Statistics and Its Interface Volume 16 (2023) 545–552

SIMEX estimation for quantile regression model with measurement error

Yiping Yang*, Peixin Zhao, and Dongsheng Wu

The quantile regression model with measurement error is considered. To deal with measurement error, we extend the simulation-extrapolation (SIMEX) method to the case of quantile regressions in the presence of covariate measurement error. The proposed SIMEX estimation corrects the bias caused by the measurement error, and not requires the equal distribution assumption of the regression error and measurement error. The asymptotic distribution of the proposed estimator is derived. The finite sample performance of the proposed method is investigated by a simulation study. A real dataset from the Framingham Heart Study is analyzed to illustrate the proposed method.

AMS 2000 subject classifications: Primary 62G05; secondary 62G20.
Keywords and phrases: Quantile regression, Measurement error, Simulation-extrapolation, Correction for attenuation.

1. INTRODUCTION

Errors-in-variables models have drawn much attention in medicines, biology, economics, finance and other fields. It is well known that the estimator can also be seriously biased for the mean regression and quantile regression if one ignores the measurement error. To correct the bias caused by the measurement error, Liang, Härdle and Carroll [1] yielded the consistent estimation by applying the so-called "correction for attenuation". Cui [2] obtained the consistent M estimators of errors-in-variables models based on the orthogonal residual when the regression error and each component of measurement errors follow the same distribution. Cui and Li [3], Cui and Chen [4] extended the orthogonal method to obtain the constrained empirical likelihood confidence region for the linear errors-in-variables models and the consistent generalized least square estimation for semilinear errors-in-variables models. Cook and Stefanski [5] proposed the SIMEX method to correct the bias due to additive measurement error. SIMEX has emerged as an important method developed to eliminate the biases of parameter estimates caused by measurement errors in a variety of models, see for example Carroll, Lombard, Küchenhoff and Stefanski [6], Carrol, Maca and Ruppert [7], Liang and Ren [8], Nolte [9], Delaigle and Hall [10] Yang, Tong and Li [11].

The above literatures are discussed in mean regression with the errors-in-variables. Quantile regression with measurement error has been investigated before, however, little research has been done in this field due to the difficulty of correcting the bias directly. He and Liang [12] extended the orthogonal method to obtain a consistent quantile estimation for linear errors-in-variables model. Jiang [13] also considered the composite quantile regression method for linear error-in-variable models based on the orthogonal residuals. The orthogonal method requires the equal distribution assumption, which is very strong and difficult to verify in practice. Hu and Schennach [14] proposed the consistent estimators of the nonparametric quantile regression model based on instrumental variables. But sometimes instrumental variables are difficult to obtain in practical application.

The SIMEX method has been widely used in mean regression with measurement error, while seldom used in quantile regression. In this paper, we apply the SIMEX method to the quantile regression model with measurement error. Our method is feasible and easy to apply without assuming the same distribution of the regression error and measurement error. The asymptotic distribution of the proposed estimator is investigated. Simulation results show that the proposed method leads to much better performance than the naive approach that ignores measurement error and the orthogonal method proposed by He and Liang [12].

The remainder of this paper proceeds as follows. Section 2 proposes the SIMEX estimator for linear quantile regression model and its asymptotic properties. In section 3, the results of Monte-Carlo experiments are reported. Section 4 discusses the application of the proposed method. The proofs are given in the Appendix.

2. SIMEX ESTIMATOR

Consider the following linear quantile regression model

$$Y_i = X_i^T \beta + \varepsilon_i, \quad i = 1, \ldots, n,$$

where Y_i is the response variable, and $X_i \in R^p$ is the vector of covariables, β is the p-dimensional unknown parameter, ε_i is the error term and satisfies $P(\varepsilon_i \leq 0 | X_i) = \tau$ with

*Corresponding author.

$\tau \in (0,1)$. Often, the covariables X are not exactly observed, we observe covariables W with additive measurement error,

$$W_i = X_i + U_i,$$

where $U_i \sim N(0, \Sigma_{uu})$ is independent of (X, Y), and Σ_{uu} is assumed known. However, the proposed method can be still be used when Σ_{uu} is estimated, e.g, by the replication experiments method in Carroll et.al [15]. It is well known that the quantile regression estimator of β can be seriously biased with ignoring the measurement error of X_i and using W_i instead of X_i. Here, what of interest is to obtain an unbiased estimate of β for a particular τ by the SIMEX method. The key idea is to add additional measurement errors to the mismeasured variables in order to estimate the effect of the estimation bias and variance of the measurement error, and then extrapolate back to the case of no measurement error (Carroll et al. [15]). In this section, we extend the SIMEX method to quantile regression (QR) model with measurement error. The proposed SIMEX algorithm can be described as follows:

1. Simulation step
   (a) Choose a grid of $\lambda = 0 < \lambda_1 < \ldots < \lambda_M$. Here λ controls how much additional independent measurement error is added to W.
   (b) For a particular τ and each λ_m,
      i. Generate a sequence of variables

$$W_{ib}(\lambda_m) = W_i + (\lambda_m \Sigma_{uu})^{1/2} U_{ib}, b = 1, \ldots, B,$$

      where $U_{ib} \sim N(0, I_p)$, I_p is a $p \times p$ identity matrix, B is a given integer. Noted that $E(W_{ib}(\lambda_m)|X_i) = X_i$ and $\mathrm{Var}(W_{ib}(\lambda_m)|X_i) = \mathrm{Var}\{(W_{ib}(\lambda_m) - X_i)^2|X_i\} = (1+\lambda_m)\mathrm{Var}(W_i|X_i)$. Hence, when $\lambda_m = -1$, $\mathrm{Var}(W_{ib}(\lambda_m)|X_i) = 0$.
      ii. Calculate the QR estimation for the simulated data $W_{ib}(\lambda_m)$,

$$\hat{\beta}_b^{(\tau)}(\lambda_m) = \arg\min_{\beta} \sum_{i=1}^{n} \rho_\tau(Y_i - W_{ib}^T(\lambda_m)\beta),$$

      where $\rho_\tau(r) = \tau r - rI(r < 0)$ is the quantile loss function.
      iii. Average the estimated values $\hat{\beta}_b^{(\tau)}(\lambda_m)$ over $b = 1, \ldots, B$,

$$\hat{\beta}^{(\tau)}(\lambda_m) = \frac{1}{B} \sum_{b=1}^{B} \hat{\beta}_b^{(\tau)}(\lambda_m).$$

