



Performance Comparison of SIMEX and Regression Calibration for Measurement Error Correction in Nonparametric Regression

* Nawal Mohamed Othman Mustafa

Department of Statistics, Faculty of Science, Al-Marj, University of Benghazi.

مقارنة أداء طريقتي SIMEX ومعايرة الانحدار (Regression Calibration) لتصحيح

خطأ القياس في الانحدار غير المعلمي

نوال محمد عثمان مصطفى

قسم الإحصاء، كلية العلوم المرج، جامعة بنغازي

nawal.mustafa@uob.edu.ly

Received: 08/9/2026 Revised: 13/9/2026 Accepted: 05/10/2026 Publication: 26/10/2026

Abstract

It is well established that measurement errors in explanatory variables can result in severe attenuation bias and over-smoothing in nonparametric kernel estimation. This study compares two prominent measurement error correction procedures Simulation-Extrapolation (SIMEX) using quadratic extrapolation and Regression Calibration (RC) against the standard, uncorrected Nadaraya–Watson (NW) estimator under classical additive error. A Monte Carlo simulation study was conducted across three sample sizes ($n = 100, 200, 500$) and three error variance levels ($\sigma^2 = 0.1, 0.3, 0.5$) applied to a non-linear target curve. Estimator performance was evaluated using Mean Squared Error (MSE), Mean Absolute Error (MAE) and Bias. The findings reveal a clear, complementary performance pattern supported by empirical evidence: SIMEX optimally reduces bias and MAE under low error variance ($\sigma^2 = 0.1$), achieving significant reductions in MSE compared to the uncorrected estimator, ranging from 18.1% at $n = 100$, to 24.0% at $n = 200$, and reaching up to 36.1% at $n = 500$. Furthermore, SIMEX reduces MAE by up to 21.0% in large sample sizes under low noise. Conversely, Regression Calibration (RC) exhibits superior stability and consistently lower overall MSE and MAE at moderate-to-severe error levels ($\sigma^2 \geq 0.3$), offering MSE reductions of 10.3% to 14.4% across sample sizes, while avoiding the variance inflation and performance degradation inherent in SIMEX's quadratic extrapolation step under high noise. These empirical results demonstrate that selecting the optimal error correction method depends directly on the measurement error magnitude and the targeted performance metric.

Keywords: Nonparametric regression; Nadaraya–Watson estimator; Classical measurement error; Simulation-Extrapolation (SIMEX); Regression Calibration (RC); Monte Carlo simulation.

الملخص

من المعروف أن أخطاء القياس في المتغيرات التفسيرية قد تؤدي إلى تحيز التقدير وتشوه شكل المنحنى المقدر، عند استخدام أساليب الانحدار غير المعلمي القائمة على النواة. تهدف هذه الدراسة إلى مقارنة أداء طريقتين بارزتين لتصحيح خطأ القياس، هما طريقة المحاكاة والاستقراء (Simulation-Extrapolation, SIMEX) وطريقة معايرة الانحدار (Regression Calibration, RC).

