Annals of the Institute of Statistical Mathematics
https://doi.org/10.1007/s10463-026-00984-x

Debiased inference in errors-in-variables problems with non-Gaussian measurement error

Nicholas Woolsey[1] · **Xianzheng Huang**[1]

Received: 2 August 2025 / Revised: 27 January 2026 / Accepted: 6 April 2026

Abstract
We consider drawing statistical inferences based on data subject to non-Gaussian measurement error. The proposed strategy exploits hypercomplex numbers to reduce bias in naive estimation that ignores non-Gaussian measurement error. We apply this new method to several widely applicable parametric regression models with error-prone covariates, and kernel density estimation using error-contaminated data. The efficacy of this method in bias reduction is demonstrated in simulation studies and a real-life application in sports analytics.

Keywords Corrected score · Hypercomplex number · M-Estimation · Tessarine

1 Introduction

Data subject to measurement error due to imprecise measuring devices or human error are ubiquitous in many applications. There is a rich collection of literature on the study of effects of measurement error on statistical inferences and strategies for addressing complications caused by measurement error. Fuller (2009); Carroll et al., (2006); Buonaccorsi (2010); Yi (2017) and Yi et al., (2021), among several other books on this subject, provide comprehensive reviews of errors-in-variables problems. The history of research on these problems has mainly focused on Gaussian measurement errors. Several landmark methods, including the corrected score method (Nakamura, 1990; Novick and Stefanski, 2002), the conditional score method (Stefanski and Carroll, 1987), and the method of simulation-and-extrapolation (Stefanski and Cook, 1995)[SIMEX], were all developed under the assumption of Gaussian measurement error. Many recent adaptations of these methods to modern statistical learning keep

✉ Nicholas Woolsey
nwoolsey@email.sc.edu

Xianzheng Huang
huang@stat.sc.edu

[1] Department of Statistics, University of South Carolina, Leconte College, 1523 Greene Street, Suite 219, Columbia, SC 29225, USA

Published online: 18 June 2026

the Gaussian assumptions on measurement errors. For example, Chen (2023) incorporated SIMEX in a boosting algorithm to perform variable selection in generalized linear models with continuous covariates contaminated by Gaussian error, and also applied the regression calibration strategy to construct the gradient function used in the boosting algorithm. Zhao and Wang (2024) developed debiased regression calibration in the presence of high-dimensional error-prone compositional covariates, where they assumed additive Gaussian error contaminating the log-contrast compositional data. More recently, Chen (2025) derived a corrected score used in ridge regression with binary, count, and continuous covariates, all prone to measurement error, where the errors contaminating the continuous covariates are assumed to be Gaussian.

However, non-Gaussian measurement errors, such as skewed or heavy-tailed errors, are common in fields including finance and actuarial science (Adcock et al., 2015), political science (Millimet and Parmeter, 2022), and environmental science (Torabi et al., 2023). There have been some attempts to account for non-Gaussian measurement errors in limited model settings, primarily in linear regression. Most existing methods along this theme assume some specific type of error such as skew-normal (Arellano-Valle et al., 2005), or assume a known distribution for the unobserved error-prone variable (Tomaya and de Castro, 2018). To impose milder distributional assumptions, some authors brought in validation or external data, based on which one infers the error distribution or models the relationship between the error-prone variable and its surrogate. For example, Cheng et al., (2023) developed mediation analysis when the continuous exposure is mismeasured, where they used validation data to infer posited models of the true exposure given its error-contaminated surrogate. Faller et al., (2025) adopted a neural network to model an unobserved image given its noisy version using validation data to increase the prediction performance of Bayesian deep errors-in-variables models. Without requiring validation data, external data, or replicate data, Delaigle and Hall (2016) used the phase function to overcome the identifiability issues in measurement error problems under the assumption that the measurement error distribution is symmetric around zero and the true covariate distribution is asymmetric. Under similar assumptions, Nghiem et al., (2020) proposed regression coefficients estimation in errors-in-covariates linear regression based on the phase function.

In this study, we propose a new strategy exploiting hypercomplex numbers to account for measurement errors not limited to Gaussian errors in a wide range of inference problems. The idea of utilizing hypercomplex quantities to effectively mitigate the adverse effects of measurement error is developed in Sect. 2. Section 3 addresses some unique issues in implementing the proposed method involving hypercomplex numbers. We apply the new method in parametric regression settings under the framework of M-estimation in Sect. 4, and then apply it to the problem of nonparametric density estimation in Sect. 5. Empirical evidence for the method's efficacy in comparison with several competing methods is presented in Sect. 6 using synthetic data. Section 7 presents a real-life data application of these considered methods. We conclude the paper with a brief discussion on the future research agenda in Sect. 8.

2 Methodology

2.1 Correcting for Gaussian measurement error using complex numbers

Let X be a random variable of interest that cannot be observed due to error contamination in a study. Suppose that the error-contaminated surrogate is W, relating to X via

$$W = X + U, \tag{1}$$

where U is the measurement error supported on the real line $\mathbb{R}$ independent of X and other random quantities in the study, such as the response Y and the random error in a regression model. In other words, U is an additive nondifferential measurement error (Carroll et al., 2006, Section2.5). A popular recipe for developing methods to draw inferences based on error-contaminated data consists of two main steps. Firstly, one adopts an inference method had X been observed. This usually amounts to choosing a statistic such as an estimator for a target parameter, a score function used in M-estimation (Stefanski and Boos, 2002), or an objective function such as a log-likelihood function or a loss function in the absence of measurement error. Let $f(X)$ generically refer to such a choice evaluated at one data point, with its potential dependence on parameters and other random variables suppressed. Secondly, acknowledging that X is not observed but W is, one seeks to "recover" $f(X)$ by constructing a sensible estimator for it based on W or replicate error-contaminated measures of X. Because $f(W)$ typically does not serve this purpose but is a naive and often biased estimator for $f(X)$, the second step sparks many proposals for estimating $f(X)$. These proposals, as well as ours, focus on functional measurement error modeling where no distributional assumption is imposed on X, unlike structural modeling that includes an assumed distribution of X (Carroll et al., 2006, Section2.1). This prolific two-step recipe has led to kernel density estimators based on data subject to measurement error (Huang and Zhou, 2020; Stefanski and Carroll, 1990), local polynomial regression with errors-in-covariates (Delaigle et al., 2009; Zhou and Huang, 2016), the corrected score method applied to various (semi-)parametric models (Nakamura, 1990; Novick and Stefanski, 2002; Song and Huang, 2005; Stefanski, 1989; Wang, 2006; Wu et al., 2015), and the corrected loss/utility method in different regression settings with error-prone covariates (Augustin, 2004; Li and Huang, 2019; Liu and Huang, 2024; Wang et al., 2012).

Along this line of development, Stefanski (1989) introduced complex errors to address the problem in the second step, where it is assumed that (i) $f(\cdot)$ is a known entire function, and (ii) the additive nondifferential error $U \sim N(0, \sigma_u^2)$, with the variance σ_u^2 known. The author then phrased the problem of estimating $f(X)$ based on W as estimating an entire function of a normal mean. The solution to the problem is inspired by the intriguing result that we reiterate in Theorem 1 for its high relevance in our methodological development.

Theorem 1 *Under the following two conditions: (i) $f(\cdot)$ is an entire function over the complex plane; (ii) $U \sim N(0, \sigma_u^2)$ in (1), one has $E\{f(X + U + iV)|X\} = f(X)$,*

where $i = (-1)^{1/2}$ *is the imaginary unit, and* $V \sim N(0, \sigma_u^2)$*, which is independent of* U *and* X*.*

The expectation of a complex random variable is a complex number of which the real and imaginary parts are equal to the expectations of the real and imaginary parts of the random variable, respectively. According to Theorem 1, under conditions (i) and (ii), $f(W + iV)$ is an unbiased estimator for $f(X)$. Essentially, by adding the complex random noise iV to W, $f(W+iV)$ corrects the naive estimator $f(W)$ for measurement error, as if it removed measurement error in the naive estimator to recover $f(X)$. This correction stems from the fact that $U + iV$ is a circularly symmetric random variable, which belongs to a type of random variables widely studied in the field of signal processing (Adali, Schreier and Scharf, 2011; Picinbono, 1994; Schreier and Scharf, 2010).

A complex random variable C is circularly symmetric if $e^{it}C$ and C follow the same distribution for any real constant t. This definition leads to Lemma 1, proved in Appendix A .1.

Lemma 1 *If a complex random variable* C *is circularly symmetric and has moments of all orders, then* $E(C^\ell) = 0$*, for any* ℓ *in the set of natural numbers* $\mathbb{N}$*.*

Now consider $C = U + iV$ introduced in Theorem 1. By construction, C follows a central complex Gaussian distribution with the probability density function $p_C(c) = (2\pi\sigma_u^2)^{-1} \exp\{-\bar{c}c/(2\sigma_u^2)\}$, where $\bar{c}$ is the conjugate of c (Goodman, 1963). Because $p_C(c)$ depends on c only via its magnitude $|c| = (\bar{c}c)^{1/2}$, rotating C does not change the distribution as the magnitude is invariant under rotation. Hence, $e^{it}C$, as a rotation transformation of C, is identically distributed as C for any $t \in \mathbb{R}$, and thus the so-constructed C is circularly symmetric. Returning to the unbiased estimator of $f(X)$ revealed in Theorem 1, we have, using a Taylor expansion around X,

$$\begin{aligned} & f(W + iV) \\ &= f(X + C) = f(X) + \sum_{\ell=1}^{\infty} \frac{f^{(\ell)}(X)}{\ell!} C^\ell, \end{aligned} \tag{2}$$

under the entirety assumption (i) for $f(\cdot)$ that guarantees existence of derivatives of all orders, $\{f^{(\ell)}(\cdot),\ \ell \in \mathbb{N}\}$. Taking the expectation of (2) conditional on X gives $E\{f(X + C)|X\} = f(X)$ since a complex Gaussian possesses moments of all orders and, by Lemma 1, $E(C^\ell) = 0$, for $\ell \in \mathbb{N}$. This concludes Theorem 1.

Denote by $\Re(c)$ the real part of a complex number c. Because, by Theorem 1, the imaginary part of $f(W + iV)$ has a mean of zero, one may use $\Re\{f(W + iV)\}$ as the final estimator for $f(X)$. The Monte Carlo corrected score method (Novick and Stefanski, 2002) amounts to using the Monte Carlo average, $\hat{f}_{\mathrm{mc}}(X) = \sum_{m=1}^{M} \Re\{f(W + iV_m)\}/M$, as an estimator of $f(X)$ based on a random sample $(V_1, \ldots, V_M)$ from $N(0, \sigma_u^2)$.

2.2 Correcting for non-Gaussian measurement error using complex numbers

Keeping assumption (i), if we relax assumption (ii) by allowing non-Gaussian U, then terms in the infinite sum in (2) may not reduce to zero for all $\ell \in \mathbb{N}$ when taking expectations. This is because it is practically infeasible to construct an independent complex random variable that adds to the non-Gaussian U to produce a circularly symmetric C with all moments existing as accomplished above. Indeed, a circularly symmetric distribution that is not complex Gaussian may not even possess a density (Ollila et al., 2012), hinting at the infeasibility of generating a circularly symmetric C with all moments well-defined when U is non-Gaussian. With circular symmetry practically out of reach in this case, a more fruitful direction to pursue is to formulate C so that some terms of the Taylor series in (2) have mean zero, in the hope of reducing, although not fully eliminating, the bias of $f(W)$. As a starter, we strive to have the first two terms of the infinite sum in (2) reduce to zero in expectation when U follows some distribution with variance σ_u^2 but not necessarily mean-zero or Gaussian. To this end, we turn to *proper* complex random variables, a notion first introduced in Neeser and Massey (1993).

A complex random variable C is said to be proper if its pseudovariance vanishes (Neeser and Massey 1993, Definition1), where the pseudovariance is defined as $\text{pvar}(C) = E[\{C - E(C)\}^2]$, in contrast to the variance $\text{var}(C) = E\{|C - E(C)|^2\}$. We establish Theorem 2 in Appendix A .2 in order to provide a way to convert a generic real U to a proper complex random variable.

Theorem 2 *If the first two moments of U exist, then $C = U - E(U) + i\{V - E(V)\}$ is proper, where V is independent of U and follows the same distribution as U.*

Clearly, the proper C introduced in Theorem 2 includes $U + iV$ appearing in Theorem 1 as a special case under condition (ii) stating that $U \sim N(0, \sigma_u^2)$. By Theorem 2, if we use in (2) $C = U - E(U) + i\{V - E(V)\}$ that follows some central complex distribution not necessarily complex Gaussian, then the first two terms of the infinite sum in (2) vanish when taking the expectation. This suggests that $f(X + C) = f(W - E(U) + i\{V - E(V)\})$, or keeping its real part, potentially gives a debiased estimator for $f(X)$ that improves upon $f(W)$. Define $\hat{f}_{\text{pc}}(X) = \Re\{f(W - E(U) - i\{V - E(V)\})\}$ as this estimator involving a proper complex random variable.

A formal justification of the debiasing effect of $\hat{f}_{\text{pc}}(X)$ requires a comparison between the bias of this estimator (conditional on X),

$$\text{Bias}\{\hat{f}_{\text{pc}}(X)\} = \sum_{\ell=3}^{\infty} \frac{f^{(\ell)}(X)}{\ell!} \Re\{E(C^\ell)\}, \tag{3}$$

and the bias of $f(W) = f(X + U) = f(X) + \sum_{\ell=1}^{\infty}\{f^{(\ell)}(X)/\ell!\}U^\ell$, which is equal to

$$\text{Bias}\{f(W)\} = \sum_{\ell=1}^{\infty} \frac{f^{(\ell)}(X)}{\ell!} E(U^\ell). \tag{4}$$

Even with $E(U) = 0$ so that the infinite sum in (4) can start at $\ell = 2$, $\hat{f}_{\text{pc}}(X)$ may still achieve some bias reduction because the infinite sum in (3) begins with $\ell = 3$, and, when the ℓ-th moment is nonzero for $\ell > 2$, $\Re\{E(C^\ell)\}$ is comparable with $E(U^\ell)$ because the highest-order moment of U that $\Re\{E(C^\ell)\}$ linearly depends on is $E(U^\ell)$.

Similar to the Monte Carlo corrected score estimator $\hat{f}_{\text{mc}}(X)$, one may use a Monte Carlo average, $\hat{f}_{\text{pc}}(X) = \sum_{m=1}^{M} \Re\{f(W - E(U) + i\{V_m - E(V)\})\}/M$, to estimate $f(X)$ provided that one can generate $\{V_m\}_{m=1}^{M}$ from the measurement error distribution. On the other hand, (3) indicates that the propriety requirement on C can be relaxed because one only needs $\Re\{E(C^\ell)\} = 0$ instead of $E(C^\ell) = 0$, for $\ell = 1, 2$, when a real-valued estimator for $f(X)$ is desired. In fact, setting $C = U - E(U) + i\sigma_u$ suffices even though it is not a proper complex random variable. This gives another estimator for $f(X)$ that depends on a complex constant besides W,

$$\hat{f}_{\text{cc}}(X) = \Re\{f(W - E(U) + i\sigma_u)\}, \tag{5}$$

with the same bias as $\hat{f}_{\text{pc}}(X)$ but freeing one from generating $\{V_m\}_{m=1}^{M}$, which in turn lowers the variability in estimation. To estimate $f(X)$ via (5) only requires the first two moments of U instead of its distribution as for $\hat{f}_{\text{mc}}(X)$ and $\hat{f}_{\text{pc}}(X)$. Certainly, in an application, one typically has to estimate these moments based on replicate measures of X or some form of external or validation data. Even in these practical settings, estimating the first two moments is much easier and subject to less uncertainty than estimating the (unspecified) distribution of U.

2.3 Correcting for non-Gaussian measurement error using tessarines

Continuing on the course pursued in Sect. 2.2, one may wonder if it is possible to construct C in (2) with vanishing $\Re\{E(C^\ell)\}$ also for ℓ above 2 to improve over $\hat{f}_{\text{cc}}(X)$. As an extension of complex numbers that we utilize to achieve $\Re\{E(C)\} = \Re\{E(C^2)\} = 0$, a natural place to look for possible solutions is the tessarine number system (Cockle, 1848). A tessarine $\mathcal{S}$ is formulated as $\mathcal{S} = a + bi + cj + dk$, where $a, b, c, d \in \mathbb{R}$, and i, j, k are the imaginary units that satisfy $i^2 = -j^2 = k^2 = -1$, $ij = ji = k$, $jk = kj = i$, and $ki = ik = -j$. These rules suggest that tessarine multiplication is commutative and that tessarines have zero divisors, that is, the product of two non-zero tessarines can equal zero.

