Published in final edited form as:
J Comput Graph Stat. 2014 October 1; 23(4): 1101–1125. doi:10.1080/10618600.2014.899237.

Bayesian Semiparametric Density Deconvolution in the Presence of Conditionally Heteroscedastic Measurement Errors

Abhra Sarkar,
Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143 USA

Bani K. Mallick,
Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143 USA

John Staudenmayer,
Department of Mathematics and Statistics, University of Massachusetts, Amherst, MA 01003-9305 USA

Debdeep Pati, and
Department of Statistics, Florida State University, Tallahassee, FL 32306-4330 USA

Raymond J. Carroll
Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143 USA

Abhra Sarkar: abhra@stat.tamu.edu; Bani K. Mallick: bmallick@stat.tamu.edu; John Staudenmayer: jstauden@math.umass.edu; Debdeep Pati: debdeep@stat.fsu.edu; Raymond J. Carroll: carroll@stat.tamu.edu

Abstract

We consider the problem of estimating the density of a random variable when precise measurements on the variable are not available, but replicated proxies contaminated with measurement error are available for sufficiently many subjects. Under the assumption of additive measurement errors this reduces to a problem of deconvolution of densities. Deconvolution methods often make restrictive and unrealistic assumptions about the density of interest and the distribution of measurement errors, e.g., normality and homoscedasticity and thus independence from the variable of interest. This article relaxes these assumptions and introduces novel Bayesian semiparametric methodology based on Dirichlet process mixture models for robust deconvolution of densities in the presence of conditionally heteroscedastic measurement errors. In particular, the models can adapt to asymmetry, heavy tails and multimodality. In simulation experiments, we show that our methods vastly outperform a recent Bayesian approach based on estimating the densities via mixtures of splines. We apply our methods to data from nutritional epidemiology.

Supplementary Materials
Results on Model Flexibility, R programs, Data: The supplementary materials, available in a single zip file (supplements.zip), contain some figures and tables referenced in the main paper, results of additional simulation experiments, and R programs for implementing our methods with default hyper-prior choices. The EATS data set analyzed in this paper reside at the National Cancer Institute (NCI, http://www.cancer.gov/) and may be obtained from NCI arranging a Material Transfer Agreement. A simulated data set representative of the actual data (i.e., generated from the fitted model) is included in the supplementary materials. A readme file (ReadMe.txt) that provides further details of the contents is also included.

Even in the special case when the measurement errors are homoscedastic, our methodology is novel and dominates other methods that have been proposed previously. Additional simulation results, instructions on getting access to the data set and R programs implementing our methods are included as part of online supplemental materials.

Some Key Words

B-spline; Conditional heteroscedasticity; Density deconvolution; Dirichlet process mixture models; Measurement errors; Skew-normal distribution; Variance function

1 Introduction

Many problems of practical importance require estimation of the unknown density of a random variable. The variable, however, may not be observed precisely, observations being subject to measurement errors. Under the assumption of additive measurement errors, the observations are generated from a convolution of the density of interest and the density of the measurement errors. The problem of estimating the density of interest from available contaminated measurements then becomes a problem of deconvolution of densities.

This article proposes novel Bayesian semiparametric approaches for robust estimation of the density of interest when the variability of the measurement errors depends on the associated unobserved value of the variable of interest through an unknown relationship. The proposed methodology is fundamentally different from existing deconvolution methods, relaxes many restrictive assumptions of existing approaches by allowing both the density of interest and the distribution of measurement errors to deviate from standard parametric laws, and significantly outperforms previous methodology.

The literature on the problem of density deconvolution is vast. Most of the early literature on density deconvolution considers scenarios when a single contaminated measurement is available for each subject and assumes that the measurement errors are independently and identically distributed according to some known probability law (often normal) with constant variance. See, for example, Carroll and Hall (1988), Liu and Taylor (1989), Devroye (1989), Fan (1991a, 1991b, 1992) and Hesse (1998) among others. Of course, in reality the distribution of measurement errors is rarely known, and the assumption of constant variance measurement errors is also often unrealistic. The difficulty of a deconvolution problem depends directly on the shape (more specifically the smoothness) of the measurement error distribution (Fan 1991a, 1991b, 1992). Misspecification of the distribution of measurement errors may therefore lead to biased and inefficient estimates of the density of interest. The focus of recent deconvolution literature has thus been on robust deconvolution methods that relax the restrictive assumptions on the error distribution, assuming the availability of replicated proxies for each unknown value of the variable of interest. See, for example, Li and Vuong (1998) and Carroll and Hall (2004) among others.

