.5) if there is no multicollinearity in instrumental variables, while if all instrumental variables measured without error, the estimator (2.5) will simplified to the traditional two stage least square estimator as

$$\hat{\boldsymbol{\beta}}_{2SLS} = (\hat{\mathbf{X}}^T \hat{\mathbf{X}})^{-1} \hat{\mathbf{X}}^T \mathbf{y}, \quad (2.9)$$

where $\hat{\mathbf{X}} = \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{X}$.

3 Asymptotic Properties

In this section, we investigate some asymptotic properties of the proposed estimator in Section 2. The following theorem states the asymptotic properties of our estimator. First, we accept some assumptions.

Assumption 1:

1. \mathbf{V} and \mathbf{Z} are stochastic and full rank matrices.
2. The $p \times p$ matrix $\mathbf{M} = \text{plim} \frac{\Phi^T \Phi}{n}$ exists and is nonsingular.
3. The limit of $n^{-1}(\mathbf{Z}^T \mathbf{Z} + \lambda \mathbf{I})$ exists.

Theorem 3.1. Under the assumption 1, $\hat{\eta}_{\text{RME}}$ has asymptotically normal distribution with mean $\mathbf{Z}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z} + \lambda \mathbf{I})^{-1} \boldsymbol{\eta}$ and variance-covariance matrix $(\mathbf{Z}^T \mathbf{Z} + \lambda \mathbf{I})^{-1} (\boldsymbol{\Lambda} (n\sigma^2 + \boldsymbol{\eta}^T \mathbf{Z}^T \mathbf{Z} \boldsymbol{\eta}) + \sigma^2 \mathbf{Z}^T \mathbf{Z}) (\mathbf{Z}^T \mathbf{Z} + \lambda \mathbf{I})^{-1}$.

Proof. We have $E^*(\Phi^T \Phi) = \mathbf{Z}^T \mathbf{Z} + n\boldsymbol{\Lambda}$, $j = 1, 2, \dots, m$, and by Fung et. al. (2003) we can write

$$\Phi^T \Phi = \mathbf{Z}^T \mathbf{Z} + n\boldsymbol{\Lambda} + O_p(n^{\frac{1}{2}}).$$

Thus, we obtain

$$n^{-1}(\Phi^T \Phi - n\boldsymbol{\Lambda} + \lambda \mathbf{I}) = n^{-1}(\mathbf{Z}^T \mathbf{Z} + \lambda \mathbf{I}) + O_p(n^{-\frac{1}{2}}),$$

from (2.5), we have

$$\begin{aligned} \sqrt{n}\hat{\eta}_r &= \left(n^{-1}(\Phi^T \Phi - n\boldsymbol{\Lambda} + \lambda \mathbf{I}) \right)^{-1} n^{-\frac{1}{2}} \Phi^T \mathbf{X} \\ &= \left(n^{-1}(\mathbf{Z}^T \mathbf{Z} + \lambda \mathbf{I} + O_p(n^{-\frac{1}{2}})) \right)^{-1} n^{-\frac{1}{2}} \Phi^T \mathbf{X} \\ &= \left(n^{-1}(\mathbf{Z}^T \mathbf{Z} + \lambda \mathbf{I} + O_p(n^{-\frac{1}{2}})) \right)^{-1} n^{-\frac{1}{2}} \Phi^T \mathbf{X} \\ &= \left(\mathbf{I} + O_p(n^{-\frac{1}{2}}) \right)^{-1} \left(n^{-1}(\mathbf{Z}^T \mathbf{Z} + \lambda \mathbf{I}) \right)^{-1} n^{-\frac{1}{2}} \Phi^T \mathbf{X} \\ &= \left(\mathbf{I} + O_p(n^{-\frac{1}{2}}) \right) \left(n^{-1}(\mathbf{Z}^T \mathbf{Z} + \lambda \mathbf{I}) \right)^{-1} n^{-\frac{1}{2}} \Phi^T \mathbf{X}, \end{aligned}$$

and, since by assumption 1.3, the limit of $\mathbf{C} = n^{-1}(\mathbf{Z}^T \mathbf{Z} + \lambda \mathbf{I})$ exist, hence,

$$\sqrt{n}\hat{\eta}_r = \mathbf{C}^{-1} \boldsymbol{\xi} + O_p(n^{-\frac{1}{2}}),$$

