Biometrics 70, 823–834
December 2014
DOI: 10.1111/biom.12197

Bayesian Semiparametric Regression in the Presence of Conditionally Heteroscedastic Measurement and Regression Errors

Abhra Sarkar, Bani K. Mallick, and Raymond J. Carroll*

Department of Statistics, Texas A&M University, College Station, Texas 77843-3143, U.S.A.
**email:* carroll@stat.tamu.edu

Summary. We consider the problem of robust estimation of the regression relationship between a response and a covariate based on sample in which precise measurements on the covariate are not available but error-prone surrogates for the unobserved covariate are available for each sampled unit. Existing methods often make restrictive and unrealistic assumptions about the density of the covariate and the densities of the regression and the measurement errors, for example, normality and, for the latter two, also homoscedasticity and thus independence from the covariate. In this article we describe Bayesian semiparametric methodology based on mixtures of B-splines and mixtures induced by Dirichlet processes that relaxes these restrictive assumptions. In particular, our models for the aforementioned densities adapt to asymmetry, heavy tails and multimodality. The models for the densities of regression and measurement errors also accommodate conditional heteroscedasticity. In simulation experiments, our method vastly outperforms existing methods. We apply our method to data from nutritional epidemiology.

Key words: B-splines; Conditional heteroscedasticity; Dirichlet process mixture models; Measurement errors; Regression with errors in covariates; Variance functions.

1. Introduction

We develop a Bayesian semiparametric approach for robust estimation of a regression function when the covariate is measured with error, the density of the covariate, the density of the measurement errors and the density of the regression errors are all unknown, and the variability of both the measurement errors and the regression errors may depend on the associated unobserved value of the covariate through unknown relationships. By "robust" we mean that we avoid restrictive assumptions common in the literature, such as homoscedasticity and normally distributed measurement and regression errors.

The literature on regression with errors in covariates is extensive. A brief review of the existing literature relevant to our problem is presented here. For a more extensive review of the state of the art see Carroll et al. (2006) and Buonaccorsi (2010).

The problem of linear regression in the presence of errors in covariates is vast, and besides the references above also includes the classic text by Fuller (1987). More complex problems have also been studied. Cheng and Riu (2006) studied linear models and considered maximum likelihood, method of moments and generalized least squares estimators for heteroscedastic normally distributed regression and measurement errors. However, they assume that the variances are known and independent of the unobserved value of the covariate. Cook and Stefanski (1994) proposed a simulation-extrapolation (SIMEX) based method that did not make any assumptions about the density of the covariate and the density of the regression errors, but assumes homoscedasticity of both regression and measurement errors: strictly speaking, the latter is assumed to be normally distributed. The SIMEX method also requires the density of the measurement errors to be known. In the presence of replicated surrogates for the unobserved covariate, Devanarayan and Stefanski (2002) relaxed the homoscedasticity assumptions of the SIMEX approach, but the measurement errors are still required to be normally distributed. Carroll, Roeder, and Wasserman (1999a) proposed a Bayesian solution to the problem for normally distributed homoscedastic regression and measurement errors. They modeled the unknown density of the covariate by a finite mixture of normals.

Our focus here is on flexible nonparametric and semiparametric models for all the components. The problem of nonparametric regression with errors in covariates when the regression and measurement errors are both homoscedastic is studied by Fan and Truong (1993), Carroll, Maca, and Ruppert (1999b), Berry, Carroll, and Ruppert (2002), Carroll and Hall (2004) among others. Fan and Truong (1993) studied deconvoluting kernel type estimators when the density of the measurement errors is known. Carroll et al. (1999b) studied SIMEX estimators for the nonparametric regression with errors in covariates problem using three different types of models for the regression function, kernel mixtures, smoothing splines, and penalized truncated polynomial splines, but assuming homoscedastic normally distributed measurement errors. Berry et al. (2002) provided a Bayesian solution to the problem in the presence of normally distributed regression and measurement errors. They also assumed normality of the covariate and modeled the regression function using smoothing splines and penalized mixtures of truncated polynomial splines. Carroll and Hall (2004) considered the problem of estimating a low order estimate of the regression function, rather than the regression function itself. Their method required

knowledge of low order moments of the density of the measurement errors that can also be estimated from replicated surrogates. Schennach (2004a, 2004b) studied least squares and Nadaraya–Watson type estimators for nonlinear and nonparametric regression problems, respectively, when the measurement error density is unknown but replicated proxies are available and the measurement error in at least one of the replicates is homoscedastic and independent of the covariate. Delaigle and Meister (2007) relaxed the homoscedasticity assumption on the measurement errors but retained it for the regression errors. They developed deconvoluting kernel type estimators for problems when replicated surrogates are available for the unobserved covariates, the density of the regression errors is unknown and homoscedastic, the density of the measurement errors is unknown and heteroscedastic but they are both independent of the covariate. Besides the very strong independence assumption, when one would expect variability to depend on the error prone covariate, their method can only use data that have at least two replicates, whereas our method makes use of all observed data.

In this article we consider the problem of nonparametric estimation of the regression function in the presence of conditionally heteroscedastic regression and measurement errors, when the densities of the covariate, the regression errors and the measurement errors are all unknown. Conditional heteroscedasticity, in particular, can be a prominent feature of the distributions of regression and measurement errors in applied problems, as we illustrate in this article. To the best of our knowledge, this general problem has not been addressed. Indeed, it is not clear how the general deconvoluting kernel approach or even the automated SIMEX approach can be extended to accommodate conditional heteroscedasticity in regression and measurement errors. On the other hand, a Bayesian hierarchical framework, as we show in this article, can provide a natural way to tackle this otherwise complicated problem where the regression function and the nuisance densities can be modeled separately through a natural hierarchy. Importantly, the absence of precise covariate information can render techniques that are successful in regression problems with accurately measured covariates inefficient, numerically unstable, computationally challenging or intractable, and such measurement error poses new modeling challenges.

We present a Bayesian semiparametric solution to the problem, catering to such issues pertaining specifically to the measurement error setup. We model the density of the covariate by a flexible location-scale mixture of normals induced by a Dirichlet process (Ferguson, 1973). We model the regression function using a flexible mixture of B-splines. For modeling conditionally heteroscedastic regression and measurement errors, we assume that they can be factored into "scaled errors" that are independent of the covariate, and "variance function" components that explain the conditional heteroscedasticity. The densities of the scaled errors are also modeled using flexible mixture models induced by Dirichlet processes, each component of the mixture being itself a two-component normal mixture with its mean restricted at zero. The variance functions are modeled by positive mixtures of B-splines. Our approach, thus, uses Dirichlet process mixtures to model the three densities and B-spline mixtures to model the regression and the two variance functions.

The article is organized as follows. Section 2 details the models. Simulation experiments that compare the performances of our method and the method of Berry et al. (2002) are presented in Section 3, showing that our new methods dominate. Section 4 presents an application in nutritional epidemiology. Implementation issues, such as automatic choice of the hyper-parameters, details of the posterior calculations, the choice of initial values, the structure of the B-splines and additional supporting simulations are discussed in the Supplementary Materials.

2. Models

We consider the problem of robust estimation of the regression relationship between a response Y and a covariate X based on sample in which direct measurements on X are not available, but replicated proxies W for the latent X are available for a subset of the observed data. Specifically, for $i = 1, 2, \ldots, n$ individuals and for $j = 1, 2, \ldots, m_i$ with $m_i >= 2$ for at least some individuals, we assume

$$Y_i = r(X_i) + U_{Y,i}, \tag{1}$$

$$W_{ij} = X_i + U_{W,ij}. \tag{2}$$

Given X_i, the regression errors $U_{Y,i}$ and the measurement errors $U_{W,ij}$ have mean zero and are conditionally independent.