(Calibration, RC)، وذلك مقارنةً بمقدّر ناداراي-واتسون (Nadaraya–Watson, NW) التقليدي غير المصحح، في ظل نموذج خطأ القياس الإضافي الكلاسيكي. أُجريت دراسة محاكاة مونت كارلو باستخدام ثلاثة أحجام للعينات ($n=100, 200, 500$) وثلاثة مستويات لتباين خطأ القياس ($\sigma^2=0.1, 0.3, 0.5$)، بهدف تقييم أداء المقدّرات في تقدير دالة انحدار غير خطية. وقد جرى تقييم الأداء باستخدام متوسط مربع الخطأ (MSE)، ومتوسط الخطأ المطلق (MAE) والتحيز (Bias). أظهرت النتائج نمطاً متكافئاً في أداء الطريقتين، إذ حققت طريقة SIMEX أداءً أفضل في خفض متوسط الخطأ المطلق عند انخفاض تباين خطأ القياس ($\sigma^2=0.1$)، كما حققت انخفاضاً ملحوظاً في متوسط مربع الخطأ مقارنةً بالمقدّر التقليدي غير المصحح، بلغت نسبته 18.1% عند حجم عينة ($n=100$)، و 24.0% عند ($n=200$)، ووصلت إلى 36.1% عند ($n=500$). كذلك، حققت طريقة SIMEX انخفاضاً في متوسط الخطأ المطلق بنسبة وصلت إلى 21.0% عند أحجام العينات الكبيرة ومستويات خطأ القياس المنخفضة. في المقابل، أظهرت طريقة معايرة الانحدار (RC) أداءً أفضل من حيث الاستقرار ومتوسط مربع الخطأ ومتوسط الخطأ المطلق عند مستويات خطأ القياس المتوسطة والمرتفعة، محققةً انخفاضاً في متوسط مربع الخطأ تراوحت نسبته بين 10.3% و 14.4% عبر أحجام العينات المختلفة، مع الحد من الزيادة في التباين وتراجع الأداء المرتبطتين بخطوة الاستقراء التربيعي في طريقة SIMEX عند ارتفاع تباين خطأ القياس. وتشير هذه النتائج إلى أن اختيار الطريقة الأنسب لتصحيح خطأ القياس يعتمد على مستوى تباين الخطأ ومقياس الأداء المستخدم في التقييم.

الكلمات المفتاحية: الانحدار غير المعلمي؛ مقدّر ناداراي-واتسون؛ خطأ القياس الكلاسيكي؛ طريقة المحاكاة والاستقراء (SIMEX)؛ معايرة الانحدار (RC)؛ محاكاة مونت كارلو.

1. Introduction

Measurement error in explanatory variables poses a fundamental challenge in statistical modeling, as using observed proxies instead of true unobserved variables attenuates estimated relationships and leads to invalid statistical inferences (Fuller, 1987; Carroll et al., 2006). This issue is particularly acute in nonparametric regression, where the underlying functional relationship between the response and the predictor is left unspecified to capture complex nonlinearities and interactions. When explanatory variables are contaminated with measurement noise, standard nonparametric kernel estimators suffer from severe attenuation bias and over-smoothing, failing to recover the target regression function accurately (Carroll, Maca, & Ruppert, 1999; Delaigle & Hall, 2008). To mitigate measurement error, various correction techniques have been developed. While deconvolution methods offer theoretical elegance, structural approaches such as Simulation-Extrapolation (SIMEX) and Regression Calibration (RC) remain widely favored in practice due to their computational accessibility and flexibility. However, their relative efficacy under nonparametric settings remains nuanced: SIMEX relies on extrapolation models that may interact unpredictably with bandwidth selection, whereas Regression Calibration imposes implicit linearity assumptions on the calibration function that may falter under strong nonlinearity. The primary objective of this study is to conduct a systematic Monte Carlo performance comparison between SIMEX (utilizing quadratic extrapolation) and Regression Calibration against the uncorrected (naive) Nadaraya–Watson estimator under classical additive error. The evaluation encompasses varying sample sizes ($n = 50, 100, 500$), error variance

magnitudes ($\sigma^2 = 0.1, 0.3, 0.5$), and nonlinear target curve configurations. Performance is assessed using Mean Squared Error (MSE), Mean Absolute Error (MAE), and Bias. Rather than seeking a universally superior method, this work aims to provide a clear framework highlighting the operational trade-offs of each procedure as a function of sample size and noise severity.

1.1 Nonparametric Regression

Unlike parametric regression models which assume a rigid, predefined functional form for the relationship between variables nonparametric regression relaxes these structural constraints. Instead, it allows the data to directly determine the shape of the regression function with minimal distributional assumptions.

A standard univariate nonparametric regression model is expressed as:

$$y = g(x) + \varepsilon \quad (1.1)$$

y_i represents the response (dependent) variable for the i the observation.

x_i represents the explanatory (independent) variable.