All the above mentioned papers still assume that the measurement errors are independent of the variable of interest. Staudenmayer, et al. (2008) further relaxed this often unrealistic assumption and considered the problem of density deconvolution in the presence of conditionally heteroscedastic measurement errors. They took a Bayesian route and modeled the density of interest by a penalized positive mixture of normalized quadratic B-splines. Measurement errors were assumed to be normally distributed but the measurement error variance was modeled as a function of the associated unknown value of the variable of interest using a penalized positive mixture of quadratic B-splines.

The focus of this article is also on deconvolution in the presence of conditionally heteroscedastic measurement errors, but the proposed Bayesian semiparametric methods are vastly different from the approach of Staudenmayer, et al. (2008), as well as from other existing methods. The density of interest is modeled by a flexible location-scale mixture of normals induced by a Dirichlet process (Ferguson, 1973; Lo, 1984). For modeling conditionally heteroscedastic measurement errors, it is assumed that the measurement errors can be factored into ‘scaled errors’ that are independent of the variable of interest and have zero mean and unit variance, and a ‘variance function’ component that explains the conditional heteroscedasticity. This multiplicative structural assumption on the measurement errors was implicit in Staudenmayer, et al. (2008), where the scaled errors were assumed to come from a standard normal distribution.

Our approach is based on a more flexible representation of the scaled errors. The density of the scaled measurement errors is modeled using an infinite mixture model induced by a Dirichlet process, each component of the mixture being itself a two-component normal mixture with mean zero. This gives us the flexibility to model other aspects of the distribution of scaled errors. This deconvolution approach, therefore, uses flexible Dirichlet process mixture models twice, first to model the density of interest and second to model the density of the scaled errors, freeing them both from restrictive parametric assumptions, while at the same time accommodating conditional heteroscedasticity through the variance function.

It is important to see that even when the measurement errors are homoscedastic, our methodology is novel and dominates other methods that have been proposed previously. Our methods apply to this problem, allowing flexibility in the density of the variable of interest, flexible representations of the density of the measurement errors, and, if desired, at the same time build modeling robustness lest there be any remaining heteroscedasticity.

The article is organized as follows. Section 2 details the models. Sections 3 discusses some model diagnostic tools. Section 4 presents extensive simulation studies comparing the proposed semiparametric methods with the method of Staudenmayer, et al. (2008) and a possible nonparametric alternative. Section 5 presents an application of the proposed methodology in estimation of the distributions of daily dietary intakes from contaminated 24 hour recalls in a nutritional epidemiologic study. Section 6 contains concluding remarks. Appendices discuss model identifiability (Appendix A), the choice of hyper-parameters (Appendix B) and details of posterior computations (Appendix C). The supplementary materials provide results of additional simulation experiments and R programs implementing our methods.

2 Density Deconvolution Models

2.1 Background

The goal is to estimate the unknown density of a random variable X. There are $i = 1, 2, \ldots, n$ subjects. Precise measurements of X are not available. Instead, for $j = 1, 2, \ldots, m_i$, replicated proxies W_{ij} contaminated with heteroscedastic measurement errors U_{ij} are available for each subject. The replicates are assumed to be generated by the model

$$W_{ij}=X_i+U_{ij}, \quad (1)$$

$$U_{ij}=v^{1/2}(X_i)\,\varepsilon_{ij}, \quad (2)$$

where X_i is the unobserved true value of X; ε_{ij} are independently and identically distributed with zero mean and unit variance and are independent of the X_i, and v is an unknown smooth variance function. Identifiability of model (1)–(2) is discussed in Appendix A, where we show that 3 replicates more than suffices. Some simple diagnostic tools that may be employed in practical applications to assess the validity of the structural assumption (2) on the measurement errors are discussed in Section 3.

Of course, a special case of our work is when the measurement errors are homoscedastic, so that $v(x)$ is constant. Even in this case, the use of Dirichlet process mixtures for both the target density and error distribution has not been considered previously.