For a tessarine random variable $C = U + a + bi + cj + dk$, we now have, following simple tessarine arithmetic (Luna-Elizarrarás et al., 2015),

$$\Re\{E(C^2)\} = E\{(U + a)^2\} - (b^2 - c^2 + d^2), \tag{6}$$

$$\Re\{E(C^3)\} = E\{(U + a)^3\} - 3(b^2 - c^2 + d^2)E(U + a) - 6bcd, \tag{7}$$

$$\begin{aligned}\Re\{E(C^4)\} = {} & E\{(U + a)^4\} - 6(b^2 - c^2 + d^2)E\{(U + a)^2\} - 24bcdE(U + a) \\ & + (b^2 - c^2 + d^2)^2 - 4\{b^2(c^2 - d^2) + c^2d^2\}.\end{aligned} \tag{8}$$

Clearly, setting $a = -E(U)$ makes $\Re\{E(C)\} = 0$. Plugging in (6)–(8) this choice of a and setting them to zeros reveal that (b, c, d) solves the following system of

equations,

$$\begin{aligned} \sigma_u^2 - b^2 + c^2 - d^2 &= 0, \\ \mu_3 - 6bcd &= 0, \\ \mu_4 - 5\sigma_u^4 - 4\{b^2(c^2 - d^2) + c^2 d^2\} &= 0, \end{aligned} \tag{9}$$

where μ_ℓ is the ℓ-th central moment of U, for $\ell \in \mathbb{N}$. And the estimator that involves the corresponding tessarine constant is

$$\hat{f}_{\text{tc}}(X) = \Re\{f(W - E(U) + bi + cj + dk)\}. \tag{10}$$

Setting $b = \sigma_u$ and $c = d = 0$ in (10) makes $\hat{f}_{\text{tc}}(X)$ reduce to $\hat{f}_{\text{cc}}(X)$ in (5). If $f^{(3)}(t) = 0$, for all $t \in \mathbb{R}$, then $\hat{f}_{\text{tc}}(X)$ has the same bias as that of $\hat{f}_{\text{cc}}(X)$ because the summand in the Taylor series with $\ell \geq 3$ in (3) vanish despite the values of $\Re\{E(C^\ell)\}$. As long as $f^{(3)}(t) \neq 0$ for some $t \in \mathbb{R}$, $\hat{f}_{\text{tc}}(X)$ can potentially improve over $\hat{f}_{\text{cc}}(X)$. Computing $\hat{f}_{\text{tc}}(X)$ requires evaluating $f(\cdot)$ at a tessarine and finding (b, c, d) that satisfies (9), which we elaborate next.

3 Implementation

3.1 Computing tessarine functions

Deriving $\Re\{E(C^\ell)\}$ when C is a tessarine in Sect. 2.3 is straightforward because computing the polynomial of a tessarine only involves the multiplication rules of tessarine imaginary units and simple tessarine arithmetic. It is however less obvious how to compute, for instance, $\exp(C)$. As being done for complex numbers and quaternions (Jacobson, 2012; Zhang, 1997), it is convenient to use a 2×2 matrix with complex-valued entries to represent a tessarine so that every operation that can be done on tessarines can translate to an operation on matrices. This alternative representation of a tessarine is based on the matrix representation of the tessarine basis elements, 1, i, j, and k, given by $\mathbb{1} = \boldsymbol{I}_2$, $\mathbb{I} = i\boldsymbol{I}_2$, $\mathbb{J} = \boldsymbol{J}_2$, and $\mathbb{K} = i\boldsymbol{J}_2$, respectively, where $\boldsymbol{I}_2$ is the 2×2 identify matrix, $\boldsymbol{J}_2$ results from swapping the columns of $\boldsymbol{I}_2$, and $i = (-1)^{1/2}$. One can easily verify that the basis elements in the matrix form, $\mathbb{1}$, $\mathbb{I}$, $\mathbb{J}$, and $\mathbb{K}$, satisfy the multiplication rules for the original basis elements, that is, $\mathbb{I}^2 = -\mathbb{J}^2 = \mathbb{K}^2 = -\mathbb{1}$, $\mathbb{I}\mathbb{J} = \mathbb{J}\mathbb{I} = \mathbb{K}$, $\mathbb{J}\mathbb{K} = \mathbb{K}\mathbb{J} = \mathbb{I}$, and $\mathbb{K}\mathbb{I} = \mathbb{I}\mathbb{K} = -\mathbb{J}$. Despite the noncommutativity of matrix multiplication in general, the commutativity of tessarine multiplication is preserved under the matrix representation of tessarines. In the sequel, we view $C = a + bi + cj + dk$ equivalent to the 2×2 complex matrix $C = a\mathbb{1} + b\mathbb{I} + c\mathbb{J} + d\mathbb{K}$, with $\Re(C) = a$ appearing in (the real parts of) the diagonal entries. With this matrix representation of C, the reciprocal of C is equivalent to the inverse of the corresponding matrix, and the product of tessarines is reflected as the product of matrices. The reason for the presence of zero divisors in the tessarine system also becomes clear since zero divisors for matrices are a common occurrence.

As an application of tessarine multiplication, the following lemma relates the tessarine power function to matrix power.

Lemma 2 *For a tessarine*

$$C = a\mathbb{1} + b\mathbb{I} + c\mathbb{J} + d\mathbb{K} = (a+bi)\boldsymbol{I}_2 + (c+di)\boldsymbol{J}_2 = \begin{bmatrix} a+bi & c+di \\ c+di & a+bi \end{bmatrix},$$

the power function evaluated at C results in another tessarine C^ℓ equal to, for $\ell \in \mathbb{N}$,

$$\frac{1}{2}\Big(\Big[\{a-c+i(b-d)\}^\ell + \{a+c+i(b+d)\}^\ell\Big]\boldsymbol{I}_2 + \Big[\{a+c+i(b+d)\}^\ell - \{a-c+i(b-d)\}^\ell\Big]\boldsymbol{J}_2\Big);$$

and consequently,

$$\begin{aligned}\Re(C^\ell) = \frac{1}{2}\Bigg[&\Big\{(a-c)^2+(b-d)^2\Big\}^{\ell/2}\cos(\ell \times arg\{a-c+i(b-d)\}) \\ &+ \Big\{(a+c)^2+(b+d)^2\Big\}^{\ell/2}\cos(\ell \times arg\{a+c+i(b+d)\})\Bigg],\end{aligned} \tag{11}$$

where $arg(t)$ is the argument of a complex number t.

The proof of Lemma 2 is provided in Appendix A .3, starting with matrix diagonalization for C. Using Lemma 2, we prove Theorem 3 in Appendix A .4, where we use the matrix exponential to derive the tessarine exponential.

Theorem 3 *For a tessarine $C = a + bi + cj + dk$,*

$$\exp(C) = \exp(a+bi)\{\cosh(c+di)\boldsymbol{I}_2 + \sinh(c+di)\boldsymbol{J}_2\},$$

where $\sinh(t) = (e^t - e^{-t})/2$ *and* $\cosh(t) = (e^t + e^{-t})/2$ *for a complex t. Thus* $\Re\{\exp(C)\} = \exp(a)\{\cos(b)\cos(d)\cosh(c) - \sin(b)\sin(d)\sinh(c)\}$.

Using Theorem 3, one can evaluate any function at a tessarine that can be expressed via exponential functions, such as trigonometric functions. If $f(\cdot)$ is not any of the polynomials, exponential, and trigonometric functions, one may resort to the Taylor series to evaluate the estimator in (10) under the entirety assumption for $f(\cdot)$, that is, $\hat{f}_{\text{tc}}(X) = \sum_{\ell=0}^{\infty}\{f^{(\ell)}(W - E(U))/\ell!\}\Re\{(bi+cj+dk)^\ell\}$, and one can use (11) with $a = 0$ to compute the summand.

3.2 Finding the tessarine

Following the resultant procedure (Bareiss, 1959), one can show that (9) is equivalent to

$$\mu_3^2 + 9(5\sigma_u^4 - \mu_4)d^2 - 36\sigma_u^2 d^4 + 36d^6 = 0, \tag{12}$$

$$\mu_3^2 + 36d^2(d^2 - \sigma_u^2)c^2 - 36d^2c^4 = 0, \tag{13}$$

$$b^2 - c^2 + d^2 - \sigma_u^2 = 0. \tag{14}$$

One can find d, c, and b sequentially, first solving (12) for d, then solving (13) for c, and lastly solving (14) for b, each step using a simple root-finding algorithm. This provides a simple algorithm for finding a real-valued solution to (9) when real solutions exist.

If there is no real solution to (9), then there is no tessarine of the form $C = U - E(U) + bi + cj + dk$ satisfying $\Re\{E(C^2)\} = \Re\{E(C^3)\} = \Re\{E(C^4)\} = 0$. To reduce the bias in (3), we now aim to find $(b, c, d) \in \mathbb{R}^3$ that minimizes these real moments in combination by minimizing the following objective function,

$$Q(b, c, d) = \left\{\frac{f_1(b, c, d)}{\sigma_u^2}\right\}^2 + \left\{\frac{f_2(b, c, d)}{3\mu_3}\right\}^2 + \left\{\frac{f_3(b, c, d)}{12\mu_4}\right\}^2, \tag{15}$$

where $f_1(b, c, d)$ is the difference between the two sides of (6) with a set at $-E(U)$, $f_2(b, c, d)$ and $f_3(b, c, d)$ are similarly defined based on (7) and (8), respectively, and the weights in (15) (proportional to $\ell!$ for $\ell = 2, 3, 4$) are motivated by the Taylor series in (3). One may employ the gradient descent algorithm to find a minimizer of (15), using multiple starting points to avoid being trapped at a saddle point or local minimum.

The search for a desirable tessarine $C = U - E(U) + bi + cj + dk$, either via solving (9) for (b, c, d) or by minimizing (15) with respect to (b, c, d), can yield multiple choices of C due to, say, the existence of multiple solutions to (9). These choices are equally viable in terms of producing (nearly) vanishing first four moments of C (in its real part).

4 Applying to M-estimation in regression analysis

4.1 Overview

Consider the generic problem of M-estimation based on data from n independent experimental units, $\{(Y_r, X_r)\}_{r=1}^n$, that one uses to fit a regression model with Θ as the parameter(s) of interest. A consistent M-estimator for Θ solves the estimating equation, $\sum_{r=1}^n \Psi(Y_r, X_r, \Theta) = 0$, where $\Psi(\cdot)$ is a score function satisfying $E\{\Psi(Y, X, \Theta^*)\} = 0$, in which Θ^* is the true value of Θ (Stefanski and Boos, 2002), and we suppress the index r in Y and X (and other random quantities later) when we refer to the data of an experimental unit generically. In the presence of measure-

ment error when $\{W_r\}_{r=1}^n$ are observed instead of $\{X_r\}_{r=1}^n$, the corrected score method (Nakamura, 1990) suggests using a different score function that is an unbiased estimator of $\Psi(Y, X, \Theta)$ given (Y, X) in the estimating equation. Using the notations in earlier sections, $f(X) = \Psi(Y, X, \Theta)$, and its unbiased estimator is referred to as a corrected score.

Finding a corrected score usually has to be done on a case-by-case basis, which is a daunting task in most regression settings, and it may not be possible to find one explicitly. Without assuming Gaussian measurement error in (1), we propose to estimate $\Psi(Y, X, \Theta)$ via

$$\tilde{\Psi}(Y, W, \Theta) = \Re\{\Psi(Y, W - E(U) + \Im, \Theta)\}, \tag{16}$$

where $\Im$ is the imaginary portion of a (hyper)complex number considered in Sect. 2. More specifically, setting $\Im = i\sigma_u$ as in (5) gives a score as an application of $\hat{f}_{\text{cc}}(X)$, and having $\Im = bi + cj + dk$ leads to another score in the spirit of $\hat{f}_{\text{tc}}(X)$, with b, c, d obtained as described in Sect. 3.2 based on assumed or estimated first four moments of U. The goal here is to abort the often impossible mission of analytically finding an unbiased estimator for the true-data score $\Psi(Y, X, \Theta)$, and instead to find a score that improves over the naive observed-data score $\Psi(Y, W, \Theta)$. We call a score function in the form of (16) a debiased score in contrast with a corrected score. We zoom in on four regression models to illustrate this new score next.

4.2 Linear regression

Consider a linear regression model with error-in-covariates,

$$Y_r = \beta_0 + X_r^\top \beta + \epsilon_r, \qquad W_r = X_r + U_r, \quad (r = 1, \ldots, n),$$

where $X_r^\top$ is the transpose of the $m \times 1$ covariate vector X_r, $\Theta = (\beta_0, \beta^\top)^\top \in \mathbb{R}^{m+1}$ contains the parameters of interest, ϵ_r is a mean-zero error, and the additive measurement error model generalizes (1) by allowing a multivariate covariate supported on $\mathbb{R}^m$, and also the m-dimensional measurement error U_r with the variance-covariance matrix denoted by Ω. In the absence of measurement error, the least square estimator of Θ is the solution to the estimating equation $\sum_{r=1}^n (Y_r - \beta_0 - X_r^\top \beta)(1, X_r^\top)^\top = 0_{m+1}$, where 0_{m+1} is the $(m + 1) \times 1$ vector of zeros. That is, the true-data score is $\Psi(Y, X, \Theta) = (Y - \beta_0 - X^\top \beta)(1, X^\top)^\top$.

In the presence of measurement error, the naive score, $\Psi(Y, W, \Theta) = (Y - \beta_0 - W^\top \beta)(1, W^\top)^\top$, is typically a biased estimator of $\Psi(Y, X, \Theta)$ given (Y, X), with the bias provided by the multivariate version of the Taylor series in (4). Because $f(X) = \Psi(Y, X, \Theta)$ is a quadratic function of X here, terms of order above two (i.e., $\ell > 2$) in the Taylor series reduce to zero, $\hat{f}_{\text{tc}}(X)$ does not achieve further bias reduction compared to $\hat{f}_{\text{cc}}(X)$. The application of $\hat{f}_{\text{cc}}(X)$ in the presence of multivariate measurement error is equivalent to setting $\Im = iB$ in (16), where $B = (b_1, \ldots, b_m)^\top \in \mathbb{R}^m$ is chosen to reduce, or ideally, eliminate the bias of $\Psi(Y, W, \Theta)$. One can generalize the arguments in Sect. 2 to adapt to an m-dimensional complex

random variable, $C = U - E(U) + iB$, and find B such that $\Re\{E(C^2)\}$ vanishes, where $C^2 = C_1^{\ell_1} \cdots C_m^{\ell_m}$, in which $C_1, \ldots, C_m$ are entries of C, and $\ell_1, \ldots, \ell_m \in \{0, 1, 2\}$ such that $\ell_1 + \ldots + \ell_m = 2$. To gain more insight on the debiased score $\tilde{\Psi}(Y, W, \Theta)$ in (16) under linear regression, we take a different route to find B via elaborating (16) next.

Define $\mathscr{T} = W - E(U) + \Im$. With the least square score function in (16), we have

$$\begin{aligned}\tilde{\Psi}(Y, W, \Theta) &= \Re\left\{(Y - \beta_0 - \mathscr{T}^\top \beta)\begin{bmatrix}1\\ \mathscr{T}\end{bmatrix}\right\} \\ &= \left[Y - \beta_0 - \{W - E(U)\}^\top \beta\right]\begin{bmatrix}1\\ W - E(U)\end{bmatrix} - \begin{bmatrix}0\\ \Re(\Im\Im^\top)\beta\end{bmatrix}. \quad (17)\end{aligned}$$

Taking the expectation of (17) conditional on (Y, X) gives

$$E\{\tilde{\Psi}(Y, W, \Theta)|Y, X\} = \begin{bmatrix}Y - \beta_0 - X^\top \beta\\ (Y - \beta_0 - X^\top \beta)X - \Omega\beta - \Re(\Im\Im^\top)\beta\end{bmatrix},$$

which is equal to the least square score $\Psi(Y, X, \Theta)$ when $\Im = iB$, with B satisfying $BB^\top = \Omega$. This gives a debiased score, $\tilde{\Psi}(Y, W, \Theta) = [Y - \beta_0\{W - E(U)\}^\top \beta](1, \{W - E(U)\}^\top)^\top + (0, \beta^\top \Omega)^\top$, which coincides with the corrected score in linear regression with errors-in-covariates, and includes $\hat{f}_{\rm cc}(X)$ in (5) as a special case when $m = 1$.