For random variables S and T, we denote the marginal density of S, the conditional density of S given T, and the joint density of (S, T) by the generic notation f_S, $f_{S|T}$ and $f_{S,T}$, respectively. The densities f_X, $f_{U_Y|X}$ and $f_{U_W|X}$ are all unknown. Given r, f_X, $f_{U_Y|X}$ and $f_{U_W|X}$, the likelihood $f_{Y,\mathbf{W}_{1:m}}$ is obtained by the convolution $f_{Y,\mathbf{W}_{1:m}}(Y, \mathbf{W}_{1:m}) = \int f_{U_Y|X=x}\{Y - r(x)\} \prod_{j=1}^{m} f_{U_{W_j}|X=x}(W_j - x) f_X(x)\, \mathrm{d}x$. In a Bayesian hierarchical framework, the problem therefore reduces to separate problems of modeling the density of the covariate f_X, modeling the regression function r, and modeling the conditional densities of the regression and the measurement errors $f_{U_Y|X}$ and $f_{U_W|X}$.

2.1. *Density of the Covariate*

We use Dirichlet process mixture models (DPMMs) (Ferguson, 1973; Escobar and West, 1995) for modeling f_X. For modeling a density f, a DPMM with concentration parameter α, base measure P_0, and mixture components coming from a parametric family $\{f_c(\cdot \mid \boldsymbol{\phi}) : \boldsymbol{\phi} \sim P_0\}$, can be specified as

$$f(\cdot) = \sum_{k=1}^{\infty} \pi_k \; f_c(\cdot \mid \boldsymbol{\phi}_k), \quad \boldsymbol{\phi}_k \sim P_0,$$

$$\pi_k = s_k \prod_{j=1}^{k-1} (1 - s_j), \quad s_k \sim \mathrm{Beta}(1, \alpha).$$

In the literature, this construction of random mixture weights $\{\pi_k\}_{k=1}^{\infty}$ (Sethuraman, 1994), is often represented as $\pi \sim \mathrm{Stick}(\alpha)$. DPMMs are therefore mixture models with a potentially infinite number of mixture components or "clusters." For a given data set of finite size, however, the number of active clusters exhibited by the data is finite and can be inferred from the data.

Choice of the parametric family $\{f_c(\cdot \mid \boldsymbol{\phi}) : \boldsymbol{\phi} \sim P_0\}$ is important. Mixtures of normal kernels are, in particular, very

popular for their flexibility and computational tractability (West, Müller, and Escobar, 1994; Escobar and West, 1995). In this article also, f_X is specified as a mixture of normal kernels, with a conjugate normal-inverse-gamma (NIG) prior on the location and scale parameters

$$f_X(X) = \sum_{k=1}^{\infty} \pi_{X,k} \text{ Normal}(X \mid \mu_k, \sigma_k^2), \quad (3)$$

$$\pi_X \sim \text{Stick}(\alpha_X), \qquad (\mu_k, \sigma_k^2) \sim \text{NIG}(\mu_0, \sigma_0^2/\nu_0, \gamma_0, \sigma_0^2). \quad (4)$$

Here Normal$(\cdot \mid \mu, \sigma^2)$ denotes a normal distribution with mean μ and standard deviation σ. In what follows, the generic notation p_0 will sometimes be used for specifying priors and hyper-priors.

2.2. *Regression Function*

The problem of flexible modeling of the regression function is addressed in this subsection. Specifically, we are interested in models that are numerically stable and lead to easy and efficient posterior computation.

Mixtures of splines (de Boor, 2000) and Gaussian processes (Rasmussen and Williams, 2006) are immensely popular and successful for regression problems with precisely measured covariates. For measurement error problems, however, Gaussian process priors are not particularly suitable since the unobserved values of X would be involved in the prior covariance matrix of the regression function and will not be conditionally independent in the posterior, rendering the method computationally complex and numerically unstable. Splines, on the other hand, do not lead to additional complications in the measurement error setup. In regression with errors in covariates, Carroll et al (1999b) and Berry et al. (2002) used penalized mixtures of truncated polynomial splines to model the regression function. In this article, we model the regression function as a mixture of B-spline basis functions with a smoothness inducing prior on the coefficients (Eilers and Marx, 1996). The B-splines are locally supported, nearly orthogonal and can be computed using a simple recursion. These properties of B-splines make them numerically more stable than polynomial splines. Later on, B-splines are again used to model conditional heteroscedasticity in regression and measurement errors. Mixture of B-splines for modeling the regression function, thus, also allows reuse of programming codes for fitting different components of the complete model.

For a given positive integer K_R, partition an interval $[A, B]$ of interest into K_R subintervals using knot points $t_{R,1} = \cdots = t_{R,q+1} = A < t_{R,q+2} < \cdots < t_{R,q+K_R+1} = \cdots = t_{R,2q+K_R+1} = B$. Using these knot points, $(q + K_R) = J_R$ B-spline bases of degree q, denoted by $\mathbf{B}_{q,J_R} = \{b_{q,1}, b_{q,2}, \ldots, b_{q,J_R}\}$, can be defined through the recursion relation given on page 90 of de Boor (2000). A flexible model for the regression function is then given by

$$r(X) = \sum_{j=1}^{J_R} b_{q,j}(X)\xi_{R,j} = \mathbf{B}_{q,J_R}(X)\boldsymbol{\xi}_R, \quad (5)$$

$$p_0(\boldsymbol{\xi}_R \mid J_R, \sigma^2_{R,\xi}) \propto \exp\{-\boldsymbol{\xi}_R^{\mathrm{T}} P_R \boldsymbol{\xi}_R/(2\sigma^2_{R,\xi})\}, \quad (6)$$

where $\boldsymbol{\xi}_R = \{\xi_{R,1}, \xi_{R,2}, \ldots, \xi_{R,J_R}\}^{\mathrm{T}}$. We choose $P_R = D_R^{\mathrm{T}} D_R$, where D_R is a $J_R \times (J_R + 2)$ matrix such that $D_R\boldsymbol{\xi}_R$ computes the second differences in $\boldsymbol{\xi}_R$. The prior $p_0(\boldsymbol{\xi}_R \mid \sigma^2_{R,\xi})$ induces smoothness in the coefficients because it penalizes $\sum_{j=1}^{J_R}(\Delta^2\xi_{R,j})^2 = \boldsymbol{\xi}_R^{\mathrm{T}} P_R \boldsymbol{\xi}_R$, the sum of squares of the second order differences in $\boldsymbol{\xi}_R$. The variance parameter $\sigma^2_{R,\xi}$ plays the role of a smoothing parameter: the smaller the value of $\sigma^2_{R,\xi}$, the stronger the penalty and the smoother the regression function.

2.3. *Conditional Densities of Regression and Measurement Errors*

The problems of flexible modeling of $f_{U_Y|X}$ and $f_{U_W|X}$ and are now addressed. Specifically, we are interested in models for $f_{U_Y|X}$ and $f_{U_W|X}$ that can capture departures from normality, accommodate conditional heteroscedasticity and allow efficient estimation even though the conditioning variable X is not measured accurately.

In the context of regression analysis with precisely measured covariates, moment constrained infinite mixture models have recently been used by Pelenis (2014) for flexible modeling of the distribution of regression errors. Pelenis considered the mixture $f_{U_Y|X}(U_Y \mid X) = \sum_{k=1}^{\infty} \pi_k(X)\{p_k \text{ Normal}(U_Y \mid \mu_{k1}, \sigma^2_{k1}) + (1 - p_k) \text{ Normal}(U_Y \mid \mu_{k2}, \sigma^2_{k2})\}$, with the moment constraint $p_k\mu_{k1} + (1 - p_k)\mu_{k2} = 0$ for all k. The zero mean constraint of the components is inherited by the mixture, so the usual assumption $E(U_Y \mid X) = 0$ is satisfied. Use of two-component mixture of normals as components also allows the mixture to capture departures from normality in other unconstrained aspects of the distribution including skewness, multimodality and heavy tails. Incorporating covariate information X in modeling the mixture probabilities, this model allows all aspects of the error distribution to vary flexibly with the covariates including the conditional variance.