The density of X is denoted by f_X. The density of ε_{ij} is denoted by f_ε. The implied conditional distributions of W_{ij} and U_{ij}, given X_i, is denoted by the generic notation $f_{W|X}$ and $f_{U|X}$, respectively. The marginal density of W_{ij} is denoted by f_W.

Model (2), along with the moment restrictions imposed on the scaled errors ε_{ij}, implies that the conditional heteroscedasticity of the measurement errors is explained completely through the variance function v, while other features of $f_{U|X}$ are derived from f_ε. In a Bayesian hierarchical framework, model (1)–(2) reduces the problem of deconvolution to three separate problems: (a) modeling the density of interest f_X; (b) modeling the variance function v, and (c) modeling the density of the scaled errors f_ε.

2.2 Modeling the Distribution of *X*

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 $\{f_c(\cdot \mid \varphi)$: $\varphi \sim P_0\}$, can be specified as

$$f(\cdot)=\sum_{k=1}^{\infty}\pi_k\, f_c(\cdot|\phi_k), \quad \phi_k\sim P_0, \quad \pi_k=s_k\prod_{j=1}^{k-1}(1-s_j), \quad s_k\sim\mathrm{Beta}(1,\alpha).$$

In the literature, this construction of random mixture weights $\{\pi_k\}_{k=1}^{\infty}$ (Sethuraman, 1994), is often represented as $\pi \sim \text{Stick}(\alpha)$. 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 $\{f_c(\cdot \mid \varphi)\colon \varphi \sim P_0\}$ is important. Mixtures of normal kernels are, in particular, very popular for their flexibility and computational tractability (Escobar and West, 1995; West, et al. 1994). 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

$$f_X(X)=\sum_{k=1}^{\infty}\pi_k \,\mathrm{Normal}(X|\mu_k,\sigma_k^2), \quad (3)$$

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

Here $\mathrm{Normal}(\cdot \mid \mu, \sigma^2)$ 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.

2.3 Modeling the Variance Function

Examples of modeling log-transformed variance functions using flexible mixtures of splines are abundant in the literature when there is no measurement error. Yau and Kohn (2003), for example, modeled $\log\{v(X)\}$ using flexible mixtures of polynomial and thin-plate splines. Liu, et al. (2006) proposed a penalized mixture of smoothing splines, whereas Chan, et al. (2006) considered mixtures of locally adaptive radial basis functions.

In this article we model the variance function as a positive mixture of B-spline basis functions with smoothness inducing priors on the coefficients. For a given positive integer K, partition an interval $[A, B]$ of interest into K subintervals using knot points $t_1 = \cdots = t_{q+1} = A < t_{q+2} < t_{q+3} < \cdots < t_{q+K} < t_{q+K+1} = \cdots = t_{2q+K+1} = B$. For $j = (q+1), \ldots, (q+K)$, define $\Delta_j = (t_{j+1} - t_j)$ and $\Delta_{\max} = \max_j \Delta_j$. It is assumed that $\Delta_{\max} \to 0$ as $K \to \infty$. Using these knot points, $(q + K) = J$ B-spline bases of degree q, denoted by $\mathbf{B}_{q,J} = \{b_{q,1}, b_{q,2}, \ldots, b_{q,J}\}$, can be defined through the recursion relation given on page 90 of de Boor (2000), see Figure S.1 in the supplementary materials. A flexible model for the variance function is

$$v(X)=\sum_{j=1}^{J}b_{q,j}(X)\exp(\xi_j)=\mathbf{B}_{q,J}(X)\exp(\boldsymbol{\xi}), \quad (5)$$

$$p_0(\boldsymbol{\xi}|J,\sigma_\xi^2)=\mathrm{MVN}_J(\mathbf{0},\sigma_\xi^{-2}P), \;\; p_0(\sigma_\xi^2)=\mathrm{IG}(a_\xi,b_\xi). \quad (6)$$