4.3 Polynomial regression

Consider an mth order polynomial regression model with errors-in-covariates,

$$Y_r = \beta_0 + \sum_{q=1}^{m} \beta_q X_r^q + \epsilon_r, \qquad W_r = X_r + U_r, \qquad (r = 1, \ldots, n),$$

where $\Theta = (\beta_0, \beta_1, \ldots, \beta_m)^\top \in \mathbb{R}^{m+1}$ is the parameter vector of interest, ϵ_r is a mean-zero error, and the measurement error model is the same as (1) with a univariate nondifferential measurement error. The least square estimator of Θ in the absence of measurement error solves the estimating equation $\sum_{r=1}^{n}(Y_r - \beta_0 - \sum_{q=1}^{m} \beta_q X_r^q)(1, X_r, X_r^2, \ldots, X_r^m)^\top = 0_{m+1}$. That is, the true-data score is $\Psi(Y, X, \Theta) = Y\tilde{X} - \tilde{X}\tilde{X}^\top\Theta$, where $\tilde{X} = (1, X, X^2, \ldots, X^m)^\top$. Similarly, let $\tilde{W} = (1, W, W^2, \ldots, W^m)^\top$.

Because the score function $f(X) = \Psi(Y, X, \Theta)$ here is a polynomial function of order $2m$ of X, $f^{(3)}(t)$ is not always zero when $m \geq 2$. According to (4), the bias of the naive score, $\Psi(Y, W, \Theta) = Y\tilde{W} - \tilde{W}\tilde{W}^\top\Theta$, depends on higher-order moments of U, μ_ℓ, for $\ell \geq 3$. This is the scenario where $\hat{f}_{\rm tc}(X)$ can improve over $\hat{f}_{\rm cc}(X)$. We thus consider a debiased score in (16) involving tessarines, that is,

$$\tilde{\Psi}(Y, W, \Theta) = \Re(Y\tilde{\mathscr{T}} - \tilde{\mathscr{T}}\tilde{\mathscr{T}}^\top\Theta) = Y\Re(\tilde{\mathscr{T}}) - \Re(\tilde{\mathscr{T}}\tilde{\mathscr{T}}^\top)\Theta, \qquad (18)$$

where $\tilde{\mathscr{T}} = (1, \mathscr{T}, \mathscr{T}^2, \ldots, \mathscr{T}^m)^\top$, in which $\mathscr{T} = W - E(U) + bi + cj + dk$. After one finds (b, c, d) following the strategies described in Sect. 3.2, one can use (11) to compute $\Re(\mathscr{T}^\ell)$, for $\ell = 2, 3, \ldots, 2m$, and then evaluates (18).

Let us take the quadratic regression as an example for further elaboration. With $m = 2$, we have $\Re(\tilde{\mathscr{T}})$ equal to

$$\Re\left(\begin{bmatrix} 1 \\ \mathscr{T} \\ \mathscr{T}^2 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ W - E(U) \\ \{W - E(U)\}^2 - (b^2 - c^2 + d^2) \end{bmatrix} = \begin{bmatrix} 1 \\ W - E(U) \\ \{W - E(U)\}^2 - \sigma_u^2 \end{bmatrix},$$

where we conclude the third entry of $\Re(\tilde{\mathscr{T}})$ by (14). The $[s, t]$ entry of $\Re(\tilde{\mathscr{T}}\tilde{\mathscr{T}}^\top)$ is $\Re(\mathscr{T}^{s+t-2})$, for $s, t \in \{1, 2, 3\}$. That is, entries of $\Re(\tilde{\mathscr{T}}\tilde{\mathscr{T}}^\top)$ include those in $\Re(\tilde{\mathscr{T}})$ and

$$\begin{aligned} \Re(\mathscr{T}^3) &= \{W - E(U)\}^3 - 3\sigma_u^2\{W - E(U)\} - \mu_3, \\ \Re(\mathscr{T}^4) &= \{W - E(U)\}^4 - 6\sigma_u^2\{W - E(U)\}^2 - 4\mu_3\{W - E(U)\} - \mu_4 + 6\sigma_u^4, \end{aligned}$$

where constraints on (b, c, d) suggested in (9) are used to simplify both results. Plugging in (18) the resulting $\Re(\tilde{\mathscr{T}})$ and $\Re(\tilde{\mathscr{T}}\tilde{\mathscr{T}}^\top)$ yields the debiased score $\tilde{\Psi}(Y, W, \Theta)$ that is identical to the corrected score provided in Cheng and Schneeweiss (1997) for a quadratic regression model. In Cheng and Schneeweiss (1997), the authors derived unbiased estimators of $Y\tilde{X}$ and $\tilde{X}\tilde{X}^\top$ to construct an unbiased estimator of $\Psi(Y, X, \Theta) = Y\tilde{X} - \tilde{X}\tilde{X}^\top\Theta$, for $m \geq 2$, without assuming ϵ and U independent or assuming Gaussian measurement error. The price they pay for this generality is the assumption that $\{\mu_\ell\}_{\ell=1}^{2m}$ and $\{E(U^\ell\epsilon)\}_{\ell=1}^{m}$ are all known. The tessarine-based debiased score $\hat{f}_{\text{tc}}(X)$ is a corrected score only when $m \leq 2$, but can still reduce bias when $m > 2$ in the naive score $f(W)$, the Monte Carlo corrected score $\hat{f}_{\text{mc}}(X)$ in the presence of non-Gaussian measurement error, and the score involving a complex constant, $\hat{f}_{\text{cc}}(X)$, while only requiring knowledge of $E(U)$ and $\{\mu_\ell\}_{\ell=2}^{4}$.

4.4 Poisson regression

Consider a Poisson regression model (Cameron, 1998) with errors-in-covariate,

$$Y_r \mid X_r \sim \text{Poisson}(\exp(\beta_0 + \beta_1 X_r)), \qquad W_r = X_r + U_r, \qquad (r = 1, \ldots, n), \tag{19}$$

where $\Theta = (\beta_0, \beta_1)^\top \in \mathbb{R}^2$ contains regression coefficients of interest. Had there been no measurement error, the maximum likelihood estimator of Θ is the solution to the estimating equation $\sum_{r=1}^{n}\{Y_r - \exp(\beta_0 + \beta_1 X_r)\}(1, X_r)^\top = 0_2$. Hence, the true-data score function is $\Psi(Y, X, \Theta) = \{Y - \exp(\beta_0 + \beta_1 X)\}(1, X)^\top$.

To reduce the bias of the naive score $\Psi(Y, W, \Theta) = \{Y - \exp(\beta_0 + \beta_1 W)\}(1, W)^\top$ in the presence of non-Gaussian measurement error, we consider the debiased score

in (16) that involves tessarine $\mathscr{T} = W - E(U) + bi + cj + dk$, that is,

$$\begin{aligned}\tilde{\Psi}(Y, W, \Theta) &= \Re\left(\{Y - \exp(\beta_0 + \beta_1 \mathscr{T})\}\begin{bmatrix}1\\ \mathscr{T}\end{bmatrix}\right)\\ &= \begin{bmatrix} Y - e^{\beta_0}\Re\{\exp(\beta_1 \mathscr{T})\}\\ Y\{W - E(U)\} - e^{\beta_0}\Re\{\exp(\beta_1 \mathscr{T})\mathscr{T}\}\end{bmatrix},\end{aligned} \tag{20}$$

where, using Theorem 3, we have

$$\begin{aligned}\Re\{\exp(\beta_1 \mathscr{T})\} = \frac{1}{2}\exp(\beta_1\{W - E(U)\})\{&e^{\beta_1 c}\cos(\beta_1(b+d))\\ &+ e^{-\beta_1 c}\cos(\beta_1(b-d))\},\end{aligned}$$

and, using the matrix representations of $\mathscr{T}$ (as in Lemma 2) and $\exp(\beta_1 \mathscr{T})$ (given in Theorem 3) to carry out the tessarine multiplication $\exp(\beta_1 \mathscr{T})\mathscr{T}$ as matrix multiplication, then extracting the real part of the first diagonal entry of the resulting matrix gives

$$\begin{aligned}\Re\{\exp(\beta_1 \mathscr{T})\mathscr{T}\} = \frac{1}{2}\exp(\beta_1\{W - E(U)\})\big(&e^{\beta_1 c}[\{W - E(U) + c\}\\ &\cos(\beta_1(b+d)) - (b+d)\sin(\beta_1(b+d))]\\ &+ e^{-\beta_1 c}[\{W - E(U) - c\}\cos(\beta_1(b-d))\\ &+ (b-d)\sin(\beta_1(b-d))]\big).\end{aligned}$$

To express the debiased score more succinctly, we assume $E(U) = 0$ and let $\xi_1 = \cos(\beta_1(b+d))$, $\xi_2 = \cos(\beta_1(b-d))$, $\eta_1 = \sin(\beta_1(b+d))$, and $\eta_1 = \sin(\beta_1(b-d))$, then

$$\tilde{\Psi}(Y, W, \Theta) = \begin{bmatrix} Y - \exp(\beta_0 + \beta_1 W)\left(\xi_1 e^{\beta_1 c} + \xi_2 e^{-\beta_1 c}\right)/2\\ \\ YW - \exp(\beta_0 + \beta_1 W)\left[\{\xi_1(W+c) - \eta_1(b+d)\}e^{\beta_1 c}\right.\\ \left.+\{\xi_2(W-c) + \eta_2(b-d)\}e^{-\beta_1 c}\right]/2\end{bmatrix}. \tag{21}$$

If U is symmetric around zero with the kurtosis greater than 5, then one can show that a solution to (9) is $(b, c, d) = (b^*, c^*, 0)$, with $b^* = \{\sigma_u^2 + (\mu_4 - 4\sigma_u^4)^{1/2}\}^{1/2}/\sqrt{2}$ and $c^* = (b^{*2} - \sigma_u^2)^{1/2}$. Using this choice of (b, c, d) in (21) yields a debiased score that partially accounts for symmetric measurement error that has a heavier tail than a Gaussian distribution,

$$\tilde{\Psi}(Y, W, \Theta) = \begin{bmatrix} Y - \exp(\beta_0 + \beta_1 W)\cos(\beta_1 b^*)\cosh(\beta_1 c^*)\\ \\ YW - \exp(\beta_0 + \beta_1 W)[\cosh(\beta_1 c^*)\{W\cos(b^*\beta_1) -\\ b^*\sin(b^*\beta_1)\} + c^*\cos(b^*\beta_1)\sin(c^*\beta_1)]\end{bmatrix}. \tag{22}$$

This new score is obviously very different from the following corrected score derived under the assumption of mean-zero Gaussian measurement error for Poisson regression (Carroll et al., 2006, Section7.4.3),

$$\Psi_{\mathrm{cs}}(Y, W, \Theta) = \begin{bmatrix} Y - \exp(\beta_0 + \beta_1 W - \beta_1^2\sigma_u^2/2) \\ YW - \exp(\beta_0 + \beta_1 W - \beta_1^2\sigma_u^2/2)(W - \beta_1\sigma_u^2) \end{bmatrix}. \tag{23}$$

4.5 Logistic regression

Having considered regression models for continuous responses and response as a count in Sects. 4.2–4.4, we shift our focus to a logistic regression model for a binary response with errors-in-covariate in this subsection,

$$Y_r \mid X_r \sim \text{Bernoulli}(\mathcal{G}(\beta_0 + \beta_1 X_r)), \qquad W_r = X_r + U_r, \qquad (r = 1, \ldots, n), \tag{24}$$

where $\Theta = (\beta_0, \beta_1)^\top \in \mathbb{R}^2$ is the parameter of interest, and $\mathcal{G}(t) = 1/\{1 + \exp(-t)\}$ is the expit function. Based on error-free data, the maximum likelihood estimator of Θ solves the estimating equation

$$\sum_{r=1}^{n} \{Y_r - \mathcal{G}(\beta_0 + \beta_1 X_r)\}(1, X_r)^\top = 0_2.$$

Unlike the true-data scores considered in earlier subsections, here we have $f(X) = \Psi(Y, X, \Theta) = \{Y - \mathcal{G}(\beta_0 + \beta_1 X)\}(1, X)^\top$, which is not an entire function of X because the expit function $\mathcal{G}(t)$ is not entire (Levin, 1996). In particular, $t = i\ell\pi$ is a singularity point of $\mathcal{G}(t)$ when ℓ is an odd integer. As a result, one cannot use the Taylor series as those in (2)–(4) to justify or inspect the debiasing effect of the score in (16) when comparing with the naive score $f(W) = \Psi(Y, W, \Theta) = \{Y - \mathcal{G}(\beta_0 + \beta_1 W)\}(1, W)^\top$. This complication due to the nonentire $f(X)$ presents the same challenge in applying the Monte Carlo corrected score $\hat{f}_{\mathrm{mc}}(X)$ in logistic regression. For simpler notations, suppose a random sample of size $M = 1$ from $N(0, 1)$ is used in constructing $\hat{f}_{\mathrm{mc}}(X)$, then this score is ((Novick and Stefanski, 2002), Section8.2)

$$\Psi_{\mathrm{mc}}(Y, W, \Theta) = \begin{bmatrix} Y - \dfrac{\exp(\beta_0 + \beta_1 W) + \cos(\beta_1\sigma_u Z)}{\exp(-\beta_0 - \beta_1 W) + \exp(\beta_0 + \beta_1 W) + 2\cos(\beta_1\sigma_u Z)} \\ YW - \dfrac{\{\exp(\beta_0 + \beta_1 W) + \cos(\beta_1\sigma_u Z)\}W - \sigma_u Z \sin(\beta_1\sigma_u Z)}{\exp(-\beta_0 - \beta_1 W) + \exp(\beta_0 + \beta_1 W) + 2\cos(\beta_1\sigma_u Z)} \end{bmatrix}, \tag{25}$$

where $Z \sim N(0, 1)$. The authors argued that, even though $\Psi_{\mathrm{mc}}(Y, W, \Theta)$ is a biased estimator for $\Psi(Y, X, \Theta)$ (thus is not a corrected score), the bias is small "for measurement error variances of the magnitudes commonly encountered in applications." They provided empirical evidence suggesting that estimates for Θ resulting from the Monte Carlo corrected score method are similar to those from the conditional score method. The conditional score is an unbiased estimator of $\Psi(Y, X, \Theta)$ (and thus is a

corrected score) when U is Gaussian; more specifically, this score is (Carroll et al., 2006)[Section7.2.2]

$$\Psi_{\text{cd}}(Y, W, \Theta) = \{Y - \mathcal{G}(\beta_0 + (W + Y\sigma_u^2\beta_1)\beta_1 - \beta_1^2\sigma_u^2/2)\}(1, W + Y\sigma_u^2\beta_1)^\top. \tag{26}$$

Inspired by these existing works on logistic regression with covariates contaminated by Gaussian measurement error, we consider the debiased score in (16) with $\mathscr{T} = W - E(U) + bi + cj + dk$ substituting X in $\Psi(Y, X, \Theta)$, leading to

$$\tilde{\Psi}(Y, W, \Theta) = \begin{bmatrix} Y - \Re\{\mathcal{G}(\beta_0 + \beta_1\mathscr{T})\} \\ Y\{W - E(U)\} - \Re\{\mathcal{G}(\beta_0 + \beta_1\mathscr{T})\mathscr{T}\} \end{bmatrix}, \tag{27}$$

with the two real parts elaborated in Appendix A .5. We will assess the debiasing effect of the corresponding estimator for Θ in the presence of non-Gaussian measurement error in simulation experiments in Sect. 6.

