

SIMEX and standard error estimation in semiparametric measurement error models

Tatiana V. Apanasovich

*Division of Biostatistics, Thomas Jefferson University
Philadelphia, PA 19107*

Raymond J. Carroll

*Department of Statistics, Texas A&M University
College Station, TX 77843*

and

Arnab Maity

*Department of Biostatistics, Harvard School of Public Health
Boston, MA 02115*

Abstract: SIMEX is a general-purpose technique for measurement error correction. There is a substantial literature on the application and theory of SIMEX for purely parametric problems, as well as for purely nonparametric regression problems, but there is neither application nor theory for semiparametric problems. Motivated by an example involving radiation dosimetry, we develop the basic theory for SIMEX in semiparametric problems using kernel-based estimation methods. This includes situations that the mismeasured variable is modeled purely parametrically, purely nonparametrically, or that the mismeasured variable has components that are modeled both parametrically and nonparametrically. Using our asymptotic expansions, easily computed standard error formulae are derived, as are the bias properties of the nonparametric estimator. The standard error method represents a new method for estimating variability of nonparametric estimators in semiparametric problems, and we show in both simulations and in our example that it improves dramatically on first order methods.

We find that for estimating the parametric part of the model, standard bandwidth choices of order $O(n^{-1/5})$ are sufficient to ensure asymptotic normality, and undersmoothing is not required. SIMEX has the property that it fits misspecified models, namely ones that ignore the measurement error. Our work thus also more generally describes the behavior of kernel-based methods in misspecified semiparametric problems.

AMS 2000 subject classifications: 62G05, 62H12, 62F12, 62J05.

Keywords and phrases: Berkson measurement errors, measurement error, misspecified models, nonparametric regression, radiation epidemiology, semiparametric models, SIMEX, simulation-extrapolation, standard error estimation, uniform expansions.

Received December 2008.

Contents

1	Introduction	319
2	SIMEX	321
3	Likelihood functions and estimators	322
4	SIMEX when X is modeled nonparametrically	323
4.1	Introduction and theoretical development	323
4.2	Standard error estimation	325
4.2.1	Estimation for the nonparametric part	325
4.2.2	Estimation for the parametric part	326
5	SIMEX when X is modeled parametrically	326
5.1	Introduction and theoretical development	326
5.2	Standard error estimation	327
5.2.1	Estimation for the nonparametric part	328
5.2.2	Estimation for the parametric part	328
5.3	When both X and Z are mismeasured	328
6	Simulations	329
6.1	The partially linear model	329
6.2	Partially linear logistic model	330
6.3	Why SIMEX?	331
7	Nevada test site thyroid disease study	332
7.1	Introduction	332
7.2	Data analysis	334
7.2.1	Various fits	334
7.2.2	Standard Errors	335
8	Discussion	337
A	Sketch of technical arguments	338
A.1	Outline of the Appendix	338
A.2	Assumptions	338
A.3	Results for general models	339
A.3.1	Nonparametric regression expansions	339
A.3.2	Semiparametric Models	340
A.4	Sketch of Result 3	340
A.5	Sketch of Result 4	342
A.6	SIMEX when X is modeled nonparametrically	342
A.6.1	Justification of (10)	343
A.7	SIMEX when X is modeled parametrically	344
	Acknowledgments	344
	References	345

1. Introduction

Regression models with measurement errors arise frequently in practice and have attracted attention in the statistics literature. A semiparametric regression model with errors in variable has been considered by several authors in the

attempt to develop a measurement error calibration when the errors are in the linear part of linear regression [17], or generalized linear regression [18]. [43] used a method of moments and deconvolution to construct the calibration for the case of partially linear models when the mismeasured covariate appears in parametric and nonparametric parts. SIMEX has considered only in the case of partially linear model and when the measurement errors are in the linear part [13]. However, we are after very general results, and are not restricting attention to simple partially linear models.

Indeed, the purpose of this paper is to derive the general theory for a popular alternative to regression-calibration, a simulation-extrapolation method (SIMEX) in the situations where the mismeasured variable is modeled purely parametrically, purely nonparametrically, or that the mismeasured variable has components that are modeled both parametrically and nonparametrically.

The SIMEX method [8, 37] is a general-purpose, widely applicable method for correcting parameter estimates for the biases induced by measurement error in covariates. It is a functional method, in the sense that it makes no assumptions about the distribution of the unobserved true covariate.

A major strength of SIMEX is that it is extremely easy to implement: it requires only a program for computing estimates in the absence of measurement error, and the ability to simulate adding additional measurement error to the process. A brief description of the method is given in Section 2: many more details are available in [4]. There is a suite a programs for implementing SIMEX in Stata, as well as in R at <http://cran.r-mirror.de/src/contrib/Descriptions/simex.html>.

SIMEX has found many applications in parametric modeling, e.g., biostatistics and epidemiology [10, 25, 26], ecology [11, 12, 14, 36], data confidentiality [16] and longitudinal modeling [19, 41, 42]. SIMEX has also found application in local polynomial nonparametric regression and in spline-based methodology [2, 3, 9, 38].

While SIMEX is well-studied in purely parametric and purely nonparametric problems, to the best of our knowledge, SIMEX was considered only in the case of partially linear model and when the measurement errors are in the linear part [13]. In this paper we derive the general theory of SIMEX in semiparametric problems and describe a simple method to estimate standard errors for both the parametric and the nonparametric parts of the model. We also describe the first order bias properties of the nonparametric estimators. We show using our example that the standard error method improves dramatically on first order methods for estimating standard errors of the nonparametric components, and is thus of importance even when there is no measurement error.

An outline of this paper is as follows. In Section 2, we briefly define the SIMEX method. Section 3 briefly describes well-known standard estimation algorithms appropriate to semiparametric estimation. Sections 4 and 5 state the main results for the cases that the variable measured with error is modeled nonparametrically and parametrically, respectively. In the latter section we also add a brief discussion of what happens when the mismeasured variable is in both parts of the model. Importantly, both these sections describe sim-

ple methods for constructing standard error estimates of both parametric and nonparametric components, ones that in our example and simulations improve greatly upon first order methods. Section 6 presents a simulation study to evaluate the performance of SIMEX itself and more importantly our approach to variance estimation in semiparametric problems. In Section 7, we describe an important, complex problem in radiation epidemiology involving measurement error. Section 8 gives concluding remarks. All technical details are sketched in an Appendix.

2. SIMEX

The SIMEX method is both well-known and straightforward. Suppose we have a problem in which the predictor X cannot be observed, but we observe $W = X + U$, where $U = \text{Normal}(0, \Sigma_u)$. Some components of U may equal zero indicating no measurement error in that component. Consider a set of values $0 = \lambda_1 < \lambda_2 \cdots < \lambda_J$.

- In the *simulation step*, additional independent measurement errors with covariance matrix $\lambda_j \Sigma_u$ are generated and added to the original W data, thereby creating data sets with successively larger measurement error variances. For the j^{th} data set, the total measurement error variance is $\Sigma_u + \lambda_j \Sigma_u = (1 + \lambda_j) \Sigma_u$.
- Next, estimates are obtained from each of the generated contaminated data sets, using an algorithm that would have been used if there were no measurement error, i.e., one's favorite method.
- The simulation and estimation steps are repeated a large number B times, and the average value of the estimate for each level of contamination λ_j is calculated. These averages are plotted against the λ values and a regression technique, for example, polynomial least squares, is used to fit an extrapolant function to the averaged, error-contaminated estimates.
- Extrapolation to the ideal case of no measurement error ($\lambda = -1$) yields the SIMEX estimate.

Remark 1 In general parametric problems, the “favorite method” will be equivalent to solving an estimating function. Because of the measurement error in X , the estimating function is biased because the underlying model is misspecified. This causes no difficulty, since it is well-known what happens to estimating function methods for misspecified models. Similarly, in nonparametric regression, the “favorite method” will be equivalent to solving a local estimating equation, and the same general remarks apply.

Remark 2 Semiparametric problems, in contrast, combine global and local estimating equations, and their properties under misspecified models are not well-understood. For example, see Ai and Chen (2007), for a sieve-based approach to this general issue. Our work can be looked at as addressing the problem is misspecified semiparametric models within the kernel framework, which is then applied to the SIMEX algorithm itself.

3. Likelihood functions and estimators

This paper considers a wide class of semiparametric problems with some covariates modeled parametrically and one covariate entering the model through a nonparametric function. Important examples of such model are the partially linear model and partially linear logistic model considered in Section 6. For example, with X measured with error and Z measured exactly, and $H(\cdot)$ being the logistic distribution function, the partially linear logistic models with response Y can take either of two forms:

$$\begin{aligned} \text{pr}(Y = 1|X, Z) &= H\{X^T\beta + \theta(Z)\}; \\ \text{pr}(Y = 1|X, Z) &= H\{Z^T\beta + \theta(X)\}. \end{aligned}$$

In the first model, the variable measured with error is modeled parametrically, while in the second model it is modeled nonparametrically. We will derive results for both cases, and of course for much more general models.

SIMEX is based on repeatedly adding more measurement error and then calculating a standard estimator on this remeasured data set. In this section, we describe the well-known standard basic profile algorithm when estimation in the no-measurement error context is based upon maximization of a criterion function, e.g., a loglikelihood function. We will phrase things in terms of loglikelihoods, but as in [20], the case of any criterion function is similar.

Let $K(\cdot)$ be a smooth symmetric density function with bounded support, let h be a bandwidth, and let $K_h(z) = h^{-1}K(z/h)$. Define $\phi_k = \int z^k K(z)dz$ and $G_h(z) = (1, z/h)^T$.

Here we follow a traditional notation from the measurement error literature, where Y denotes the response, X denotes the covariate(s) measured with error and Z denotes the other covariate(s). Likelihood problems take two forms, depending on whether X is modeled parametrically or nonparametrically. In the former case, the loglikelihood is $\mathcal{L}\{Y, X, \theta(Z), \beta\}$, while in the latter the loglikelihood is $\mathcal{L}\{Y, Z, \theta(X), \beta\}$.

To handle both of these cases, we define generic random variables $(\mathcal{A}, \mathcal{D})$, and consider the loglikelihood as $\mathcal{L}\{Y, \mathcal{A}, \theta(\mathcal{D}), \beta\}$, where \mathcal{D} is scalar. For our notation about derivatives, make the following definitions:

$$\begin{aligned} \mathcal{L}_\theta\{Y, \mathcal{A}, \theta(\mathcal{D}), \beta\} &= (\partial/\partial v)\mathcal{L}(Y, \mathcal{A}, v, \beta)|_{v=\theta(\mathcal{D})}; \\ \mathcal{L}_\beta\{Y, \mathcal{A}, \theta(\mathcal{D}), \beta\} &= (\partial/\partial\beta)\mathcal{L}\{Y, \mathcal{A}, \theta(\mathcal{D}), \beta\}; \\ \mathcal{L}_{\theta\theta}\{Y, \mathcal{A}, \theta(\mathcal{D}), \beta\} &= (\partial^2/\partial v^2)\mathcal{L}(Y, \mathcal{A}, v, \beta)|_{v=\theta(\mathcal{D})}, \end{aligned}$$

etc. Then, for any β_* , we estimate $\theta(\cdot)$ by $\hat{\theta}(z_0, \beta_*)$ via standard local likelihood estimation [5]. We consider a local linear estimator, which is $(1, 0) \times \alpha$, where α solves the local loglikelihood score equation

$$0 = n^{-1} \sum_{i=1}^n K_h(\mathcal{D}_i - z_0)G_h^T(\mathcal{D}_i - z_0)\mathcal{L}_\theta\{Y_i, \mathcal{A}_i, G_h^T(\mathcal{D}_i - z_0)\alpha, \beta_*\}. \quad (1)$$

In general, $\widehat{\theta}(z_0, \beta_*)$ converges to $\theta(z_0, \beta_*)$, where

$$0 = E[\mathcal{L}_\theta\{Y, \mathcal{A}, \theta(z_0, \beta_*), \beta_*\} | \mathcal{D} = z_0]. \tag{2}$$