where $\boldsymbol{\xi} = n^{-\frac{1}{2}} \Phi^T \mathbf{X}$ is asymptotically normal with mean $n^{-\frac{1}{2}} \mathbf{Z}^T \mathbf{Z} \boldsymbol{\eta}$ (See for example, Fung et al. (2003) and Ghapani and Babadi (2016)). So, we readily conclude that

$$\sqrt{n}(\hat{\eta}_r - \mathbf{Z}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z} + \lambda \mathbf{I})^{-1} \boldsymbol{\eta}) = \mathbf{C}^{-1} (\boldsymbol{\xi} - E(\boldsymbol{\xi})) + O_p(n^{-\frac{1}{2}}).$$

Consequently, $\sqrt{n}(\hat{\eta}_r - \mathbf{Z}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z} + \lambda \mathbf{I})^{-1} \boldsymbol{\eta})$ has asymptotically normal distribution with mean zero.

Also, we have $AVar(\sqrt{n}\hat{\eta}_r) = \mathbf{C}^{-1}Var(\boldsymbol{\xi})\mathbf{C}^{-1}$. The variance-covariance matrix of $\boldsymbol{\xi}$ can be obtained as

$$\begin{aligned} Var(\boldsymbol{\xi}) &= E_{\mathbf{X}}(Var(\boldsymbol{\xi}|\mathbf{X})) + Var_{\mathbf{X}}(E(\boldsymbol{\xi}|\mathbf{X})) = n^{-1}E_{\mathbf{X}}(\mathbf{X}^T \mathbf{y} \boldsymbol{\Lambda}) + n^{-1}Var_{\mathbf{X}}(\mathbf{Z}^T \mathbf{X}) \\ &= n^{-1}E_{\mathbf{X}}(\mathbf{X}^T \mathbf{X} \boldsymbol{\Lambda}) + n^{-1}Var_{\mathbf{X}}(\mathbf{Z}^T \mathbf{X}) = n^{-1}\boldsymbol{\Lambda}(n\sigma^2 + \boldsymbol{\eta}^T \mathbf{Z}^T \mathbf{Z} \boldsymbol{\eta}) + n^{-1}\sigma^2 \mathbf{Z}^T \mathbf{Z} \\ &= n^{-1}[\boldsymbol{\Lambda}(n\sigma^2 + \boldsymbol{\eta}^T \mathbf{Z}^T \mathbf{Z} \boldsymbol{\eta}) + \sigma^2 \mathbf{Z}^T \mathbf{Z}]. \end{aligned}$$

Thus,

$$AVar(\hat{\eta}_r) = (\mathbf{Z}^T \mathbf{Z} + \lambda \mathbf{I})^{-1} \left(\boldsymbol{\Lambda}(n\sigma^2 + \boldsymbol{\eta}^T \mathbf{Z}^T \mathbf{Z} \boldsymbol{\eta}) + \sigma^2 \mathbf{Z}^T \mathbf{Z} \right) (\mathbf{Z}^T \mathbf{Z} + \lambda \mathbf{I})^{-1},$$

which completes the proof. \square

In addition to assumptions (1), we need the following assumptions to prove the next theorems.

Assumption 2

1. \mathbf{X} is an stochastic and full rank matrix.
2. The $p \times p$ matrix $\hat{\mathbf{X}}_{RME}^T \hat{\mathbf{X}}_{RME}$ of (2.5) exists and is nonsingular.
3. $E(\mathbf{Z}\boldsymbol{\epsilon}) = \mathbf{0}$.
4. $\frac{\sum_{i=1}^n Z_i \epsilon_i}{n} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Phi})$, for some $p \times p$ symmetric, finite and nonsingular matrix $\boldsymbol{\Phi}$.
5. The matrices $\mathbf{Q}_{xv} = E(\mathbf{X}^T \boldsymbol{\Phi})$ and $\mathbf{Q}_{zz} = \mathbf{Z}^T \mathbf{Z}$ are symmetric, finite and of full rank.
6. The matrix $\mathbf{Q}_g = \boldsymbol{\Phi}^T \boldsymbol{\Phi} - n\boldsymbol{\Lambda} + \lambda \mathbf{I}$ is finite and singular.