4.6 Asymptotic bias analysis

Following the inspection of the debiased score in (16) applied to four regression models, we now study the corresponding debiased estimator in general. Suppose there exists a solution (b, c, d) to (9) so that $\Re\{E(C^\ell)\} = 0$ for $\ell = 1, 2, 3, 4$, with $C = U - E(U) + bi + cj + dk$. Denote by $\hat{\Theta}_{\text{tc}}$ the debiased estimator involving tessarine random errors, $\{C_r = U_r - E(U) + bi + cj + dk\}_{r=1}^n$, that satisfies $n^{-1}\sum_{r=1}^n \Re\{\Psi(Y_r, X_r + C_r, \hat{Theta}_{\text{tc}})\} = 0$. Assuming $\Psi(y, x, \theta)$ infinitely differentiable with respect to x for each (y, θ), then the estimating equation is equivalent to, following the Taylor expansion of $\Psi(\cdot, X_r + C_r, \cdot)$ around X_r,

$$\frac{1}{n}\sum_{r=1}^n \left\{ \Psi(Y_r, X_r, \hat{\Theta}_{\text{tc}}) + \sum_{\ell=1}^\infty \Psi_2^{(\ell)}(Y_r, X_r, \hat{\Theta}_{\text{tc}})\Re\left(C_r^\ell\right)/\ell! \right\} = 0, \tag{28}$$

where $\Psi_2^{(\ell)}(Y_r, x, \hat{\Theta}_{\text{tc}}) = (\partial^\ell/\partial x^\ell)\Psi(Y_r, x, \hat{\Theta}_{\text{tc}})$, with the subscript "2" signifying that it is the ℓ-th order partial derivative of $\Psi(y, x, \theta)$ with respect to the second argument. A first-order Taylor expansion of (28) around the truth Θ^* gives

$$\begin{aligned} 0 = &\frac{1}{n}\sum_{r=1}^n \left\{ \Psi(Y_r, X_r, \Theta^*) + \sum_{\ell=1}^\infty \Psi_2^{(\ell)}(Y_r, X_r, \Theta^*)\Re\left(C_r^\ell\right)/\ell! \right\} + \\ &\frac{1}{n}\sum_{r=1}^n \left\{ \Psi_3^{(1)}(Y_r, X_r, \Theta^*) + \sum_{\ell=1}^\infty \Psi_{2,3}^{(\ell,1)}(Y_r, X_r, \Theta^*)\Re\left(C_r^\ell\right)/\ell! \right\}(\hat{\Theta}_{\text{tc}} - \Theta^*) \\ &+ \mathcal{R}^*, \end{aligned} \tag{29}$$

where $\Psi_3^{(1)}(y, x, \Theta^*)$ is $(\partial/\partial\theta)\Psi(y, x, \theta)$ evaluated at $\theta = \Theta^*$, with the subscript "3" stressing that it is the partial derivative with respect to the third argument of

$\Psi(y, x, \theta)$; similarly, $\Psi_{2,3}^{(\ell,1)}(y, X_r, \Theta^*)$ is equal to $(\partial/\partial\theta)\{(\partial^\ell/\partial x^\ell)\Psi(y, x, \theta)\}$ evaluated at $(x, \theta) = (X_r, \Theta^*)$; lastly, $\mathcal{R}^*$ is the remainder term of the Taylor expansion that depends on functionals (as derivatives) of $\Psi(y, x, \theta)$ evaluated at (Y_r, X_r)'s and $\tilde{\Theta}$ that lies between $\hat{\Theta}_{\text{tc}}$ and Θ^*.

By (29), one has, assuming the invertibility of the first term below,

$$
\begin{aligned}
&\hat{\Theta}_{\text{tc}} - \Theta^* \\
&= -\left\{ \frac{1}{n}\sum_{r=1}^{n} \Psi_3^{(1)}(Y_r, X_r, \Theta^*) + \sum_{\ell=1}^{\infty} \frac{1}{n}\sum_{r=1}^{n} \Psi_{2,3}^{(\ell,1)}(Y_r, X_r, \Theta^*) \Re\left(C_r^\ell\right)/\ell! \right\}^{-1} \\
&\times \left\{ \frac{1}{n}\sum_{r=1}^{n} \Psi(Y_r, X_r, \Theta^*) + \sum_{\ell=1}^{\infty} \frac{1}{n}\sum_{r=1}^{n} \Psi_2^{(\ell)}(Y_r, X_r, \Theta^*) \Re\left(C_r^\ell\right)/\ell! + \mathcal{R}^* \right\} \qquad (30) \\
&= -\left[E\left\{\Psi_3^{(1)}(Y, X, \Theta^*)\right\} + \sum_{\ell=5}^{\infty} E\left\{\Psi_{2,3}^{(\ell,1)}(Y, X, \Theta^*)\right\} \Re\left\{E\left(C^\ell\right)\right\}/\ell! + o_p(1) \right]^{-1} \\
&\times \left[\sum_{\ell=5}^{\infty} E\left\{\Psi_2^{(\ell)}(Y, X, \Theta^*)\right\} \Re\left\{E\left(C^\ell\right)\right\}/\ell! + \mathcal{R}^* + o_p(1) \right], \qquad (31)
\end{aligned}
$$

under regularity conditions imposed on the score function $\Psi(y, x, \theta)$ (e.g, see Boos and Stefanski (2013) Theorem 7.1), such as $E\{\Psi(Y, X, \Theta^*)\} = 0$ and moment conditions required for the empirical means in (30) converging in probability to the corresponding expectations in (31) by the weak law of large numbers. According to (31), the dominating bias of $\hat{\Theta}_{\text{tc}}$ mainly depends on the partial derivatives of $\Psi(y, x, \theta)$ with respect to x of order above four. A sufficient condition for this dominating term to vanish, besides all the aforementioned conditions leading to (31), is to have these high-order derivatives vanish. The practical implication of this is that the bias of $\hat{\Theta}_{\text{tc}}$ tends to be more negligible when $\Psi(y, x, \theta)$, as a function of x, can be better approximated by some fourth-order polynomial for each (y, θ).

5 Kernel density estimation

After considering in Sect. 4 four parametric regression models with error-prone covariates, we turn to a nonparametric problem of estimating the probability density function of X, $p_X(x)$. In the absence of measurement error, an estimator for $p_X(x)$ based on a random sample $\{X_r\}_{r=1}^n$ is

$$\hat{p}_X(x) = (1/n)\sum_{r=1}^{n} K_h(X_r - x),$$

where $K_h(\cdot) = (1/h)K(\cdot/h)$, $K(\cdot)$ is a kernel and $h > 0$ is the bandwidth (Parzen, 1962; Rosenblatt, 1956).

In the notations established in Sect. 2, here we have $f(X) = K_h(X - x)$, with the naive counterpart $f(W) = K_h(W - x)$ appearing in the naive density estimator of $p_X(x)$ based on the error-contaminated data $\{W_r\}_{r=1}^n$, $\hat{p}_{\text{naive}}(x) = (1/n)\sum_{r=1}^n K_h(W_r - x)$. The proposed debiased estimator of $f(X)$ in (10) is then $\hat{f}_{\text{tc}}(X) = \Re\{K_h(W - x - E(U) + bi + cj + dk)\}$ for some (b, c, d) obtained according to Sect. 3.2. This yields a new estimator of $p_X(x)$ that can improve over $\hat{p}_{\text{naive}}(x)$ in the presence of non-Gaussian measurement error,

$$\hat{p}_{\text{tc}}(x) = \frac{1}{n}\sum_{r=1}^{n} \Re\left\{K_h(W_r - x - E(U) + bi + cj + dk)\right\}. \tag{32}$$

We use the logistic kernel, $K(t) = 1/(2 + e^{-t} + e^{t})$, for its entirety and ease in evaluation at tessarines. As shown in Appendix A .6, evaluating the logistic kernel at a tessarine $C = a + bi + cj + dk$ gives

$$\Re\{K(C)\} = \frac{2\xi(a, b, c, d)}{\{\xi(a, b, c, d)\}^2 + \{\psi(a, b, c, d)\}^2}, \tag{33}$$

with $\xi(a, b, c, d) = 2\{2 + \cosh(a + c)\cos(b + d) + \cosh(a - c)\cos(b - d)\}$ and $\psi(a, b, c, d) = \sinh(a + c)\sin(b + d) + \sinh(a - c)\sin(b - d)$. Other kernel choices are possible provided that the kernel is defined over the entire tessarine ring.

Stefanski and Carroll (1990) proposed to estimate $f(X) = K_h(X - x)$ via the deconvoluting kernel, $L_h(W - x)$, where $L_h(\cdot) = (1/h)L(\cdot/h)$, and

$$L(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx}\frac{\phi_K(t)}{\phi_U(-t/h)}dt, \tag{34}$$

in which $\phi_K(\cdot)$ is the Fourier transform of $K(\cdot)$, and $\phi_U(\cdot)$ is the characteristic function of U. Under certain conditions to guarantee a well-defined integral in (34), it has been shown that $E\{L_h(W - x) \mid X\} = K_h(X - x)$, and thus $L_h(W - x)$ is an unbiased estimator of $f(X)$. These conditions are imposed on $\phi_K(t)$ depending on the rate of convergence of $\phi_U(-t/h)$ towards zero as h tends to zero. Using the deconvoluting kernel, one attains an estimator of $p_X(x)$, referred to as the deconvoluting kernel density estimator, $\hat{p}_{\text{dk}}(x) = (1/n)\sum_{r=1}^n L_h(W_r - x)$. This strategy of density estimation was later exploited in Chen and Yi (2022), with different assumptions on U, to estimate the functional distance correlation to facilitate feature screening. The estimator $\hat{p}_{\text{dk}}(x)$ requires the knowledge of $\phi_U(\cdot)$, whereas the debiased estimator $\hat{p}_{\text{tc}}(x)$ in (32) only needs the first four moments of U. Moreover, computing $\hat{p}_{\text{dk}}(x)$ requires a great deal of caution when evaluating (34) with the untamed integrand. In comparison, using (33), computing $\hat{p}_{\text{tc}}(x)$ is a much simpler numerical task. As witnessed in our extensive simulation study presented in Sect. 6, this simplicity dramatically reduces computation time at a small cost to accuracy.

Besides the simulation study on kernel density estimation in Sect. 6.3, also included in the simulation experiments in other subsections of Sect. 6 are the M-estimation in Poisson regression and logistic regression, which are analytically less transparent when

it comes to comparing different estimators than the comparisons in linear regression and polynomial regression.

6 Simulation study

6.1 Experiments in Poisson regression

Using simulated data from a Poisson regression model with errors-in-covariate, we compare three estimators for regression coefficients $\Theta = (\beta_0, \beta_1)^\top$ using the maximum likelihood estimator based on error-free data as a reference estimator, denoted by $\hat{\Theta}$. The three estimators include the naive estimator based on error-contaminated data, denoted by $\hat{\Theta}_{\text{naive}}$, the corrected score estimator based on the score $\Psi_{\text{cs}}(Y, W, \Theta)$ in (23), denoted by $\hat{\Theta}_{\text{cs}}$, and the tessarine-based estimator corresponding to the debiased score $\tilde{\Psi}(Y, W, \Theta)$ in (21), denoted by $\hat{\Theta}_{\text{tc}}$.

Define $\lambda = \text{var}(X)/\{\text{var}(X) + \sigma_u^2\}$ as the reliability ratio (Carroll et al., 2006)[Section3.2.1]. After generating a random sample of size $n \in \{250, 500\}$ from $N(0, 1)$ as the error-free covariate data $\{X_r\}_{r=1}^n$, we generate $\{(Y_r, W_r)\}_{r=1}^n$ according to (19), with $\beta_0 = 1$ and $\beta_1 = -1$, where $\{U_r\}_{r=1}^n$ is a random sample from a gamma distribution with the shape parameter equal to $4(1-\lambda)/\lambda$ and a scale parameter of 0.5, in which $\lambda \in \{0.8, 0.9, 1\}$. This design of measurement error distribution gives $\mu_2 = \sigma_u^2 = (1-\lambda)/\lambda$ and a kurtosis of $3[1 + \{\lambda/(1-\lambda)\}^{1/2}]$. Moreover, under this design, real-valued solutions to (9) exist. At each simulation setting specified by the levels of n and λ, we use 1000 Monte Carlo replicates to compare $\hat{\Theta}_{\text{naive}}$, $\hat{\Theta}_{\text{cs}}$, and $\hat{\Theta}_{\text{tc}}$. We acknowledge that, among the three estimators, only $\hat{\Theta}_{\text{tc}}$ is developed without assuming mean-zero measurement error. To make $\hat{\Theta}_{\text{naive}}$ and $\hat{\Theta}_{\text{cs}}$ relatively comparable with $\hat{\Theta}_{\text{tc}}$, we compute $\hat{\Theta}_{\text{naive}}$ and $\hat{\Theta}_{\text{cs}}$ based on covariate data contaminated by the centered measurement error $\{U_r - 2(1-\lambda)/\lambda\}_{r=1}^n$. Similar treatments on data used by competing methods are applied in all our empirical study.

Three metrics are used to assess the performance of an estimator. Take the slope parameter β_1 as an example and denote by $\hat{\beta}_{1,q}$ an estimate of it resulting from a considered estimation method applied to the qth Monte Carlo replicate dataset. Two of the metrics are the empirical bias defined as $(1000)^{-1}\sum_{q=1}^{1000}(\hat{\beta}_{1,q} - \beta_1)$, and the empirical mean squared error defined as $(1000)^{-1}\sum_{q=1}^{1000}(\hat{\beta}_{1,q} - \beta_1)^2$. A third metric is the empirical interquartile range of the estimates $\{\hat{\beta}_{1,q}\}_{q=1}^{1000}$. Fig. 1 depicts these metrics for $\hat{\Theta}$, $\hat{\Theta}_{\text{naive}}$, $\hat{\Theta}_{\text{cs}}$, and $\hat{\Theta}_{\text{tc}}$ as λ varies when $n = 250$. The counterpart results when $n = 500$ are provided in Appendix A .8. These graphical summaries show that, when bias and mean squared error are concerned, $\hat{\Theta}_{\text{tc}}$ is the closest to the reference estimator $\hat{\Theta}$ that is unattainable in the presence of measurement error. Although $\hat{\Theta}_{\text{cs}}$ slightly improves upon $\hat{\Theta}_{\text{naive}}$ in terms of bias, it is much less satisfactory than $\hat{\Theta}_{\text{tc}}$ and is subject to high variability. This illustrates the adverse effects of the departure from normality on the existing corrected score estimator. Admittedly, $\hat{\Theta}_{\text{tc}}$ tends to be more variable than $\hat{\Theta}_{\text{naive}}$, but the bias reduction achieved by $\hat{\Theta}_{\text{tc}}$ outweighs the inflation in variability, resulting in a much lower mean squared error than that of $\hat{\Theta}_{\text{naive}}$.

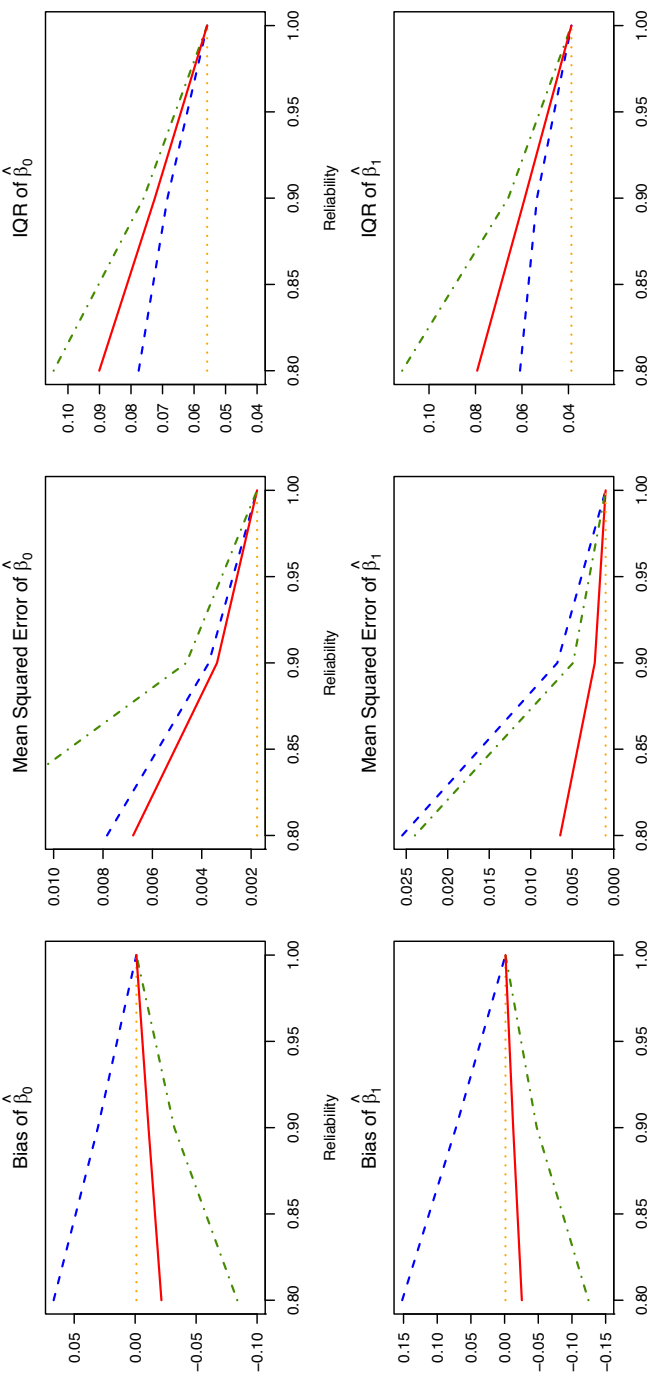

Fig. 1 Empirical bias, mean squared errors, and interquartile ranges (IQR) of four estimators for β_0 (top panels) and for β_1 (bottom panels) as the regression coefficients $\Theta = (\beta_0, \beta_1)^\top$ in the Poisson regression model when $n = 250$: $\hat{\Theta}$ based on error-free data (dotted lines), $\hat{\Theta}_{\text{naive}}$ (dashed lines), $\hat{\Theta}_{\text{cs}}$ (dot-dashed lines), and $\hat{\Theta}_{\text{lc}}$ (solid lines)

When the departure from normality of U is less severe than designed above, $\hat{\Theta}_{\text{cs}}$ may not suffer as much as witnessed in Fig. 1, and can even surpass $\hat{\Theta}_{\text{tc}}$. To illustrate this phenomenon, we repeat the above experiment but, instead of a gamma measurement error, we let $U = \sqrt{(1-\lambda)/(1.4\lambda)} \times t_7$, where t_7 is a random variable following a t distribution with 7 degrees of freedom, so that σ_u^2 is at the value leading to $\lambda \in \{0.8, 0.9, 1\}$ as before. Figure 2 presents the comparison between $\hat{\Theta}_{\text{naive}}$, $\hat{\Theta}_{\text{cs}}$, and $\hat{\Theta}_{\text{tc}}$ in this setting when $n = 500$, where $\hat{\Theta}_{\text{tc}}$ corresponds to the debiased score in (22) that (21) simplifies to when U is mean-zero and symmetric as in this context.