By differentiation of (2), the derivative of $\theta(z_0, \beta_*)$ with respect to β_* , call it $\theta_\beta(z_0, \beta_*)$, satisfies

$$\theta_\beta(z_0, \beta_*) = -\frac{E[\mathcal{L}_{\theta\beta}\{Y, \mathcal{A}, \theta(z_0, \beta_*), \beta_*\} | \mathcal{D} = z_0]}{E[\mathcal{L}_{\theta\theta}\{Y, \mathcal{A}, \theta(z_0, \beta_*), \beta_*\} | \mathcal{D} = z_0]}. \tag{3}$$

The profile method maximizes $\sum_{i=1}^n \mathcal{L}\{Y_i, \mathcal{A}_i, \widehat{\theta}(\mathcal{D}_i, \beta), \beta\}$ in β . Let $\widehat{\theta}_\beta(z_0, \beta)$ be the derivative of $\widehat{\theta}(z_0, \beta)$ with respect to β . Then profiling is equivalent to solving the equation

$$0 = n^{-1} \sum_{i=1}^n \left[\mathcal{L}_\beta\{Y_i, \mathcal{A}_i, \widehat{\theta}(\mathcal{D}_i, \beta), \beta\} + \widehat{\theta}_\beta(\mathcal{D}_i, \beta) \mathcal{L}_\theta\{Y_i, \mathcal{A}_i, \widehat{\theta}(\mathcal{D}_i, \beta), \beta\} \right].$$

The solution, $\widehat{\beta}$, converges to β_{PF} , which satisfies

$$0 = E[\mathcal{L}_\beta\{Y, \mathcal{A}, \theta(\mathcal{D}, \beta_{PF}), \beta_{PF}\} + \theta_\beta(\mathcal{D}, \beta_{PF}) \mathcal{L}_\theta\{Y, \mathcal{A}, \theta(\mathcal{D}, \beta_{PF}), \beta_{PF}\}],$$

and because of (2), this means that β_{PF} , satisfies

$$0 = E[\mathcal{L}_\beta\{Y, \mathcal{A}, \theta(\mathcal{D}, \beta_{PF}), \beta_{PF}\}]. \tag{4}$$

Remark 3 For estimating β_{PF} , it is known that the profile estimate of β_{PF} is asymptotically normally distributed when the bandwidth $h \propto n^{-a}$ for $1/5 \leq a \leq 1/3$, and that the optimal rate for estimating $\theta(\cdot)$ is $h \propto n^{-1/5}$. See [20] for a definition and discussion of an alternative backfitting approach.

4. SIMEX when X is modeled nonparametrically

4.1. Introduction and theoretical development

In the SIMEX problem, we have a response Y , predictors Z measured without error and a predictor X measured with error. The loglikelihood or criterion function for the case considered in this section is $\mathcal{L}\{Y, Z, \theta(X), \beta\}$. In what follows, we will simply state the results, which follow from Results 3 and 4 in the Appendix, see also Sections A.6 and A.7. The basic conditions for these results are stated in Section A.2.

In SIMEX, for $b = 1, \dots, B$ with B fixed but large, we observe $W_{ib}(\lambda) = W_i + \lambda^{1/2}U_{ib}$, where $W_i = X_i + U_i$, $U_i = \text{Normal}(0, \Sigma_u)$ and the simulated random variables $U_{ib} = \text{Normal}(0, \Sigma_u)$. We then apply profile likelihood for each b and each λ where we replace X_i by $W_{ib}(\lambda)$.

Let $f_W(x_0, \lambda)$ be the density function of $W_b(\lambda)$. Define $\beta(\lambda)$ and $\theta\{x_0, \beta(\lambda), \lambda\}$ as the solutions to

$$\begin{aligned} 0 &= E(\mathcal{L}_\beta[Y, Z, \theta\{W_b(\lambda), \beta(\lambda), \lambda\}, \beta(\lambda)]); \\ 0 &= E(\mathcal{L}_\theta[Y, Z, \theta\{x_0, \beta(\lambda), \lambda\}, \beta(\lambda)] | W_b(\lambda) = x_0). \end{aligned} \tag{5}$$

In the SIMEX algorithm, for any λ and for $b = 1, \dots, B$, let the parameter and function estimates be denoted as $\widehat{\beta}_b(\lambda)$ and $\widehat{\theta}_b(\cdot, \widehat{\beta}_b(\lambda), \lambda)$, respectively. Then average these estimators to form $\widehat{\beta}(\lambda) = B^{-1} \sum_{b=1}^B \widehat{\beta}_b(\lambda)$ and, in a slight abuse of notation, $\widehat{\theta}\{\cdot, \widehat{\beta}(\lambda), \lambda\} = B^{-1} \sum_{b=1}^B \widehat{\theta}_b\{\cdot, \widehat{\beta}_b(\lambda), \lambda\}$.

For convenience, we phrase the result as based on polynomial extrapolation, with more general cases requiring only changes in notation. In this case, we set up a grid of λ values, say $(0 = \lambda_1 < \lambda_2 < \dots < \lambda_J)$, and it is obvious that extrapolating a polynomial regression back to -1 makes the SIMEX estimators take the form

$$\widehat{\beta}_{\text{simex}} = \sum_{j=1}^J s_j \widehat{\beta}(\lambda_j); \quad \widehat{\theta}_{\text{simex}}(x_0) = \sum_{j=1}^J s_j \widehat{\theta}\{x_0, \widehat{\beta}(\lambda_j), \lambda_j\} \quad (6)$$

for some fixed constants s_j , and these have limiting values

$$\beta_{\text{simex}} = \sum_{j=1}^J s_j \beta(\lambda_j); \quad \theta_{\text{simex}}(x_0) = \sum_{j=1}^J s_j \theta\{x_0, \beta(\lambda_j), \lambda_j\}. \quad (7)$$

For example, suppose that $J = 5$ and that the λ -values are $(0.0, 0.5, 1.0, 1.5, 2.0)$. If we fit a quadratic model $\widehat{\beta}(\lambda) = a_0 + a_1\lambda + a_2\lambda^2$, the SIMEX extrapolant at $\lambda = -1$ is $a_0 - a_1 + a_2$, and the values $(s_1, \dots, s_J) = (3.0, -0.4, -1.8, -1.2, 1.4)$.

Also define

$$\begin{aligned} \theta_{\beta}\{x_0, \beta(\lambda), \lambda\} &= -\frac{E(\mathcal{L}_{\theta\beta}[Y, Z, \theta\{x_0, \beta(\lambda), \lambda\}, \beta(\lambda)]|W_b(\lambda) = x_0)}{E(\mathcal{L}_{\theta\theta}[Y, Z, \theta\{x_0, \beta(\lambda), \lambda\}, \beta(\lambda)]|W_b(\lambda) = x_0)}; \\ \Sigma(\lambda) &= E[E\{\mathcal{L}_{\beta\beta}(\bullet)|W_b(\lambda)\} \\ &\quad + E\{\mathcal{L}_{\beta\theta}(\bullet)|W_b(\lambda)\}\theta_{\beta}^T\{W_b(\lambda), \beta(\lambda), \lambda\}]; \\ \Omega_W\{x_0, \beta(\lambda), \lambda\} &= E\{\mathcal{L}_{\theta\theta}(\bullet)|W_b(\lambda) = x_0\}, \end{aligned}$$

where here the argument $(\bullet) = [Y, Z, \theta\{W_b(\lambda), \beta(\lambda), \lambda\}, \beta(\lambda)]$. Make the further definitions

$$\begin{aligned} \{\bullet_{ib}(\lambda)\} &= [Y_i, Z_i, \theta\{W_{ib}(\lambda), \beta(\lambda), \lambda\}, \beta(\lambda)]; \\ \xi_{i,B}(\lambda) &= -B^{-1} \sum_{b=1}^B [\mathcal{L}_{\beta}\{\bullet_{ib}(\lambda)\} + \theta_{\beta}\{W_{ib}(\lambda), \beta(\lambda), \lambda\}\mathcal{L}_{\theta}\{\bullet_{ib}(\lambda)\}]; \\ \chi_{i,B}(\lambda) &= B^{-1} \sum_{b=1}^B \mathcal{L}_{\theta}\{\bullet_{ib}(\lambda)\}. \end{aligned}$$

Result 1 For fixed B and as $n \rightarrow \infty$, we have the asymptotic expansion

$$\widehat{\beta}_{\text{simex}} = \beta_{\text{simex}} - n^{-1} \sum_{i=1}^n \sum_{j=1}^J s_j \Sigma^{-1}(\lambda_j) \xi_{i,B}(\lambda_j) + o_p(n^{-1/2}). \quad (8)$$

The nonparametric estimator differs from $\theta_{\text{simex}}(x_0)$ with large sample bias given as

$$\text{bias} = (\phi_2 h^2 / 2) \sum_{j=1}^J s_j \theta^{(2)}\{x_0, \beta(\lambda_j), \lambda_j\}.$$

Finally, since B can be made as large as we want, terms of order $O\{(Bnh)^{-1} + n^{-1}\}$ are much smaller than terms of order $O\{(nh)^{-1}\}$. Ignoring these terms, the limiting variance is given as

$$\text{variance} = \frac{\int K^2(z) dz}{nh f_W(x_0, 0)} s_1^2 \text{var} \left[\frac{\chi_{i,B}(0)}{\Omega_W\{x_0, \beta(0), 0\}} \Big| W = x_0 \right]. \tag{9}$$

4.2. Standard error estimation

4.2.1. Estimation for the nonparametric part

At first glance, estimating the variance in (9) is quite easy, because its form only depends on the case that $\lambda = 0$ through s_1^2 , and standard techniques apply. This unique result is a generalization of a result found by [3] for the Gaussian nonparametric case. Asymptotically, the variance of the SIMEX estimate is just a factor s_1^2 larger than the variance of nonparametric regression that ignores measurement error. As described above, with quadratic extrapolation and the λ -values (0.0, 0.5, 1.0, 1.5, 2.0), $s_1 = 3$ and the SIMEX estimate is asymptotically 9 times more variable than the estimate that ignores measurement error.

However, as we demonstrate numerically in Sections 6 and 7, this asymptotic result, while interesting, is not useful in finite samples.

In the case of classical nonparametric regression, [38] recognized that the SIMEX estimator is linear in the (Y_i) , so that an explicit expression of its variance is available if $\text{var}\{Y|W_b(\lambda)\}$ is known: they then estimate $\text{var}\{Y|W_b(\lambda)\}$. In our case, not only do we have the issue of the parametric component, but also there is no closed-form formula for the variance of the SIMEX estimator.

In order to get roughly believable standard errors, we turn to the theory, particularly to equations (8) and (A.6) of the appendix, the former being an asymptotic expansion for the parametric part, the latter for the nonparametric part. The estimate of the nonparametric function is just

$$\hat{\theta}_{\text{simex}}(x_0) = \sum_j s_j B^{-1} \sum_{b=1}^B \hat{\theta}_b\{x_0, \hat{\beta}_b(\lambda_j), \lambda_j\},$$

and to create reasonable standard error estimates we need to account for the variability in $\hat{\beta}_b(\lambda_j)$ via (8), the variability in $\hat{\theta}_b\{x_0, \beta(\lambda_j), \lambda_j\}$ via (A.6) and the fact that the number of SIMEX steps B is finite. Define $\hat{\beta}(\lambda_j) = B^{-1} \sum_{b=1}^B \hat{\beta}_b(\lambda_j)$. Using a Taylor series involving $\hat{\beta}_b(\lambda_j)$, these formulae imply that for fixed B ,

a consistent estimate of the variance of the estimator $\widehat{\theta}_{\text{simex}}(x_0)$ (see appendix A.6.1 for details) is n^{-1} times the sample variance of the terms

$$\sum_{j=1}^J s_j \frac{B^{-1} \sum_{b=1}^B K_h\{W_{ib}(\lambda_j) - x_0\} \mathcal{L}_{\theta}(\widehat{\bullet}_{ib})}{(nB)^{-1} \sum_{r=1}^n \sum_{b=1}^B K_h\{W_{rb}(\lambda_j) - x_0\} \mathcal{L}_{\theta}(\widehat{\bullet}_{rb})} \quad (10)$$

$$+ \sum_{j=1}^J s_j \widehat{\theta}_{\beta}^T(x_0, \widehat{\beta}(\lambda_j), \lambda_j) \widehat{\Sigma}^{-1}(\lambda_j) \widehat{\xi}_{i,B}(\lambda_j),$$

where $\widehat{\bullet}_{ib}$ is the estimated version of \bullet_{ib} and Σ is estimated by its sample counterpart $\widehat{\Sigma}$. Asymptotically, if $B = \infty$ and there are only a finite number of λ -values, only the first term with $j = 1$ contributes to the first order variance, but we found in both the simulations of Section 6 and the empirical example in Section 7 that ignoring the other terms causes an underestimation of standard error by as much as 50%. In our numerical work we estimate the variance as an average of squares of terms from (10).