Theorem 3.2. Under assumptions 2,

$$\hat{\boldsymbol{\beta}}_{2SLS-RME} \xrightarrow{p} (\mathbf{Q}_{xv} \mathbf{Q}_g^{-1} \mathbf{Q}_{zz} \mathbf{Q}_g^{-1} \mathbf{Q}_{vx})^{-1} \mathbf{Q}_{xv} \mathbf{Q}_g^{-1} \boldsymbol{\beta}.$$

Proof. From (2.7), we readily have

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{2SLS-RME} &= \left(\hat{\mathbf{X}}_{RME}^T \hat{\mathbf{X}}_{RME} \right)^{-1} \hat{\mathbf{X}}_{RME}^T \mathbf{y} \\ &= \left(\mathbf{X}^T \boldsymbol{\Phi} \mathbf{Q}_g^{-1} \mathbf{Z}^T \mathbf{Z} \mathbf{Q}_g^{-1} \boldsymbol{\Phi}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \boldsymbol{\Phi} \mathbf{Q}_g^{-1} \mathbf{Z}^T \mathbf{y} \\ &= \left[\left(\frac{\mathbf{X}^T \boldsymbol{\Phi}}{n} \right) n \mathbf{Q}_g^{-1} \left(\frac{\mathbf{Z}^T \mathbf{Z}}{n} \right) n \mathbf{Q}_g^{-1} \left(\frac{\boldsymbol{\Phi}^T \mathbf{X}}{n} \right) \right]^{-1} \left(\frac{\mathbf{X}^T \boldsymbol{\Phi}}{n} \right) n \mathbf{Q}_g^{-1} \left(\frac{\mathbf{Z}^T \mathbf{y}}{n} \right). \end{aligned}$$

Using (2.1a) and defining $\mathbf{W} = \mathbf{M}^{-1}\mathbf{H}$, $\mathbf{M} = (\mathbf{X}^T \Phi \mathbf{Q}_g^{-1} \mathbf{Z}^T \mathbf{Z} \mathbf{Q}_g^{-1} \Phi^T \mathbf{X})$ and $\mathbf{H} = \mathbf{X}^T \Phi \mathbf{Q}_g^{-1} \mathbf{Z}^T \mathbf{X}$, we have

$$\begin{aligned} \hat{\beta}_{2SLS-RME} - \mathbf{W}\beta &= \mathbf{W}\epsilon \\ &= \left[\left(\frac{\mathbf{X}^T \Phi}{n} \right) n \mathbf{Q}_g^{-1} \left(\frac{\mathbf{Z}^T \mathbf{Z}}{n} \right) n \mathbf{Q}_g^{-1} \left(\frac{\Phi^T \mathbf{X}}{n} \right) \right]^{-1} \left(\frac{\mathbf{X}^T \Phi}{n} \right) n \mathbf{Q}_g^{-1} \left(\frac{\mathbf{Z}^T \epsilon}{n} \right) \\ &= \left[\hat{\mathbf{Q}}_{xv} \hat{\mathbf{Q}}_g^{-1} \hat{\mathbf{Q}}_{zz} \hat{\mathbf{Q}}_g^{-1} \hat{\mathbf{Q}}_{xv}^T \right]^{-1} \hat{\mathbf{Q}}_{xv} \hat{\mathbf{Q}}_g^{-1} \frac{\mathbf{Z}^T \epsilon}{n}, \end{aligned}$$

where $\hat{\mathbf{Q}}_{xv}^T = \hat{\mathbf{Q}}_{xv} = \frac{\mathbf{X}^T \Phi}{n} = n^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{v}_i^T$, $\hat{\mathbf{Q}}_{zz} = \frac{\mathbf{Z}^T \mathbf{Z}}{n} = n^{-1} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T$ and $\hat{\mathbf{Q}}_g = n^{-1} \mathbf{Q}_g$.

By WLLN, we have $\hat{\mathbf{Q}}_{xv} \xrightarrow{p} \mathbf{Q}_{xv}$, $\hat{\mathbf{Q}}_g \xrightarrow{p} \mathbf{Q}_g$, $\hat{\mathbf{Q}}_{zz} \xrightarrow{p} \mathbf{Q}_{zz}$ and $\frac{\mathbf{Z}^T \epsilon}{n} \xrightarrow{p} E(Z_i \epsilon_i) = 0$ and, since \mathbf{Q}_{xv} and \mathbf{Q}_{zz} are full rank matrices and \mathbf{Q}_g is nonsingular, so $\mathbf{Q}_{xv} \mathbf{Q}_g^{-1} \mathbf{Q}_{zz} \mathbf{Q}_g^{-1} \mathbf{Q}_{xv}^T$ is a symmetric and nonsingular matrix and, by the continuity theorem, we deduce that