1.2 Nadaraya-Watson Estimator

The Nadaraya–Watson kernel estimator (Nadaraya, 1964; Watson, 1964) is a fundamental non-parametric technique used to estimate the conditional expectation function $m(x) = \{E\}[Y \mid X = x]$ without imposing structural parametric assumptions on the regression curve. Conceptually, the estimator constructs a locally weighted average of the observed response values y_i ,

assigning larger weights to observations whose corresponding covariates x_i lie closer to the evaluation point x .

Mathematically the Nadaraya–Watson estimator is expressed as:

$$\hat{m}_{NW}(x) = \frac{\sum_{i=1}^n K\left(\frac{X - x_i}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)} \quad (1.2)$$

where:

$K(\bullet)$ represents a symmetric kernel function operating as a probability density, satisfying

$h > 0$ is the bandwidth (smoothing parameter) that controls the window width around x .

In the presence of measurement error in x_i , the standard Nadaraya–Watson estimator becomes inconsistent and suffers from severe attenuation bias, as the

noise contaminates the distance metrics $K\left(\frac{X_i - x}{h}\right)$ and artificially flattens the estimated curve $\hat{m}_{NW}(x)$.

1.3 Classical Measurement Error Model

In many practical regression settings, the true explanatory variable X cannot be measured directly due to instrument noise, self-reporting bias, or physical limitations. Instead, an uncorrupted surrogate variable W is observed. Under the

classical additive measurement error model, the observed covariate W_i is related to the unobserved true covariate X_i via the formulation:

$$W_i = X_i + U_i \quad i=1,2,\dots,n \quad (1.3)$$

where U_i represents the random measurement error component.

The classical measurement error framework relies on the following standard statistical assumptions:

Zero Conditional Mean: $E[U_i / X_i] = 0$, implying that the observed variable W_i is an unbiased estimator of X_i ,

Independence from True Predictor: $COV(X_i, U_i) = 0$ indicating that the magnitude of error is uncorrelated with the true value of the variable

When measurement error is ignored and W_i is directly substituted for X_i into the standard Nadaraya–Watson estimator, the distance calculations in the kernel function become contaminated. This contamination causes severe attenuation bias (flattening of the estimated curve) and artificial over-smoothing, rendering standard non-parametric inference invalid.

1.4 Simulation–Extrapolation (SIMEX) Method

The Simulation–Extrapolation (SIMEX) method (Cook & Stefanski, 1994; Sevilimedu & Yu, 2022) is a computer-intensive, simulation-based technique designed to mitigate measurement error bias when the error variance σ_u^2 is known or reliably estimated. The core mechanism of SIMEX relies on parametrically inflating the measurement error to establish a trend between the estimator's bias and the variance of the error, which is subsequently extrapolated back to the zero-error scenario.

The procedure operates through two primary stages:

Simulation Step: For a grid of user-defined error-inflation parameters $\lambda \in \{0, \lambda_1, \lambda_2, \dots, \lambda_M\}$ (where $\lambda = 0$ corresponds to the original contaminated data), pseudo-datasets are generated. For each observation $i (i = 1, \dots, n)$ and simulation replication $b (b = 1, \dots, B)$

$$W_{i,b}(\lambda) = W_i + \sqrt{\lambda} U_{i,b}$$

where $U_{i,b} \approx N(0, \sigma_u^2)$ are independent synthetic errors. The variance of the pseudo-covariate $W_{i,b}(\lambda)$ conditional on x_i becomes $(1 + \lambda)\sigma_u^2$. For each value of λ the Nadaraya–Watson estimator is computed on each pseudo-dataset, and the average estimate across the B replications is obtained:

$$\hat{m}(x, \lambda) = \frac{1}{B} \sum \hat{m}_{NW,b}(X, \lambda)$$

Extrapolation Step: The averaged estimates $\hat{m}(x, \lambda)$ are modeled as a parametric function of λ capture potential non-linear interactions under nonparametric smoothing, a quadratic extrapolation function is adopted:

$$\mathfrak{Z}(\lambda) = \gamma_0 + \gamma_1 \lambda + \gamma_2 \lambda^2$$

The coefficients $(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2)$ are fitted via ordinary least squares using the evaluated points $\{\lambda, \hat{m}(x, \lambda)\}$. The final SIMEX estimator is evaluated at $\lambda = -1$ which theoretically cancels out the measurement error variance $(1 + (-1))\sigma_u^2 = 0$

$$\hat{m}_{SIMEX}(x) = \eta(-1) = \hat{\gamma}_0 - \hat{\gamma}_1 + \hat{\gamma}_2 \quad (1.4)$$