2. Extrapolation step
   (a) Use the extrapolant function $\mathcal{G}(\lambda, \Gamma)$ to fit the data $\{\hat{\beta}^{(\tau)}(\lambda_m), \lambda_m, m = 1, \ldots, M\}$. The extrapolation function is usually unknown. The following extrapolate function tends to be the most widely used: the quadratic function $\mathcal{G}(\lambda, \Gamma) = \gamma_0 + \gamma_1\lambda + \gamma_2\lambda^2$ with $\Gamma = (\gamma_0, \gamma_1, \gamma_2)^T$ (see Liang and Ren [8] and Lin and Carroll [16]). Assume that the quadratic function $\mathcal{G}(\lambda, \Gamma)$ serves as a good approximation to the relationship for this data. Then, for $j = 1, \ldots, p$,

$$\hat{\Gamma}_j = \arg\min_{\Gamma_j} \sum_{m=1}^{M} \Big(\hat{\beta}_j^{(\tau)}(\lambda_m) - \mathcal{G}(\lambda_m, \Gamma_j)\Big)^2,$$

      where $\Gamma_j = (\gamma_{0j}, \gamma_{1j}, \gamma_{2j})^T$, $\hat{\Gamma}_j = (\hat{\gamma}_{0j}, \hat{\gamma}_{1j}, \hat{\gamma}_{2j})^T$ and $\hat{\beta}_j^{(\tau)}(\lambda_m)$ is the jth component of $\hat{\beta}^{(\tau)}(\lambda_m)$.
   (b) Extrapolate to the case of no measurement error to obtain the SIMEX estimator

$$\hat{\beta}_{\mathrm{SIMEX},j}^{(\tau)} = \mathcal{G}(-1, \hat{\Gamma}_j).$$

      where $\hat{\beta}_{\mathrm{SIMEX},j}^{(\tau)}$ is the jth component of $\hat{\beta}_{\mathrm{SIMEX}}^{(\tau)}$.

Remark 2.1. When $\lambda = 0$, the SIMEX estimator reduces to the naive estimator, $\hat{\beta}_{\mathrm{Naive},j}^{(\tau)} = \mathcal{G}(0, \hat{\Gamma}_j)$, which ignores the measurement error and directly uses W instead of X.

To establish the asymptotic properties of $\hat{\beta}_{\mathrm{SIMEX}}^{(\tau)}$, we first list the following regularity conditions:

(C1) The matrix $\Omega(\lambda) = E\{W_{ib}(\lambda)W_{ib}^T(\lambda)\}$ is positive definite matrix for $\lambda \in \Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_M\}$.

(C2) The conditional distribution of Y given $W_b(\lambda)$ is absolutely continuous, the corresponding density function $f(\cdot)$ is bounded away from zero and finity at the τ conditional quantiles.

Let $\hat{\beta}^{(\tau)}(\Lambda) = (\hat{\beta}^{(\tau)T}(\lambda_1), \ldots, \hat{\beta}^{(\tau)T}(\lambda_M))^T$, $\mathbf{\Gamma} = (\Gamma_1^T, \ldots, \Gamma_p^T)^T$, where Γ_j is the parameter vector estimated in the extrapolation step for the jth component of $\hat{\beta}^{(\tau)}(\lambda)$ with $j = 1, \ldots, p$. Write $\mathcal{G}(\Lambda, \mathbf{\Gamma}) = \mathrm{vec}\{\mathcal{G}(\lambda_m, \Gamma_j), j = 1, \ldots, p, m = 1, \ldots, M\}$, $\mathrm{Res}(\mathbf{\Gamma}) = \hat{\beta}^{(\tau)}(\Lambda) - \mathcal{G}(\Lambda, \mathbf{\Gamma})$, $s(\mathbf{\Gamma}) = \{\partial/\partial(\mathbf{\Gamma})\}\mathrm{Res}(\mathbf{\Gamma})$, $D(\mathbf{\Gamma}) = s(\mathbf{\Gamma})s^T(\mathbf{\Gamma})$, $\varepsilon_{ib}^* = Y_i - W_{ib}^T(\lambda)\beta^{(\tau)}(\lambda)$,

$$\eta_{iB}(\lambda, \tau) = \frac{1}{f(0)} \frac{1}{B} \sum_{b=1}^{B} W_{ib}(\lambda) \Big[I(\varepsilon_{ib}^* \le 0) - \tau\Big],$$

$$\Psi_{iB}\Big\{\Lambda, \tau\Big\} = \mathrm{vec}\{\eta_{iB}(\lambda, \tau), \lambda \in \Lambda\},$$

$$\mathcal{A}_{11}\Big(\Lambda\Big) = \mathrm{diag}\Big\{\Omega(\lambda), \lambda \in \Lambda\Big\}$$

and

$$\Sigma(\Lambda, \tau) = \mathcal{A}_{11}^{-1}\Big(\Lambda\Big) C_{11}\Big\{\Lambda, \tau\Big\}\Big\{\mathcal{A}_{11}^{-1}\Big(\Lambda\Big)\Big\}^T,$$

where

$$C_{11}\Big\{\Lambda, \tau\Big\} = \mathrm{cov}\Big(\Psi_{iB}\Big\{\Lambda, \tau\Big\}\Big).$$