Here $\boldsymbol{\xi} = \{\xi_1, \xi_2, \ldots, \xi_J\}^{\mathrm{T}}$; $\exp(\xi) = \{\exp(xi;_1), \exp(xi;_2), \ldots, \exp(xi;_J)\}^{\mathrm{T}}$, $\mathrm{MVN}_J(\boldsymbol{\mu}, \Sigma)$ denotes a J-variate normal distribution with mean $\boldsymbol{\mu}$ and positive semidefinite covariance matrix Σ, and IG(a, b) denotes an inverse-Gamma distribution with shape parameter a and scale parameter b. We choose $P = D^{T}D$, where D is a $J \times (J + 2)$ matrix such that $D\xi$ computes the second differences in ξ. The prior $p_0(\boldsymbol{\xi}|\sigma_\xi^2)$ induces smoothness in the coefficients because it penalizes $\sum_{j=1}^{J}(\Delta^2\xi_j)^2=\boldsymbol{\xi}^{\mathrm{T}}P\boldsymbol{\xi}$, the sum of squares of the second order differences in ξ (Eilers and Marx, 1996). The variance parameter σ_ξ^2 plays the role of smoothing parameter - the smaller the value of σ_ξ^2, the stronger the penalty and the smoother the variance function. The inverse-Gamma hyper-prior on σ_ξ^2 allows the data to have strong influence on the posterior smoothness and makes the approach data adaptive.

2.4 Modeling the Distribution of the Scaled Errors

Three different approaches of modeling the density of the scaled errors f_ε are considered here, successively relaxing the model assumptions as we progress.

2.4.1 Model-I: Normal Distribution—We first consider the case where the scaled errors are assumed to follow a standard normal distribution

$$f_\varepsilon(\varepsilon)=\mathrm{Normal}(\varepsilon|0,1). \quad (7)$$

This implies that the conditional density of measurement errors is given by $f_{U|X}(U \mid X) = \mathrm{Normal}\{U \mid 0, v(X)\}$. Such an assumption was made by Staudenmayer, et al. (2008).

2.4.2 Model-II: Skew-Normal Distribution—The strong parametric assumption of normality of measurement errors may be restrictive and inappropriate for many practical applications. As a first step towards modeling departures from normality, we propose a novel use of skew-normal distributions (Azzalini, 1985) to model the distribution of scaled errors. A random variable Z following a skew-normal distribution with location ξ, scale ω and shape parameter λ has the density $f(Z) = (2/\omega)\varphi\{(Z - xi;)/\omega\}\Phi\{\lambda(Z - \xi)/\omega\}$. Here φ and Φ denote the probability density function and cumulative density function of a standard normal distribution, respectively. Positive and negative values of λ result in right and left skewed distributions, respectively. The Normal$(\cdot \mid \mu, \sigma^2)$ distribution is obtained as special cases with $\lambda = 0$, whereas the folded normal or half-normal distributions are obtained as limiting cases with $\lambda \to \pm\infty$, see Figure S.2 in the supplementary materials. With $\delta = \lambda/(1 + \lambda^2)^{1/2}$, the mean and the variance of this density are given by $\mu = \xi + \omega\delta(2/\pi)^{1/2}$ and $\sigma^2 = \omega^2(1 - 2\delta^2/\pi)$, respectively. Although the above parametrization is more constructive and intuitive in revealing the relationship with the normal family, we consider a different parametrization in terms of μ, σ^2 and λ, denoted by SN$(\cdot \mid \mu, \sigma^2, \lambda)$, that is more useful for specifying distributions with moment constraints, namely $f(Z) = (2\zeta_2/\sigma)\varphi\{\zeta_1 + \zeta_2(Z - \mu)/\sigma\}\Phi[\lambda\{\zeta_1 + \zeta_2(Z - \mu)/\sigma\}]$, where $\zeta_1 = \delta(2/\pi)^{1/2}$ and $\zeta_2 = (1 - 2\delta^2/\pi)^{1/2}$. For specifying the distribution of the scaled errors we now let

$$f_\varepsilon(\varepsilon)=\mathrm{SN}(\varepsilon|0,1,\lambda), \quad (8)$$

$$p_0(\lambda)=\mathrm{Normal}(\lambda|\mu_{0\lambda},\sigma^2_{0\lambda}). \quad (9)$$

The implied conditionally heteroscedastic, unimodal and possibly asymmetric distribution for the measurement errors is given by $f_{U|X}(U \mid X) = \mathrm{SN}\{U \mid 0, v(X), \lambda\}$.