4.2.2. Estimation for the parametric part

Define $\mathcal{T}_{jk} = \text{cov}\{\xi_{i,B}(\lambda_j), \xi_{i,B}(\lambda_k)\}$. Then the asymptotic covariance matrix for $\widehat{\beta}_{\text{simex}}$ is just $n^{-1} \sum_{j,k=1}^J s_j s_k \Sigma^{-1}(\lambda_j) \mathcal{T}_{jk} \Sigma^{-1}(\lambda_k)$, so that we need to estimate $\Sigma(\lambda)$ and \mathcal{T}_{jk} . Estimation of these two terms is relatively straightforward via plug-in methods. For example, constituent terms such as $E\{\mathcal{L}_{\beta\beta}(\bullet_{ib})|W_{ib}(\lambda)\}$ can be estimated by pooling across $b = 1, \dots, B$ and using nonparametric regression of $\mathcal{L}_{\beta\beta}[Y_i, Z_i, \theta\{W_{ib}(\lambda), \widehat{\beta}(\lambda), \lambda\}, \widehat{\beta}(\lambda)]$ on the $W_{ib}(\lambda)$. In practice, to save storage requirements, one can use only a few of the SIMEX samples, do the parametric regression on a wide grid, and then interpolate for the other SIMEX samples. Having estimated the terms $\xi_{i,B}(\lambda_j)$, then \mathcal{T}_{jk} can be estimated via the sample covariance of $\xi_{i,B}(\lambda_j)$ and $\xi_{i,B}(\lambda_k)$.

In practice, there is often little need to adjust the standard errors for the estimation of the nonparametric component. The reason is that the semiparametric profile algorithm involves a projection similar to that which occurs in parametric problems.

5. SIMEX when X is modeled parametrically

5.1. Introduction and theoretical development

Here we consider the case that X is modeled parametrically and scalar Z is modeled nonparametrically. The loglikelihood or criterion function then takes the form $\mathcal{L}\{Y, X, \theta(Z), \beta\}$.

For any λ , define $\beta(\lambda)$ and $\theta\{z_0, \beta(\lambda), \lambda\}$ as the solutions to

$$0 = E(\mathcal{L}_{\beta}[Y, W_b(\lambda), \theta\{Z, \beta(\lambda), \lambda\}, \beta(\lambda)]);$$

$$0 = E(\mathcal{L}_{\theta}[Y, W_b(\lambda), \theta\{z_0, \beta(\lambda), \lambda\}, \beta(\lambda)]|Z = z_0),$$

the values to which the profile likelihood estimates converge for given λ . Also define

$$\begin{aligned} \theta_\beta\{z_0, \beta(\lambda), \lambda\} &= -\frac{E\{\mathcal{L}_{\theta\beta}(\bullet)|Z = z_0\}}{E\{\mathcal{L}_{\theta\theta}(\bullet)|Z = z_0\}}, \\ \Sigma(\lambda) &= E[E\{\mathcal{L}_{\beta\beta}(\bullet)|Z\} + E\{\mathcal{L}_{\beta\theta}(\bullet)|Z\}\theta_\beta^T\{Z, \beta(\lambda), \lambda\}]; \\ \Omega_Z\{z_0, \beta(\lambda), \lambda\} &= E\{\mathcal{L}_{\theta\theta}(\bullet)|Z = z_0\}, \end{aligned}$$

where here the argument $(\bullet) = [Y, W_b(\lambda), \theta\{Z, \beta(\lambda), \lambda\}, \beta(\lambda)]$. Further, make the definitions

$$\begin{aligned} \{\bullet_{ib}(\lambda)\} &= [Y_i, W_{ib}(\lambda), \theta\{Z_i, \beta(\lambda), \lambda\}, \beta(\lambda)]; \\ \xi_{i,B}(\lambda) &= -B^{-1} \sum_{b=1}^B [\mathcal{L}_\beta\{\bullet_{ib}(\lambda)\} + \theta_\beta\{Z_i, \beta(\lambda), \lambda\}\mathcal{L}_\theta\{\bullet_{ib}(\lambda)\}]; \\ \chi_{i,B}(\lambda) &= B^{-1} \sum_{b=1}^B \mathcal{L}_\theta\{\bullet_{ib}(\lambda)\}. \end{aligned}$$

By definition, $\xi_{i,B}(\lambda)$ has mean zero, while $\chi_{i,B}(\lambda)$ has mean zero conditional on Z_i .

Let the estimators for any b and any λ be denoted as $\hat{\beta}_b(\lambda)$ and $\hat{\theta}_b\{z_0, \hat{\beta}_b(\lambda), \lambda\}$, respectively. Average these estimators across $b = 1, \dots, B$ to form $\hat{\beta}(\lambda) = B^{-1} \sum_{b=1}^B \hat{\beta}_b(\lambda)$ and $\hat{\theta}\{z_0, \hat{\beta}(\lambda), \lambda\} = B^{-1} \sum_{b=1}^B \hat{\theta}_b\{z_0, \hat{\beta}_b(\lambda), \lambda\}$. The estimators and their limit values again take the forms (6) and (7), respectively.

Result 2 For fixed B and as $n \rightarrow \infty$, we have the asymptotic expansion

$$\hat{\beta}_{\text{simex}} = \beta_{\text{simex}} - n^{-1} \sum_{i=1}^n \sum_{j=1}^J s_j \Sigma^{-1}(\lambda_j) \xi_{i,B}(\lambda_j) + o_p(n^{-1/2}). \tag{11}$$

The nonparametric estimator differs from $\theta_{\text{simex}}(z_0)$ with large sample bias and variance

$$\begin{aligned} \text{bias} &= (\phi_2 h^2 / 2) \sum_{j=1}^J s_j \theta^{(2)}\{z_0, \beta(\lambda_j), \lambda_j\}; \\ \text{variance} &= \frac{\int K^2(z) dz}{nh f_Z(z_0)} \text{var} \left[\sum_{j=1}^J s_j \frac{\chi_{i,B}(\lambda_j)}{\Omega_Z\{z_0, \beta(\lambda_j), \lambda_j\}} \middle| Z = z_0 \right]. \tag{12} \end{aligned}$$

5.2. Standard error estimation

Result 2 and the expansions given in the Appendix make it easy to construct asymptotically correct variance and standard error estimates.

5.2.1. Estimation for the nonparametric part

Calculation of an asymptotic variance estimate for $\theta_{\text{simex}}(z_0)$ follows a similar strategy as in Section 4.2.1. In the form of the variance given in (12), the density function $f_Z(\cdot)$ can be estimated by any kernel density estimate. To estimate $\Omega_Z\{z_0, \beta(\lambda), \lambda\}$, first form the averages over $b = 1, \dots, B$ of the terms $\mathcal{L}_{\theta\theta}[Y_i, W_{ib}(\lambda), \theta\{Z_i, \hat{\beta}(\lambda), \lambda\}, \hat{\beta}(\lambda)]$ and then regress these averages on Z via nonparametric regression. Let $\hat{\chi}_{i,B}(\lambda)$ being the obvious estimate of $\chi_{i,B}(\lambda)$.

As in the technical arguments in Section A.6, the analogue of Section 4.2.1 is to estimate the variance of $\hat{\theta}_{\text{simex}}(z_0)$, while taking into account the estimation of the parametric part, as the sample variance of the following terms:

$$\begin{aligned} & \sum_{j=1}^J s_j \frac{K_h(Z_i - z_0) \hat{\chi}_{i,B}(\lambda_j)}{n^{-1} \sum_{p=1}^n K_h(Z_p - z_0) \hat{\Omega}_Z\{z_0, \hat{\beta}(\lambda_j), \lambda_j\}} \\ & + \sum_{j=1}^J s_j \hat{\theta}_{\beta}^T\{z_0, \hat{\beta}(\lambda_j), \lambda_j\} \hat{\Sigma}^{-1}(\lambda_j) B^{-1} \\ & \times \sum_{b=1}^B \left[\mathcal{L}_{\beta}\{\hat{\bullet}_{ib}(\lambda_j)\} + \hat{\theta}_{\beta}\{z_0, \hat{\beta}(\lambda_j), \lambda_j\} \mathcal{L}_{\theta}\{\hat{\bullet}_{ib}(\lambda_j)\} \right]. \end{aligned}$$

5.2.2. Estimation for the parametric part

First consider $\hat{\beta}_{\text{simex}}$ and define $\mathcal{T}_{jk} = \text{cov}\{\xi_{i,B}(\lambda_j), \xi_{i,B}(\lambda_k)\}$. Then the asymptotic covariance matrix for $\hat{\beta}_{\text{simex}}$ is just $n^{-1} \sum_{j,k=1}^J s_j s_k \Sigma^{-1}(\lambda_j) \mathcal{T}_{jk} \Sigma^{-1}(\lambda_k)$, so that we need to estimate $\Sigma(\lambda)$ and \mathcal{T}_{jk} . The former can be done by simple plug-in methods which we describe next. For any λ , use the averages of the B SIMEX replicates to form $\mathcal{L}_{\beta\beta}(\bullet)$, $\mathcal{L}_{\beta\theta}(\bullet)$ and $\mathcal{L}_{\theta\theta}(\bullet)$, then do nonparametric regression on each to form $E\{\mathcal{L}_{\beta\beta}(\bullet)|Z\}$, $E\{\mathcal{L}_{\beta\theta}(\bullet)|Z\}$ and $\theta_{\beta}\{Z, \beta(\lambda), \lambda\}$, form $E\{\mathcal{L}_{\beta\beta}(\bullet)|Z\} - E\{\mathcal{L}_{\beta\theta}(\bullet)|Z\} \theta_{\beta}^T\{Z, \beta(\lambda), \lambda\}$, and average over Z to get the estimate of $\Sigma(\lambda)$.

The terms \mathcal{T}_{jk} are also readily estimated via plug-in rules. Let $\hat{\xi}_{i,B}(\lambda_j)$ is the obvious estimate of $\xi_{i,B}(\lambda_j)$. Then \mathcal{T}_{jk} can be estimated as the sample covariance matrix between $\hat{\xi}_{i,B}(\lambda_j)$ and $\hat{\xi}_{i,B}(\lambda_k)$.

5.3. When both X and Z are mismeasured

Our results also apply to the case that both X and Z are subject to measurement error. Consider the problem in Section 4 that Z is modeled parametrically, and that X is modeled nonparametrically. Suppose that instead of observing X we observe $W = X + U$, instead of Z we observe $\mathcal{V} = Z + V$, and that $(V^T, U)^T = \text{Normal}(0, \Sigma_{vu})$. Let $\{\mathcal{V}_{ib}^T(\lambda), W_{ib}(\lambda)\} = (\mathcal{V}_i^T, W_i) + \lambda^{1/2}(V_{ib}^T, U_{ib})$, where $(V_{ib}^T, U_{ib})^T = \text{Normal}(0, \Sigma_{vu})$. Then all the results in Section 4, including

standard error estimation, go through if Z is replaced everywhere by $\mathcal{V}_b(\lambda)$. Thus, for example, we replace (5) by

$$0 = E(\mathcal{L}_\beta[Y, \mathcal{V}_b(\lambda), \theta\{W_b(\lambda), \beta(\lambda), \lambda\}, \beta(\lambda)]).$$

Similar substitutions are made everywhere else.