1.5 Regression Calibration (RC) Method

Regression Calibration (RC) (Carroll et al., 2006; Keogh et al., 2020) is an intuitive, computationally efficient correction procedure that addresses measurement error by substituting the unobserved true covariate X_i with its conditional expectation given the observed contaminated surrogate W_i .

Under the classical additive measurement error model ($W_i = X_i + U_i$) with $U_i \approx N(0, \sigma_u^2)$ and assuming $X_i \approx N(\mu_x, \sigma_x^2)$ the conditional distribution of X_i given W_i is normal. The calibration function determining the predicted uncorrupted covariate $\hat{X}_i$ is derived as:

$$\hat{X}_i = E[X_i | W_i] = \mu_x + \gamma(W_i - \mu_x) = (1 - \gamma)\mu_x + \gamma W_i$$

where γ denotes the reliability ratio (attenuation factor)

$$\gamma = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} = \frac{\text{var}(X)}{\text{var}(W)}$$

In practical implementations, population parameters are generally unknown and must be estimated using appropriate sample information. Estimation of the reliability ratio requires sufficient information about the variance of the true covariate and the measurement error variance. Regression Calibration can provide appropriate bias correction in linear regression models under suitable assumptions. However, in nonlinear regression, it generally provides an approximation because replacing the true covariate with its conditional expectation does not necessarily preserve the original regression relationship.

2. The Simulation Study

A Monte Carlo simulation study was conducted to evaluate the performance of the Nadaraya–Watson (NW) estimator and its measurement-error-corrected versions using Simulation–Extrapolation (NW-SIMEX) and Regression Calibration (NW-RC). The study was designed to investigate how sample size and measurement error affect the estimation accuracy of these methods under different regression settings. Three sample sizes were considered: $n = 100$, $n = 200$, and $n = 500$. Three measurement error levels were examined: 0.1, 0.3, and 0.5. The Gaussian kernel was used for all kernel-based estimators to ensure consistency in the comparison. For each combination of experimental conditions, 500 Monte Carlo replications were performed. The performance of the estimators was assessed using bias, mean squared error (MSE), and mean absolute error (MAE) to evaluate estimation accuracy and the extent to which each method mitigates measurement error. The experimental factors and their corresponding levels are summarized in Table 1.

Performance Comparison of SIMEX and Regression Calibration for Measurement Error Correction in Nonparametric Regression

Table1: The Examples some Regression Functions used simulation study

Simulation Factor	Levels/ Values
Sample size (n)	100, 200, 500
True Covariate x	$x \approx N(0,1)$
Measurement Error Variance σ_u^2	0.1,0.3,0.5
Error Model Type	Classical Additive Model: $W = X + U$, $u \approx N(0, \sigma_u^2)$
True Regression Function $m(x)$	$m(x) = \sin(\pi x) + 0.5x$
Model Noise ε	$\varepsilon \approx N(0, 0.25)$
Estimation Methods	Uncorrected NW, NW-SIMEX, NW-RC
Kernel function	Gaussian kernel
Bandwidth Selection (\$h\$)	Cross-Validation (CV) / Adaptive CV
Monte Carlo replications	500 per cell
Performance Measures	MSE, MAE, Integrated Bias

2.1 Numerical Summary of the Simulation Results

The numerical results of the simulation study are summarized in Tables 2–4 for different sample sizes ($n = 100, 200, 500$) and measurement error levels 0.1, 0.3, and 0.5. The performance of the Nadaraya–Watson (NW) estimator and its corrected versions, NW-SIMEX and NW-RC, was evaluated using Bias, Mean Squared Error (MSE), and Mean Absolute Error (MAE). Lower values of MSE and MAE indicate superior accuracy, whereas Bias values closer to zero reflect minimal systematic estimation error. This simulation design enables a comparative assessment of the conventional NW estimator and the two correction approaches across varying sample sizes and measurement error conditions.