Although still substantially improves over $\hat{\Theta}_{\text{naive}}$, $\hat{\Theta}_{\text{tc}}$ does not outperform $\hat{\Theta}_{\text{cs}}$ now that the latter is less compromised by the (mild) violation of Gaussianity of U. Using the solution to (9) appearing in (22), $(b^*, c^*, 0)$, in the tessarine C, which makes $\Re\{E(C^\ell)\} = 0$ for $\ell = 2, 3, 4, 5$, we derive the dominating bias of $\tilde{\Psi}(Y, W, \Theta)$, as an estimator for $\Psi(Y, X, \Theta)$. This dominating bias depends on $\Re\{E(C^6)\}$. Subtracting an estimator of this dominating bias from $\tilde{\Psi}(Y, W, \Theta)$ yields another score that leads to another tessarine-based debiased estimator, denoted by $\hat{\Theta}_{\text{tc2}}$, that we include in this simulation experiment and in Fig. 2. This new tessarine-based estimator improves upon $\hat{\Theta}_{\text{tc}}$ and is comparable with $\hat{\Theta}_{\text{cs}}$ in this setting. These results suggest that, when compared with the corrected score estimator, the real benefit of the tessarine-based debiasing strategy mostly lies in correcting skewed measurement error. Appendix A .7 presents the detailed derivation of the further debiased strategy using generic notations as those in Sect. 2, as this improvement is not limited to Poisson regression or M-estimation.

6.2 Experiments in logistic regression

In the context of logistic regression, we compare four estimators for the regression coefficients $\Theta = (\beta_0, \beta_1)^\top$, using the maximum likelihood estimator based on error-free data, $\hat{\Theta}$, as a reference estimator. The four considered estimators are the naive estimator $\hat{\Theta}_{\text{naive}}$, the Monte Carlo corrected score estimator $\hat{\Theta}_{\text{mc}}$ based on the limiting version of $\Psi_{\text{mc}}(Y, W, \Theta)$ in (25) as $M \to \infty$ (Novick and Stefanski, 2002)[see Equation (16),], the conditional score estimator $\hat{\Theta}_{\text{cd}}$ based on the score $\Psi_{\text{cd}}(Y, W, \Theta)$ in (26), and the estimator $\hat{\Theta}_{\text{tc}}$ based on the debiased score $\tilde{\Psi}_{\text{tc}}(Y, W, \Theta)$ in (27).

After generating error-free covariate data $\{X_r\}_{r=1}^n$ from $N(1, 1)$, we generate the binary responses and error-contaminated covariate data $\{(Y_r, W_r)\}_{r=1}^n$ according to (24), with $\beta_0 = 5$, $\beta_1 = -5$, and $\{U_r\}_{r=1}^n$ as described in Sect. 6.1. At each simulation setting specified by $n \in \{500, 1000\}$ and $\lambda \in \{0.8, 0.9, 1\}$, we estimate Θ using the four estimators. Figure 3 presents the empirical bias, mean squared error, and interquartile range of each estimator across 1000 Monte Carlo replicates when $n = 500$. The counterpart results when $n = 1000$ are provided in Appendix A .8. Even though the debiasing effect of the proposed estimator $\hat{\Theta}_{\text{tc}}$ cannot be justified via an infinite Taylor series due to the nonentirety of the expit function, such effect is evident in the actual estimate when compared with the naive estimate. The proposed estimator also substantially outperforms the other two non-naive estimators developed under the Gaussian measurement error assumption.

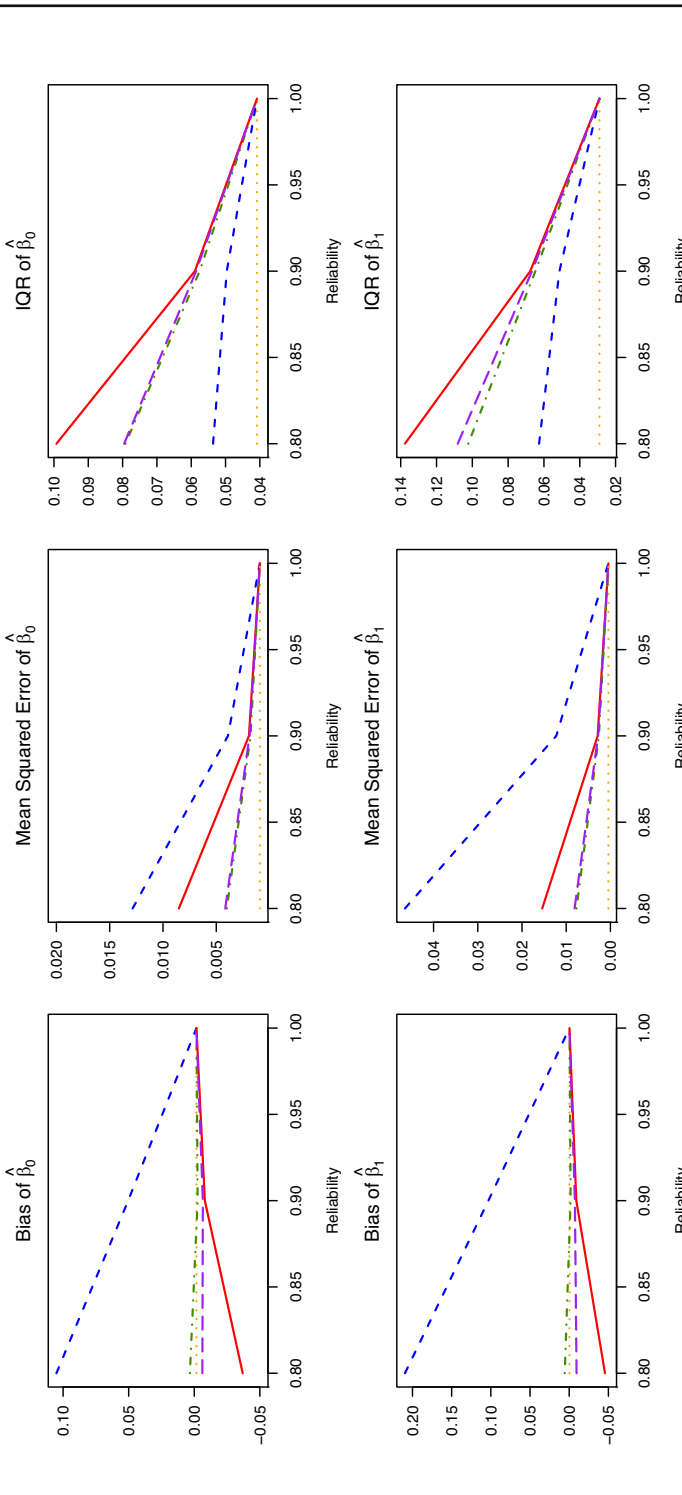

Fig. 2 Empirical bias, mean squared errors, and interquartile ranges (IQR) of five estimators for β_0 (top panels) and for β_1 (bottom panels) as the regression coefficients $\Theta = (\beta_0, \beta_1)^\top$ in the Poisson regression model with mean-zero symmetric measurement errors when $n = 500$: the maximum likelihood estimator $\hat{\Theta}$ based on error-free data (dotted lines), the naive estimator $\hat{\Theta}_{\text{naive}}$ (dashed lines), the corrected score estimator $\hat{\Theta}_{\text{cs}}$ (dot-dashed lines), the original tessarine-based estimator $\hat{\Theta}_{\text{tc}}$ (solid lines) and the further-debiased tessarine-based estimator $\hat{\Theta}_{\text{tc2}}$ (long dashed lines)

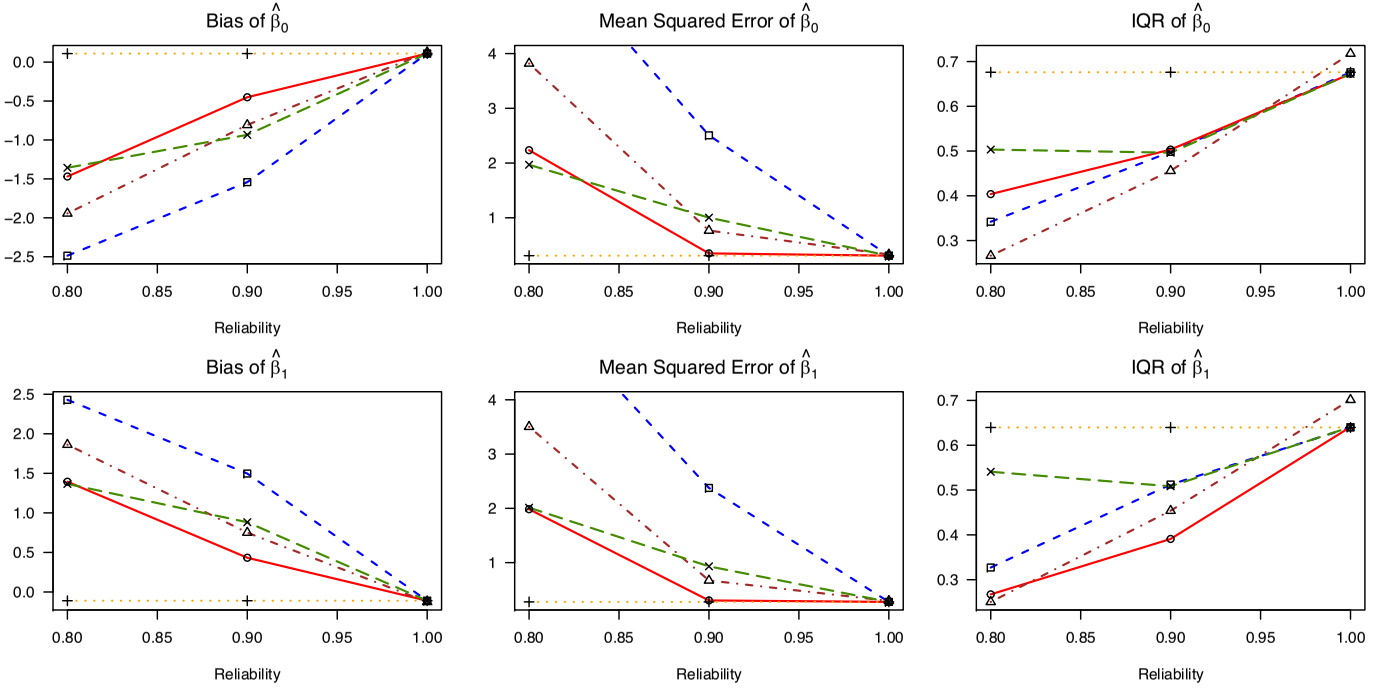

Fig. 3 Empirical bias, mean squared errors, and interquartile ranges (IQR) of five estimators for β_0 (top panels) and for β_1 (bottom panels) as the regression coefficients $\Theta = (\beta_0, \beta_1)^\top$ in the logistic regression when $n = 500$: $\hat{\Theta}$ (dotted lines, cross points), $\hat{\Theta}_{\text{naive}}$ (short dashed lines, square points), $\hat{\Theta}_{\text{mc}}$ (dot-dashed lines, triangle points), $\hat{\Theta}_{\text{cd}}$ (long dashed lines, $\times$ points), and $\hat{\Theta}_{\text{tc}}$ (solid lines, circle points)

6.3 Experiments in density estimation

With the target of inferences being the probability density function of X, $p_X(x)$, we use the kernel density estimator based on error-free data $\{X_r\}_{r=1}^n$, $\hat{p}_X(x)$, as a reference estimator to compare three density estimators based on error-contaminated data $\{W_r\}_{r=1}^n$. These include the naive estimator $\hat{p}_{\text{naive}}(x)$, the deconvoluting kernel density estimator $\hat{p}_{\text{dk}}(x)$, and the tessarine-based debiased estimator $\hat{p}_{\text{tc}}(x)$. In all considered estimators, the logistic kernel is used as $K(\cdot)$; and, for $\hat{p}_{\text{dk}}(x)$, $\phi_U(\cdot)$ in (34) is taken as the Laplace characteristic function.

At each of 500 Monte Carlo replicates at a simulation setting, the error-free data $\{X_r\}_{r=1}^n$ are generated from a two-component mixture distribution, $0.5N(0, 1) + 0.5N(5, 1)$, for $n \in \{50, 75\}$, following which we generate $\{W_r = X_r + U_r\}_{r=1}^n$, where $\{U_r\}_{r=1}^n$ is a random sample from a lognormal distribution with a logarithm of scale parameter $\sigma = \log[\{T(T-1)+1\}/2T]$, with $T = 2^{1/3}\sigma_u^2/[(\sigma_u^2[1+2\sigma_u^2] + \sigma_u^2(1+4\sigma_u^2)^{1/2}]^{1/3}$, and a logarithm of location $\mu = 1/2(\log[\sigma^2/(\exp[\sigma^2-1])])$, so that it yields $\mu_2 = \sigma_u^2$ as λ varies from 0.75 to 1 at increments of 0.025. Under this design, real-valued solutions to (9) do not exist. The metric used to assess the efficacy of a density estimator is the empirical mean integrated square error based on grid points $\{x_g\}_{g=1}^{100}$ equally spaced in $[-2, 8]$. For example, let $\hat{p}_{\text{naive},q}(x)$ be the naive estimate from the qth Monte Carlo replicate, the empirical mean integrated square error of $\hat{p}_{\text{naive}}(x)$ is $(500 \times 100)^{-1} \sum_{q=1}^{500} \sum_{g=1}^{100} \{\hat{p}_{\text{naive},q}(x_g) - p_X(x_g)\}^2$.

Bandwidth selection is an integral part of kernel density estimation. This problem in the error-free case has been well-studied (Hastie et al., 2009). In cases with error-contaminated data, Delaigle and Gijbels (2004) provided several strategies for choosing the bandwidth in $\hat{p}_{\text{dk}}(x)$. These methods are tailored for estimation based on the deconvoluting kernel that involves a Fourier transform, and thus cannot be easily revised to select a bandwidth for $\hat{p}_{\text{tc}}(x)$. Deriving an objective function, such as the mean integrated squared error, that is practically useful for selecting a bandwidth in $\hat{p}_{\text{tc}}(x)$ is much less straightforward. Furthermore, such an objective function depends on higher-order derivatives of $p_X(x)$ that need to be estimated, preferably using a similar strategy that leads to $\hat{p}_{\text{tc}}(x)$, which are topics beyond the scope of the current study. To empirically compare different density estimators here, we choose the bandwidth for each method that minimizes the corresponding empirical mean integrated squared error. To choose a bandwidth in the naive estimator $\hat{p}_{\text{naive}}(x)$, the true density of W is used in place of $p_X(\cdot)$ in the empirical mean integrated squared error so that the method is still considered naive.

Figure 4 presents the empirical mean integrated error for the three considered estimators compared with the reference estimator $\hat{p}_X(x)$. Even though the deconvoluting kernel density estimator $\hat{p}_{\text{dk}}(x)$ outperforms the debiased estimator $\hat{p}_{\text{tc}}(x)$ in terms of the mean integrated squared error, $\hat{p}_{\text{tc}}(x)$ significantly improves upon the naive estimator $\hat{p}_{\text{naive}}(x)$ and is less variable than $\hat{p}_{\text{dk}}(x)$ despite the lack of real-valued solutions to (9). Moreover, the computation time of $\hat{p}_{\text{tc}}(x)$ is dramatically shorter than that of $\hat{p}_{\text{dk}}(x)$. At a sample size of $n = 50$, it takes around 9 s to compute $\hat{p}_{\text{dk}}(x)$, whereas $\hat{p}_{\text{tc}}(x)$ takes 0.11 s to compute on a computer with an Intel Core i5 12600KF running at 4500 Mhz.