However, in the context of measurement error problems, particularly when the number of surrogates for each unobserved X is small and the measurement errors have significant variability, the mixture probabilities π_k's, which are functions of X and other parameters, become numerically unstable, making simultaneous learning of X and other parameters of the model difficult and the residuals become noninformative about the true error distribution. For these reasons the model of Pelenis (2014) may not be efficient for modeling $f_{U_W|X}$. In addition, the problem of modeling the conditional distribution of regression errors is even harder. First, there are usually multiple proxies but only a single response available for each unknown X. Hence, there is substantially less data available for modeling $f_{U_Y|X}$. Second, the conditional mean of the surrogates, given X, is simply X, so the residuals can be readily calculated. In contrast, to calculate the residuals for regression errors the unknown regression function also needs to be estimated, and hence the regression residuals are much less informative about the truth.

Measurement error problems therefore require semiparametric alternatives, that is, models that may be less flexible than that of Pelenis (2014) but still allow us to break free from many unrealistic and restrictive parametric assumptions of the existing literature. In this article one such

semiparametric route is taken that can be used for modeling both $f_{U_W|X}$ and $f_{U_Y|X}$. To avoid unnecessary repetition, the subscripts Y and W are dropped and the generic notation U is used to refer to both U_Y and U_W. In the sections that follow the subscripts reappear as and when necessary. The same convention is followed for different components of the models and the parameters involved. We assume

$$[U \mid X] = v^{1/2}(X)\epsilon, \tag{7}$$

where ϵ, referred to as the "scaled errors" henceforth, are independently and identically distributed with zero mean and are also independent of X. The problem of modeling $f_{U|X}$ now reduces to the problem of modeling two separate functions, v and f_ϵ. Instances of modeling conditional heteroscedasticity by making the structural assumption (7) on the errors are abundant in the literature. See, for example, Crainiceanu et al. (2007), Liu, Tong, and Wang (2006), Chan et al. (2006), and Staudenmayer, Ruppert, and Buonaccorsi (2008). Invariably these authors assumed $f_\epsilon = \text{Normal}(0, 1)$ and focused only on modeling the variance function. In this article, the problem of flexible modeling of f_ϵ is also addressed.

In the context of our problem, model (7) can be motivated as follows. In the real world applications we are interested in, conditional heteroscedasticity is a very prominent feature of the regression and measurement errors, see Section 4. Model (7) captures this through the "variance function" $v(X)$. Other features of $f_{U|X}$, including skewness, multimodality and heavy tails, are derived from f_ϵ. For reasons described above, variations in these features of $f_{U|X}$ for varying values of X, if any, are extremely difficult to detect when X is measured imprecisely. Model (7) therefore assumes the scaled errors to be independent of X. The model is, however, fairly robust to departures from the assumed multiplicative structure. Numerical evidence to support the robustness are presented in the Supplementary Materials.

Since $v(X)^{1/2}\epsilon = \{c\ v(X)^{1/2}\}(\epsilon/c)$ for any $c > 0$, the representation of U given by (7) is not unique. However, for inference on the regression function r, the variance function v and the density of the scaled errors f_ϵ need not be separately identifiable. Conditional variability of U may simply be obtained as $\text{var}(U \mid X) = v(X)\text{var}(\epsilon)$, and to aid in comparison, versions of f_ϵ adjusted to have unit variance may be retained for each MCMC iteration.

To model f_ϵ, we can now use a DPMM with mixture components as specified in Pelenis (2014). That is, we let $f_\epsilon(\epsilon) = \sum_{k=1}^{\infty} \pi_{\epsilon k} f_{c\epsilon}(\epsilon \mid p_k, \mu_{k1}, \mu_{k2}, \sigma_{k1}^2, \sigma_{k2}^2)$, $\pi_\epsilon \sim \text{Stick}(\alpha_\epsilon)$, where $f_{c\epsilon}(\epsilon \mid p, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = \{p\ \text{Normal}(\epsilon \mid \mu_1, \sigma_1^2) + (1-p)\ \text{Normal}(\epsilon \mid \mu_2, \sigma_2^2)\}$, subject to the moment constraint $p\mu_1 + (1-p)\mu_2 = 0$. The moment constraint of zero mean implies that each component density can be described by four parameters. One such parametrization that facilitates prior specification is in terms of parameters $(p, \tilde{\mu}, \sigma_1^2, \sigma_2^2)$, where (μ_1, μ_2) can be retrieved from $\tilde{\mu}$ as $\mu_1 = c_1\tilde{\mu}$, $\mu_2 = c_2\tilde{\mu}$, where $c_1 = (1-p)/\{p^2 + (1-p)^2\}^{1/2}$ and $c_2 = -p/\{p^2 + (1-p)^2\}^{1/2}$. Clearly the zero mean constraint is satisfied, since $p\mu_1 + (1-p)\mu_2 = \{pc_1 + (1-p)c_2\}\tilde{\mu} = 0$. The family includes normal densities as special cases with $(p, \tilde{\mu}) = (0.5, 0)$ or $(0, 0)$ or $(1, 0)$. Symmetric component densities are obtained as special cases when $p = 0.5$ or $\tilde{\mu} = 0$. The mixture is symmetric when the all components are as well. Specification of the prior for f_ϵ is completed by assuming noninformative priors for $(p, \tilde{\mu}, \sigma_1^2, \sigma_2^2)$. Let $\text{Unif}(\ell, u)$ denote a uniform distribution on the interval (ℓ, u), and $\text{IG}(a, b)$ denote an inverse-Gamma distribution with shape parameter a and scale parameter b. The DPMM prior on f_ϵ can then be specified as

$$f_\epsilon(\epsilon) = \sum_{k=1}^{\infty} \pi_{\epsilon k} f_{c\epsilon}(\epsilon \mid p_k, \tilde{\mu}_k, \sigma_{k1}^2, \sigma_{k2}^2), \quad \pi_\epsilon \sim \text{Stick}(\alpha_\epsilon), \tag{8}$$

$$(p_k, \tilde{\mu}_k, \sigma_{k1}^2, \sigma_{k2}^2) \sim \text{Unif}(0, 1)\text{Normal}(0, \sigma_{\tilde{\mu}}^2)\text{IG}(a_\epsilon, b_\epsilon)\text{IG}(a_\epsilon, b_\epsilon). \tag{9}$$

To model v flexibly, similar to (5), we use a mixture of B-spline basis functions with smoothness inducing priors on the coefficients. The coefficients of (5) are now exponentiated to ensure the function is positive.

$$v(X) = \sum_{j=1}^{J} b_{q,j}(X)\exp(\xi_j) = \mathbf{B}_{q,J}(X)\exp(\boldsymbol{\xi}), \tag{10}$$

$$p_0(\boldsymbol{\xi} \mid J, \sigma_\xi^2) \propto \exp\{-\boldsymbol{\xi}_R^{\mathrm{T}} P_R \boldsymbol{\xi}_R/(2\sigma_\xi^2)\}. \tag{11}$$

In the Supplementary Materials, we discuss implementation issues, including automatic choice of the hyperparameters, and details of the posterior calculations, the choice of initial values, the structure of the B-splines used and additional supporting simulations.

3. Simulation Experiments

3.1. *Basics*

Based on M samples $\boldsymbol{\xi}_R^{(m)}$, $m = 1, \ldots, M$, drawn from the posterior, a Monte Carlo estimate $\widehat{r}(X)$ can be obtained as $\widehat{r}(X) = M^{-1}\sum_{m=1}^{M} \mathbf{B}_{q,J_R}(X)\boldsymbol{\xi}_R^{(m)}$. The integrated squared error of estimation of the regression function $r(\cdot)$ by the estimator $\widehat{r}(\cdot)$ is defined as $\text{ISE} = \int\{r(X) - \widehat{r}(X)\}^2\,\mathrm{d}X$. Based on B simulated data sets, a Monte Carlo estimate of the mean integrated squared error (MISE) is given by $\text{MISE}_{\text{est}} = B^{-1}\sum_{b=1}^{B}\sum_{i=1}^{N}\{r(X_i^\Delta) - \widehat{r}^{(b)}(X_i^\Delta)\}^2\Delta_i$, where $\{X_i^\Delta\}_{i=0}^N$ are a set of grid points on the range of X and $\Delta_i = (X_i^\Delta - X_{i-1}^\Delta)$ for all i.