Theorem 2.1. *Suppose that conditions (C1)and (C2) hold, then*

$$\sqrt{n}(\hat{\beta}^{(\tau)}_{\text{SIMEX}} - \beta^{(\tau)}) \xrightarrow{\mathcal{L}} N\{0, \mathcal{G}_{\mathbf{\Gamma}}(-1, \mathbf{\Gamma})\Sigma(\mathbf{\Gamma})\{\mathcal{G}_{\mathbf{\Gamma}}(-1, \mathbf{\Gamma})\}^T\},$$

where

$$\mathcal{G}_{\mathbf{\Gamma}}(\lambda, \mathbf{\Gamma}) = \{\partial/\partial(\mathbf{\Gamma})\}\mathcal{G}(\lambda, \mathbf{\Gamma}),$$

$$\Sigma(\mathbf{\Gamma}) = D^{-1}(\mathbf{\Gamma})s(\mathbf{\Gamma})\Sigma(\Lambda, \tau)s^T(\mathbf{\Gamma})D^{-1}(\mathbf{\Gamma}).$$

3. SIMULATION

In this section, we discuss the finite sample performance of our proposed estimator (SIMEXQR) by simulation studies. Consider the following model

$$\begin{cases} Y_i = X_{i1}\beta_1 + X_{i2}\beta_2 + (\varepsilon_i - F_\varepsilon^{-1}(\tau)), \\ W_i = X_i + U_i, \qquad i = 1, \ldots, n, \end{cases}$$

where $F_\varepsilon(\cdot)$ is the distribution function of ε_i, $X_{i1}, X_{i2} \sim N(0,1)$, $(\beta_1, \beta_2) = (1,2)$, U_i is generated from $N(0, \text{diag}(0.4^2, 0.4^2))$.The random error variables ε_i are taken to be $N(0, 0.4^2)$, $0.1{*}t(1)$ and $0.2{*}\text{Cauchy}(0,1)$ distribution. We compare the SIMEXQR method with the naive quantile regression (NQR) proposed by Koenker and Bassett [17], which ignores the measurement error and directly uses W_i instead of X_i, and the orthogonal QR method (ORQR) proposed by He and Liang [12]. In each simulation, we run 500 times to assess the performance with $n = 150$. In the SIMEX algorithm, we take $\lambda = 0, 0.2, \ldots, 2$ and $B = 100$. We compute the biases and standard errors (SE) for the three different types of distributions of the random error with different quantile levels.

First, the Q-Q plots of the SIMEXQR estimators of β_1 and β_2 for the Cauchy random error with $\tau = 0.1, 0.5, 0.9$ are plotted in Figure 1 and other cases are similar. Figure 1 shows that empirically these estimators are asymptotically normal.

To evaluate the performance of the fitness of the quadratic extrapolate function in extrapolation process, we plot the trace of the extrapolation step of the SIMEX algorithm for the Cauchy random error with $\tau = 0.1, 0.5, 0.9$ for one run in Figure 2, and other cases are similar. Figure 2 shows the quadratic extrapolate function fits the data $(\hat{\beta}^{(\tau)}(\lambda), \lambda)$ well.

Next, we compare the SIMEXQR method with the NQR method and the ORQR method. Table 1, Table 2, Table 3 and Table 4 present the results for the three different random error variables, respectively. From Table 1, Table 2, Table 3 and Table 4, we can see the following results:

(1) When the regression and measurement error are normally distributed with the equal variance $\sigma_\varepsilon^2 = \sigma_{uu}^2 = 0.4^2$ (see Table 1), the SIMEXQR and ORQR estimators have smaller biases than the NQR estimator. Hence, the NQR method is biased. The SE of SIMEXQR estimator is slightly bigger than the ORQR estimator, but the bias of SIMEXQR estimator is smaller than the ORQR estimator in most cases. Furthermore, we provide the mean squared error (MSE) results to better compare the SIMEXQR method and the ORQR method in Table 2. From the Table 2, we can see that the MSE of SIMEXQR estimator is slightly bigger than the ORQR estimator.

(2) For the heavy-tailed error distributions with $t(1)$ and Cauchy$(0,1)$ (see Table 3 and Table 4), the SIMEXQR es-