2.4.3 Model-III: Infinite Mixture Models—

While skew-normal distributions can capture moderate skewness, they are still quite limited in their capacity to model more severe departures from normality. They can not, for example, model multimodality or heavy tails. In the context of regression analysis when there is no measurement error, moment constrained infinite mixture models have recently been used by Pelenis (2014) (see also the references therein) for flexible modeling of error distributions that can capture multimodality and heavy tails. They considered the mixture $f_{U|X}(U|X)=\sum_{k=1}^{\infty}\pi_k(X)\{p_k\,\mathrm{Normal}(U|\mu_{k1},\sigma^2_{k1})+(1-p_k)\,\mathrm{Normal}(U|\mu_{k2},\sigma^2_{k2})\}$, with the moment constraint $p_k\mu_{k1} + (1-p_k)\mu_{k2} = 0$ for all k. Use of a two-component mixture of normals as components with each component constrained to have mean zero restricts the mean of the mixture to be zero while allowing the mixture to model other unconstrained aspects of the error distribution. Incorporating covariate information X in modeling the mixture probabilities, this model allows all aspects of the error distribution, other than the mean, to vary nonparametrically with the covariates, not just the conditional variance. Designed for regression problems, these nonparametric models, however, assume that this covariate information is precise. If X is measured with error, as is the case with deconvolution problems, the subject specific residuals may not be informative enough, particularly when the number of replicates per subject is small and the measurement errors have high conditional variability, making simultaneous learning of X and other parameters of the model difficult.

In this article, we take a different semiparametric middle path. The multiplicative structural assumption (2) on the measurement errors that reduces the problem of modeling $f_{U|X}$ to the two separate problems of modeling (a) a variance function and (b) modeling an error distribution independent of the variable of interest is retained. The difficult problem of flexible modeling of an error distribution with zero mean and unit variance moment restrictions is avoided through a simple reformulation of model (2) that replaces the unit variance identifiability restriction on the scaled errors by a similar constraint on the variance function. Model (2) is rewritten as

$$U_{ij}=v^{1/2}(X_i)\,\varepsilon_{ij}=\frac{v^{1/2}(X_i)}{v^{1/2}(X_0)}v^{1/2}(X_0)\varepsilon_{ij}=\tilde{v}^{1/2}(X_i)\,\tilde{\varepsilon}_{ij}, \quad (10)$$

where X_0 is arbitrary but fixed point, $\tilde{v}(X_i) = v(X_i)/v(X_0)$, and $\tilde{\varepsilon}_{ij} = v^{1/2}(X_0)\varepsilon_{ij}$. With this specification, $\tilde{v}(X_0) = 1$, $\mathrm{var}(\tilde{\varepsilon}_{ij}) = v(X_0)$ and $\mathrm{var}(U \mid X) = v(X_0)\,\tilde{v}(X)$. The problem of modeling the unrestricted variance function v has now been replaced by the problem of modeling $\tilde{v}$ restricted to have value 1 at X_0. The problem of modeling the density of ε with zero mean and unit variance moment constraints has also been replaced by the easier problem of modeling the density of $\tilde{\varepsilon}_{ij}$ with only a single moment constraint of zero mean.

The conditional variance of the measurement errors is now a scalar multiple of $\tilde{v}$. So $\tilde{v}$ can still be referred to as the 'variance function'. The variance of $\widetilde{\varepsilon}_{ij}$, however, does not equal unity, but is, in fact, unrestricted. With some abuse of nomenclature, $\widetilde{\varepsilon}_{ij}$ is still referred to as the 'scaled errors'. For notational convenience $\widetilde{\varepsilon}_{ij}$ is denoted simply by ε_{ij}.