We performed simulation experiments to compare the MISE performance of our method with that of Berry et al. (2002), referred to as the BCR method henceforth, a naive method, and a deconvoluting kernel based estimator, referred to as the DKE method henceforth. The naive method ignores the measurement errors and treats the subject specific means as the true covariates but accommodates conditional heteroscedasticity in the regression errors. The DKE method is implemented using the "DeconNpr" function from the R package "decon" (Wang and Wang, 2011) allowing subject specific heteroscedasticity. We compared the methods over a wide range of possibilities. The reported estimated MISEs are

all based on a grid of 500 equidistant points on $[-2, 2]$ for $B = 200$ simulated data sets. In each case 10,000 MCMC iterations were run and the initial 5,000 iterations were discarded as burn-in. To reduce autocorrelation among the sampled values, the post burn-in samples were thinned by a thinning interval of length 5.

3.2. *Setup 1: Homoscedasticity and Normally Distributed X*

We mimic simulation experiment setups from Berry et al. (2002). We let $f_X(X) = \text{Normal}(X \mid 0, 1)$, $r(X) = \sin(\pi X/2)/[1 + 2X^2\{\text{sign}(X) + 1\}]$, $U_W \sim \text{Normal}(0, 0.8^2)$, $\text{var}(U_Y) = 0.3^2$ and compare the methods over a factorial combination of (i) two sample sizes $n = 500, 1000$; (ii) two choices for the number of surrogates per subject $m = 2, 3$; and (iii) five different distributions for the scaled regression errors (three light-tailed densities including the Normal density and two heavy-tailed densities, see Table 1 and Figure 1).

The results are presented in Table 2. The results show that the MISE performance of our method is better than the performance of the BCR method in all 20 cases considered, including the case of normally distributed regression errors, when the parametric assumptions of the BCR method are all satisfied. Results produced by our method and the BCR method for this special case are summarized in Figure 2. The BCR method uses truncated polynomial splines (P-splines), while we are using B-splines. As opposed to P-splines, B-splines are locally supported and nearly orthogonal, and

Table 1

The distributions used to generate the scaled errors in the simulation experiments of Section 3. SMRTCN($K, \boldsymbol{\pi}_\epsilon, \mathbf{p}, \widetilde{\boldsymbol{\mu}}, \boldsymbol{\sigma}_1^2, \boldsymbol{\sigma}_2^2$) denotes the scaled version of a K component mixture of moment restricted two-component normals: $\sum_{k=1}^{K} \pi_{\epsilon k} f_{c\epsilon}(\cdot \mid p_k, \tilde{\mu}_k, \sigma_{k1}^2, \sigma_{k2}^2)$, scaled to have variance one. Laplace(μ, b) denotes a Laplace distribution with location μ and scale b. With μ_k denoting the kth order central moments of the scaled errors, the skewness and excess kurtosis of the distribution of scaled errors are measured by the coefficients $\gamma_1 = \mu_3$ and $\gamma_2 = \mu_4 - 3$, respectively. The shapes of these densities are illustrated in Figure 1.

Distribution of scaled errors	Skewness (γ_1)	Excess kurtosis (γ_2)
(a) Normal(0,1)	0	0
(b) SMRTCN(1,1,0.4,2,2,1)	0.499	−0.966
(c) SMRTCN(1,1,0.5,2,1,1)	0	−1.760
(d) SMRTCN{2,(0.8,0.2), (0.5,0.5),(0,0),(0.25,5),(0.25,5)}	0	7.524
(f) Laplace($0,2^{-1/2}$)	0	3

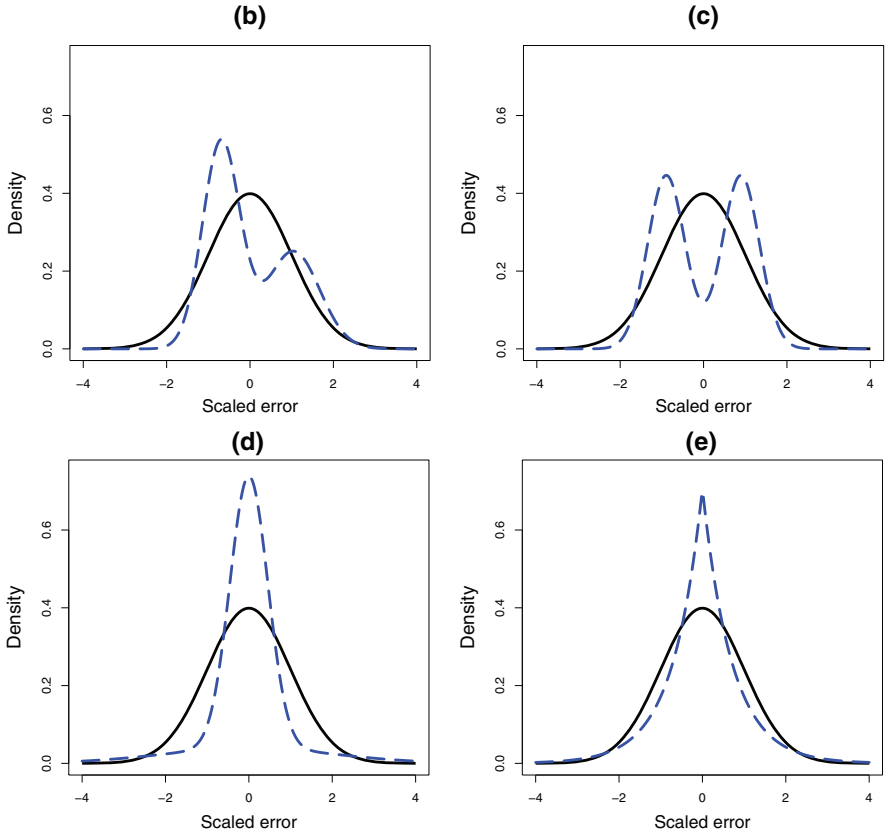

Figure 1. The distributions used to generate the scaled regression and measurement errors in simulation experiments, superimposed over a standard normal density—(a) standard normal (solid lines in each panel, not shown separately), (b) asymmetric bimodal, (c) symmetric bimodal, (d) symmetric heavy-tailed, and (e) symmetric heavy-tailed with a sharp peak at zero. The figure appears in color in the electronic version of this article.

Table 2
Mean integrated square error (MISE) performance of our model (BSP) compared to the model of Berry et al. (2002) (BCR), a naive model that ignores measurement errors (Naive), and a deconvolution kernel estimator (DKE) for homoscedastic simulation experiments in Section 3.2, with $X \sim Normal(0,1)$, $r(X) = \sin(\pi X/2)/[1+2X^2\{sign(X)+1\}]$, $U_W \sim Normal(0, 0.8^2)$ *and five different densities for the scaled regression errors (three light-tailed and two heavy-tailed, see Table 1 and Figure 1 for details) with* $var(U_Y) = 0.3^2$. *Our method allows non-normality of X and heteroscedasticity.*

True error distribution	Sample size	Number of replicates	MISE × 100 BCR	BSP	Naive	DKE
Normal	500	2	4.98	**2.84**	16.66	24.85
		3	4.09	**1.82**	11.97	22.84
	1000	2	3.11	**1.53**	18.05	20.21
		3	2.42	**0.96**	10.88	16.64
Light-tailed bimodal skewed	500	2	4.73	**2.20**	17.75	26.05
		3	4.23	**1.61**	12.29	26.73
	1000	2	3.12	**1.30**	18.45	22.63
		3	2.50	**0.92**	10.80	19.67
Light-tailed bimodal symmetric	500	2	4.83	**3.49**	18.60	26.08
		3	4.30	**1.78**	12.50	24.05
	1000	2	3.25	**2.64**	18.60	22.24
		3	2.53	**0.85**	11.13	18.67
Heavy-tailed symmetric 1	500	2	4.78	**1.82**	17.75	21.69
		3	4.09	**1.38**	11.37	19.22
	1000	2	2.87	**1.10**	19.25	16.42
		3	2.38	**0.76**	11.08	15.90
Heavy-tailed symmetric 2	500	2	4.77	**2.34**	18.72	24.28
		3	4.14	**1.77**	11.75	23.76
	1000	2	2.99	**1.24**	18.22	17.68
		3	2.41	**0.92**	10.69	16.66

are therefore numerically more stable than P-splines. This increased numerical stability of our model results in better performance even in situations when the parametric assumption of the BCR model are satisfied. Additional simulation results that support these findings are presented in the Supplementary Materials.