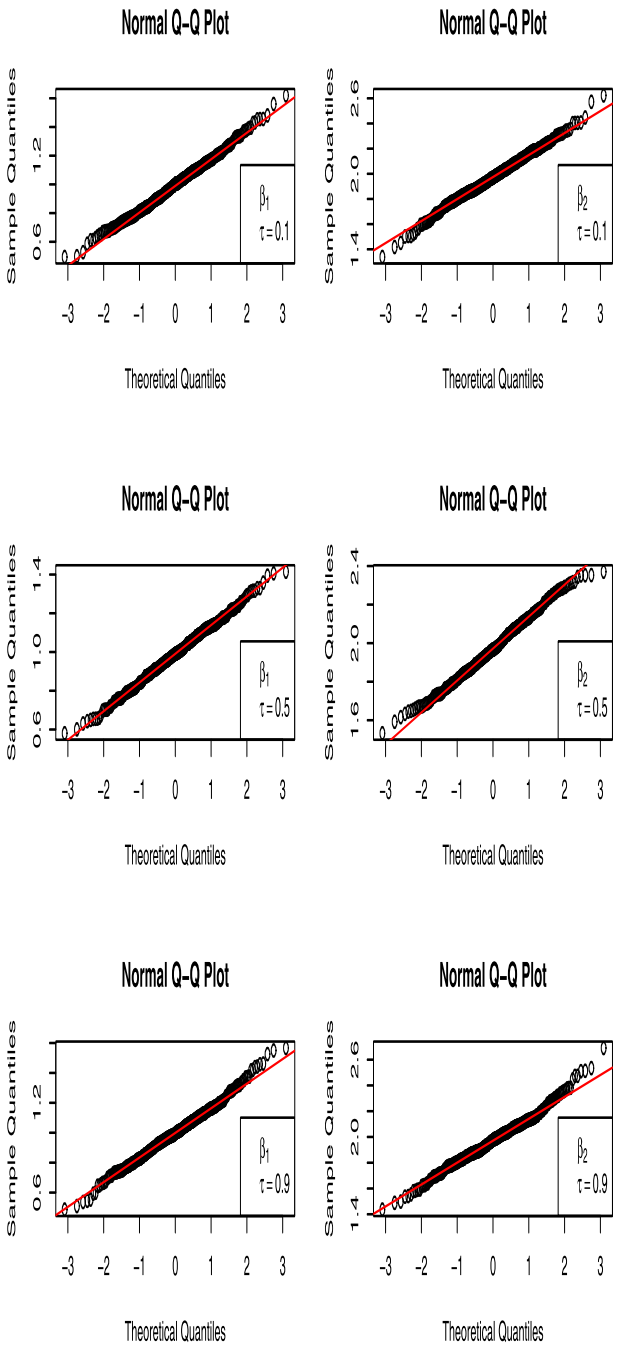

Figure 1. The Q-Q plots of the SIMEXQR estimates of every parameter for the Cauchy random error with $\tau = 0.1, 0.5, 0.9$.

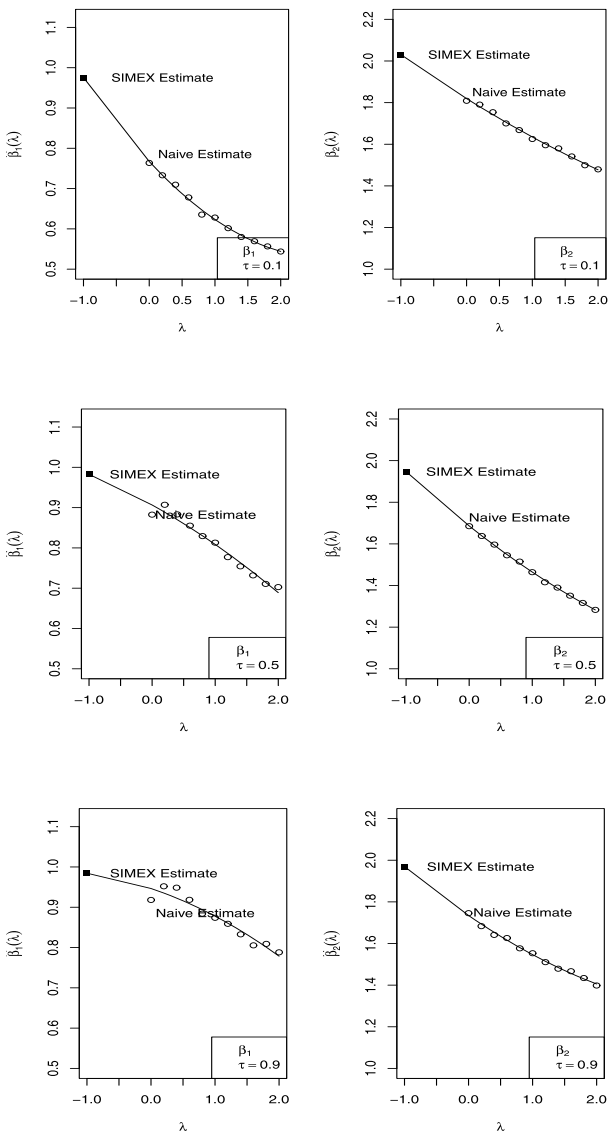

Figure 2. The trace of the extrapolation step of the SIMEX algorithm for the Cauchy random error with $\tau = 0.1, 0.5, 0.9$ for one run. The simulated estimates $(\hat{\beta}^{(\tau)}(\lambda), \lambda)$ are plotted (dots), and the fitted quadratic function (solid lines) is extrapolated to $\lambda = -1$. The extrapolation results are the SIMEX estimates (squares), the naive estimates correspond to 0 on the horizontal axis.

timator still performs well, but the NQR method is biased because of ignoring the measurement error. The bias and SE of the ORQR estimator are much bigger than the SIMEXQR estimator, the principal reason is that the equal distribution assumption of the regression error and measurement error is violated.

(3) Note that the SE of the NQR estimator is smaller than the SIMEXQR estimator, it is because that the naive estimator doesn't consider the extra variance caused by the measurement error.

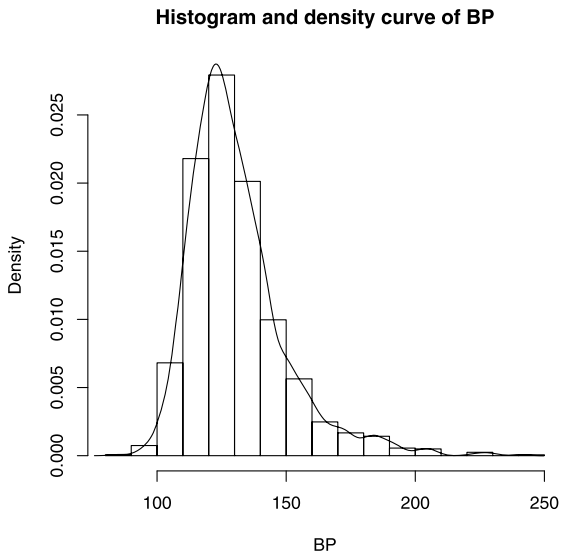

Figure 3. The histogram and density curve of the average blood pressure (BP) in a fixed two-year period for the Framingham Heart Study.