$$\left[\hat{\mathbf{Q}}_{xv} \hat{\mathbf{Q}}_g^{-1} \hat{\mathbf{Q}}_{zz} \hat{\mathbf{Q}}_g^{-1} \hat{\mathbf{Q}}_{xv}^T \right]^{-1} \xrightarrow{p} \left[\mathbf{Q}_{xv} \mathbf{Q}_g^{-1} \mathbf{Q}_{zz} \mathbf{Q}_g^{-1} \mathbf{Q}_{xv}^T \right]^{-1},$$

consequently, we have

$$\hat{\beta}_{2SLS-RME} - \mathbf{W}\beta \xrightarrow{p} \left[\mathbf{Q}_{xv} \mathbf{Q}_g^{-1} \mathbf{Q}_{zz} \mathbf{Q}_g^{-1} \mathbf{Q}_{xv}^T \right]^{-1} \mathbf{Q}_{xv} \mathbf{Q}_g^{-1} \cdot 0 = \mathbf{0}.$$

□

Theorem 3.3. Under assumptions 2,

$$\sqrt{n}(\hat{\beta}_{2SLS-RME} - \mathbf{W}\beta) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Omega}), \quad (3.1)$$

where \mathbf{W} is as in theorem 3.1 and $\mathbf{\Omega} = \mathbf{M}^{-1} \mathbf{Q}_{xv} \mathbf{Q}_g^{-1} \Phi \mathbf{Q}_g^{-1} \mathbf{Q}_{xv}^T \mathbf{M}^{-1}$.

Proof. Using the notations of theorem 3.1, we have

$$\sqrt{n}(\hat{\beta}_{RME} - \mathbf{W}\beta) = \hat{\mathbf{M}}^{-1} \hat{\mathbf{Q}}_{xv} \hat{\mathbf{Q}}_g^{-1} \frac{\mathbf{Z}^T \epsilon}{\sqrt{n}} = \hat{\mathbf{A}} \frac{\mathbf{Z}^T \epsilon}{\sqrt{n}},$$

where $\hat{\mathbf{M}} = \hat{\mathbf{Q}}_{xv} \hat{\mathbf{Q}}_g^{-1} \hat{\mathbf{Q}}_{zz} \hat{\mathbf{Q}}_g^{-1} \hat{\mathbf{Q}}_{xv}^T$. Now, by the CLT theorem (assumption 2.4), we have

$$\frac{\mathbf{Z}^T \epsilon}{\sqrt{n}} = \frac{\sum_{i=1}^n Z_i \epsilon_i}{\sqrt{n}} \xrightarrow{d} N(\mathbf{0}, \Phi),$$

where Φ is a finite and nonsingular matrix. Denoting the random vector $H \sim N(\mathbf{0}, \Phi)$ and applying the Slutsky Theorem, we have

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{RME} - W\beta) &\xrightarrow{d} M^{-1}Q_{xv}Q_g^{-1}.N(\mathbf{0}, \Phi) \\ &\sim N(\mathbf{0}, M^{-1}Q_{xv}Q_g^{-1}\Phi Q_g^{-1}Q_{xv}^T M^{-1}) \\ &\sim N(\mathbf{0}, \Omega), \end{aligned}$$

where Ω is as (3.1). □

The following corollaries are the direct consequences of the above Theorems. The first corollary yields by letting $\lambda = 0$ and the second one holds by considering the diagonal elements of Λ equal zero.

Corollary 3.1. *If there are only some measurement errors in the instrumental variables then*

$$\sqrt{n}(\hat{\beta}_{2SLS-ME} - W\beta) \xrightarrow{d} N(\mathbf{0}, \Omega),$$

where W and Ω are as in Theorem 3.3 with $Q_g = \Phi^T \Phi - n\Lambda$.

Corollary 3.2. *If there are no measurement errors and multicollinearity in the instrumental variables, then*

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} N(\mathbf{0}, \sigma_u^2 \mathbf{X}^T \mathbf{X}),$$

where $\hat{\beta}_{2SLS}$ is the 2SLS estimator (2.9).