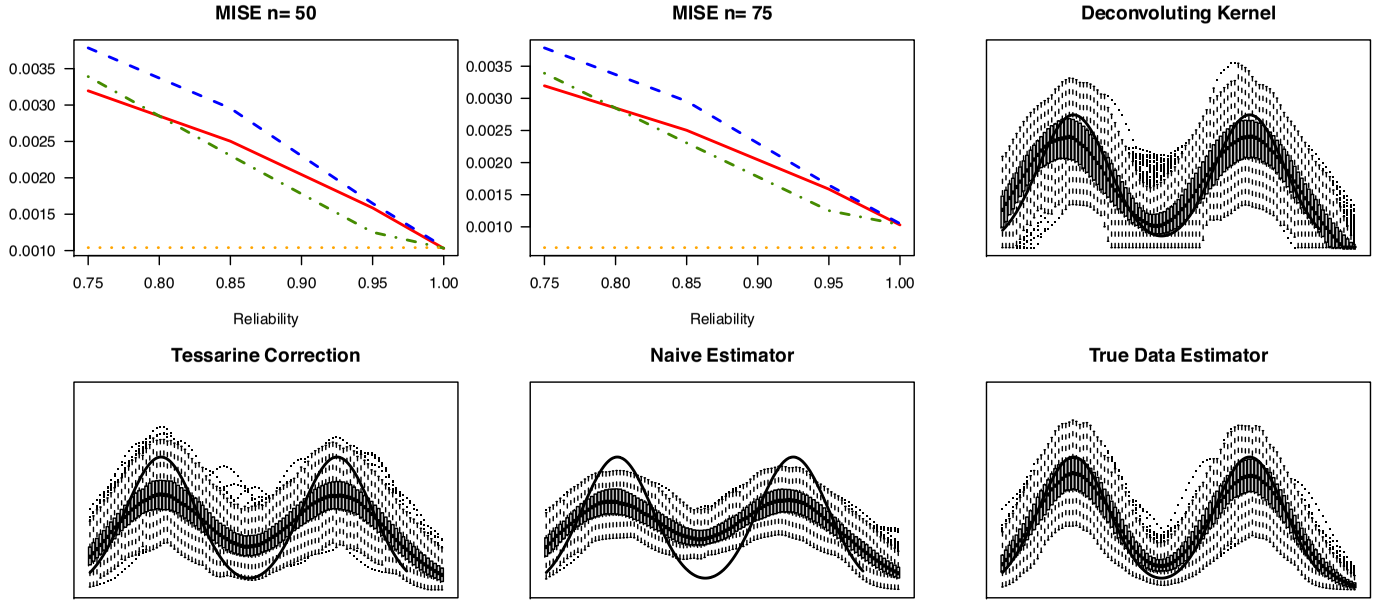

Fig. 4 Empirical mean integrated square errors (MISE) of four density estimators when $n = 50$ (top left panel) and $n = 75$ (top middle panel): $\hat{p}_X(x)$ (dotted lines), $\hat{p}_{\text{naive}}(x)$ (dashed lines), $\hat{p}_{\text{dk}}(x)$ (dot-dashed lines), and $\hat{p}_{\text{tc}}(x)$ (solid lines), along with boxplots of estimates from each method when $n = 50$ and $\lambda = 0.9$ superimposed on the true density curve (solid lines)

6.4 Sensitivity analysis

Because the proposed tessarine-based debiased estimator depends on the first four moments of U, a legitimate question is the impact of moment misspecification on the proposed estimator. To empirically inspect the influence of moment misspecification on different estimators, we conduct a sensitivity analysis on the estimation of regression coefficients in a Poisson regression model in a simulation setting similar to that in Sect. 6.1.

In particular, we generate a dataset of size $n = 250$, $\{(X_r, Y_r, W_r)\}_{r=1}^{n}$, following the data generating mechanism described for (X, Y, W) in Sect. 6.1, where U follows a gamma distribution with the shape parameter equal to 4/9 and the scale parameter equal to 0.5, gamma(4/9, 0.5), resulting in $\lambda = 0.9$. Then, instead of using the first four moments of gamma(4/9, 0.5) to find (b, c, d) in the tessarine-based estimator $\hat{\Theta}_{\text{tc}}$, we assume the second, third, and fourth central moments of U equal to $\mathcal{E}_1\sigma_u^2$, $\mathcal{E}_2\mu_3$, and $\mathcal{E}_3\mu_4$, respectively, where σ_u^2, μ_3, and μ_4 are the true moments of gamma(4/9, 0.5), $\mathcal{E}_1$, $\mathcal{E}_2$, and $\mathcal{E}_3$ are random numbers generated independently from a uniform distribution supported on $[1-\delta,\ 1+\delta]$, with δ varying from 0 to 0.6 at increments of 0.02. At each level of δ, we repeatedly compute $\hat{\Theta}_{\text{tc}}$ under the wrong (when $\delta > 0$) assumption of the measurement error moments using 5,000 Monte Carlo replicates. For comparison, we also compute the naive estimate $\hat{\Theta}_{\text{naive}}$, which is not affected by δ, and the corrected score estimate $\hat{\Theta}_{\text{cs}}$, which depends on the variance of U, according to (23), and thus is affected by δ.

Figure 5 provides the empirical bias, mean squared errors, and interquartile ranges of the four considered sets of estimates across 5,000 Monte Carlo replicates. Interestingly $\hat{\Theta}_{\text{tc}}$ and $\hat{\Theta}_{\text{cs}}$ are similar in the rate of deterioration as moments misspecification becomes more severe, despite the fact that $\hat{\Theta}_{\text{tc}}$ involves more misspecified moments that $\hat{\Theta}_{\text{cs}}$ does. Even though $\hat{\Theta}_{\text{tc}}$ is compromised by the misspecification, its performance remains competitive over a wide range of δ, that is, even when the misspecification is moderate. We thus conclude that, in the presence of non-Gaussian measurement error, even when the first four moments of U are estimated and thus subject to misspecification, the tessarine-based debiased estimator can still outperform the naive estimator, and is more preferable than the corrected score estimator developed under the assumption of Gaussian measurement error.

To estimate the central moments, μ_2, μ_3, and μ_4, of U when replicate data are available, one may follow the rationale behind the variance estimation in Carroll et al., (2006, Equation(4.3)), built upon sample moments with bias correction according to Fisher (1930). More specifically, suppose for $r = 1, \ldots, n$ there are $n_r (> 3)$ replicate measurements of X_r in $\{W_{r,s}\}_{s=1}^{n_r}$ contaminated by measurement errors in $\{U_{r,s}\}_{s=1}^{n_r}$. Let $\overline{W}_{r\cdot} = n_r^{-1}\sum_{s=1}^{n_r} W_{r,s}$, then an empirical counterpart of μ_ℓ is $\hat{\mu}_{\ell,r} = n_r^{-1}\sum_{s=1}^{n_r}(W_{r,s} - \overline{W}_{r\cdot})^\ell$, for $\ell = 2, 3, 4$, since, although U is unobservable, $W_{r,s} - \overline{W}_{r\cdot} = U_{r,s} - \overline{U}_{r,\cdot}$. Using all $\sum_{r=1}^{n} n_r$ observed surrogate covariate values, we have

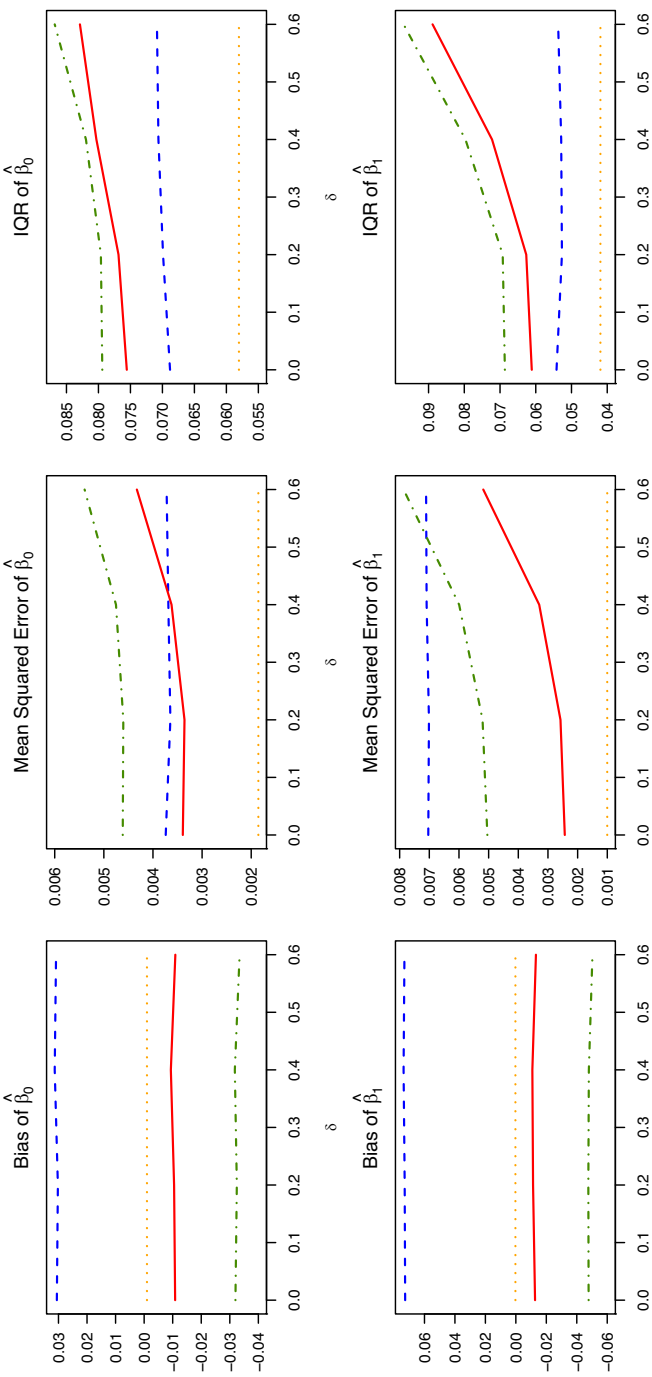

Fig. 5 Empirical bias, mean squared errors, and interquartile ranges (IQR) for β_0 (top panels) and for β_1 (bottom panels) as the regression coefficients $\Theta = (\beta_0, \beta_1)^\top$ in a Poisson regression model across 5,000 Monte Carlo replicates in the sensitivity analysis: $\hat{\Theta}$ (dotted lines), $\hat{\Theta}_{\text{naive}}$ (dashed lines), $\hat{\Theta}_{\text{cs}}$ (dot-dashed lines), and $\hat{\Theta}_{\text{tc}}$ (solid lines)

bias-corrected estimators for μ_2, μ_3, and μ_4 given by

$$\hat{\mu}_2 = \frac{1}{n}\sum_{r=1}^{n} \frac{n_r}{n_r - 1}\hat{\mu}_{2,r}, \quad \hat{\mu}_3 = \frac{1}{n}\sum_{r=1}^{n} \frac{n_r^2}{(n_r - 1)(n_r - 2)}\hat{\mu}_{3,r}, \text{ and}$$

$$\hat{\mu}_4 = \frac{1}{n}\sum_{r=1}^{n} \frac{n_r\left\{(n_r^2 - 2n_r + 3)\hat{\mu}_{4,r} - 3(2n_r - 3)\hat{\mu}_{2,r}^2\right\}}{(n_r - 1)(n_r - 2)(n_r - 3)}, \text{ respectively.}$$

The mean of U is the only moment that the tessarine-based debiased method needs, yet cannot be estimated by such replicate data, and thus validation or external data are needed to achieve identifiability.

7 Application in hockey data

An important metric used to assess the effectiveness of hockey players in the National Hockey League is the expected goals. Sports analysts use factors such as shot quality, shot location, and speed to calculate the number of expected goals a player would have scored on average during a game. This metric has a natural equivalent for the goaltender (as the player attempting to stop shots), which is the expected goals allowed. In this study, we consider a Poisson model in (19) for the number of wins for each goaltender (Y), regressing on the number of goals allowed (X), based on a dataset consisting of $n = 88$ records of win data from Hockey Reference (https://www.hockey-reference.com/leagues/NHL_2022_goalies.html) and expected goals data from Moneypuck.com (https://moneypuck.com/data.htm).

To demonstrate inferential methods for errors-in-variables problems, we create a surrogate of X by adding random noise from gamma$(1, 20)$ to the expected goals recorded in this dataset. The so-obtained surrogate covariate W corresponds to the measurement error $U = W - X$, with an estimated skewness of 2.3 and an estimated kurtosis of 12.06. Figure 6 presents in the left panel the normal quantile-quantile plot of the measurement errors. This plot, along with the estimated skewness and kurtosis, clearly suggests a deviation from normality, especially in the tails that appear to be heavier than that for Gaussian errors. These observations warrant regression analysis which accounts for non-Gaussian measurement errors that may not have mean zero.

Figure 6 shows in the right panel five estimates for the regression function $m(x) = E(Y|X = x) = \exp(\beta_0 + \beta_1 x)$, including the estimate based on the maximum likelihood estimate of $\Theta = (\beta_0, \beta_1)^\top$ using error-free data $\{(Y_r, X_r)\}_{r=1}^{88}$, denoted by $\hat{m}(x)$, its naive counterpart based on the error-contaminated data $\{(Y_r, W_r)\}_{r=1}^{88}$, denoted by $\hat{m}_{\text{naive}}(x)$, the corrected score estimate $\hat{m}_{\text{cs}}(x)$, the estimate involving complex constants in the spirit of $\hat{f}_{\text{cc}}(x)$, denoted by $\hat{m}_{\text{cc}}(x)$, and the tessarine-based debiased estimate $\hat{m}_{\text{tc}}(x)$. Evidently, $\hat{m}_{\text{tc}}(x)$ resembles $\hat{m}(x)$ most closely among the four estimated regression functions based on error-contaminated data. Both $\hat{m}_{\text{naive}}(x)$ and $\hat{m}_{\text{cs}}(x)$ seem to suggest a much weaker association between the number of wins and the number of goals allowed than what $\hat{m}(x)$ indicates, and $\hat{m}_{\text{cc}}(x)$ is similar to $\hat{m}_{\text{cs}}(x)$. Table 1 provides the estimates for Θ corresponding to the five estimated

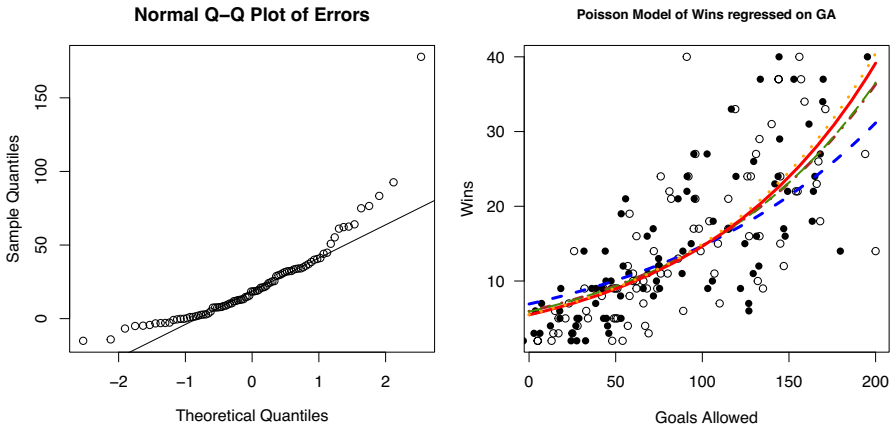

Fig. 6 The normal quantile-quantile plot (left panel) for expected goals−actual goals, and five estimates for the regression function (right panel): $\hat{m}(x)$ (dotted line), $\hat{m}_{\text{naive}}(x)$ (dashed line), $\hat{m}_{\text{cs}}(x)$ (long dashed line), $\hat{m}_{\text{cc}}(x)$ (dot-dashed line), and $\hat{m}_{\text{tc}}(x)$ (solid line), superimposed on error-free data (empty circles) and error-contaminated data (solid circles)

Table 1 Regression coefficients estimates in the estimated regression functions based on the hockey data, with the corresponding estimated standard errors in parentheses

Regression function	$\hat{\beta}_0$	$\hat{\beta}_1$
$\hat{m}(x)$	1.7089 (0.0687)	0.0100 (0.0006)
$\hat{m}_{\text{naive}}(x)$	1.9350 (0.0659)	0.0075 (0.0005)
$\hat{m}_{\text{cs}}(x)$	1.7747 (0.2092)	0.0091 (0.0022)
$\hat{m}_{\text{cc}}(x)$	1.7842 (0.2052)	0.0090 (0.0021)
$\hat{m}_{\text{tc}}(x)$	1.7008 (0.0921)	0.0098 (0.0010)

regression functions, along with the estimated standard errors based on the sandwich variance estimator (Boos and Stefanski, 2013)[Section7.2] that can be straightforwardly constructed for all five estimators based on their scores. All estimated covariate effects based on error-contaminated data are attenuated when compared to the estimate based on error-free data. But the debiased estimate $\hat{\beta}_{1,\text{tc}}(x)$ is least attenuated, and less variable than the other two non-naive estimations using the same data.