4. APPLICATION

In this section, we analyze a data set from the Framingham Heart Study to illustrate the proposed procedure. The dataset contains 5 variables with 1615 males. Liang, Härdle and Carroll [1] used the partially linear errors-in-variables model to analyze the relationship among the age, the logarithm of serum cholesterol level and the blood pressure. What we're interested in is how the serum cholesterol affects the blood pressure. The response variable Y is their average blood pressure in a fixed two-year period and W is the standardized variable for the logarithm of the serum cholesterol level (log(SC)). Similar to Liang, Härdle and Carroll [1], W is measured with error. σ_{uu}^2 is estimated to be 0.2632 by two replicates experiments. Figure 3 shows the density curve of Y. From Figure 3, we see that the distribution of Y is non-normal. Further, the p-value of the Kolmogorov-Smirnov normal test is $1.954e-14$. From the test, we can also see that the distribution Y is non-normal. Therefore, it may be more reasonable to use the quantile regression to analyze the dataset than the mean regression. The estimators and standard errors of β based on NQR, SIMEXQR and ORQR are reported in Table 5. Noted that we can obtain the standard error by estimating the asymptotic variance in Theorem 2.1, but the asymptotic variance in Theorem 2.1 is very complex. In order to avoid estimating the asymptotic variance, we use the Bootstrap method to compute the standard errors. We sample with replacement 200 times from the original 1615 data set and compute $\hat{\beta}^{(\tau)}$, then repeat the above step a number of times, 1000, to come up with estimators $\hat{\beta}_1^{(\tau)}, \ldots, \hat{\beta}_{1000}^{(\tau)}$. The stand deviation of the values $\hat{\beta}_1^{(\tau)}, \ldots, \hat{\beta}_{1000}^{(\tau)}$ is our estimator of the standard error

Table 1. The biases and standard errors (SE) of the parameters β_1 and β_2 obtained by the NQR, SIMEXQR and ORQR methods for different quantile levels with $\varepsilon_i \sim N(0, 0.4^2)$.

	NQR				SIMEXQR				ORQR			
	β_1		β_2		β_1		β_2		β_1		β_2	
τ	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
0.1	−0.1401	0.1014	−0.2801	0.1030	−0.0194	0.1547	−0.0287	0.1613	−0.0427	0.1165	−0.0931	0.1095
0.25	−0.1348	0.0934	−0.2805	0.0908	−0.0076	0.1490	−0.0300	0.1448	−0.0118	0.1060	−0.0344	0.1051
0.5	−0.1394	0.0879	−0.2798	0.0883	−0.0152	0.1422	−0.0273	0.1449	−0.0039	0.1118	−0.0056	0.1068
0.75	−0.1398	0.0905	−0.749	0.0893	−0.0104	0.1383	−0.0224	0.1400	−0.0171	0.1083	−0.0250	0.1072
0.9	−0.1412	0.0986	−0.2755	0.0999	−0.0126	0.1535	−0.0209	0.1572	−0.0475	0.1088	−0.0839	0.1074

Table 2. The mean squared errors (MSE) of the parameters β_1 and β_2 obtained by the SIMEXQR and ORQR methods for different quantile levels with $\varepsilon_i \sim N(0, 0.4^2)$

Method	ß	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.9$
SIMEXQR	β_1	0.0243	0.0223	0.0204	0.0192	0.0237
	β_2	0.0268	0.0219	0.0217	0.0201	0.0251
ORQR	β_1	0.0154	0.0114	0.0125	0.0120	0.0141
	β_2	0.0207	0.0122	0.0114	0.0121	0.0186

Table 3. The biases and standard errors (SE) of the parameters β_1 and β_2 obtained by the NQR, SIMEXQR and ORQR methods for different quantile levels with $\varepsilon_i \sim 0.1 * t(1)$

	NQR				SIMEXQR				ORQR			
	β_1		β_2		β_1		β_2		β_1		β_2	
τ	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
0.1	−0.1376	0.1056	−0.2730	0.1067	−0.0144	0.1516	−0.00136	0.1569	0.5918	3.0447	0.6010	3.0412
0.25	−0.1368	0.0973	−0.2772	0.0956	−0.0108	0.1447	−0.0218	0.1468	0.6419	3.2551	0.8456	3.5285
0.5	−0.1392	0.0821	−0.2801	0.0875	−0.0157	0.1293	−0.0211	0.1425	0.7457	3.3828	1.0568	3.8368
0.75	−0.1357	0.1014	−0.2717	0.0915	−0.0089	0.1427	−0.0235	0.1424	1.0920	4.1982	1.0600	3.9107
0.9	−0.1251	0.1001	−0.2813	0.1079	−0.0109	0.1529	−0.0299	0.1562	0.9406	3.9840	0.8293	3.5379

Table 4. The biases and standard errors (SE) of the parameters β_1 and β_2 obtained by the NQR, SIMEXQR and ORQR methods for different quantile levels with $\varepsilon_i \sim 0.2 * \text{Cauchy}(0, 1)$