6. Simulations

6.1. The partially linear model

We generated data following the partially linear model

$$Y_i = Z_i^T \beta_0 + \theta_0(X_i) + \epsilon_i.$$

We take the true values as $\beta_0 = (1, 1)^T$ and $\theta_0(x) = 0.5 * \cos(2 * x)$. In addition $\epsilon = \text{Normal}(0, 1)$ and $X = \text{Uniform}(0, \pi)$. We generated $Z = (Z_1, Z_2)^T$ from a Normal distribution with identity covariance matrix and conditional mean $E(Z_1|X) = E(Z_2|X) = X$. We assumed that the nonparametrically modeled variable X is measured with error, that is, we observe $W = X + U$ with $U = \text{Normal}(0, \sigma_u^2)$ with $\sigma_u^2 = 0.16$. The sample size was $n = 200$. Nonparametric regression was performed using the Epanechnikov kernel with bandwidth $h = \hat{\sigma}_w n^{-1/5}$, where $\hat{\sigma}_w$ is the standard deviation of the observed predictors W . We generated 1,000 data sets and for each of the generated data sets, we used $B = 200$ SIMEX simulated data sets and used $(\lambda_1, \dots, \lambda_5) = (0.0, 0.5, \dots, 2.0)$. We considered both quadratic and cubic extrapolants with similar results: only the quadratic extrapolant results are given here as the same general conclusions applied in the cubic case.

Of course, in this context the SIMEX estimates are linear in the responses, and it is possible to create estimated standard errors for the nonparametric component estimation using this fact. However, here the purpose is to test the general method described in Section 4.2.1, and not to see whether ad hoc methods work well.

The estimated pointwise standard errors of SIMEX fits for $\theta(x)$ are given in Figure 1 for quadratic extrapolation. The solid line is our new method of standard error estimation, the dot-dashed line is the usual pointwise standard error estimates using our method to account for finite B but ignoring the asymptotically negligible variability of the parametric part (that is, the sample variance of the first term of (10)), and the dashed line represents the pointwise standard error across 1,000 simulations. The conclusion is clear: ignoring the variability of the parametric part leads to a badly biased estimated standard error for the SIMEX procedure. In contrast, our new method of estimating standard error gives results more in line with reality.

We have not displayed the “large- B , known β ” asymptotic formula (9). However, that method gave standard error estimates whose mean was grossly too large, more than double the actual standard errors. See Figure 5 for an example of this over-estimation in the empirical context of Section 7.

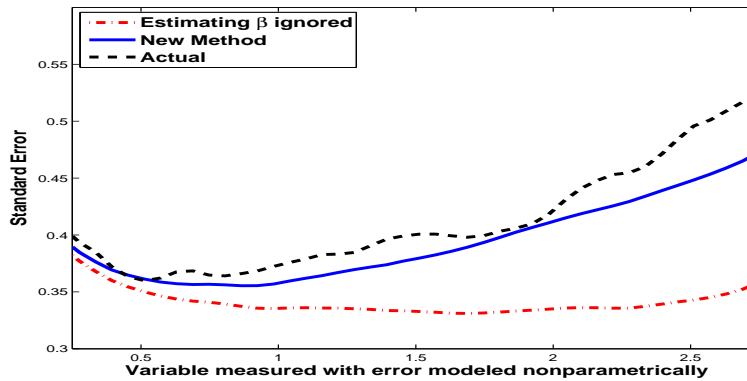


FIG 1. Results from the simulation study in Section 6.1 for the partially linear model with measurement error in the covariate modeled nonparametrically. Displayed are the pointwise estimated standard errors of SIMEX estimate of $\theta(x)$ using quadratic extrapolation. Solid line: our method as described in Section 4.2.1. Dot-dashed line: standard errors when the variability of the parametric part is not taken into account. Dashed line: pointwise standard errors of $\hat{\theta}(z)$ over 1,000 simulations.

In results not reported here, we also considered standard error estimation for the parametric part. Our standard error estimates were in very close agreement with the actual standard deviations from the simulation, and coverage probabilities for the SIMEX limiting value of β also gave very close to nominal coverage probabilities.

6.2. Partially linear logistic model

We generated data following the logistic model

$$\text{pr}(Y = 1|X, Z) = H\{Z\beta_1 + Z^2\beta_2 + \theta_0(X)\},$$

where $H(\cdot)$ denotes the logistic distribution function. We took the true values of $\beta_1 = \beta_2 = 0.5$ and $\theta_0(x) = 0.5 * \cos(2 * x) - 1$. We generated X from $\text{Uniform}(0, \pi)$ and Z from a Normal distribution with variance 1.0 and conditional mean $E(Z|X) = -X$. We assume that the nonparametrically modeled variable X is measured with error, that is, we observe $W = X + U$ with $U = \text{Normal}(0, \sigma_u^2)$ with $\sigma_u^2 = 0.16$. We use sample size $n = 500$. Nonparametric regression was performed using the Epanechnikov kernel and we used bandwidth $h = \hat{\sigma}_w n^{-1/5}$, where $\hat{\sigma}_w$ is the standard deviation of the observed predictors W . We generated 2,000 data sets and for each of the generated data sets, we used $B = 200$ SIMEX simulated data sets and took $(\lambda_1, \dots, \lambda_5) = (0.0, 0.5, \dots, 2.0)$. We considered both quadratic and cubic extrapolants: only the quadratic results are reported here as the same general conclusions applied in the cubic case.

The estimated pointwise standard errors of SIMEX fits for $\theta(x)$ are given in Figure 2 for quadratic extrapolation. The solid line is our new method of standard error estimation, the dot-dashed line is the usual pointwise standard error

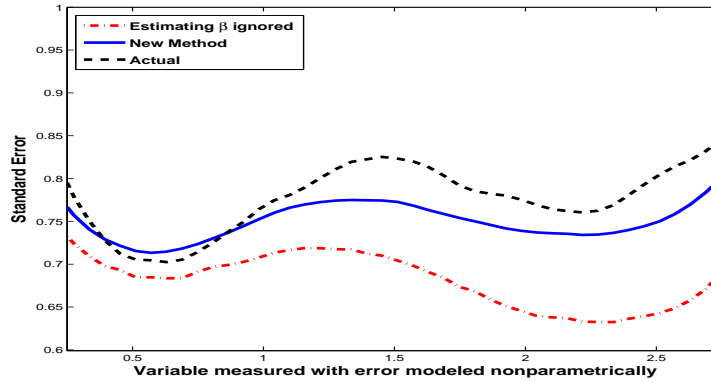


FIG 2. Results from the simulation study in Section 6.2 for the logistic partially linear model with measurement error in the covariate modeled nonparametrically. Displayed are the pointwise estimated standard errors of SIMEX estimate of $\theta(x)$ using quadratic extrapolation. Solid line: our method as described in Section 4.2.1. Dot-dashed line: standard errors when the variability of the parametric part is not taken into account. Dashed line: pointwise standard errors of $\hat{\theta}(z)$ over 2,000 simulations.

estimates ignoring the variability of the parametric part (that is, the sample variance of the first term of (10)), and the dashed line represents the pointwise standard error across 1,000 simulations. The conclusion is clear: ignoring the variability of the parametric part leads to a noticeable biased estimated standard error for the SIMEX procedure, although the bias is not as serious as in the linear case. In contrast, our new method of estimating standard error gives results more in line with reality.

In results not reported here, we also considered standard error estimation for the parametric part. Our standard error estimates were again in very close agreement with the actual standard deviations from the simulation, and coverage probabilities for the SIMEX limiting value of β also gave very close to nominal coverage probabilities.

6.3. Why SIMEX?

This section gives some brief numerical evidence of the power of SIMEX in semiparametric problems. We performed a simulation of a semiparametric model in which the variable measured with error, X , was modeled parametrically. In the parametric simulation, we had $Z = \text{Uniform}[0, \pi]$, $X = \text{Normal}(-1.0, 1.0)$, $n = 500$, $\beta_1 = \beta_2 = 0.7$, $\sigma_u^2 = 0.16$, $\text{var}(Y|X, Z) = 0.5$ and the model was that $E(Y|X, Z) = g(X, \beta_1, \beta_2) + \theta(Z) = -1 + \beta_1 X + \beta_2 X^2 + 0.5 * \cos(2 * Z)$. We used $(\lambda_1, \dots, \lambda_5) = (0.0, 0.5, 1.0, 1.5, 2.0)$, with $B = 100$ SIMEX simulated data sets for each λ . Both the quadratic and the cubic extrapolant were considered. For purposes of mean squared error analysis, we evaluated the estimated function

TABLE 1

Results of the simulation for the model

$E(Y|X, Z) = g(X, \beta_1, \beta_2) + \theta(Z) = -1 + \beta_1 X + \beta_2 X^2 + 0.5 * \cos(2 * Z)$, where X is measured with error. Here Quadratic and Cubic refer to SIMEX extrapolation via the quadratic and cubic functions. MSE Efficiency is mean squared error efficiency of the methods relative to the naive method that ignores measurement error

		β_1	β_2	$\theta(\cdot)$
Mean Absolute Bias	Naive	0.250	0.184	0.039
	Quadratic	0.030	0.024	0.023
	Cubic	0.002	0.004	0.025
MSE Efficiency	Quadratic	5.26	10.08	0.85
	Cubic	9.69	9.98	0.65

on an equally spaced grid from 0.10 to $\pi - 0.10$. Nonparametric regression was performed using the Epanechnikov kernel function. There were 200 simulated data sets.

For bandwidth selection, we used two algorithms. First, at each calculation of the nonparametric function for fixed (β_1, β_2) , we used the DPI bandwidth selection method of Ruppert, Sheather and Wand (1995). We also did a separate analysis with the bandwidth fixed globally at $\hat{\sigma}_w n^{-1/5}$, where $\hat{\sigma}_w$ is the standard deviation of the observed error-prone predictors W . The results were in rough agreement, with the former reported here.

Because in this simulation X and the measurement error U are normally distributed, and X is independent of Z , it follows that the distribution of X given (Z, W) is also normal, i.e., $\text{Normal}(\alpha_0 + \alpha_1 W, \sigma_{x|w}^2)$, where with $a = 1/1.16$, $\alpha_0 = -(1 - a)$, $\alpha_1 = a$ and $\sigma_{x|w}^2 = 0.16a$, see [4](Chapter 3). It then follows that the observed data in (W, Z) have the mean $E(Y|W, Z) = \beta_1(\alpha_1 + 2\alpha_0\alpha_1\sigma_{x|w}^2)W + \beta_2\alpha_1^2W^2 + \beta_2\sigma_{x|w}^2 + \theta(Z)$. Thus, the naive estimate that ignores measurement error should be biased downward by the amount 0.22 for β_1 and by 0.18 for β_2 . The bias in the estimated function is the constant $\beta_2\sigma_{x|w}^2 = 0.10$.

Table 1 basically confirms what the theory says. The simulated mean bias of the naive estimator for both β_1 and β_2 is very nearly the large sample value, and both the quadratic and cubic extrapolants largely remove this bias. With $n = 500$, this translates into a sizeable gain in mean squared error efficiency. For estimating $\theta(\cdot)$, there is little bias even theoretically in the naive estimator, relative to the variability of nonparametric regression kernel estimators, and we see little difference between the naive and SIMEX estimators.

7. Nevada test site thyroid disease study

7.1. Introduction

An important example of the types of semiparametric problems that SIMEX is able to address is the following. In the 1950's, the United States conducted above-ground nuclear testing, and in the 1980's the University of Utah conducted the Nevada Test Site (NTS) Thyroid Disease Study. In the Nevada study,

2,491 individuals who were exposed to radiation as children were examined for thyroid disease. The primary radiation exposure to the thyroid glands of these children came from the ingestion of milk and vegetables contaminated with radioactive isotopes of iodine. The idea of the study was to relate various thyroid disease outcomes to radiation exposure to the thyroid. The original version of this study was described by [15, 39] and [34]. Recently, the dosimetry for the study was redone [35], and the study results were reported [22].