3.3. *Setup 2: Homoscedasticity and Non-Normally Distributed X*

Next we keep the error variances constant at $\text{var}(U_Y) = 0.3^2$ and $\text{var}(U_W) = 0.8^2$ and consider the same regression function $r(X) = \sin(\pi X/2)/[1+2X^2\{\text{sign}(X)+1\}]$ as before, but allow all the densities f_X, f_{U_Y}, and f_{U_W} to differ from Normality. We now let $f_X(X) = 0.8\,\text{Normal}(X \mid -1, 0.5) + 0.2\,\text{Normal}(X \mid 1, 0.5)$ and compare the methods over a factorial combination of (i) two sample sizes $n = 500, 1000$; (ii) two choices for the number of surrogates per subject $m = 2, 3$; and (iii) five different distributions for the scaled errors (three light-tailed and two heavy-tailed, see Table 1 and Figure 1). The results are presented in Table 3.

3.4. *Setup 3: Heteroscedasticity and Non-Normally Distributed X*

Finally we consider conditionally heteroscedastic errors and let $v_Y(X) = (0.3 + X/8)^2$ and $v_W(X) = (0.8 + X/4)^2$. As before we let $f_X(X) = 0.8\,\text{Normal}(X \mid -1, 0.5) + 0.2\,\text{Normal}(X \mid 1, 0.5)$, $r(X) = \sin(\pi X/2)/[1+2X^2\{\text{sign}(X)+1\}]$ and compare the methods over a factorial combination of (i) two sample sizes $n = 500, 1000$; (ii) two choices for the number of surrogates per subject $m = 2, 3$; (iii) and five different distributions for the scaled errors (three light-tailed and two heavy-tailed, see Table 1 and Figure 1). The results are presented in Table 4. Results for the heavy tailed error distribution (d) are summarized in Figure 3.

Results presented in Tables 3 and 4 show that our method vastly out-performed the BCR model in all 40 cases considered. For example, in Table 3, for the symmetric heavy-tailed error distribution (d) with $n = 1000$, the improvement in MISE over the BCR model is $18.17/1.21 \approx 15$ times when there are 2 surrogates per unit and $14.50/0.94 \approx 15$ times when there are 3 surrogates per unit. Similarly, in Table 4, for the error distribution (d) with $n = 1000$, the improvement in MISE is $23.89/1.49 \approx 16$ times for 2 surrogates per unit and $15.42/1.05 \approx 15$ times for 3 surrogates per unit.

3.5. *Additional Simulations*

The use of B-splines in our model, as opposed to P-splines used in the BCR model, can explain the somewhat surprising results of Section 3.2, where our method was shown to outperform the BCR method even when the parametric assumptions of the BCR method were satisfied. Additional simulation experiments that support this claim are presented in the Supplementary Materials, where we compared our method with an improved version of the BCR method, referred to as the BCRB method, that makes the same parametric assumptions as the BCR model but uses B-splines, not P-splines, to model the regression function. We considered two subcases from each of the three scenarios considered above. When the parametric assumptions of the BCR model were true, the BCRB method outperformed our method. In all other cases, our method outperformed the BCRB method.

Additional simulation experiments were also performed to assess the MISE performance of our method when the true error generating densities depart from the multiplicative structural assumption (7). Results that suggest our model is fairly robust to such departures are presented in the Supplementary Materials.

The results of these additional simulation experiments emphasize the importance of using flexible but numerically stable components for building measurement error models.

4. Example

As an illustration of our methodology, we analyze data collected in the Eating at America's Table (EATS) study (Subar et al., 2001), a large scale epidemiologic study conducted by the National Cancer Institute (NCI) to assess the role of diet in the etiology and prevention of diseases.

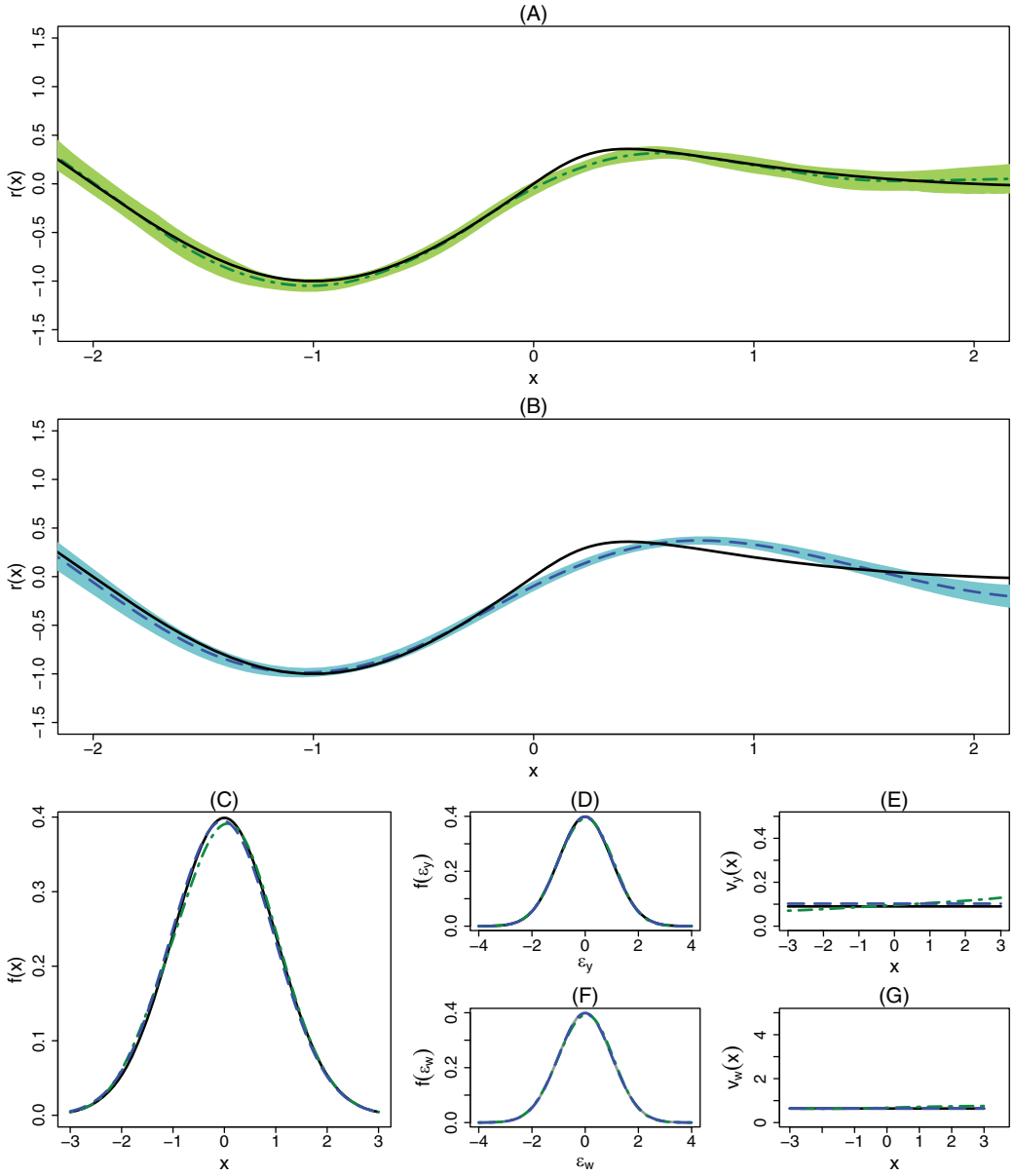

Figure 2. Results for our method corresponding to the median MISE in the simulation of Section 3.2 when the parametric assumptions of Berry et al. (2002) are satisfied—X is Normal, the regression errors and the measurement errors are Normal and homoscedastic. Sample size n = 1000 and $m = 3$ replicates per subject. In all panels the solid lines represent the truth, the dot-dashed lines represent the estimates obtained by our method and the dashed lines represent the estimates obtained by the method of Berry et al. (2002) (BCR). (A) The regression function estimated by our method and (B) the regression function estimated by the BCR method. They are presented separately for clarity. In (A) and (B), the gray dots represent estimated posterior mean of the covariate values (x-axis) and the observed responses (y-axis), and the bands represent point wise 90% credible intervals. (C) The density of the covariate. The truth is standard normal. (D) The density of the scaled regression errors. The truth is standard normal. (E) The variance function of the regression errors. The truth is constant. (F) The density of the scaled measurement errors. The truth is standard normal. (G) The variance function of the measurement errors. The truth is constant. The gray dots represent subject-specific sample means (x-axis) and variances (y-axis) of the surrogates. The figure appears in color in the electronic version of this article.