	NQR				SIMEXQR				ORQR			
	β_1		β_2		β_1		β_2		β_1		β_2	
τ	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
0.1	−0.1345	0.1168	−0.2757	0.1251	−0.0029	0.1821	−0.0279	0.1848	0.9272	3.5626	1.3069	3.9905
0.25	−0.1365	0.1014	−0.2788	0.1074	−0.0123	0.1511	−0.0293	0.1593	1.8678	5.2401	1.9404	4.8803
0.5	−0.1363	0.0931	−0.2785	0.1006	−0.0113	0.1461	−0.0277	0.1572	1.9441	5.2781	2.2721	5.2483
0.75	−0.1399	0.0992	−0.2735	0.1054	−0.0175	0.1494	−0.0165	0.1582	1.6731	4.9010	2.0222	5.0782
0.9	−0.1292	0.1120	−0.2802	0.1165	−0.0052	0.1722	−0.0274	0.1789	1.1818	4.2165	1.2621	4.2187

of $\hat{\beta}^{(\tau)}$. The traces of the extrapolation step for the SIMEX algorithm are presented with $\tau = 0.1, 0.5$ and 0.9 in Figure 4. As can be seen from Figure 4, it is reasonable to use the quadratic extrapolation function. Compared with the ORQR method, the SIMEXQR estimators of β_2 is larger. The estimators of β_2 based on the NQR and SIMEXQR method are not significantly different at $\tau = 0.9$ quantile level. The SIMEXQR estimator of β_2 is larger than the NQR estimator at $\tau = 0.1, 0.25, 0.5, 0.75$. This means that the blood pressure and the serum cholesterol are more pos-

Table 5. *The estimators (ES) and standard errors (SE) of the parameters β_1 and β_2 obtained by the NQR, SIMEXQR and ORQR methods for different quantile levels in the Framingham Heart Study*

	NQR				SIMEXQR				ORQR			
	β_1		β_2		β_1		β_2		β_1		β_2	
τ	ES	SE	ES	SE	ES	SE	ES	SE	ES	SE	ES	SE
0.1	111.8502	0.3275	1.8971	0.4396	111.8631	0.3775	2.3050	0.5463	112.9550	0.3282	1.8596	0.4771
0.25	118.8131	0.2962	1.6674	0.3546	118.8616	0.3371	1.9239	0.4753	120.2340	0.2305	1.7033	0.3371
0.5	127.1026	0.3271	2.5329	0.4081	127.0436	0.3374	2.9517	0.5629	129.0969	0.4011	2.7682	0.3407
0.75	138.4845	0.5887	3.0920	0.5825	138.4643	0.6654	3.8900	0.7302	141.1629	0.6775	3.1428	0.7071
0.9	153.6328	1.0732	1.4826	0.9951	153.3928	1.1904	1.4204	1.2850	158.9772	1.8104	1.5424	0.5262

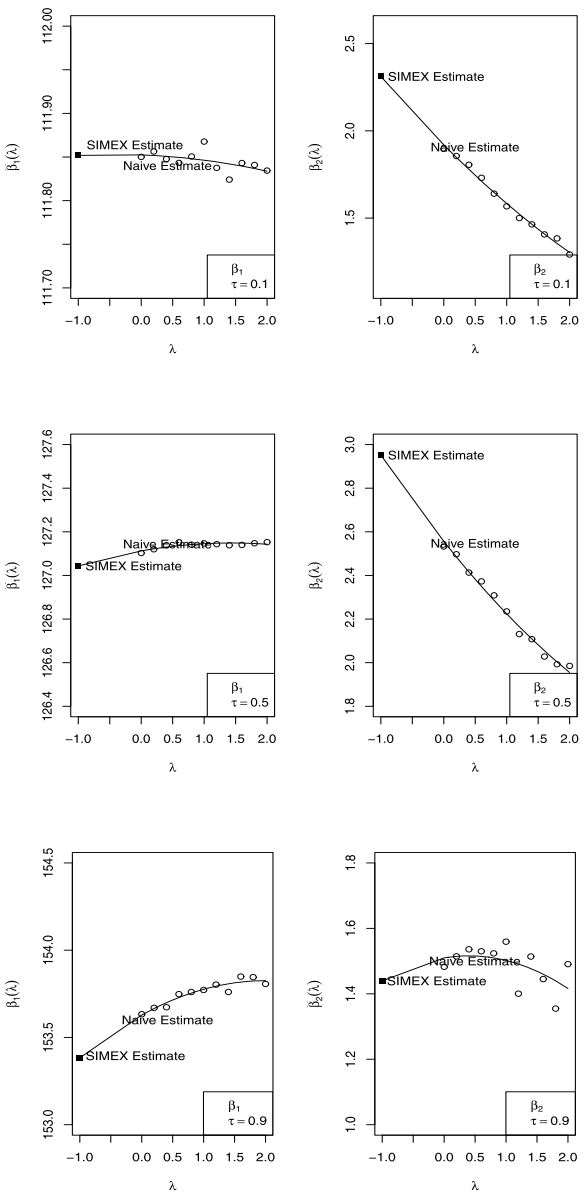

Figure 4. The traces of the extrapolation step for the SIMEX algorithm at $\tau = 0.1, 0.5, 0.9$ for the Framingham Heart Study.

itively correlated when the measurement errors are taken into account. The results are similar to those in [1].

5. CONCLUSION

In this paper, we use the SIMEX method to obtain the consistent parameter estimation for quantile regression models in the presence of covairate measurement error. Our method is easy to implement, and avoids the equal distribution assumption of the regression error and measurement error. The proposed method can be easily extended to various regression models, linear or nonlinear. The simulation results show that the proposed method reduces the bias due to the measurement error compared to the naive one. Compared with the method of He and Liang [12], the SIMEX method performs better when the equal distribution assumption of the regression and measurement error is violated.