The estimation of individual radiation dose that occurred 50 years in the past is well-known to be subject to large uncertainties, especially when mathematical models are employed due to the absence of direct measurements of the concentrations of radioactivity in foods or within the thyroid gland of individuals. There are many references on this subject, a good introduction to which is given by [29] and the many statistical papers in that volume. Various statistical papers [21, 24, 26, 28, 32, 33, 40] describe measurement error properties and analysis in this context. What is typical in these studies is to build a large dosimetry model that attempts to convert the known data, e.g., about the above-ground nuclear tests, to radiation actually absorbed into the thyroid. Dosimetry calculations for individual subjects were based on age at exposure, gender, residence history, whether as a child the individual was breast-fed, and a diet questionnaire filled out by the parent focusing on milk consumption and vegetables. The data were then input into a complex model and for each individual, the point estimate of thyroid dose (the arithmetic mean of a lognormal distribution of dose estimates) and an associated measurement error term (the geometric standard deviation) were reported.

Generally, authors engaged in dose reconstruction using mathematical models conclude that radiation doses are estimated with a combination of Berkson measurement error and the classical type of measurement error. This type of model, see [28], in the log-scale of dose says that true log-dose T is related to observed or calculated log-dose W by a latent intermediate X via

$$T = X + U_{\text{berk}}; \quad (13)$$

$$W = X + U_{\text{class}}, \quad (14)$$

where U_{berk} is the Berkson uncertainty with variance $\sigma_{u,\text{berk}}^2$ depending on the individual, and U_{class} is the classical uncertainty with variance $\sigma_{u,\text{class}}^2$ depending on the individual. In the NTS data, the total uncertainty $\sigma_{u,\text{class}}^2 + \sigma_{u,\text{berk}}^2$ is known, but not the relative contributions. Various of the references given above do sensitivity analysis with different amounts of relative contributions: in our illustration, we will consider two situations, one with 50% of the total uncertainty being Berkson and the other with all the uncertainty being classical.

It is typical to assume that the Berkson error U_{berk} is normally distributed. If X could be observed, it is typical to use the total mean dose $\exp(X + \sigma_{u,\text{berk}}^2/2)$ as the main predictor of risk, and we take this as the target.

Let Y , the response, be the incidence of thyroiditis (inflammation of the thyroid gland), and consider in addition a variable measured without error, Z , the sex of the patient. A typical model relating total mean dose and gender to

disease is the excess relative risk model

$$\text{pr}(Y = 1|X, Z) = H[\beta_0 + \beta_1 Z + \log\{1 + \gamma \exp(X + \sigma_{u,\text{berk}}^2/2)\}], \quad (15)$$

where $H(\cdot)$ is the logistic distribution function and γ is called the excess relative risk. Because the amount of Berkson uncertainty depends upon the individual, the term $\sigma_{u,\text{berk}}^2/2$ in the excess relative risk model may be thought of as an offset.

There is some doubt as to what the correct dose-response model should be, and indeed some researchers use linear dose-response models. It thus makes sense to consider a semiparametric model that allows the dose-response to be flexible. The semiparametric model of interest here is the partially linear logistic model:

$$\text{pr}(Y = 1|X, Z) = H\{\beta Z + \theta(X + \sigma_{u,\text{berk}}^2/2)\}, \quad (16)$$

where $\theta(\cdot)$ is an unknown function. This is an example when the variable that is known exactly, Z , is modeled parametrically but the variable measured with error, X , is modeled nonparametrically.

7.2. Data analysis

In our analysis, we fit models (15) and (16). We used quadratic extrapolation, set the λ -values as (0.0, 0.5, 1.0, 1.5, 2.0), used $B = 100$ SIMEX simulated data sets, and employed the Epanechnikov kernel function. We assumed 50% of the measurement error was Berkson and the rest was of the usual additive form. We also redid the analysis with all the measurement error being of the classical form. The bandwidth chosen was 1.5, but similar results were obtained for 1.0 and 2.0.

Profiling was used in the semiparametric model calculations: it was simple to implement because the nonparametric regressions done used standard software for logistic regression with weights and an offset. That is, in model (16), for any given β , $\theta(\cdot)$ can be estimated by logistic regression with the offset $Z\beta$ and with the weights given by the kernel weights.

Standard errors were estimated as described in Section 4.2.1. In addition, we ran 1000 bootstrap samples, and recalculated the estimates.

7.2.1. Various fits

Because X and Z are essentially independent in the data, the estimate of β_1 is not much affected by the measurement error. Indeed, in either the mixture of Berkson and classical errors, or when there is no Berkson error, and using either the excess relative risk model (15), a naive semiparametric fit that ignores the measurement error, or using the SIMEX semiparametric fit, the estimate of $\beta_1 \approx 1.75$ with a standard error (s.e.) ≈ 0.25 . There were only very slight

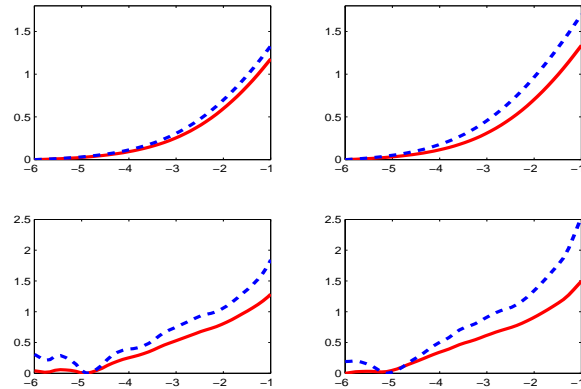


FIG 3. Logits of the effects of dose in the Nevada Test Site thyroiditis data. Solid lines are the fits that ignore measurement error, dashed as those that account for measurement error. Left two panels: the mixture of Berkson errors model using parametric (top) and semiparametric (bottom) fits. Right two panels: the no Berkson error model using parametric (top) and semiparametric (bottom) fits. In the bottom two panels, the dashed lines are the SIMEX fits.

differences in any of these cases. The result is plausible: β_1 indicates that women are at higher risk for thyroiditis.

Figure 3 shows the results of various fitting methods. In all cases, the solid lines are the fits that ignore measurement error, while the dashed lines account for measurement error. The left two panels give results for the mixture of Berkson errors model using parametric (top) and semiparametric (bottom) fits. The right two panels give fits when no Berkson errors is assumed using parametric (top) and semiparametric (bottom) fits. In the bottom two panels, the dashed lines are the SIMEX fits. The overall conclusion is that there is some change in the logits when the classical error is accounted for, and if all the error is classical, a substantial change is seen. The excess relative risk parameter estimates for the naive and measurement error analysis are, respectively, 6.24 and 7.75 for the mixture error model, and 7.83 and 12.60 for the no Berkson error model.

7.2.2. Standard Errors

Figures 4 and 5 show the power of our asymptotic theory for estimating standard errors. All analyses concern the mixture of Berkson and classical errors.

In Figure 4, we consider the naive semiparametric fits, i.e., $s_1 = 1$ and $s_j = 0$ for $j > 1$. The solid line is the usual estimated pointwise standard errors that ignore the estimation of β_1 : these estimates ignore the second term in (10). The dashed line is our new method based upon equation (10), while the dot-dashed line is the pointwise standard error from 1000 bootstrap simulations. The result is clear: ignoring the estimation of β_1 produces standard errors for the nonparametric fit that are badly biased.

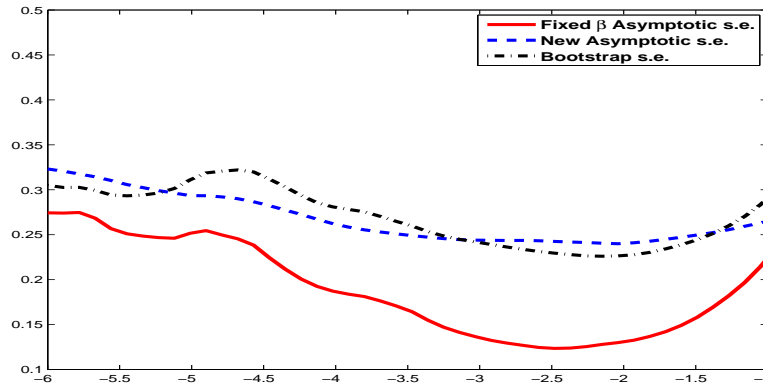


FIG 4. Estimated standard errors of the nonparametric fits in the Nevada Test Site thyroiditis data ignoring measurement error when the errors are a mixture of Berkson and Classical errors. Solid line: the ordinary estimated pointwise standard errors when the variability of the parametric part is not taken into account. Dashed line: our new method for estimating standard errors as in equation (10). Dot-dashed line: bootstrap standard errors based on 1,000 bootstrap replicates.

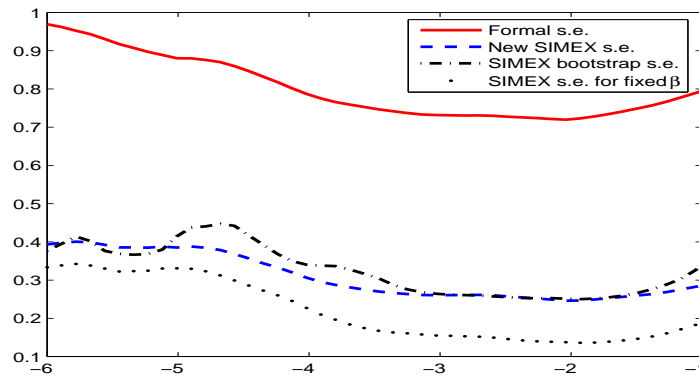


FIG 5. Estimated standard errors of the nonparametric SIMEX fits in the Nevada Test Site thyroiditis data when the errors are a mixture of Berkson and Classical errors. Solid line: the formal estimated pointwise standard errors using the formal asymptotic theory of equation (9). Dashed line: our new method for estimating standard errors as in equation (10). Dot-dashed line: bootstrap standard errors based on 1,000 bootstrap replicates. Dotted line: pointwise standard errors when the variability of the parametric part is not taken into account.

In Figure 5, we consider the SIMEX semiparametric fits. The solid line is the usual estimated pointwise standard errors based upon the formal theory in Result 1 and equation (9). The dashed line is our new method based upon equation (10), while the dot-dashed line is the pointwise standard error from 1000 bootstrap simulations. Also, the dotted line is estimated pointwise standard

errors when the variability of the parametric part is not taken into account. The result is also clear: the formal asymptotic theory is not very useful, and it is only our asymptotic expansions that give sensible results.

Remark 4 For SIMEX, it is worth pointing out that bootstrapping is particularly time-consuming, and thus having workable standard errors via (10) is practically very important. With $B = 100$ and four non-zero λ values, our standard error formulae requires $B = 401$ semiparametric fits. The bootstrap with 1000 bootstrap samples requires 401,000 semiparametric fits.

8. Discussion

The main point of this paper is to derive the limiting distribution of SIMEX in semiparametric problems, when the variable X subject to measurement error is modeled parametrically, nonparametrically or a combination of both, and to give computable asymptotically correct standard error estimates that improve upon first-order methods in empirical applications. In the case that the variable measured with error is modeled nonparametrically, our standard error formula is new, and was vastly superior to the asymptotic formula in both simulations and in our empirical analysis. Our method of estimating standard errors has not previously been discussed in the literature, as far as we know, and was facilitated by our asymptotic expansions.

It is useful to point out that the standard error estimates are also applicable to cases that there is no measurement error. Indeed, in simulations that are the no measurement error versions of the problems in Sections 6.1 and 6.2, the same basic phenomenon happens, namely that the standard errors for the nonparametric function obtained ignoring the asymptotically negligible variability in the parametric parts considerably underestimated the actual simulated standard errors, while our methods were able to reproduce them rather well.

There are two issues that we have not addressed in detail.