Table 3

Mean integrated square error (MISE) performance of our model (BSP) compared to the model of Berry et al. (2002) (BCR), a naive model that ignores measurement errors (Naive), and a deconvolution kernel estimator (DKE) for the simulation experiments in Section 3.3, with $X \sim 0.8\ Normal(-1, 0.5) + 0.2\ Normal(1, 0.5)$, $r(X) = \sin(\pi X/2)/[1 + 2X^2\{sign(X) + 1\}]$ *and five different densities for the scaled errors (three light-tailed and two heavy-tailed, see Table 1 and Figure 1 for details) with* $var(U_Y) = 0.3^2$ *and* $var(U_W) = 0.8^2$. *Our method allows non-normality of X and heteroscedasticity.*

True error distribution	Sample size	Number of replicates	MISE × 100			
			BCR	BSP	Naive	DKE
Normal	500	2	20.30	**6.97**	30.80	43.58
		3	17.29	**3.64**	21.15	37.08
	1000	2	15.85	**3.44**	31.82	37.87
		3	13.18	**2.31**	21.11	32.14
Light-tailed bimodal skewed	500	2	19.52	**4.66**	34.44	46.67
		3	16.20	**2.84**	23.86	38.36
	1000	2	14.01	**2.61**	33.57	37.64
		3	11.79	**1.55**	23.22	33.30
Light-tailed bimodal symmetric	500	2	20.18	**5.09**	34.67	45.97
		3	17.15	**3.20**	24.08	37.54
	1000	2	15.73	**2.67**	31.61	38.95
		3	13.01	**1.87**	22.52	32.56
Heavy-tailed symmetric 1	500	2	24.02	**2.19**	20.39	37.08
		3	18.98	**1.76**	16.76	33.49
	1000	2	18.17	**1.21**	21.16	32.00
		3	14.50	**0.94**	17.56	28.96
Heavy-tailed symmetric 2	500	2	21.74	**4.64**	26.76	40.84
		3	18.25	**3.20**	19.96	37.54
	1000	2	16.99	**2.32**	25.27	33.90
		3	13.48	**1.67**	19.40	29.60

Table 4

Mean integrated square error (MISE) performance of our model (BSP) compared to the model of Berry et al. (2002) (BCR), a naive model that ignores measurement errors (Naive), and a deconvolution kernel estimator (DKE) for the simulation experiments in Section 3.4, with $X \sim 0.8\ Normal(-1, 0.5) + 0.2\ Normal(1, 0.5)$, $r(X) = \sin(\pi X/2)/[1 + 2X^2\{sign(X) + 1\}]$, $v_Y(X) = (0.3 + X/8)^2$, $v_W(X) = (0.8 + X/4)^2$, *and five different densities for the scaled errors (three light-tailed and two heavy-tailed, see Table 1 and Figure 1 for details).*

True error distribution	Sample size	Number of replicates	MISE × 100			
			BCR	BSP	Naive	DKE
Normal	500	2	30.85	**8.47**	21.59	55.00
		3	24.74	**5.21**	15.44	41.57
	1000	2	35.09	**5.80**	18.44	48.03
		3	27.36	**3.93**	11.58	36.14
Light-tailed bimodal skewed	500	2	43.92	**7.54**	19.74	52.35
		3	31.47	**2.57**	12.25	39.36
	1000	2	41.53	**2.55**	18.07	46.41
		3	30.10	**1.67**	12.69	34.53
Light-tailed bimodal symmetric	500	2	33.72	**4.15**	20.48	57.02
		3	29.08	**2.57**	13.17	40.98
	1000	2	33.50	**2.25**	17.42	51.66
		3	26.84	**1.25**	11.50	35.59
Heavy-tailed symmetric 1	500	2	26.44	**2.93**	12.84	65.74
		3	15.80	**2.07**	9.56	43.57
	1000	2	23.89	**1.49**	13.40	60.53
		3	15.42	**1.05**	10.05	39.29
Heavy-tailed symmetric 2	500	2	28.58	**6.11**	16.38	51.92
		3	20.01	**3.89**	11.65	40.99
	1000	2	26.73	**3.44**	15.16	47.53
		3	18.57	**2.31**	10.45	35.83

The most practical and economical method for collection of dietary data in large epidemiologic studies is the food frequency questionnaire (FFQ). In most studies the respondents receive the FFQs by mail and are instructed to complete the questionnaires independently and return them in postage paid return envelopes. For obvious reasons the data collected by FFQs on dietary intakes typically have a considerable measurement error, and need to be validated prior to or as part of dietary research. Improved methods can provide a better idea about the relationship between reported FFQs and the true unobserved dietary intakes. The study of the relationship between reported FFQs and the true dietary intakes is therefore of great importance in nutritional epidemiology. Other approaches of data collection include 24 hour dietary recalls, where participants are interviewed and their responses recorded by trained professionals. Compared to FFQs, 24 hour recalls are therefore much more expensive but the data collected are also more accurate and detailed and can be used to validate the FFQs.

In the EATS study, $n = 965$ participants returned FFQs (Y_i). They were interviewed $m_i = 4$ times over the course of a year and their 24 hour dietary recalls (W_{ij}) were recorded. The true long term average dietary intakes (X_i) are unobserved.

This is a non-standard setting in that Y is not a health outcome, but rather is also a surrogate for X. Ideally one would thus expect both W and Y to be unbiased for X, that is, $E(W \mid X) = E(Y \mid X) = X$. While Y and W are both proxies for X, the 24 hour recalls W_{ij} are recorded by trained personnel after thoroughly conducted interviews, whereas the FFQs Y_i are merely self-reported. As compared to the 24 hour recalls W, the FFQs Y are therefore much less reliable surrogates for the unobserved X, and some departure from the ideal relationship $E(Y \mid X) = X$ may be suspected. Our goal therefore is to estimate the relationship between reported FFQs and the