8 Discussion

A tessarine is a four-dimensional hypercomplex number. To continue on the path of bias reduction achieved by the proposed tessarine-based estimator $\hat{f}_{\text{tc}}(x)$, one can extend the debiased estimator construction based on a higher dimensional hypercomplex number, such as octonions that are eight-dimensional (Cayley, 1845). However, algebraic manipulations quickly become more complicated when using higher dimensional hypercomplex numbers in constructing an estimator, and the bias reduction beyond what is accomplished by $\hat{f}_{\text{tc}}(x)$ is expected to be lessened. This is because,

under certain assumptions, including the entirety of $f(x)$ and moment conditions on U, higher-order bias in (3)–(4) is dominated by terms of lower orders.

Referees of this paper brought up the potential direction of a simulation-based debiasing strategy similar to SIMEX utilizing tessarines. Although there exists a connection between the Monte Carlo corrected score estimator $\hat{f}_{\text{mc}}(X)$ and SIMEX (Carroll et al., 2006)[Section7.4.4], the connection between $\hat{f}_{\text{tc}}(X)$ and SIMEX is much less obvious mainly because we do not introduce additional random quantities in $\hat{f}_{\text{tc}}(X)$. The added complex random variable iV in $\hat{f}_{\text{mc}}(X)$ is the limit of the real random quantity that SIMEX simulates; and this complex random variable also dictates the target of extrapolation in SIMEX. A tessarine-based SIMEX may be possible after revising $\hat{f}_{\text{tc}}(X)$ to construct a debiased estimator involving a tessarine random quantity.

Except for the example of linear regression in Sect. 4.2, we have focused on errors-in-variables problems with a scalar error-prone variable, allowing us to focus on methodology development mostly involving univariate hypercomplex numbers. Generalization of this line of development adapted to more general errors-in-variables problems with multiple variables subject to correlated non-Gaussian measurement error contamination is an interesting topic for future research. This generalization is necessary to adapt the tessarine-based error correction strategy to variable selection problems and other statistical learning tasks involving multiple error-prone covariates.

Appendix

A .1 Proof of Lemma 1

Proof By the definition of circular symmetry, $e^{it}C$ and the circularly symmetric C follow the same distribution for an arbitrary constant t on the real line $\mathbb{R}$. Hence, assuming moments of C of all orders exist,

$$E(e^{i\ell t}C^{\ell}) = E(C^{\ell}), \text{ for all} t \in \mathbb{R} \text{and} \ell = 1, 2, \ldots,$$

which is equivalent to

$$E(C^{\ell})(1 - e^{i\ell t}) = 0, \text{ for all} t \in \mathbb{R} \text{and} \ell = 1, 2, \ldots.$$

By the arbitrariness of t, the above identity holds if and only if $E(C^{\ell}) = 0$, for $\ell = 1, 2, \ldots$. This completes the proof. □

A .2 Proof of Theorem 2

Proof With $C = U - E(U) + i\{V - E(V)\}, E(C) = E\{U - E(U)\} + iE\{V - E(V)\} = 0$. Hence, the pseudovariance is $\text{pvar}(C) = E(C^2)$. Because

$$C^2 = \{U - E(U)\}^2 - \{V - E(V)\}^2 + 2i\{U - E(U)\}\{V - E(V)\},$$

we have $\mathrm{pvar}(C) = \mathrm{var}(U) - \mathrm{var}(V) + 2iE\{U - E(U)\}E\{V - E(V)\} = 0$, since U and V are independent and identically distributed, thus have the same variance. Because C has a vanishing pseudovariance, by definition, C is proper. This completes the proof. □

A .3 Proof of Lemma 2

Proof Diagonalizing the matrix-version C yields

$$C = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} a - c + i(b-d) & 0 \\ 0 & a + c + i(b+d) \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Because $(-1/\sqrt{2}, 1/\sqrt{2})^\top$ and $(1/\sqrt{2}, 1/\sqrt{2})^\top$ are orthonormal, we have

$$\begin{aligned} & C^\ell \\ &= \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \{a - c + i(b-d)\}^\ell & 0 \\ 0 & \{a + c + i(b+d)\}^\ell \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \\ &= \frac{1}{2}\left(\left[\{a - c + i(b-d)\}^\ell + \{a + c + i(b+d)\}^\ell \right] \boldsymbol{I}_2 \right. \\ &\quad \left. + \left[\{a + c + i(b+d)\}^\ell - \{a - c + i(b-d)\}^\ell \right] \boldsymbol{J}_2 \right). \end{aligned}$$

The diagonal entry of C^ℓ is thus, by Euler's formula,

$$\begin{aligned} & \frac{1}{2}\left(\left\{ (a-c)^2 + (b-d)^2 \right\}^{\ell/2} \exp[i\ell \times \arg\{a - c + i(b-d)\}] \right. \\ & \left. + \left\{ (a+c)^2 + (b+d)^2 \right\}^{\ell/2} \exp[i\ell \times \arg\{a + c + i(b+d)\}] \right), \end{aligned}$$

where $\arg(t)$ is the argument of a complex number t. Extracting the real part of the above expression gives

$$\begin{aligned} \Re(C^\ell) = \frac{1}{2}\Big[& \left\{ (a-c)^2 + (b-d)^2 \right\}^{\ell/2} \cos(\ell \times \arg\{a - c + i(b-d)\}) \\ & + \left\{ (a+c)^2 + (b+d)^2 \right\}^{\ell/2} \cos(\ell \times \arg\{a + c + i(b+d)\}) \Big]. \end{aligned}$$

This completes the proof. □

A .4 Proof of Theorem 3

Proof For a square real or complex matrix C, the matrix exponential is $\exp(C) = \sum_{\ell=0}^{\infty} C^{\ell}/\ell!$ (Hall, 2013).

Expressing the tessarine $C = a + bi + cj + dk$ as a 2×2 complex matrix and using Lemma 2, we have

$$\exp(C) = \frac{1}{2}\begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{B} & \mathcal{A} \end{bmatrix}, \tag{35}$$

where

$$\begin{aligned} \mathcal{A} &= \sum_{\ell=0}^{\infty} \frac{\{a - c + i(b - d)\}^{\ell}}{\ell!} + \sum_{\ell=0}^{\infty} \frac{\{a + c + i(b + d)\}^{\ell}}{\ell!} \\ &= \exp\{a - c + i(b - d)\} + \exp\{a + c + i(b + d)\} \\ &= 2\exp(a + ib)\cosh(c + id), \end{aligned}$$

and

$$\begin{aligned} \mathcal{B} &= \sum_{\ell=0}^{\infty} \frac{\{a + c + i(b + d)\}^{\ell}}{\ell!} - \sum_{\ell=0}^{\infty} \frac{\{a - c + i(b - d)\}^{\ell}}{\ell!} \\ &= \exp\{a + c + i(b + d)\} - \exp\{a - c + i(b - d)\} \\ &= 2\exp(a + ib)\sinh(c + id). \end{aligned}$$

Therefore,

$$\begin{aligned} \exp(C) &= \exp(a + bi)\begin{bmatrix} \cosh(c + di) & \sinh(c + di) \\ \sinh(c + di) & \cosh(c + di) \end{bmatrix} \\ &= \exp(a + bi)\{\cosh(c + di)\boldsymbol{I}_2 + \sinh(c + di)\boldsymbol{J}_2\}. \end{aligned}$$

Extracting the real part of the diagonal entry of $\exp(C)$ gives

$$\begin{aligned} \Re\{\exp(C)\} &= \Re\{\exp(a + bi)\cosh(c + di)\} \\ &= \exp(a)\{\cos(b)\cos(d)\cosh(c) - \sin(b)\sin(d)\sinh(c)\}. \end{aligned}$$

This completes the proof. □

A .5 Derivations of the real parts in the debiased score in logistic regression

The debiased score $\tilde{\Psi}(Y, W, \Theta)$ involves $\Re\{\mathcal{G}(\beta_0 + \beta_1 \mathscr{T})\}$ and $\Re\{\mathcal{G}(\beta_0 + \beta_1 \mathscr{T})\mathscr{T}\}$, where $\mathcal{G}(t) = 1/\{1 + \exp(-t)\}$ and $\mathscr{T} = W - E(U) + bi + cj + dk$. To derive these real parts, we first derive the matrix version of the expit function evaluated at a tessarine $t = R_t + I_t i + J_t j + K_t k$, where $R_t, I_t, J_t, K_t \in \mathbb{R}$ are the real part and the imaginary parts of t.

Firstly, the matrix form of the tessarine basis element 1 is $\mathbb{1} = \boldsymbol{I}_2$, that is, the 2×2 identity matrix. Secondly, by Theorem 3,

$$\begin{aligned}\exp(-t) &= \exp(-R_t - I_t i)\begin{bmatrix}\cosh(-J_t - K_t i) & \sinh(-J_t - K_t i)\\ \sinh(-J_t - K_t i) & \cosh(-J_t - K_t i)\end{bmatrix}\\ &= \exp(-R_t - I_t i)\begin{bmatrix}\cosh(J_t + K_t i) & -\sinh(J_t + K_t i)\\ -\sinh(J_t + K_t i) & \cosh(J_t + K_t i)\end{bmatrix}.\end{aligned}$$

Summing these two matrices gives

$$\begin{aligned}&D(t) = \mathbb{1} + \exp(-t)\\ &= \begin{bmatrix}1 + \exp(-R_t - I_t i)\cosh(J_t + K_t i) & -\exp(-R_t - I_t i)\sinh(J_t + K_t i)\\ -\exp(-R_t - I_t i)\sinh(J_t + K_t i) & 1 + \exp(-R_t - I_t i)\cosh(J_t + K_t i)\end{bmatrix}.\end{aligned}$$

Thirdly, expressing the reciprocal of a tessarine as the inverse of a complex matrix, we have $\mathcal{G}(t) = \{D(t)\}^{-1}$. Lastly, $\Re\{\mathcal{G}(t)\}$ is equal to the real part of the first diagonal entry of $\mathcal{G}(t)$. Given $D(t)$ above, we have the first diagonal entry of $\mathcal{G}(t)$ as follows,

$$\begin{aligned}&\mathcal{G}(t)[1,1] = \frac{D(t)[2,2]}{\det\{D(t)\}}\\ &= \frac{1 + \exp(-R_t - I_t i)\cosh(J_t + K_t i)}{\{1 + \exp(-R_t - I_t i)\cosh(J_t + K_t i)\}^2 - \{\exp(-R_t - I_t i)\sinh(J_t + K_t i)\}^2}.\end{aligned}$$

Extracting the real part of $\mathcal{G}(t)[1,1]$ gives $\Re\{\mathcal{G}(t)\}$. Setting $t = \beta_0 + \beta_1 \mathscr{T}$ gives what we need as $\Re\{\mathcal{G}(\beta_0 + \beta_1 \mathscr{T})\}$ in the debiased score.

Consider another tessarine $s = R_s + I_s i + J_s j + K_s k$. Then

$$\mathcal{G}(t)s = \{D(t)\}^{-1}\begin{bmatrix}R_s + I_s i & J_s + K_s i\\ J_s + K_s i & R_s + I_s i\end{bmatrix},$$

and its first diagonal entry is

$$\begin{aligned}&\{\mathcal{G}(t)s\}[1,1]\\ &= \frac{D(t)[2,2](R_s + I_s i) - D(t)[1,2](J_s + K_s i)}{\det\{D(t)\}}\\ &= \frac{\{1 + \exp(-R_t - I_t i)\cosh(J_t + K_t i)\}(R_s + I_s i) + \exp(-R_t - I_t i)\sinh(J_t + K_t i)(J_s + K_s i)}{\{1 + \exp(-R_t - I_t i)\cosh(J_t + K_t i)\}^2 - \{\exp(-R_t - I_t i)\sinh(J_t + K_t i)\}^2}.\end{aligned}$$

Extracting the real part of the above expression gives $\Re(\mathcal{G}(t)s\}[1,1])$, and setting $t = \beta_0 + \beta_1\mathscr{T}$ and $s = \mathscr{T}$ gives what we need as $\Re\{\mathcal{G}(\beta_0 + \beta_1\mathscr{T})\mathscr{T}\}$ in the debiased score.

A .6 Evaluation of the logistic kernel at a tessarine

For $C = a\mathbb{1} + b\mathbb{I} + c\mathbb{J} + d\mathbb{K}$, by Theorem 3, we have

$$\begin{aligned}
&2 + \exp(C) + \exp(-C)\\
&= 2\boldsymbol{I}_2 + \exp(a+bi)\{\cosh(c+di)\boldsymbol{I}_2 + \sinh(c+di)\boldsymbol{J}_2\}\\
&\quad + \exp(-a-bi)\{\cosh(c+di)\boldsymbol{I}_2 - \sinh(c+di)\boldsymbol{J}_2\}\\
&= 2\{1 + \cosh(a+bi)\cosh(c+di)\}\boldsymbol{I}_2 + 2\sinh(a+bi)\sinh(c+di)\boldsymbol{J}_2.
\end{aligned}$$

Expressing $K(C) = 1/(2 + e^C + e^{-C})$ as the inverse of the above 2×2 complex matrix, and using the result on the inverse of the sum of two matrices in Miller (1981), we have

$$\begin{aligned}
&K(C)\\
&= \frac{1}{2\{1 + \cosh(a+bi)\cosh(c+di)\}}\left\{\boldsymbol{I}_2 - \frac{\sinh(a+bi)\sinh(c+di)}{1 + \cosh(a+bi)\cosh(c+di)}\boldsymbol{J}_2\right\}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\Re\{K(C)\} &= \Re\{K(C)[1,1]\}\\
&= \Re\left[\frac{1}{2\{1 + \cosh(a+bi)\cosh(c+di)\}}\right]\\
&= \frac{2\xi(a,b,c,d)}{\xi(a,b,c,d)^2 + \psi(a,b,c,d)^2},
\end{aligned}$$

with $\xi(a,b,c,d) = 2\{2 + \cosh(a+c)\cos(b+d) + \cosh(a-c)\cos(b-d)\}$ and $\psi(a,b,c,d) = \sinh(a+c)\sin(b+d) + \sinh(a-c)\sin(b-d)$.

A .7 Another tessarine-based estimator for $f(X)$ when U is symmetric

When the measurement error distribution is symmetric, the derivation of (b,c,d) that leads to $\Re\{E(C^\ell)\} = 0$, for $\ell = 2,3,4$, is vastly simplified. More specifically, in this case, a solution to (9) is

$$b = \sqrt{\frac{\sigma_u^2 + \sqrt{\mu_4 - 4\sigma_u^4}}{2}},\quad c = \sqrt{b^2 - \sigma_u^2},\quad d = 0, \tag{36}$$

provided that the kurtosis $\mu_4/\sigma_u^4 > 5$. Denote the solution for b and c in (36) by b^* and c^*, respectively, which are the same as b^* and c^* used in (22). Moreover, one can show that $\mathrm{Re}\{E(C^5)\}$ is proportional to bcd when U is symmetric, and thus the tessarine $C = U - E(U) + b^*i + c^*j$ has $\mathrm{Re}\{E(C^\ell)\} = 0$ for $\ell = 1, 2, 3, 4, 5$. As a result, the tessarine-based estimator for $f(X)$, $\hat{f}_{\mathrm{tc}}(X)$, given in (10) has its bias given by, similar to (3),

$$\mathrm{Bias}\{\hat{f}_{\mathrm{tc}}(X)\} = \sum_{\ell=6}^{\infty} \frac{f^{(\ell)}(X)}{\ell!} \Re\{E(C^\ell)\}. \tag{37}$$

The dominating bias of $\hat{f}_{\mathrm{tc}}(X)$ now is $\{f^{(6)}(X)/6!\}\Re\{E(C^6)\}$ according to (37), where $\Re\{E(C^6)\} = \mu_6 - 27\mu_4\sigma_u^2 + 74\sigma_u^6$ for the tessarine C defined above. Using $f^{(6)}(W)$ to estimate $f^{(6)}(X)$, we propose to estimate the dominating bias in (37) via $\{f^{(6)}(W)/6!\}\Re\{E(C^6)\}$ and to subtract this estimated dominating bias from $\hat{f}_{\mathrm{tc}}(X)$. This leads to the estimator for $f(X)$ given by

$$\begin{aligned}\hat{f}_{\mathrm{tcs}}(X) &= \hat{f}_{\mathrm{tc}}(X) - \frac{f^{(6)}(W)}{6!}\Re\{E(C^6)\} \\ &= \Re\{f(W - E(U) + b^*i + c^*i)\} - \frac{f^{(6)}(W)}{6!}\left(\mu_6 - 27\mu_4\sigma_u^2 + 74\sigma_u^6\right),\end{aligned}$$

which has the potential to reduce bias in $\hat{f}_{\mathrm{tc}}(X)$ in the presence of symmetric U.