4 Numerical Study

4.1 The Monte Carlo Simulation

In this section, we employ the Monte Carlo simulation to compare the performance of estimators in the previous section with respect to their absolute relative bias (R.Bias) and estimated mean square errors (EMSE) under several degrees of multicollinearities. First, we generate data in the form of the vector $W = (x, \Phi, \epsilon)^T$, which has a multivariate normal distribution with mean zero and the covariance matrix in such a way that data have multicollinearity and endogeneity problem. For multicollinearity, we assume the correlation between the instruments is equal to 0.9 and for endogeneity, we assume

Table 1: Estimated R.Bias and MSE values of the 2SLS, 2SLS-ME and 2SLS-RME

		$\rho=0.50$		
Coefficients	β	$\hat{\beta}_{2SLS}$	$\hat{\beta}_{2SLS-ME}$	$\hat{\beta}_{2SLS-RME}$
β_1	0.32812	0.16076	0.34104	0.33984
β_2	0.13582	0.05206	0.12873	0.12973
β_3	0.08603	0.46956	0.07289	0.07370
R.Bias		2.17141	2.08083	2.06481
EMSE		0.29705	0.40886	0.40172

		$\rho=0.6$		
Coefficients	β	$\hat{\beta}_{2SLS}$	$\hat{\beta}_{2SLS-ME}$	$\hat{\beta}_{2SLS-RME}$
β_1	0.32812	0.14819	0.34423	0.34308
β_2	0.13582	0.08804	0.12143	0.12238
β_3	0.08603	0.44588	0.07675	0.07747
R.Bias		2.10336	2.07305	2.0575
EMSE		0.28895	0.39604	0.38951

		$\rho=0.70$		
Coefficients	β	$\hat{\beta}_{2SLS}$	$\hat{\beta}_{2SLS-ME}$	$\hat{\beta}_{2SLS-RME}$
β_1	0.32812	0.12852	0.32751	0.32656
β_2	0.13582	0.13298	0.13501	0.13592
β_3	0.08603	0.42284	0.08776	0.08841
R.Bias		2.07841	2.03842	2.02331
EMSE		0.29659	0.37786	0.37164

		$\rho=0.8$		
Coefficients	β	$\hat{\beta}_{2SLS}$	$\hat{\beta}_{2SLS-ME}$	$\hat{\beta}_{2SLS-RME}$
β_1	0.32812	0.10366	0.33902	0.33783
β_2	0.13582	0.20520	0.13127	0.13218
β_3	0.08603	0.37456	0.07394	0.07473
R.Bias		2.03210	2.03191	2.02116
EMSE		0.30733	0.37711	0.37108

the different levels of the endogeneity corresponding to $\rho = 0.5, 0.6, 0.7, 0.8$. Also, we assume the correlation between the instruments and covariates equals to 0.3, which means that the instruments are relevant. Moreover, we choose $p = 3, n = 100$ and $\Lambda = \text{diag}(0.5, 0.5, 0.5)$. For each set of instrumental variables, we considered largest eigen vector of $Z^T Z$ as the coefficient vector, η , and construct $X = Z\eta + U$, where U follows a multivariate normal with an identity matrix of correlation. We then, construct $y = \hat{X}\beta + \epsilon$ where β is the largest eigen vector of $\hat{X}^T \hat{X}$. Once a set of explanatory and dependent variables is constructed, all variables are standardized and the estimates are determined using the standardized variables. The experiment is replicated 5000 times by generating new vector W . We use Wu-Hausman test of endogeneity to check if it is

necessary to perform the instrumental variable analysis. (see Wu (1973) and Hausman (1978)).

The absolute relative bias R.Bias and EMSE for any estimator are calculated respectively as follows:

$$R.Bias(\hat{\beta}) = \frac{1}{5000} \sum_{l=1}^{5000} \sum_{j=1}^p \left| \frac{\hat{\beta}_{jl} - \beta_j}{\beta_j} \right|, \quad EMSE(\hat{\beta}) = \frac{1}{5000} \sum_{l=1}^{5000} \sum_{j=1}^p (\hat{\beta}_{jl} - \beta_j)^2,$$

where $\hat{\beta}_{jl}$ is the estimation of β_j in the l th replication of the simulation and β_j is the j -th element of the vector of coefficients, β . Intact, our criterion for goodness-of-fit is the standardized version of the difference between the estimated values and real values of coefficients. We use the R software version 3.4.1 and all source codes are available from the first author upon request. The results are summarized in Table 1, which displays the estimation of parameters for the various values of ρ .