- We have not discussed the issue of bandwidth selection in detail, with our numerical results using ad hoc bandwidth choices. However, based on our results, at least in principle this is straightforward. As in [38], but modified to the semiparametric case, the idea is to use our expansions and EBBS [30] to estimate the bias, while our new variance estimates of the nonparametric part given in Section 4.2.1 and 5.2.1 are used to derive more precise estimated standard errors, from which the estimated mean squared errors can then be calculated and minimized.
- It is seemingly straightforward although notationally tedious and algebraically complex to extend the results of Section 4–5 to the case that the covariance matrix of the measurement errors is unknown and needs to be estimated. In the case that the mismeasured X is modeled parametrically, this is a relatively simple combination of our work and that of [3]. The more difficult case occurs when the mismeasured X is modeled nonparametrically: the technical issue is that the arguments in the kernel weights will now depend on the estimated measurement error variance.

Appendix A: Sketch of technical arguments

A.1. Outline of the Appendix

As in Section 3, we first deal with generic random variables \mathcal{A} and \mathcal{D} in a possibly misspecified model context, so that only equations (2), (3) and (4) are used. In Section A.2 we give the main assumptions. In Section A.3 we give asymptotic expansions for the generic problem. These results are then applied to SIMEX in Sections A.6 and A.7.

A.2. Assumptions

Our calculations require little more than uniform expansions in z , to order $o_p(n^{-1/2})$, of local linear kernel regressions. These kinds of expansions are well known, see for example [7] and [6]. Except for smoothness and boundedness conditions that are necessary to justify Taylor expansions, conditions that we simply assume although details can be given, there are really only four essential conditions.

- (C.1) The kernel function K is a symmetric, continuously differentiable pdf on $[-1, 1]$ taking on the value zero at the boundaries. The design density $f(\cdot)$ of \mathcal{D} is differentiable on $B = [b_1, b_2]$, the derivative is continuous, and $\inf_{z \in B} f(z) > 0$. The function $\theta(\cdot, \beta)$ has 2 continuous derivatives on B and is also thrice differentiable with respect to β . We also assume that the same calculations done by [6] can be applied to our context.
- (C.2) The value β_{PF} is the unique maximizer of $E[\mathcal{L}\{Y, \mathcal{A}, \theta(\mathcal{D}, \beta), \beta\}]$: this is the misspecified model version of the idea of a least favorable curve. In addition, if $d/d\beta$ denotes the total derivative, then uniformly in a neighborhood of β_{PF} ,

$$-E \left[\frac{d^2}{d\beta d\beta^T} \mathcal{L}\{Y, \mathcal{A}, \theta(\mathcal{D}, \beta), \beta\} \right]$$

is positive definite. Finally, the parameter β is restricted to a compact set.

- (C.3) We can apply the results of [7] (CVK) as needed. In particular, their assumptions imply that uniformly in z_0 , for random variables R_i possessing sufficient moments, then if

$$C_n = n^{-1} \sum_{i=1}^n K_h(\mathcal{D}_i - z_0)(\mathcal{D}_i - z_0)^j R_i \{ \theta(z_0) + (\mathcal{D}_i - z_0)\theta^{(1)}(z_0) \};$$

$$D_n = n^{-1} \sum_{i=1}^n K_h(\mathcal{D}_i - z_0)(\mathcal{D}_i - z_0)^j R_i \{ \theta(\mathcal{D}_i) \},$$

then

$$\sup_{z_0} |C_n - E(C_n)| = O_p[h^j \{ \log(n)/(nh) \}^{1/2}];$$

$$\sup_{z_0} |D_n - E(D_n)| = O_p[h^j \{ \log(n)/(nh) \}^{1/2}].$$

(C.4) The bandwidth $h \sim n^{-a}$ with $1/3 \leq h \leq 1/5$, this latter meaning that $\lambda > 3$ in condition (R2) of CVK.

A.3. Results for general models

The basic issue here is that measurement error induces a misspecified model for the observed data, and the general topic of properties of misspecified semiparametric models has only recently been a topic of research, see for example Ai and Chen (2007) who make exactly this point. Our paper thus can be thought of as a general description of what happens with misspecified semiparametric models in the kernel regression area, with an application to SIMEX.

This section states the main results for misspecified models. In what follows, we focus on the profile estimator. Sketches of proofs follow later in the Appendix.

A.3.1. Nonparametric regression expansions

Standard large-sample approximations for nonparametric regression using estimating equations are given in [5], with rigorous theory about rates of convergence given by [7].

Consider the local linear estimator $\widehat{\theta}(z_0, \beta_*)$ defined in (1). Let $\theta^{(j)}(\cdot)$ be the j^{th} derivative of $\theta(\cdot)$ with respect to z_0 . Let $f_{\mathcal{D}}(z_0)$ be the density function of \mathcal{D} . Make the definitions

$$\begin{aligned} \Omega(z_0, \beta_*) &= E[\mathcal{L}_{\theta\theta}\{Y, \mathcal{A}, \theta(\mathcal{D}, \beta_*), \beta_*\} | \mathcal{D} = z_0]; \\ \mathcal{T}(z_0, \beta_*) &= f_{\mathcal{D}}(z_0)\Omega(z_0, \beta_*), \quad \phi_k = \int v^k K(v)dv. \end{aligned}$$

Result 3 Use the short-hand notation $\mathcal{L}_{i\theta}(c, \beta_*) = \mathcal{L}_{\theta}(Y_i, \mathcal{A}_i, c, \beta_*)$. Under the conditions stated in Section A.2, uniformly in z_0 , for any β_* ,

$$\begin{aligned} &\widehat{\theta}(z_0, \beta_*) - \theta(z_0, \beta_*) \\ &= (\phi_2 h^2 / 2) \theta^{(2)}(z_0, \beta_*) - n^{-1} \sum_{i=1}^n K_h(\mathcal{D}_i - z_0) \mathcal{L}_{i\theta}\{\theta(\mathcal{D}_i, \beta_*), \beta_*\} / \mathcal{T}(z_0, \beta_*) \\ &\quad + O_p\{h^4 + \log(n)/(nh)\}. \end{aligned} \tag{A.1}$$

In addition,

$$\begin{aligned} &h\{\widehat{\theta}^{(1)}(z_0, \beta_*) - \theta^{(1)}(z_0, \beta_*)\} \\ &= -n^{-1} \sum_{i=1}^n K_h(\mathcal{D}_i - z_0) \mathcal{L}_{i\theta}\{\theta(\mathcal{D}_i, \beta_*), \beta_*\} \{(\mathcal{D}_i - z_0)/h\} / \mathcal{T}(z_0, \beta_*) \\ &\quad + \phi_4 h^3 \mathcal{T}^{(1)}(z_0, \beta_*) \{\theta^{(2)}(z_0, \beta_*)/2 + \theta^{(3)}(z_0, \beta_*)/6\} / \mathcal{T}(z_0, \beta_*) \\ &\quad + O_p\{h^4 + \log(n)/(nh)\}. \end{aligned} \tag{A.2}$$

A.3.2. Semiparametric Models

We now turn to the estimation of β_{PF} . Define

$$\begin{aligned}\Sigma &= E[\mathcal{L}_{\beta\beta}\{Y, \mathcal{A}, \theta(\mathcal{D}, \beta_{\text{PF}}), \beta_{\text{PF}}\} + \mathcal{L}_{\beta\theta}\{Y, \mathcal{A}, \theta(\mathcal{D}, \beta_{\text{PF}}), \beta_{\text{PF}}\}\theta_{\beta}^{\text{T}}(\mathcal{D}, \beta_{\text{PF}})]; \\ \mathcal{G} &= \text{cov}[\mathcal{L}_{\beta}\{Y, \mathcal{A}, \theta(\mathcal{D}, \beta_{\text{PF}}), \beta_{\text{PF}}\} + \mathcal{L}_{\theta}\{Y, \mathcal{A}, \theta(\mathcal{D}, \beta_{\text{PF}}), \beta_{\text{PF}}\}\theta_{\beta}(\mathcal{D}, \beta_{\text{PF}})].\end{aligned}$$

Result 4 *Make the assumptions stated in Section A.2. If the bandwidth is of order $h \propto n^{-a}$ with $1/5 \leq a \leq 1/3$, then the profiled estimate $\hat{\beta}$ satisfies*

$$-\Sigma \left\{ n^{1/2}(\hat{\beta} - \beta_{\text{PF}}) \right\} = n^{-1/2} \sum_{i=1}^n \{ \mathcal{L}_{\beta}(\bullet_i) + \theta_{\beta}(\mathcal{D}_i, \beta_{\text{PF}}) \mathcal{L}_{\theta}(\bullet_i) \} + o_p(1), \quad (\text{A.3})$$

so that $n^{1/2}(\hat{\beta} - \beta_{\text{PF}}) \Rightarrow \text{Normal}\{0, \Sigma^{-1}\mathcal{G}(\Sigma^{\text{T}})^{-1}\}$.

A.4. Sketch of Result 3

For notational simplicity we will drop the dependence on β_* . Using Theorem 2.1 of CVK, for $j = 0, 1$,

$$\sup_{z_0} h^j |\hat{\theta}^{(j)}(z_0) - \theta^{(j)}(z_0)| = O_p \left[h^2 + \{\log(n)/(nh)\}^{1/2} \right].$$

Then a second order Taylor series of (1) shows that

$$\begin{aligned}0 &= n^{-1} \sum_{i=1}^n K_h(\mathcal{D}_i - z_0) G_h(\mathcal{D}_i - z_0) \mathcal{L}_{i\theta} \{ \theta(z_0) + \theta^{(1)}(z_0)(\mathcal{D}_i - z_0) \} \\ &\quad + n^{-1} \sum_{i=1}^n K_h(\mathcal{D}_i - z_0) G_h(\mathcal{D}_i - z_0) G_h^{\text{T}}(\mathcal{D}_i - z_0) \\ &\quad \times \mathcal{L}_{i\theta\theta} \{ \theta(z_0) + \theta^{(1)}(z_0)(\mathcal{D}_i - z_0) \} \left[\begin{array}{c} \hat{\theta}(z_0) - \theta(z_0) \\ h \{ \hat{\theta}^{(1)}(z_0) - \theta^{(1)}(z_0) \} \end{array} \right] \\ &\quad + O_p \left[h^4 + \{\log(n)/(nh)\} \right],\end{aligned}$$

uniformly in z_0 . Define

$$\begin{aligned}\mathcal{S}_n(z_0) &= n^{-1} \sum_{i=1}^n K_h(\mathcal{D}_i - z_0) G_h(\mathcal{D}_i - z_0) G_h^{\text{T}}(\mathcal{D}_i - z_0) \\ &\quad \times \mathcal{L}_{i\theta\theta} \{ \theta(z_0) + \theta^{(1)}(z_0)(\mathcal{D}_i - z_0) \}.\end{aligned}$$

Using calculations similar to equation (5.1) of CVK,

$$\sup_{z_0} |\mathcal{S}_n(z_0) - E\{\mathcal{S}_n(z_0)\}| = O_p \left[h^2 + \{\log(n)/(nh)\}^{1/2} \right].$$

Then, via standard calculations, $\sup_{z_0} |E\{\mathcal{S}_n(z_0)\} - \text{diag}(1, \phi_2)\mathcal{T}(z_0)| = O(h^2)$. The result is that uniformly in z_0 ,

$$\begin{aligned} & -\mathcal{T}(z_0) \left\{ \widehat{\theta}(z_0) - \theta(z_0) \right\} \\ &= n^{-1} \sum_{i=1}^n K_h(\mathcal{D}_i - z_0) \mathcal{L}_{i\theta} \{ \theta(z_0) + \theta^{(1)}(z_0)(\mathcal{D}_i - z_0) \} \\ & \quad + O_p \left[h^4 + \{ \log(n)/(nh) \} \right] \\ &= n^{-1} \sum_{i=1}^n K_h(\mathcal{D}_i - z_0) \mathcal{L}_{i\theta} \{ \theta(\mathcal{D}_i) \} \\ & \quad - n^{-1} \sum_{i=1}^n K_h(\mathcal{D}_i - z_0) \left[\mathcal{L}_{i\theta} \{ \theta(\mathcal{D}_i) \} - \mathcal{L}_{i\theta} \{ \theta(z_0) + \theta^{(1)}(z_0)(\mathcal{D}_i - z_0) \} \right] \\ & \quad + O_p \left[h^4 + \{ \log(n)/(nh) \} \right]. \end{aligned}$$

Using tedious, detailed Taylor series expansions, it is straightforward to show that the second term in the last equation $O_p[h^4 + \{ \log(n)/(nh) \}]$ so that uniformly in z_0 , (A.1) holds.