A .8 Additional simulation results

Figure 7 is the counterpart of Fig. 1 with a larger sample size at $n = 500$ in the context of Poisson regression with gamma measurement error.

Figure 8 is parallel to Fig. 3 with a large sample size of $n = 1000$ in the context of logistic regression.

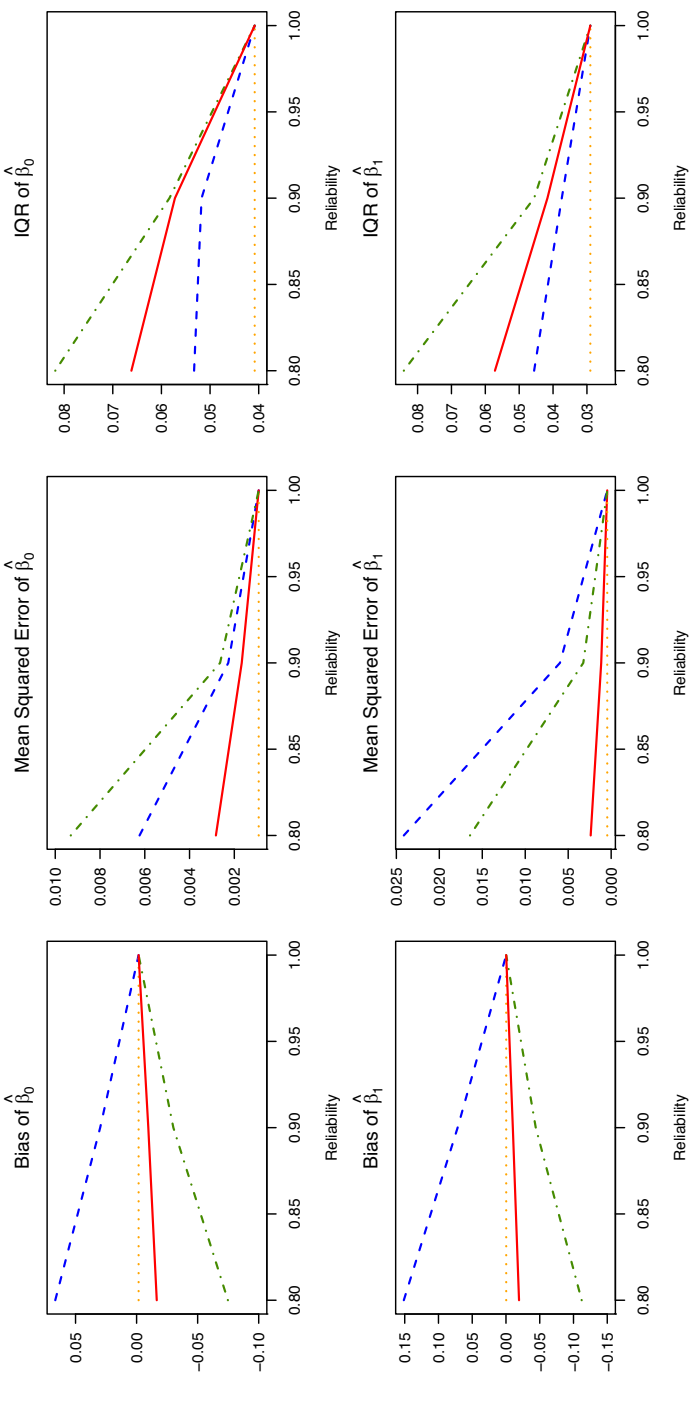

Fig. 7 Empirical bias, mean squared errors, and interquartile ranges (IQR) of four estimators for β_0 (top panels) and for β_1 (bottom panels) as the regression coefficients $\Theta = (\beta_0, \beta_1)^\top$ in the Poisson regression model when $n = 500$ and gamma measurement error: the maximum likelihood estimator $\hat{\Theta}$ based on error-free data (dotted lines), the naive estimator $\hat{\Theta}_{\text{naive}}$ (dashed lines), the corrected score estimator $\hat{\Theta}_{\text{cs}}$ (dot-dashed lines), and the tessarine-based estimator $\hat{\Theta}_{\text{tc}}$ (solid lines)

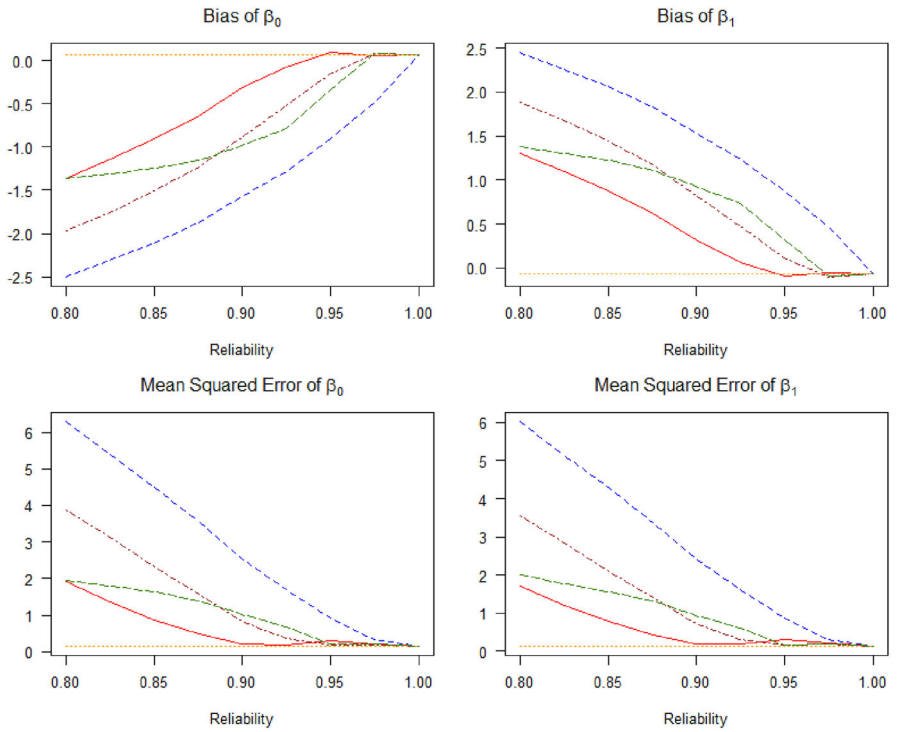

Fig. 8 Empirical bias, mean squared errors, and interquartile ranges (IQR) of five estimators for β_0 (top panels) and for β_1 (bottom panels) as the regression coefficients $\Theta = (\beta_0, \beta_1)^\top$ in the logistic regression when $n = 1000$: the maximum likelihood estimator $\hat{\Theta}$ based on error-free data (dotted lines, cross points), the naive estimator $\hat{\Theta}_{\text{naive}}$ (short dashed lines, square points), the Monte Carlo corrected score estimator $\hat{\Theta}_{\text{mc}}$ (dot-dashed lines, triangle points), the conditional score estimator $\hat{\Theta}_{\text{cd}}$ (long dashed lines, x points), and the tessarine-based estimator $\hat{\Theta}_{\text{tc}}$ (solid lines, circle points)

Funding Open access funding provided by the Carolinas Consortium.

References

Adali, T., Schreier, P. J., Scharf, L. L. (2011). Complex-valued signal processing: The proper way to deal with impropriety. *IEEE Transactions on Signal Processing, 59*(11), 5101–5125.

Adcock, C., Eling, M., Loperfido, N. (2015). Skewed distributions in finance and actuarial science: A review. *The European Journal of Finance, 21*(13–14), 1253–1281.

Arellano-Valle, R., Ozan, S., Bolfarine, H., Lachos, V. (2005). Skew normal measurement error models. *Journal of Multivariate Analysis, 96*(2), 265–281.
Augustin, T. (2004). An exact corrected log-likelihood function for cox's proportional hazards model under measurement error and some extensions. *Scandinavian Journal of Statistics, 31*(1), 43–50.
Bareiss, E. (1959). *Resultant procedure: A method for finding the zeroes of real polynomials*. New York: Argonne National Laboratory.
Boos, D. D., Stefanski, L. A. (2013). *Essential statistical inference: Theory and methods* (Vol. 591). New York: Springer.
Buonaccorsi, J. P. (2010). *Measurement error: Models, methods, and applications*. Boca Raton: CRC Press.
Cameron, A. (1998). *Regression analysis of count data*. Cambridge: Cambering University Press.
Carroll, R. J., Ruppert, D., Stefanski, L. A., Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: A modern perspective*. London: CRC Press.
Cayley, A. (1845). Xxviii. On Jacobi's elliptic functions, in reply to the Rev. Brice Bronwin and on quaternions: To the editors of the philosophical magazine and journal. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 26*(172), 208–211.
Chen, L.-P. (2023). De-noising boosting methods for variable selection and estimation subject to error-prone variables. *Statistics and Computing, 33*(2), 38.
Chen, L.-P. (2025). Variable selection via penalized ridge regression with error-prone variables. *Annals of the Institute of Statistical Mathematics, 78*, 225–261.
Chen, L.-P. and Yi, G. Y. (2022). *Robust feature screening for ultrahigh-dimensional censored data subject to measurement error* (pp. 23–53). Cham: Springer International Publishing.
Cheng, C., Spiegelman, D., Li, F. (2023). Mediation analysis in the presence of continuous exposure measurement error. *Statistics in medicine,42*(11), 1669–1686.
Cheng, C.-L., Schneeweiss, H. (1997). Polynomial regression with errors in variables. *Journal of the Royal Statistical Society: Series B, 60*, 189–199.
Cockle, J. (1848). On certain functions resembling quaternions and on a new imaginary in algebra. *Philosophical Magazine, 33*, 435–439.
Delaigle, A., Fan, J., Carroll, R. J. (2009). A design-adaptive local polynomial estimator for the errors-in-variables problem. *Journal of the American Statistical Association, 104*(485), 348–359.
Delaigle, A., Gijbels, I. (2004). Practical bandwidth selection in deconvolution kernel density estimation. *Computational Statistics & Data Analysis, 45*, 249–267.
Delaigle, A., Hall, P. (2016). Methodology for non-parametric deconvolution when the error distribution is unknown. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 78*(1), 231–252.
Faller, J., Martin, J., Elster, C. (2025). Deep errors-in-variables using a diffusion model. *Machine Learning,114*(4), 107.
Fisher, R. A. (1930). Moments and product moments of sampling distributions. *Proceedings of the London Mathematical Society, 2*(1), 199–238.
Fuller, W. A. (2009). *Measurement error models*. New York: John Wiley & Sons.
Goodman, N. R. (1963). Statistical analysis based on a certain multivariate complex gaussian distribution (an introduction). *The Annals of Mathematical Statistics, 34*(1), 152–177.
Hall, B. C. (2013). *Lie groups, Lie algebras, and representations*. New York: Springer.
Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Tibshirani, R., Friedman, J. (2009). Kernel smoothing methods. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (pp. 191–218). New York: Springer.
Huang, X., Zhou, H. (2020). Conditional density estimation with covariate measurement error. *Electronic Journal of Statistics, 14*, 970–1023.
Jacobson, N. (2012). *Basic algebra I*. New York: Courier Corporation.
Levin, B. Y. (1996). *Lectures on entire functions* (Vol. 150). New York: American Mathematical Soc.
Li, X., Huang, X. (2019). Linear mode regression with covariate measurement error. *Canadian Journal of Statistics, 47*(2), 262–280.
Liu, Q., Huang, X. (2024). Parametric modal regression with error in covariates. *Biometrical Journal, 66*(1), 2200348.
Luna-Elizarrarás, M. E., Shapiro, M., Struppa, D. C., Vajiac, A. (2015). *Bicomplex holomorphic functions: The algebra, geometry and analysis of bicomplex numbers*. Basel: Birkhäuser.
Miller, K. S. (1981). On the inverse of the sum of matrices. *Mathematics Magazine, 54*(2), 67–72.
Millimet, D. L., Parmeter, C. F. (2022). Accounting for skewed or one-sided measurement error in the dependent variable. *Political Analysis, 30*(1), 66–88.

Nakamura, T. (1990). Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika, 77*(1), 127–137.
Neeser, F. D., Massey, J. L. (1993). Proper complex random processes with applications to information theory. *IEEE Transactions on Information Theory, 39*(4), 1293–1302.
Nghiem, L. H., Byrd, M. C., Potgieter, C. J. (2020). Estimation in linear errors-in-variables models with unknown error distribution. *Biometrika, 107*(4), 841–856.
Novick, S., Stefanski, L. (2002). Corrected score estimation via complex variable simulation extrapolation. *Journal of the American Statistical Association,97*(458), 472–481.
Ollila, E., Tyler, D. E., Koivunen, V., Poor, H. V. (2012). Complex elliptically symmetric distributions: Survey, new results and applications. *IEEE Transactions on Signal Processing, 60*(11), 5597–5625.
Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics, 33*(3), 1065–1076.
Picinbono, B. (1994). On circularity. *IEEE Transactions on Signal Procession, 42*(12), 3473–3482.
Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics, 27*(3), 832–837.
Schreier, P. J., Scharf, L. L. (2010). *Statistical signal processing of complex-valued data: The theory of improper and noncircular signals*. Cambridge: Cambridge University Press.
Song, X., Huang, Y. (2005). On corrected score approach for proportional hazards model with covariate measurement error. *Biometrics, 61*(3), 702–714.
Stefanski, L., Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement- error models. *Biometrika, 74*(4), 703–716.
Stefanski, L., Cook, J. (1995). Simulation-extrapolation: The measurement error jackknife. *Journal of the American Statistical Association, 90*(432), 1247–1256.
Stefanski, L. A. (1989). Unbiased estimation of a nonlinear function a normal mean with application to measurement-error models. *Communications in Statistics-Theory and Methods, 18*(12), 4335–4358.
Stefanski, L. A., Boos, D. D. (2002). The calculus of M-estimation. *The American Statistician, 56*(1), 29–38.
Stefanski, L. A., Carroll, R. J. (1990). Deconvolving kernel density estimators. *Statistics, 21*(2), 169–184.
Tomaya, L. C., de Castro, M. (2018). A heteroscedastic measurement error model based on skew and heavy-tailed distributions with known error variances. *Journal of Statistical Computation and Simulation, 88*(11), 2185–2200.
Torabi, M., Ghosh, M., Myung, J., Steel, M. (2023). Measurement error in linear regression models with fat tails and skewed errors. *Communications in Statistics - Theory and Methods, 52*(15), 5407–5426.
Wang, C. (2006). Corrected score estimator for joint modeling of longitudinal and failure time data. *Statistica Sinica, 16*, 235–253.
Wang, H. J., Stefanski, L. A., Zhu, Z. (2012). Corrected-loss estimation for quantile regression with covariate measurement errors. *Biometrika, 99*(2), 405–421.
Wu, Y., Ma, Y., Yin, G. (2015). Smoothed and corrected score approach to censored quantile regression with measurement errors. *Journal of the American Statistical Association, 110*(512), 1670–1683.
Yi, G. Y. (2017). *Statistical analysis with measurement error or misclassification: Strategy method and application*. New York: Springer.
Yi, G. Y., Delaigle, A., Gustafson, P. (2021). *Handbook of measurement error models*. Boca Raton: CRC Press.
Zhang, F. (1997). Quaternions and matrices of quaternions. *Linear Algebra and Its Applications, 251*, 21–57.
Zhao, H., Wang, T. (2024). Debiased high-dimensional regression calibration for errors-in-variables log-contrast models. *Biometrics, 80*(4), ujad153.
Zhou, H., Huang, X. (2016). Nonparametric modal regression in the presence of measurement error. *Electronic Journal of Statistics, 10*, 3579–3620.