In this table, $\hat{\beta}_{2SLS}$ is estimated using the 2SLS method in equation $X = Z\eta + U$ by ignoring the multicollinearity and measurement errors in the instrumental variables Z , while two stage least square estimators $\hat{\beta}_{2SLS-ME}$ and $\hat{\beta}_{2SLS-RME}$ are estimated using equations (2.5) and (2.8), respectively. The estimated value of the ridge parameter is obtained as $\hat{\lambda} = 0.042$ via the generalized cross validation technique, which is introduced by Golub et al. (1979), see Roozbeh et al. (2020) and Liu and Jiang (2012) and Akdeniz and Roozbeh (2017) for more information.

From the measurement error point of view, when some predictors prone to error-in measurement, the most ill-condition situation may arises is the endogeneity and 2SLS-ME estimator can overcome this problem, see, e.g., Sheikhi et al. (2020). Comparing 2SLS-ME and 2SLS-RME, we observe that for these simulated data, estimation of the vector of coefficients using 2SLS-ME method is more accurate than 2SLS, while 2SLS-RME is better than 2SLS-ME. This is because the estimated value of λ satisfies the condition $\lambda < 2\sigma_u^2/\eta^T\eta$. Although the estimated mean square error in 2SLS is less than the other two estimators, the absolute relative bias values of 2SLS-ME and 2SLS-RME is smaller than 2SLS. Moreover, as can be seen in this table, comparing 2SLS-ME and 2SLS-RME in both R.Bias and EMSE, we conclude that in all cases of ρ , 2SLS-RME behaves better than 2SLS-ME.

4.2 Health and Retirement Study data

Several authors have reported the presence or absence of an association between the genetic variant(s) and the outcome as the primary analysis result, rather than a causal effect estimate, (Didelez V, Sheehan (2007) and Burgess and Dylan (2016)). In this

section, we use our method to overcome multicollinearity in the data set of Health and Retirement Study (HRS), which is consisted of some variables that are contaminated with errors in measurement. This data contain genotype and phenotype variant which may effect some outcomes such as educational attainment. (see e.g., Rietveld et al. (2013) and Okbay et al. (2016)).

We are involved in running multiple regression to predict educational attainment (EA) using major depressive disorder (MDD), schizophrenia polygenic score (SPS) and subjective well-being (SWB) and, since EA follows a normal distribution (with sig=0.318 of Lilliefors test for normality), we carry out a multiple regression between MDD, SPS and SWB and observed that the residual of regression were correlated with those exploratory variables, we found that the *relevant* variables as instrumental variables are general cognition score (GCS), main arterial pressure (MAP) and pulse pressure (PP) and Wu-Hausman test of endogeneity confirmed this instrumental variables analysis. The instrumental variables were extremely correlated (such as $\hat{\rho} = 0.992$ between MAP and PP). Also, there exist high values of variance inflation factors in the regression equation between these exposure variables, for example VIFs 69.79, 66.32 and 29.72 for MAP, PP and SBP, redpectively¹). So, we used a cross validation technique and obtain the ridge parameter $k = 0.008$. Hence, using equation (2.7) we can estimate model parameters. Also, It is relatively common in epidemiological studies to use indirect measures of cognitive functions, which cover the areas of initiative, inhibition, shifting, emotional control, etc., and is prone to measurement error (see e. g. Harris et al. (2016)). We consider GCS, MAP and PP respectively as V_1, V_2 and V_3 and assume that MAP and PP are measured without errors. From (2.2), we will have $(V_1, V_2, V_3)^T = (Z_1, Z_2, Z_3)^T + (\delta_1, 0, 0)^T$. Similar to Rasekh (2006), in the first step, we estimate the estimated value of $\hat{\eta}$, $\hat{\sigma}_u^2$, $\hat{\lambda}$ and \hat{Z} . An estimator of Z can be obtained using $\hat{Z} = \Phi + \hat{\sigma}_w^{-2} \hat{w} \hat{\eta}^T \Lambda$, where $\hat{w} = X - \Phi \hat{\eta}$ and $\hat{\sigma}_w = \hat{\sigma}_u + \hat{\eta}^T \Lambda \hat{\eta}$ (see e.g., Ghapani and Babadi, 2016). Then, we derive the second stage estimated values of the coefficients of MDD, SPS and SWB using (2.8). The Root Mean Square Error (RMSE) was our criterion to compare these methods. It is well known that RMSE can be obtained as