Table 2: Mean Squared Error (MSE) comparison for Naive NW, NW-SIMEX, and NW-RC estimators across sample sizes and measurement error variances

	n	σ_u	NW	NW- SIMEX	NW- RC
MSE	100	0.1	0.3043983	0.2493202	0.2766506
		0.3	0.4147756	0.3975704	0.3720363
		0.5	0.4598613	0.4712974	0.4233402
	200	0.1	0.2657454	0.2019226	0.2374981
		0.3	0.3979477	0.3689312	0.3468815
		0.5	0.4527376	0.4487917	0.4135301
	500	0.1	0.2133797	0.1363458	0.1816858
		0.3	0.3639640	0.3167369	0.3115065
		0.5	0.4308797	0.4291027	0.3825567

Table 3: Mean Absolute Error (MAE) comparison for Naive NW, NW-SIMEX, and NW-RC estimators across sample sizes and measurement error variances

Performance Comparison of SIMEX and Regression Calibration for Measurement Error Correction in Nonparametric Regression

	n	σ_u	NW	NW- SIMEX	NW- RC
MAE	100	0.1	0.4918887	0.4378543	0.4670834
		0.3	0.5742939	0.5531027	0.5409443
		0.5	0.6008520	0.6024326	0.5746172
	200	0.1	0.4630921	0.3983907	0.4364489
		0.3	0.5658981	0.5390888	0.5264235
		0.5	0.6000976	0.5951722	0.5742707
	500	0.1	0.4173945	0.3295282	0.3836742
		0.3	0.5438031	0.5025024	0.5012744
		0.5	0.5888821	0.5869247	0.5552251

Table 4: Bias comparison for Naive NW, NW-SIMEX, and NW-RC estimators across sample sizes and measurement error variances

	n	σ_u	NW	NW- SIMEX	NW- RC
Bias	100	0.1	0.0187494138	0.0216473097	0.01883330
		0.3	0.0020498445	0.0021915957	0.00141259
		0.5	0.0020761483	0.0094241256	0.00265094
	200	0.1	0.00051174-	0.0011714235	0.00053819
		0.3	0.00370058-	0.00130020-	0.0036218-
		0.5	0.01014680-	0.00546610-	0.0035279-
	500	0.1	0.0040427351	0.0030693898	0.00361878
		0.3	0.0040187931	0.0034209951	0.00137751
		0.5	0.0056071685	0.0094623100	0.00217561

Discussion of Simulation Results

Discussion of Table 2: Mean Squared Error (MSE)

The MSE results presented in Table 2 indicate that the relative performance of the estimators depends on the measurement error variance and sample size. Under mild measurement error $\sigma^2=0.1$, NW-SIMEX consistently achieves the lowest MSE across all sample sizes, with a reduction of approximately 36.1% relative to the naive NW estimator at $n=500$. In contrast, under moderate and high measurement error variances $\sigma^2=0.3$ and 0.5 , NW-RC achieves the lowest MSE across all considered sample sizes. These findings suggest that the effectiveness of measurement error correction procedures varies with the magnitude of measurement error. Although NW-SIMEX performs particularly well under mild measurement error, its MSE is slightly higher than that of the naive NW estimator when $n=100$ and $\sigma^2=0.5$. This behavior may reflect increased variability associated with the SIMEX extrapolation procedure under substantial measurement error. Overall, the results indicate that NW-SIMEX is more effective under mild measurement error, whereas NW-RC provides better estimation accuracy under moderate and high measurement error in the scenarios considered.