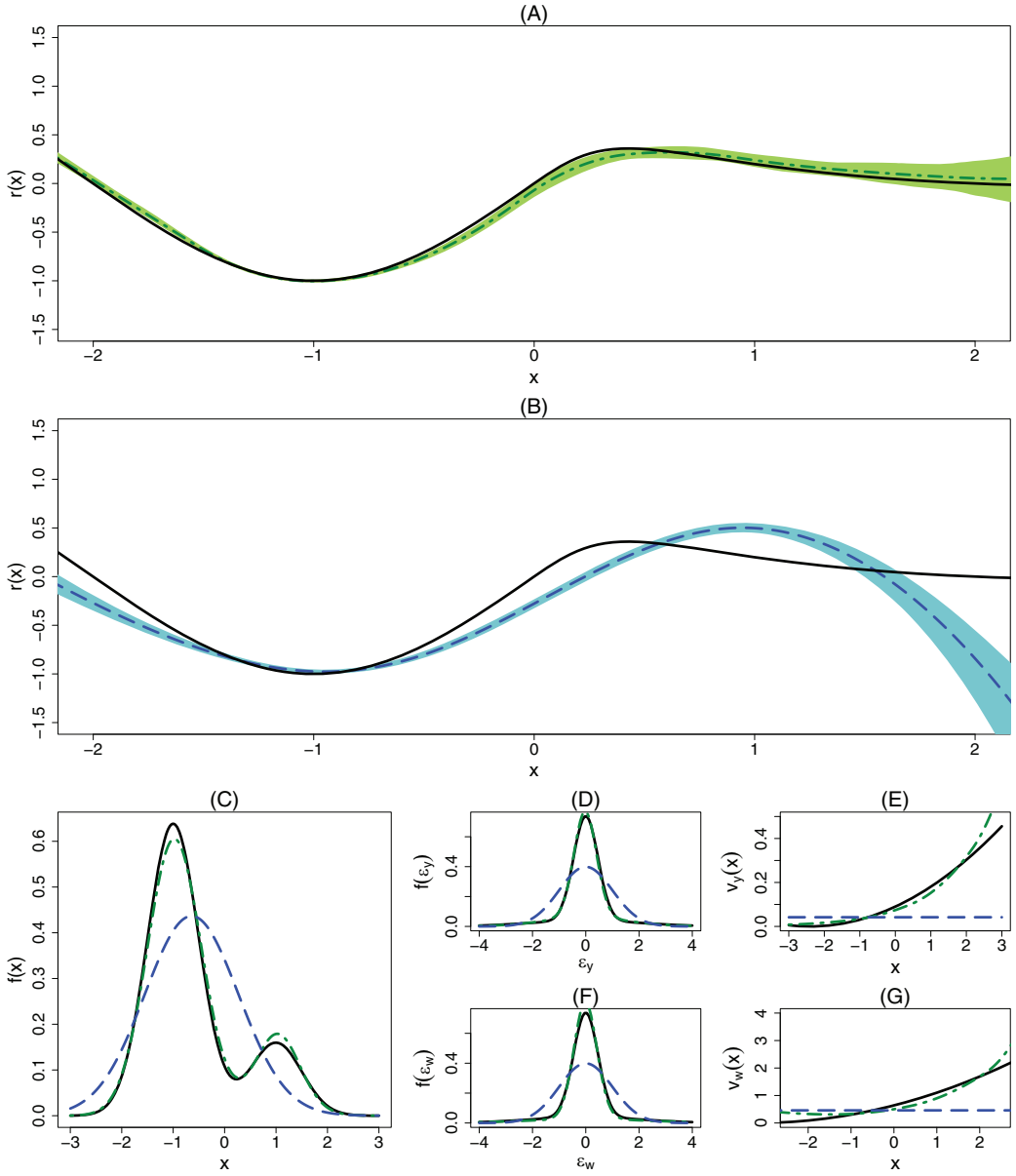

Figure 3. Results for heavy-tailed error distribution (d), sample size n = 1000 and $m = 3$ replicates per subject corresponding to the median MISEs in the simulation of Section 3.4 when X is not Normally distributed, the regression errors and the measurement errors are conditionally heteroscedastic and non-Normal. In all panels the solid lines represent the truth, the dot-dashed lines represent the estimates obtained by our method and the dashed lines represent the estimates obtained by the method of Berry et al. (2002) (BCR). (A) The regression function estimated by our method and (B) the regression function estimated by the BCR method. They are presented separately for clarity. In (A) and (B), the gray dots represent estimated posterior mean of the covariate values (x-axis) and the observed responses (y-axis), and the bands represent point wise 90% credible intervals. (C) The density of the covariate. (D) The density of the scaled regression errors. (E) The variance function of the regression errors. (F) The density of the scaled measurement errors. (G) The variance function of the measurement errors. The gray dots represent subject-specific sample means (x-axis) and variances (y-axis) of the surrogates. The figure appears in color in the electronic version of this article.

true dietary intakes through a flexible regression relationship $E(Y \mid X) = r(X)$, treating the 24 hour recalls W_{ij} as unbiased proxies.

Results for daily intakes of sodium produced by our method and the method by Berry et al. (2002) (BCR) are summarized in Figure 4. Conditional heteroscedasticity of measurements errors is one salient feature of the proxies W_{ij}, clearly identifiable from the plot of subject-specific means versus subject-specific variances. Since Y_i is essentially also a surrogate for X_i, a similar conditional heteroscedasticity pattern is expected

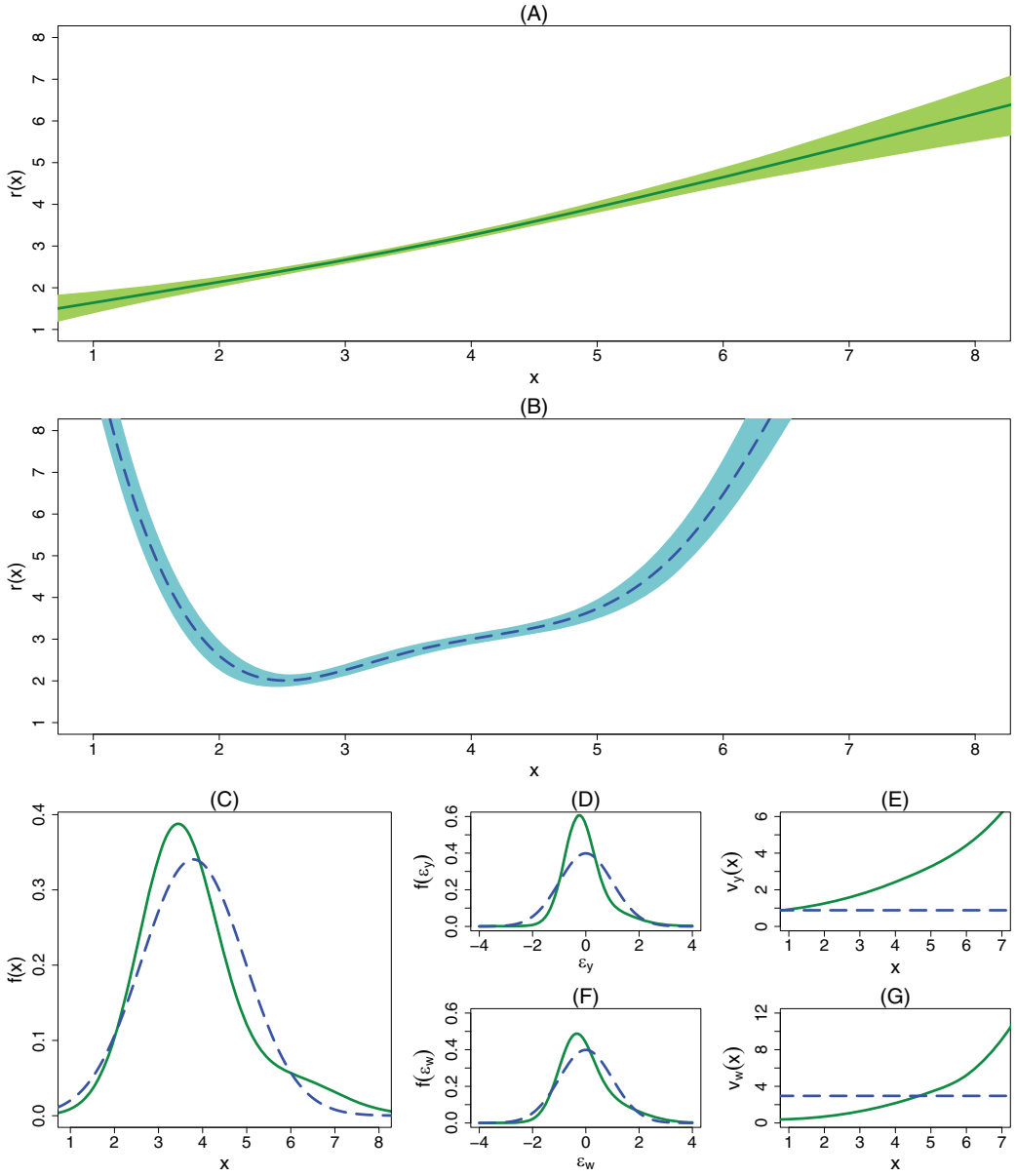

Figure 4. Results for sodium from the EATS data set. In all panels the solid lines represent the estimates obtained by our method and the dashed lines represent the estimates obtained by the method of Berry et al. (2002). (A) The regression function estimated by our method and (B) the regression function estimated by the BCR method. They are presented separately for clarity. In (A) and (B), the gray dots represent estimated posterior mean of the covariate values (x-axis) and the observed responses (y-axis), and the bands represent point wise 90% credible intervals. (C) The density of the covariate. (D) The density of the scaled regression errors. (E) The variance function of the regression errors. (F) The density of the scaled measurement errors. (G) The variance function of the measurement errors. The gray dots represent subject-specific sample means (x-axis) and variances (y-axis) of the surrogates. The figure appears in color in the electronic version of this article.