However, there remain the further research topics. We get the asymptotic properties of $\hat{\beta}^{(\tau)}_{\text{SIMEX}}$ in Theorem 2.1. Based on the Theorem 2.1, it would be worth discussing the following linear hypothesis:

$$H_0 : R\beta^{(\tau)} = r \quad \text{versus} \quad H_1 : R\beta^{(\tau)} \neq r,$$

where R is a given $q \times p$ full rank matrix, which implies $q \leq p$, r is a q-dimensional vector. Such testing problems have been widely studied in the literature, see, for example, Fan and Huang [18] and Zhu and Zhao [19]. When the equal distribution assumption is satisfied, it would be also desirable to consider some theoretical results for comparing the asymptotic variances of the SIMEXQR and ORQR methods. Furthermore, Nghiem and Potgieter [20] proposed the simulation-selection-extrapolation (SIMSELEX) algorithm with a variable selection step based on the group lasso in high-dimensional errors-in-variables models. It would be a very interesting topic of further research to apply the SIMSELEX method to in high-dimensional quantile regression models with measurement error.

APPENDIX

Proof of Theorem 2.1: Assume $\beta^{(\tau)}(\lambda), \lambda \in \Lambda$ is the true value based on the quantile model

$$Q_Y(\tau|W_b(\lambda)) = W_b^T(\lambda)\beta^{(\tau)}(\lambda).$$

Let $\sqrt{n}(\hat{\beta}_b^{(\tau)}(\lambda) - \beta^{(\tau)}(\lambda)) = u_n$. Then u_n is the minimizer of the following criterion

$$L_n = \sum_{i=1}^{n} \left\{ \rho_\tau \left(\varepsilon_{ib}^* - \frac{W_{ib}^T(\lambda)u_n}{\sqrt{n}} \right) - \rho_\tau(\varepsilon_{ib}^*) \right\}.$$

By applying the identity in Knight [21], we have

$$\rho_\tau(r,s) - \rho_\tau(r) = s\,(I(r \le 0) - \tau) + \int_0^s \{I(r \le t) - I(r \le 0)\}dt.$$

Thus we write L_n as follows

$$L_n = \sum_{i=1}^{n} \frac{W_{ib}^T(\lambda)u_n}{\sqrt{n}}[I(\varepsilon_{ib}^* \le 0) - \tau] + B_{nb}(\lambda),$$

where $B_{nb}(\lambda) = \sum_{i=1}^{n} \int_0^{W_{ib}^T(\lambda)u_n/\sqrt{n}} [I(\varepsilon_{ib}^* \le t) - I(\varepsilon_{ib}^* \le 0)]$. Moreover, we have

$$\begin{aligned} E(B_{nb}(\lambda)) &= \sum_{i=1}^{n} \int_0^{W_{ib}^T(\lambda)u_n/\sqrt{n}} [F(t) - F(0)]dt \\ &= \frac{1}{n}\sum_{i=1}^{n} \int_0^{W_{ib}^T(\lambda)u_n/\sqrt{n}} \sqrt{n}\left[F(t/\sqrt{n}) - F(0)\right] dt \\ &\to \frac{1}{2} f(0) u_n^T \Omega(\lambda) u_n. \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(B_{nb}(\lambda)) &= \sum_{i=1}^{n} E\left\{ \int_0^{W_{ib}^T(\lambda)u_n/\sqrt{n}} [I(\varepsilon_{ib}^* \le t) - I(\varepsilon_{ib}^* \le 0) \right. \\ &\quad \left. - (F(t) - F(0))]dt \right\}^2 \\ &\le \sum_{i=1}^{n}\sum_{i=1}^{n} E\left[\left| \int_0^{W_{ib}^T(\lambda)u_n/\sqrt{n}} [I(\varepsilon_{ib}^* \le t) \right. \right. \\ &\quad \left. \left. - I(\varepsilon_{ib}^* \le 0) - (F(t) - F(0))]dt \right| \right] \\ &\quad \times 2\left| \frac{W_{ib}^T(\lambda)u_n}{\sqrt{n}} \right| \\ &\le 4E(B_{nb}(\lambda)) \frac{\max_{1\le i\le n} \left| W_{ib}^T(\lambda)u_n \right|}{\sqrt{n}} \to 0. \end{aligned}$$

Then, it follows that

$$L_n = \sum_{i=1}^{n} \frac{W_{ib}^T(\lambda)u_n}{\sqrt{n}} [I(\varepsilon_{ib}^* \le 0) - \tau] + \frac{1}{2} f(0) u_n^T \Omega(\lambda) u_n + o_p(1).$$

Since L_n is a convex function, then following Knight [21] and Koenker [22], we have

$$\begin{aligned} \sqrt{n}(\hat{\beta}_b^{(\tau)}(\lambda) - \beta^{(\tau)}(\lambda)) &= -\frac{\Omega^{-1}(\lambda)}{f(0)} \sum_{i=1}^{n} n^{-1/2} W_{ib}(\lambda) \\ &\quad \times [I(\varepsilon_{ib}^* \le 0) - \tau] + o_p(1). \end{aligned}$$

According to the definition of $\hat{\beta}^{(\tau)}(\lambda)$, we have

(*A*.1)

$$\sqrt{n}(\hat{\beta}^{(\tau)}(\lambda) - \beta^{(\tau)}(\lambda)) = \Omega^{-1}(\lambda) n^{-1/2} \sum_{i=1}^{n} \eta_{iB}(\lambda, \tau) + o_p(1).$$

By (A.1), the limit distribution of $\sqrt{n}(\hat{\beta}^{(\tau)}(\Lambda) - \beta^{(\tau)}(\Lambda))$ is multivariate normal distribution $N(0, \Sigma)$.