By the same arguments, uniformly to z_0 ,

$$\begin{aligned} & \mathcal{T}(z_0) h \{ \widehat{\theta}^{(1)}(z_0) - \theta^{(1)}(z_0) \} \\ &= -n^{-1} \sum_{i=1}^n K_h(\mathcal{D}_i - z_0) \{ (\mathcal{D}_i - z_0)/h \} \mathcal{L}_{i\theta} \{ \theta(z_0) + (\mathcal{D}_i - z_0)\theta^{(1)}(z_0) \} \\ & \quad + O_p \{ h^4 + \log(n)/(nh) \} \\ &= -n^{-1} \sum_{i=1}^n K_h(\mathcal{D}_i - z_0) \{ (\mathcal{D}_i - z_0)/h \} \mathcal{L}_{i\theta} \{ \theta(\mathcal{D}_i) \} \\ & \quad + n^{-1} \sum_{i=1}^n K_h(\mathcal{D}_i - z_0) \{ (\mathcal{D}_i - z_0)/h \} \\ & \quad \times \left[\mathcal{L}_{i\theta} \{ \theta(\mathcal{D}_i) \} - \mathcal{L}_{i\theta} \{ \theta(z_0) + (\mathcal{D}_i - z_0)\theta^{(1)}(z_0) \} \right] \\ & \quad + O_p \{ h^4 + \log(n)/(nh) \}. \end{aligned}$$

By techniques similar to those used to show (A.1), the last term is

$$\begin{aligned} &= n^{-1} \sum_{i=1}^n K_h(\mathcal{D}_i - z_0) \{ (\mathcal{D}_i - z_0)/h \} \mathcal{L}_{i\theta\theta} \{ \theta(\mathcal{D}_i) \} \theta^{(2)}(z_0) (\mathcal{D}_i - z_0)^2 / 2 \\ & \quad + n^{-1} \sum_{i=1}^n K_h(\mathcal{D}_i - z_0) \{ (\mathcal{D}_i - z_0)/h \} \mathcal{L}_{i\theta\theta\theta} \{ \theta(\mathcal{D}_i) \} \theta^{(3)}(z_0) (\mathcal{D}_i - z_0)^3 / 6 \\ & \quad + O_p \{ h^4 + \log(n)/(nh) \}. \end{aligned}$$

These two terms, minus their expectations, are of order $O_p\{h^4 + \log(n)/(nh)\}$. Their expectations are clearly $(\phi_4 h^3/2)\mathcal{T}^{(1)}(z_0)\theta^{(2)}(z_0)$ and $(\phi_4 h^3/6)\mathcal{T}^{(1)}(z_0) \times \theta^{(3)}(z_0)$, completing the argument for (A.2)

A.5. Sketch of Result 4

Let $d/d\beta$ denote the total derivative. Then, we have that

$$\begin{aligned} 0 &= \frac{d}{d\beta} n^{-1} \sum_{i=1}^n \mathcal{L}\{Y_i, \mathcal{A}_i, \widehat{\theta}(\mathcal{D}_i, \widehat{\beta}), \widehat{\beta}\} \\ &= \frac{d}{d\beta} n^{-1} \sum_{i=1}^n \mathcal{L}\{Y_i, \mathcal{A}_i, \widehat{\theta}(\mathcal{D}_i, \beta_{PF}), \beta_{PF}\} \\ &\quad + \frac{d^2}{d\beta d\beta^T} n^{-1} \sum_{i=1}^n \mathcal{L}\{Y_i, \mathcal{A}_i, \widehat{\theta}(\mathcal{D}_i, \beta_*), \beta_*\} (\widehat{\beta} - \beta_{PF}), \end{aligned} \tag{A.4}$$

where β_* is between $\widehat{\beta}$ and β_{PF} . It is easily shown that the last term in (A.4) is $\Sigma(\widehat{\beta} - \beta_{PF}) + o_p(n^{-1/2})$. Thus, we must show that

$$\frac{d}{d\beta} n^{-1/2} \sum_{i=1}^n \left[\mathcal{L}\{Y_i, \mathcal{A}_i, \widehat{\theta}(\mathcal{D}_i, \beta_{PF}), \beta_{PF}\} - \mathcal{L}\{Y_i, \mathcal{A}_i, \theta(\mathcal{D}_i, \beta_{PF}), \beta_{PF}\} \right] = o_p(1).$$

Using Result 3 repeatedly, this is a straightforward but extremely detailed calculation, along the same lines as in [20].

A.6. SIMEX when X is modeled nonparametrically

In order to use the general theory of Section A.3 in the Appendix, we make the identifications $\mathcal{D} \rightsquigarrow X$ and $\mathcal{A} \rightsquigarrow Z$.

For any given b and λ , call the estimators for $\widehat{\beta}_b(\lambda)$ and $\widehat{\theta}_b\{x_0, \beta(\lambda), \lambda\}$, and call the average of the latter over b as $\widehat{\theta}\{x_0, \beta(\lambda), \lambda\}$. First, again from Result 3,

$$\begin{aligned} \widehat{\theta}_b\{x_0, \beta(\lambda), \lambda\} &= \theta\{x_0, \beta(\lambda), \lambda\} + (\phi_2 h^2 / 2) \theta^{(2)}\{x_0, \beta(\lambda), \lambda\} \\ &\quad - n^{-1} \sum_{i=1}^n K_h^T\{W_{ib}(\lambda) - x_0\} \frac{\mathcal{L}_\theta[Y_i, Z_i, \theta\{W_{ib}(\lambda), \beta(\lambda), \lambda\}, \beta(\lambda)]}{f_W(x_0, \lambda) \Omega_W\{x_0, \beta(\lambda), \lambda\}} \\ &\quad + o_p(n^{-1/2}). \end{aligned}$$

Also from Result 4,

$$\begin{aligned} &-\Sigma(\lambda) n^{1/2} \{\widehat{\beta}_b - \beta(\lambda)\} \\ &= n^{-1/2} \sum_{i=1}^n [\mathcal{L}_\beta\{\bullet_{ib}(\lambda)\} + \theta_\beta\{Z_i, \beta(\lambda), \lambda\} \mathcal{L}_\theta\{\bullet_{ib}(\lambda)\}] + o_p(1). \end{aligned}$$

Since B is fixed, this implies that

$$n^{1/2} \{\widehat{\beta}(\lambda) - \beta(\lambda)\} = -\Sigma^{-1}(\lambda) n^{-1/2} \sum_{i=1}^n \xi_{i,B}(\lambda) + o_p(1). \tag{A.5}$$

The first part of Result 1 follows from (A.5). We also have that

$$\begin{aligned} &\widehat{\theta}\{x_0, \beta(\lambda), \lambda\} \\ &= \theta\{x_0, \beta(\lambda), \lambda\} + (\phi_2 h^2/2)\theta^{(2)}\{x_0, \beta(\lambda), \lambda\} \\ &\quad - n^{-1} \sum_{i=1}^n B^{-1} \sum_{b=1}^B K_h\{W_{ib}(\lambda) - x_0\} \frac{\mathcal{L}_\theta\{\bullet_{ib}(\lambda)\}}{[2pt]} f_W(x_0, \lambda) \Omega_W\{x_0, \beta(\lambda), \lambda\} \\ &\quad + o_p(n^{-1/2}). \end{aligned} \tag{A.6}$$

As in [3], from (A.6), for $\lambda > 0$,

$$\text{var}[\widehat{\theta}\{x_0, \widehat{\beta}(\lambda), \lambda\}] = O\{(nBh)^{-1} + n^{-1}\},$$

while for $\lambda = 0$, $\text{var}[\widehat{\theta}\{x_0, \widehat{\beta}(0), 0\}] = O\{(nh)^{-1}\}$. Thus, for B sufficiently large, the variability of $\widehat{\theta}(\cdot, \lambda)$ is negligible for $\lambda > 0$ compared to $\lambda = 0$, completing the argument.

A.6.1. Justification of (10)

Since we are concerned here strictly with variability, we ignore the bias term in (A.6). Note then that the SIMEX estimator satisfies

$$\begin{aligned} &\widehat{\theta}_{\text{simex}}(x_0) - \theta_{\text{simex}}(x_0) \\ &= \sum_{j=1}^J s_j B^{-1} \sum_{b=1}^B \left[\widehat{\theta}_b\{x_0, \widehat{\beta}_b(\lambda_j), \lambda_j\} - \theta\{x_0, \beta(\lambda_j), \lambda_j\} \right] \\ &= \sum_{j=1}^J s_j B^{-1} \sum_{b=1}^B \left[\widehat{\theta}_b\{x_0, \beta(\lambda_j), \lambda_j\} - \theta\{x_0, \beta(\lambda_j), \lambda_j\} \right. \\ &\quad \left. + \theta_\beta^T\{x_0, \beta(\lambda_j), \lambda_j\} \{\widehat{\beta}_b(\lambda_j) - \beta(\lambda_j)\} \right] + o_p(n^{-1/2}) \\ &= - \sum_{j=1}^J s_j n^{-1} \sum_{i=1}^n B^{-1} \sum_{b=1}^B K_h\{W_{ib}(\lambda_j) - x_0\} \frac{\mathcal{L}_\theta\{\bullet_{ib}(\lambda_j)\}}{f_W(x_0, \lambda_j) \Omega_W\{x_0, \beta(\lambda_j), \lambda_j\}} \\ &\quad + \theta_\beta^T\{x_0, \beta(\lambda_j), \lambda_j\} \sum_{j=1}^J s_j \{\widehat{\beta}(\lambda_j) - \beta(\lambda_j)\} + o_p(n^{-1/2}) \\ &= -n^{-1} \sum_{i=1}^n \sum_{j=1}^J s_j B^{-1} \sum_{b=1}^B K_h\{W_{ib}(\lambda_j) - x_0\} \frac{\mathcal{L}_\theta\{\bullet_{ib}(\lambda_j)\}}{f_W(x_0, \lambda_j) \Omega_W\{x_0, \beta(\lambda_j), \lambda_j\}} \\ &\quad - n^{-1} \sum_{i=1}^n \sum_{j=1}^J s_j \theta_\beta^T\{x_0, \beta(\lambda_j), \lambda_j\} \Sigma^{-1}(\lambda) \xi_{i,B}(\lambda) + o_p(n^{-1/2}). \end{aligned}$$

The term $f_W(x_0, \lambda_j)\Omega_W\{x_0, \beta(\lambda_j), \lambda_j\}$ can be estimated as $(nB)^{-1} \sum_{r=1}^n \times \sum_{b=1}^B K_h\{W_{rb}(\lambda_j) - x_0\} \mathcal{L}_{\theta\theta}\{\widehat{\bullet}_{rb}(\lambda_j)\}$ leading to our suggested variance estimator.