in the errors in the reported FFQs. The BCR method assumes homoscedasticity and normality for the true intakes and regression and measurement errors. Our method, on the other hand, accommodates conditional heteroscedasticity in both regression and measurement errors and also captures departures from normality in their densities and the density of the true intakes, while providing a more robust and realistic estimate of the regression relationship. The results produced by our method indicate that the FFQs are over-reported for low true intakes and are under-reported for high true intakes. The results also indicate that the density of the true sodium intakes and the densities of the regression and measurement errors are all positively skewed. As expected the estimated conditional heteroscedasticity patterns in the 24 recalls and the FFQs are also very similar. On the other hand, although some departure from the ideal relationship

$E(Y \mid X) = X$ is suspected, the regression function estimated by the BCR method is clearly unrealistic. This is not surprising, particularly in view of the strong parametric assumptions made by the BCR method. This example vividly illustrates the importance of the problem we addressed and methodology we described.

5. Discussion

We considered the problem of robust estimation of a regression function in the presence of conditionally heteroscedastic regression and measurement errors. The problem, though extremely important for real world applications, had never been addressed before in the literature. The methodology we described therefore makes important contributions to the measurement error literature. Efficiency of the models in estimating the true regression function was illustrated through simulation experiments for a variety of situations. In particular, we showed that our method vastly dominates the method of Berry et al. (2002).

The proposed methodology being based on Bayesian hierarchical framework, different components can be separately reused to adapt to a wide variety of scenarios. For instance, if one suspects the covariate, the regression and the measurement errors to deviate from normality and homoscedasticity but thinks that there is a quadratic relationship between the latent covariate and the response, one might fit a quadratic regression function while still using other flexible components of our model to capture the suspected deviations. Our methods can also be easily adapted to other special subcases, for example, homoscedasticity of the regression error and/or the measurement errors, normality of the measurement errors, etc. In simulations not reported here, we have observed that some gain in efficiency can be obtained if one restricts our model to these special cases. However, the efficiency gain is generally not enormous.

In some settings attempts are made to transform Y and W so that an ideal model is achievable: constant variance of transformed Y given transformed X, transformed W is unbiased for the same transformation of X, and in both cases, the transformed versions of X, U_Y and U_W are all homoscedastic and normally distributed, that is, the ideal model of Berry et al. (2002) obtains. Of course, such transformation may not make sense: as in our example, nutritionists will be more interested in the regression of FFQ-measured sodium on true sodium intake than they will be in the regression of the logarithm of FFQ sodium on the logarithm of true sodium intake. More to the point though, transformations to the ideal model of Berry et al. (2002) are generally not possible. In that case, our methodology provides robustness against violation of distributional assumptions and assumptions of homoscedasticity.

6. Supplementary Materials

The Supplementary Materials referenced in Section 2 and Section 3 are available in a ZIP file under the Paper Information link at the *Biometrics* website on Wiley Online Library. Implementation issues such as automatic choice of the hyperparameters, details of the posterior calculations, the choice of initial values, the structure of the B-splines and results of additional supporting simulations are discussed in the Supplementary Materials. The sodium data analyzed in Section 4 can be accessed from National Cancer Institute after signing a Material Transfer Agreement. A simulated data set representative of the actual data, generated from the fitted model, is included in the Supplementary Materials. R programs implementing our method are available with this paper at the *Biometrics* web site on Wiley Online Library.

Acknowledgements

Carroll's research was supported in part by a grant R37-CA057030 from the National Cancer Institute. Mallick's research was supported in part by NSF grant DMS-0914951. The authors acknowledge the Texas A&M University Brazos HPC cluster that contributed to the research reported here. We are grateful to the associate editor and the referees for their constructive and insightful comments that led to an improved version of the article.

References

Berry, S. A., Carroll, R. J., and Ruppert, D. (2002). Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association* **97**, 160–169.

Buonaccorsi, J. P. (2010). *Measurement Error: Models, Methods and Applications*. New York: Chapman and Hall/CRC.

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.

Carroll, R. J., Maca, J. D., and Ruppert, D. (1999b). Nonparametric regression in presence of measurement error. *Biometrika* **86**, 541–554.

Carroll, R. J., Roeder, K., and Wasserman, L. (1999a). Flexible parametric measurement error models. *Biometrics* **55**, 44–54.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models*, 2nd edition. Boca Raton, FL: Chapman and Hall/CRC.

Chan, D., Kohn, R., Nott, D., and Kirby, C. (2006). Locally adaptive semiparametric estimation of the mean and variance functions in regression models. *Journal of Computational and Graphical Statistics* **15**, 915–936.

Cheng, C. L. and Riu, J. (2006). On estimating linear relationships when both variables are subject to heteroscedastic measurement errors. *Technometrics* **48**, 511–519.

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.

Crainiceanu, C. M., Ruppert, D., Carroll, R. J., Joshi, A., and Goodner, B. (2007). Spatially adaptive Bayesian penalized splines with heteroscedastic errors. *Journal of Computational and Graphical Statistics* **16**, 265–288.

de Boor, C. (2000). *A Practical Guide to Splines*. New York: Springer.

Delaigle, A. and Meister, M. (2007). Nonparametric regression estimation in the heteroskedastic errors-in-variables problem. *Journal of the American Statistical Association* **102**, 1416–1426.

Devanarayan, V. and Stefanski, L. A. (2002). Empirical simulation extrapolation for measurement error models with replicate measurements. *Statistics and Probability Letters* **59**, 219–225.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11**, 89–121.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**, 577–588.

Fan, J. and Truong, Y. (1993). Nonparametric regression with errors in variables. *Annals of Statistics* **21**, 1900–1925.

Ferguson, T. F. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**, 209–230.

Fuller, W. A. (1987). *Measurement Error Models*. New York: Wiley.

Liu, A., Tong, T., and Wang, Y. (2006). Smoothing spline estimation of variance functions. *Journal of Computational and Graphical Statistics* **16**, 312–329.

Pelenis, J. (2014). Bayesian regression with heteroscedastic error density and parametric mean function. *Journal of Econometrics* **178**, 624–638.

Rasmussen and Williams. (2006). *Gaussian Processes for Machine Learning*. Massachusetts, Cambridge: MIT Press.

Schennach, S. M. (2004a). Estimation of nonlinear models with measurement error. *Econometrica* **72**, 33–75.

Schennach, S. M. (2004b). Nonparametric regression in the presence of measurement error. *Econometric Theory* **20**, 1046–1093.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* **4**, 639–650.

Staudenmayer, J., Ruppert, D., and Buonaccorsi, J. P. (2008). Density estimation in the presence of heteroscedastic measurement error. *Journal of the American Statistical Association* **103**, 726–736.

Subar, A. F., Thompson, F. E., Kipnis, V., Midthune, D., Hurwitz, P. McNutt, S., McIntosh, A., and Rosenfeld, S. (2001). Comparative validation of the block, Willet, and National Cancer Institute food frequency questionnaires. *American Journal of Epidemiology* **154**, 1089–1099.

Wang, X. and Wang, B. (2011). Deconvolution estimation in measurement error models: The R package decon. *Journal of Statistical Software* **39**, 1–24.

West, M., Müller, P., and Escobar, M. D. (1994). Hierarchical priors and mixture models, with application in regression and density estimation. In *Aspects of Uncertainty: A Tribute to D. V. Lindley*, A. F. M. Smith and P. Freeman (eds), 363–386. New York: Wiley.

Received October 2013. *Revised April* 2014. *Accepted April* 2014.