Extrapolation step yields

$$\hat{\mathbf{\Gamma}} = \arg\min_{\mathbf{\Gamma}} \mathrm{Res}^T(\mathbf{\Gamma})\mathrm{Res}(\mathbf{\Gamma}).$$

The estimating equation for $\hat{\mathbf{\Gamma}}$ is

$$s(\hat{\mathbf{\Gamma}})\mathrm{Res}(\hat{\mathbf{\Gamma}}) = 0.$$

Then

$$\sqrt{n}(\hat{\mathbf{\Gamma}} - \mathbf{\Gamma}) \xrightarrow{\mathcal{L}} N(0, \Sigma(\mathbf{\Gamma})).$$

Because $\hat{\beta}_{\mathrm{SIMEX}}^{(\tau)} = \mathcal{G}(-1, \hat{\mathbf{\Gamma}})$, the limit distribution of $\sqrt{n}(\hat{\beta}_{\mathrm{SIMEX}}^{(\tau)} - \beta^{(\tau)})$ is multivariate normal distribution with mean zero and covariance

$$\mathcal{G}_{\mathbf{\Gamma}}(-1, \mathbf{\Gamma})\Sigma(\mathbf{\Gamma})\{\mathcal{G}_{\mathbf{\Gamma}}(-1, \mathbf{\Gamma})\}^T.$$

□

ACKNOWLEDGEMENTS

The authors are grateful to the Editor, an Associate Editor, and two referees for their valuable suggestions and comments that greatly improved the manuscript. Yiping Yang's research was supported by Chongqing Natural Science Foundation (cstc2021jcyj-msxmX0079) and Humanities and Social Sciences Program of Chongqing Education Commission (21SIGH118), Peixin Zhao's research was supported by Chongqing Natural Science Foundation (cstc2020jcyj-msxmX0006) and the National Social Science Foundation of China (18BTJ035).

Received 21 December 2021

REFERENCES

[1] Liang, H., Härdle, W., Carroll, R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *The Annals of Statistics* **27** 1519–1535. MR1742498

[2] Cui, H. J. (1997). Asymptotic normality of M-estimates in the EV model. *Journal of Systems Science and Complexity* **10** 225–236. MR1469182

[3] Cui, H. J., Li, R. C. (1998). On parameter estimation for semi-linear errors-in-variables models. *Journal of Multivariate Analysis* **64** 1–24. MR1619970
[4] Cui, H. J., Chen, S. X. (2003). Empirical likelihood confidence region for parameter in the errors-in-variables models. *Journal of Multivariate Analysis* **84** 101–115. MR1965825
[5] Cook, J., Stefanski, L. A. (1994). Simulation-extrapolation method in parametric measurement error models. *Journal of the American Statistical Association* **89** 1314–1328. MR1379467
[6] Carroll, R. J., Lombard, F., Küchenhoff, H., Stefanski, L. A. (1996). Asymptotics for the SIMEX estimator in structural measurement error models. *Journal of the American Statistical Association* **91** 242–250. MR1394078
[7] Carroll, R. J., Maca, J. and Ruppert, D. (1999). Nonparametric regression in the presence of measurement error. *Biometrika* **86** 541–554. MR1723777
[8] Liang, H., Ren, H. (2005). Generalized partially linear measurement error models. *Journal of Computational and Graphical Statistics* **14** 237–250. MR2137900
[9] Nolte, S. (2007). The multiplicative simulation extrapolation approach. *Center for Quantitative Methods and Survey Research, University of Konstanz*, Working Paper.
[10] Delaigle, A., Hall, P. (2008). Using SIMEX for smoothing parameter choice in errors-in-variables problems. *Journal of the American Statistical Association* **130** 280–287. MR2394636
[11] Yang, Y. P., Tong, T. J., Li, G. R. (2019). SIMEX estimation for single-index model with covariate measurement error. *AStA Advances in Statistical Analysis* **103** 137–161. MR3922272
[12] He, X., Liang, H. (2000). Quantile regression estimates for a class of linear and partially linear error-in-variable models. *Statistica Sinica* **10** 129–140. MR1742104
[13] Jiang, R. (2015). Composite quantile regression for linear errors-in-variables models. *Hacettepe Journal of Mathematical and Statistics* **44** 707–713. MR3410873
[14] Hu, Y., and Schennach, S. M. (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica* **76** 195–216. MR2374986
[15] Carroll, R. J., Ruppert, D., Crainiceanu, C. M., Stefanski, L. A. (2006). Measurement error in nonlinear models: a modern perspective. *Chapman and Hall/CRC*. MR2243417
[16] Lin, X., Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association* **95** 520–534. MR1803170
[17] Koenker, R., Bassett, Jr. G. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644
[18] Fan, J., Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* **11** 1031–1057. MR2189080
[19] Zhu, S., Zhao, P. (2019). Tests for the linear hypothesis in semi-functional partial linear regression models. *Metrika* **82** 125–148. MR3922524
[20] Nghiem, L., Potgieter, C. (2019). Simulation-selection-extrapolation: estimation in high-dimensional errors-in-variables Models. *Biometrics* **75** 1133–1144. MR4041817
[21] Knight, K. (1998). Limiting distributions for L1 regression estimators under general conditions. *The Annals of Statistics* **26** 755–770. MR1626024
[22] Koenker, R. (2005). Quantile Regression. *Cambridge University Press*. MR2268657

Yiping Yang
School of Mathematics and Statistics
Chongqing Technology and Business University
Chongqing Key Laboratory of Social Economic and Applied Statistics
Chongqing
China
E-mail address: yeepingyang@foxmail.com

Peixin Zhao
School of Mathematics and Statistics
Chongqing Technology and Business University
Chongqing
China
E-mail address: zpx81@163.com

Dongsheng Wu
School of Mathematics and Statistics
Chongqing Technology and Business University
Chongqing
China
E-mail address: 1006725462@qq.com