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2},$$

where in this case, $y_i, i = 1, 2, \dots, n$, are real values of educational attainment (EA) and \hat{y}_i are their corresponding estimated values using 2SLS, 2SLS-ME and regularized 2SLS-RME methods. Table 2 shows the estimated coefficients as well as the RMSEs of these methods, which reveals that in the presence of measurement errors and

¹using faraway package in R

multicollinearity in instruments, the regularized 2SLS-RME performs better than 2SLS-ME.

Table 2: Comparison of 2SLS, 2SLS-ME and 2SLS-RME estimations in real data

Estimation	2SLS	2SLS-ME	2SLS-RME
$\hat{\beta}_1$	-23.1059	-24.7859	-23.8733
$\hat{\beta}_2$	13.9614	13.8706	14.0619
$\hat{\beta}_3$	8.8859	9.1312	9.0957
RMSE	0.94139	0.9707	0.9590

5 Conclusion

In this work, we use the ridge regression method to combat multicollinearity in the estimation of model parameters in a 2SLS estimator, where there are some measure errors in instrumental variables. Some large sample properties of our estimator are derived and compared with some other estimators using a simulation study and a real data analysis. We hope to do more works with correlated error assumption in the future. Also, we are trying to use the Lasso and Elastic net methods to overcome the sparsity in an instrumental variable analysis.

Acknowledgements

We wish to thank the anonymous reviewers for their valuable comments and suggestions that did lead to a significant improvement of the work.

References

- Akdeniz, F., and Roozbeh, M. (2017), Efficiency of the generalized-difference-based weighted mixed almost unbiased two-parameter estimator in partially linear model. *Communications in Statistics - Theory and Methods*, **46**(24), 12259-12280.
- Arashi, M., Tabatabaey, S. M. M., and Hassanzadeh Bashtian, M. (2014), Shrinkage ridge estimators in linear regression. *Communication in Statistics-Simulation and Computation*, **43**, 871-904.

- Bowden, R., and Turkington, D. (1984), Instrumental variables. Cambridge University Press, NY.
- Burgess, S., Small, D. S., and Thompson, S. G. (2017), A review of instrumental variable estimators for Mendelian randomization. *Statistical Methods in Medical Research*, **26**(5), 2333-2355.
- Burgess, S. and Dylan, S. (2016), Predicting the Direction of Causal Effect Based on an Instrumental Variable Analysis: A Cautionary Tale. *J. Causal Infer.*, **4**(1), 49-59.
- Cao, C. Z., Lin, J. G., Shi, J. Q., Wang, W., and Zhang, X. Y. (2015), Multivariate measurement error models for replicated data under heavy-tailed distributions. *Chemometrics*, **8**, 457-466.
- Carrasco, M. (2012). A regularization approach to the many instruments problem. *Journal of Econometrics*, **170**(2), 383-398.
- Carrasco, M., and Tchuente, G. (2015), Regularized LIML for many instruments. *Journal of Econometrics*, **186**(2), 427-442.
- Carroll, R., Ruppert, J. D., Stefanski, L. A., and Crainiceanu, C. M. (2006), *Measurement Error in Nonlinear Models: A Modern Perspective*, CRC Press, Boca Raton, Fla, 2nd edition, pp. 485.
- Chan, L. K., and Mak, T. K. (1979), On the maximum likelihood estimation of a linear structural relationship when the intercept is known. *J. Multivar. Anal.*, **9**, 304-313.
- Devanarayana, V., and Stefanski, L. A. (2002), Empirical Simulation Extrapolation for Measurement Error Models With Replicate Measurements, *Statistics and Probability Letters*, **59**, 219-225.
- Didelez, V., and Sheehan, N. A. (2007), Mendelian randomization as an instrumental variable approach to causal inference. *Stat. Meth. Med. Res.*, **16**, 309-330.
- Ebbes, P. (2004), *Latent Instrumental Variables - A New Approach to Solve for Endogeneity*. Ph.D.thesis, University of Groningen Economics and Business.
- Golub, G. H., Heath, M., and Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, **21**, 215-223.
- Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N., and Smith, G. D. (2008), Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine*, **27**(8), 1133-1163.