Discussion of Table 3: Mean Absolute Error (MAE)

The MAE results reported in Table 3 broadly confirm the patterns observed in the MSE analysis. Under mild measurement error $\sigma^2=0.1$, NW-SIMEX achieves the lowest MAE across all sample sizes, with an improvement of approximately 21.05% relative to the naive NW estimator at $n=500$. For moderate and high measurement error variances $\sigma^2=0.3$ and 0.5 , NW-RC consistently yields the lowest MAE across the considered sample sizes. At $n=500$ and $\sigma^2=0.3$, the MAE of NW-RC is approximately 7.82% lower than that of the naive NW estimator.

The agreement between the MSE and MAE rankings provides complementary evidence of the differences in estimation accuracy among the three methods. Specifically, NW-SIMEX performs best under mild measurement error, whereas NW-RC achieves lower estimation errors under moderate and high measurement error. These results suggest that the choice of correction procedure should account for the magnitude of measurement error rather than assume the general superiority of a single estimator.

Discussion of Table 4: Bias

The bias results presented in Table 4 show that the signed average bias across the evaluation grid remains numerically small, with absolute values below 0.022 in all considered scenarios. However, these values should be interpreted cautiously because positive and negative local biases may offset one another when averaged across the evaluation grid. Therefore, a small signed average bias does not necessarily indicate uniformly small bias across the entire range of the predictor. The relative performance of the estimators also varies according to sample size and measurement error variance. For example, at $n=200$ and $\sigma^2=0.5$, NW-RC has the smallest absolute signed average bias among the three estimators. Similarly, at $n=500$ and $\sigma^2=0.3$, NW-RC produces the smallest absolute value. Nevertheless, the bias rankings do not always coincide with those obtained from MSE and MAE, since these measures evaluate different aspects of estimator performance. Taken together, the results indicate that the relative effectiveness of NW, NW-SIMEX, and NW-RC depends on the performance criterion and simulation conditions. Further examination of point wise bias or integrated absolute bias could provide additional insight into the behavior of the estimators across the evaluation grid.

Conclusion

Overall, the simulation results demonstrate that the relative performance of the three estimators depends on the measurement error variance and the evaluation criterion. NW-SIMEX achieves the lowest MSE and MAE under mild measurement error, whereas NW-RC performs best under moderate and high measurement error across the sample sizes considered. The bias results provide a complementary perspective but should not be interpreted in isolation, because small signed average bias may conceal local estimation errors. The consistent rankings observed for MSE and MAE strengthen the evidence that the two correction procedures have different advantages under different measurement error conditions. Accordingly, NW-SIMEX appears more suitable for the mild-error scenarios examined, while NW-RC is preferable in terms of MSE and MAE under the moderate- and high-error scenarios. These conclusions are specific to the simulation settings considered and do not establish universal superiority or statistical significance without further inferential analysis.

References

- Carroll, R. J., Delaigle, A., & Hall, P. (2009). Nonparametric prediction in measurement error models. *Journal of the American Statistical Association*, 104(487), 993–1003. <https://doi.org/10.1198/jasa.2009.tm07543>
- Carroll, R. J., Maca, J. D., & Ruppert, D. (1999). Nonparametric regression in the presence of measurement error. *Biometrika*, 86(3), 541–554. <https://doi.org/10.1093/biomet/86.3.541>
- Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: A modern perspective* (2nd ed.). Chapman & Hall/CRC.

- Cook, J. R., & Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, 89(428), 1314–1328. <https://doi.org/10.1080/01621459.1994.10476871>
- Fuller, W. A. (1987). *Measurement error models*. Wiley.
- Keogh, R. H., Shaw, P. A., Gustafson, P., Carroll, R. J., Deffner, V., Dodd, K. W., Küchenhoff, H., Tooze, J. A., Wallace, M. P., Kipnis, V., & Freedman, L. S. (2020). STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 1—Basic theory and simple methods of adjustment. *Statistics in Medicine*, 39(16), 2197–2231. <https://doi.org/10.1002/sim.8532>
- Sevilimedu, V., & Yu, L. (2022). Simulation extrapolation method for measurement error: A review. *Statistical Methods in Medical Research*, 31(8), 1617–1636. <https://doi.org/10.1177/0962280221102619>