A.7. SIMEX when X is modeled parametrically

In order to use the general theory of Section A.3 in the Appendix, we make the identifications $\mathcal{A} \rightsquigarrow X$ and $\mathcal{D} \rightsquigarrow Z$. We have the expansions for $\widehat{\beta}_{b,\text{PF}}(\lambda)$ and $\widehat{\theta}_b\{z_0, \beta(\lambda), \lambda\}$ as follows. First, from Result 3,

$$\begin{aligned} \widehat{\theta}_b\{z_0, \beta(\lambda), \lambda\} &= \theta\{z_0, \beta(\lambda), \lambda\} + (\phi_2 h^2 / 2) \theta^{(2)}\{z_0, \beta(\lambda), \lambda\} \\ &\quad - n^{-1} \sum_{i=1}^n K_h^T(Z_i - z_0) \frac{\mathcal{L}_\theta\{Y_i, W_{ib}(\lambda), \theta(Z_i, \beta(\lambda), \lambda), \beta(\lambda)\}}{f_Z(z_0, \lambda)\Omega_Z\{z_0, \beta(\lambda), \lambda\}} \\ &\quad + o_p(n^{-1/2}), \end{aligned}$$

and, from Result 4,

$$\begin{aligned} &-\Sigma(\lambda) n^{1/2} \{\widehat{\beta}_{b,\text{PF}} - \beta(\lambda)\} \\ &= n^{-1/2} \sum_{i=1}^n [\mathcal{L}_\beta\{\bullet_{ib}(\lambda)\} + \theta_\beta\{Z_i, \beta(\lambda), \lambda\} \mathcal{L}_\theta\{\bullet_{ib}(\lambda)\}] + o_p(1), \end{aligned}$$

where $\{\bullet_{ib}(\lambda)\} = [Y_i, W_{ib}(\lambda), \theta\{Z_i, \beta(\lambda), \lambda\}, \beta(\lambda)]$. Since B is fixed, this implies that

$$n^{1/2} \{\widehat{\beta}(\lambda) - \beta(\lambda)\} = -\Sigma^{-1}(\lambda) n^{-1/2} \sum_{i=1}^n \xi_{i,B}(\lambda) + o_p(1). \tag{A.7}$$

Also,

$$\begin{aligned} \widehat{\theta}\{z_0, \beta(\lambda), \lambda\} &= \theta\{z_0, \beta(\lambda), \lambda\} + (\phi_2 h^2 / 2) \theta^{(2)}\{z_0, \beta(\lambda), \lambda\} \\ &\quad - n^{-1} \sum_{i=1}^n K_h(Z_i - z_0) \frac{\chi_{i,B}(\lambda)}{f_Z(z_0, \lambda)\Omega_Z\{z_0, \beta(\lambda), \lambda\}} \\ &\quad + o_p(n^{-1/2}). \end{aligned} \tag{A.8}$$

Result 2 follows immediately from (A.7) and (A.8).

Justification of the standard error formulae in Section 5.2.1 is similar to that given in Section A.6.1.

Acknowledgments

Our research was supported by grants from the National Cancer Institute (CA57030, CA104620). Part of Carroll’s research was supported by award number KUS-CI-016-04 made by the King Abdullah University of Science and Technology (KAUST). Apanasovich’s research was supported by a grant from the

National Science Foundation #0707106. The work of Apanasovich and Carroll partially occurred during a visit to and with the support of the Department of Mathematics and Statistics at the University of Melbourne, and is gratefully acknowledged.

References

- [1] AI, C. and CHEN, X. (2007). Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. *Journal of Econometrics*, 141, 5–43. [MR2411735](#)
- [2] CARROLL, R.J. and HALL, P. (2004). Low order approximations in deconvolution and regression with errors in variables. *Journal of the Royal Statistical Society, Series B*, 66, 31–46. [MR2035757](#)
- [3] CARROLL, R.J., MACA, J.D. and RUPPERT, D. (1999). Nonparametric regression with errors in covariates. *Biometrika*, 86, 541–554. [MR1723777](#)
- [4] CARROLL, R.J., RUPPERT, D., STEFANSKI, L.A. and CRAINICEANU, C.M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, Second Edition. Chapman and Hall CRC Press. [MR2243417](#)
- [5] CARROLL, R.J., RUPPERT, D. and WELSH, A. (1998). Local estimating equations. *Journal of the American Statistical Association*, 93, 214–227. [MR1614624](#)
- [6] CLAESKENS, G. and CARROLL, R.J. (2007). Post-model selection inference in semiparametric models. *Biometrika*, 94, 249–265. [MR2331485](#)
- [7] CLAESKENS, G. and VAN KEILEGOM, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Annals of Statistics*, 31, 1852–1884. [MR2036392](#)
- [8] COOK, J.R. and STEFANSKI, L.A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, 89, 1314–1328.
- [9] DELAIGLE, A. and HALL, P. (2008). Using SIMEX for smoothing parameter choice in errors-in-variables problems. *Journal of the American Statistical Association*, 130, 280 – 287. [MR2394636](#)
- [10] GREENE, W.F. and CAI, J. (2004). Measurement error in covariates in the marginal hazards model for multivariate failure time data. *Biometrics*, 60, 987–996. [MR2133551](#)
- [11] GOULD, W.R., STEFANSKI, L.A. and POLLOCK, K.H. (1997). Effects of measurement error on catch-effort estimation. *Canadian Journal of Fisheries and Aquatic Science*, 54, 898–906.
- [12] HWANG, W.H. and HUANG, S.Y.H. (2003). Estimation in capture-recapture models when covariates are subject to measurement errors. *Biometrics*, 59, 1113–1122. [MR2025137](#)
- [13] JEONG, M. and KIM, C. (2003). Some properties of SIMEX estimator in partially linear measurement error model. *Journal of the Korean Statistical Society*, 32, 85–92. [MR1984043](#)

- [14] KANGAS, A.S. (1998). Effect of errors-in-variables on coefficients of a growth model and on prediction of growth. *Forest Ecology And Management*, 102, 203–212.
- [15] KERBER, R.L., TILL, J.E., SIMON, S.L., LYON, J.L. THOMAS, D.C., PRESTON-MARTIN, S., ROLLISON, M.L., LLOYD, R.D. and STEVENS, W. (1993). A cohort study of thyroid disease in relation to fallout from nuclear weapons testing. *Journal of the American Medical Association*, 270, 2076–2083.
- [16] LECHNER, S. and POHLMEIER, W. (2004). To blank or not to blank? A comparison of the effects of disclosure limitation methods on nonlinear regression estimates. *Annals of the New York Academy of Sciences*, 3050, 187–200.
- [17] LIANG H., HARDLE, W. and CARROLL, R.J. (1999). Estimation in a semi-parametric partially linear errors-in-variables model. *Annals of Statistics*. 27, 1519–1535. [MR1742498](#)
- [18] LIANG, H. and REN, H.B. (2005). Generalized partially linear measurement error models. *Journal of Computational and Graphical Statistics*, 14, 237–250. [MR2137900](#)
- [19] LIN, X. and CARROLL, R.J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, 95, 520–534. [MR1803170](#)
- [20] LIN, X. and CARROLL, R.J. (2006). Semiparametric estimation in general repeated measures problems. *Journal of the Royal Statistical Society, Series B*, 68, 68–88. [MR2212575](#)
- [21] LUBIN, J.H., SCHAFER, D.W. RON, E., STOVALL, M. and CARROLL, R.J. (2004). A reanalysis of thyroid neoplasms in the Israeli tinea capitis study accounting for dose uncertainties. *Radiation Research*, 161, 359–368.
- [22] LYON, J.L., ALDER, S.C., STONE, M.B., SCHOLL, A., READING, J.C. HOLUBKOV, R., SHENG, X. WHITE, G.L., HEGMANN, K.T., ANSPAUGH, L., HOFFMAN, F.O., SIMON, S.L., THOMAS, B., CARROLL, R.J. and MEIKLE, A.W. (2006). Thyroid disease associated with exposure to the Nevada Test Site radiation: a reevaluation based on corrected dosimetry and examination data. *Epidemiology*, 17, 604–614.
- [23] MAITY, A., MA, Y. and CARROLL, R.J. (2007). Efficient estimation of population-Level summaries in general semiparametric regression models with missing response. *Journal of the American Statistical Association*, 102(477), 123–139. [MR2293305](#)
- [24] MALLICK, B., HOFFMAN, F.O. and CARROLL, R.J. (2002). Semiparametric regression modeling with mixtures of Berkson and classical error, with application to fallout from the Nevada Test Site. *Biometrics*, 58, 13–20. [MR1891038](#)
- [25] MARCUS, A.H. and ELIAS, R.W. (1998) Some useful statistical methods for model validation. *Environmental Health Perspectives*, 106, 1541–1550.
- [26] MARSCHNER, I.C., EMBERSON, J., IRWIG, L. and WALTER, S.D. (2004). The number needed to treat (NNT) can be adjusted for bias when the

- outcome is measured with error. *Journal of Clinical Epidemiology*, 57, 1244–1252.
- [27] PIERCE, D.A. and KELLERER, A. (2004). Adjusting for covariate errors with nonparametric assessment of the true covariate distribution. *Biometrika*, 91, 863–876. [MR2126038](#)
- [28] REEVES, G.K., COX, D.R., DARBY, S.C. and WHITLEY, E. (1998). Some aspects of measurement error in explanatory variables for continuous and binary regression models. *Statistics in Medicine*, 17, 2157–2177.
- [29] RON, E. and HOFFMAN, F.O. (1999). *Uncertainties in Radiation Dosimetry and Their Impact on Dose response Analysis*. National Cancer Institute Press.
- [30] RUPPERT, D. (1997) Empirical-bias bandwidths for local polynomial non-parametric regression and density estimation. *Journal of the American Statistical Association*, 92, 1049–1062. [MR1482136](#)
- [31] RUPPERT, D., SHEATHER, S.J. and WAND, M.P. (1995). An effective bandwidth selector for local least squares regression (Corr: 96V91 p1380). *Journal of the American Statistical Association*, 90, 1257–1270. [MR1379468](#)
- [32] SCHAFER, D.W. and GILBERT, E.S. (2006). Some statistical implications of does uncertainty in radiation dose-response analyses. *Radiation Research*, 166, 303–312.
- [33] SCHAFER, D.W., LUBIN, J.H., RON, E., STOVALL, M. and CARROLL, R.J. (2001). Thyroid cancer following scalp irradiation: a reanalysis accounting for uncertainty in dosimetry. *Biometrics*, 57, 689–697. [MR1859805](#)
- [34] SIMON, S.L., TILL, J.E., LLOYD, R.D., KERBER, R.L., THOMAS, D.C., PRESTON–MARTIN, S., LYON, J.L. and STEVENS, W. (1995). The Utah Leukemia case–control study: dosimetry methodology and results. *Health Physics*, 68, 460–471.
- [35] SIMON, S.L., ANSPAUGH, L.R., HOFFMAN, F.O., ET AL. (2006). 2004 update of dosimetry for the Utah Thyroid Cohort Study. *Radiation Research*, 165, 208–222.
- [36] SOLOW, A.R. (1998). On fitting a population model in the presence of observation error. *Ecology*, 79, 1463–1466.
- [37] STEFANSKI, L.A. and COOK, J. (1995). Simulation extrapolation: the measurement error jackknife. *Journal of the American Statistical Association*, 90, 1247–1256. [MR1379467](#)
- [38] STAUDENMEYER, J. and RUPPERT, D. (2004). Local polynomial regression and simulation-extrapolation. *Journal of the Royal Statistical Society, Series B*, 66, 17–30. [MR2035756](#)
- [39] STEVENS, W., TILL, J.E., THOMAS, D.C., ET AL. (1992). Assessment of leukemia and thyroid disease in relation to fallout in Utah: report of a cohort study of thyroid disease and radioactive fallout from the Nevada test site. University of Utah.
- [40] STRAM, D.O. and KOPECKY, K.J. (2003). Power and uncertainty analysis of epidemiological studies of radiation-related disease risk in which dose estimates are based on a complex dosimetry system: some observations. *Radiation Research*, 160, 408–417.

- [41] WANG, N., LIN, X., GUTIERREZ, R.G. and CARROLL, R.J. (1998). Generalized linear mixed measurement error models. *Journal of the American Statistical Association*, 93, 249–261. [MR1614636](#)
- [42] WANG, C.Y., WANG, N. and WANG, S. (2000). Regression analysis when covariates are regression parameters of a random effects model for observed longitudinal measurements. *Biometrics*, 56, 487–495.
- [43] ZHU, L. and CUI, H. (2003). A semiparametric regression model with errors in variables. *Scandinavian Journal of Statistics*. 30, 429–442. [MR1983135](#)