- Liu, X. Q., and Jiang, H. Y. (2012), Optimal generalized ridge estimator under the generalized cross-validation criterion in linear regression. *Linear Algebra and its Applications*, **436**(5), 1238-1245.
- Fallah, R., Arashi, M., and Tabatabaey, S. M. M. (2017), On the ridge regression estimator with sub- space restriction, *Communication in Statistics – Theory and Methods*, **46**(23), 11854-11865.
- Ghapani, F., Babadi, B. (2016), A New Ridge Estimator in Linear Measurement Error Model with Stochastic Linear Restrictions. *JIRSS*, **15**(2), 87-103.
- Hausman, J. A. (1978), Specification tests in econometrics. *Econometrica*, **46**, 1251-1271.
- Hansen, C., and Kozbur, D. (2014), Instrumental Variables Estimation with Many Weak Instruments Using Regularized JIVE. *Journal of Econometrics*, **182**(2), 290-308.
- Harris, M. H., Gold, D. R., Rifas-Shiman, S. L., Melly, S.J., Zanolotti, A., Coul, B. A., Schwartz, J. D., Gryparis ,A., Kloog, I., and Koutrakis, P. (2016), Prenatal and childhood traffic-related air pollution exposure and childhood executive function and behavior. *Neurotoxicol. Teratol.*, **57**, 60–70.
- Horel, A. E., and Kennard, R. W. (1970), Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55- 83.
- Isogawa, Y. (1985), Estimating a multivariate linear structural relationship with replication. *J. R. Statist. Soc. B*, **47**, 211-215.
- Lachos, B. G., and Bolfarine, H. (2016), Heavy tailed calibration model with Berkson measurement errors for replicated data. *Chemometrics and Intelligent Laboratory systems*, **156**, 21-35.
- Lousdal, M. L. (2018), An introduction to instrumental variable assumptions, validation and estimation. *Emerging themes in epidemiology*, **15**(1), 1-7.
- Nawarathna, L. S., and Choudhary, P. K. (2015), A Heteroscedastic Measurement Error Model for Method Comparison Data With Replicate Measurements. *Statistics in Medicine*, **34**(7), 1242-1258.
- Okbay, A. et al. (2016), Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*.

- Roosbeh, M., Hesamian, G., Akbari, M. G., (2020), Ridge estimation in semi-parametric regression models under the stochastic restriction and correlated elliptically contoured errors. *J Comput. Appl. Math.*, **378**, 112940.
- Rasekh, A. R. (2006), Local influence in measurement error models with ridge estimate. *comput. Stat. Data Anal*, **50**, 2822-2834.
- Rasekh, A. R., Fieller, N. R. J. (2003), Influence functions in functional measurement error models with replicated data. *Statistics*, **37**(2), 169–178.
- Rietveld, C. A. et al. (2013), GWAS of 126,559 Individuals identifies genetic variants associated with educational attainment. *Science*, **340**(6139), 1467-1471.
- Saleh, A. M. E., Arashi, M., and Kibria, G. (2019), *Theory of Ridge Regression Estimation with Applications*. John Wiley, USA.
- Saleh, E., and Shalabh (2014), Ridge regression estimation approach to measurement error model, *J. Multivariate Analysis.*, **123**, 68-84.
- Sheikhi, A., Bahador, F., Arashi, M. (2020), On a generalization of the test of endogeneity in a two stage least squares estimation. *Journal of Applied Statistics*, DOI: 10.1080/02664763.2020.1837084.
- Singh, S., Jain, K., Sharma, S. (2014), Replicated measurement error model under exact linear restrictions. *Statist Pap.*, **55**, 253-274.
- Martens, E. P., Pestman, W. R., de Boer, A., Belitser, S. V., and Klungel, O. H. (2006), Instrumental variables: application and limitations. *Epidemiology*, **17**, 260–267.
- Shalabh, C. M., and Paudel, N. K. (2009), Consistent estimation of regression parameters under replicated ultrastructural model with non normal errors. *J. Stat Comp Sim*, **79**(3), 251-274.
- Wooldridge, J. M. (2016), *Introductory Econometrics: A Modern Approach 6th Edition*. Cengage Learning, Boston, MA.
- Wu, D. M. (1973), Alternative tests of independence between stochastic regressors and disturbances. *Econometrica*, **42**, 529-546.