The problem of flexibly modeling $\tilde{v}$ is now addressed. For any X, (i) $b_{q,j}(X) \geq 0 \ \forall j$, ($ii$) $\sum_{j=1}^{J} b_{q,j}(X)=1$, ($iii$) $b_{q,j}$ is positive only inside the interval $[t_j, t_{j+q+1}]$, (iv) for $j \in \{(q + 1), (q+2), \ldots, (q+K)\}$, for any $X \in (t_j, t_{j+1})$, only $(q+1)$ B-splines $b_{q,j-q}(X), b_{q,j-q+1}(X), \ldots, b_{q,j}(X)$ are positive, and (v) when $X = t_j$, $b_{q,j}(X) = 0$. We let $\tilde{v}(X) = \mathbf{B}_{q,J}(X) \exp(\boldsymbol{\xi})$, as before, and we use the above mentioned local support properties of the B-spline bases to propose a flexible model for $\tilde{v}$ subject to $\tilde{v}(X_0) = 1$. When $X_0 \in (t_j, t_{j+1})$, properties (ii) and (iv) cause the constraint to be simply $\tilde{v}(X_0)=\sum_{\ell=(q-j)}^{j} b_{q,\ell}(X_0)\exp(\xi_j)=1$. This is a restriction on only $(q + 1)$ of the ξ_j's, and the coefficients of the remaining B-splines remain unrestricted which makes the model for $\tilde{v}$ very flexible. In a Bayesian framework, the restriction $\tilde{v}(X_0) = 1$ can be imposed by restricting the support of the prior on $\boldsymbol{\xi}$ to the set $\{\boldsymbol{\xi}: \sum_{\ell=(q-j)}^{j} b_{q,\ell}(X_0)\exp(\xi_j)=1\}$. Choosing $X_0 = t_{j0}$ for some $j_0 \in \{(q + 1), \ldots, (q + K)\}$, we further have $b_{j0}(t_{j0}) = 0$, and the complete model for $\tilde{v}$ is given by

$$\tilde{v}(X)=\mathbf{B}_{q,J}(X)\exp(\boldsymbol{\xi}), \quad (11)$$

$$p_0(\boldsymbol{\xi}|J,\sigma_{\xi}^2)=\mathrm{MVN}_J(\mathbf{0},\sigma_{\xi}^{-2}P) \times I\left\{\sum_{j=(j_0-q)}^{(j_0-1)} b_{q,j}(t_{j0})\exp(\xi_j)=1\right\}, \quad (12)$$

$$p_0(\sigma_{\xi}^2)=\mathrm{IG}(a_{\xi},b_{\xi}), \qquad K \sim p_0(K), \quad (13)$$

where $I(\cdot)$ denotes the indicator function.

Now that the variance of ε_{ij} has become unrestricted and only a single moment constraint of zero mean is required, a DPMM with mixture components as specified in Pelenis (2014) can be used to model f_{ε}. That is, we let $f_{\varepsilon}(\varepsilon)=\sum_{k=1}^{\infty}\pi_{\varepsilon k}f_{c\varepsilon}(\varepsilon|p_k,\mu_{k1},\mu_{k2},\sigma_{k1}^2,\sigma_{k2}^2)$, $\pi_{\varepsilon} \sim$ Stick(α_{ε}), where $f_{c\varepsilon}(\varepsilon|p,\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)=\{p\,\mathrm{Normal}(\varepsilon|\mu_1,\sigma_1^2)+(1-p)\,\mathrm{Normal}(\varepsilon|\mu_2,\sigma_2^2)\}$, 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 $(p, \widetilde{\mu}, \sigma_1^2, \sigma_2^2)$, where (μ_1, μ_2) can be retrieved from $\widetilde{\mu}$ as $\mu_1 = c_1\widetilde{\mu}$, $\mu_2 = c_2\widetilde{\mu}$, 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 $p\mu_1 + (1 - p)\mu_2 = \{pc_1 + (1 - p)c_2\}\widetilde{\mu} = 0$. The family includes normal densities as special cases with $(p, \widetilde{\mu}) = (0.5, 0)$ or $(0, 0)$ or $(1, 0)$. Symmetric component densities are obtained as special cases when $p = 0.5$ or $\widetilde{\mu} = 0$. The mixture is symmetric when the all components are as well. Specification of the prior for f_{ε} is completed assuming non-informative priors for (p,

$\tilde{\mu}, \sigma_1^2, \sigma_2^2$). Letting Unif($\ell$, u) denote a uniform distribution on the interval (ℓ, u), the complete DPMM prior on f_ε can then be specified as

$$f_\varepsilon(\varepsilon) = \sum\nolimits_{k=1}^{\infty} \pi_{\varepsilon k}\, f_{c\varepsilon}(\varepsilon | p_k, \tilde{\mu}_k, \sigma_{k1}^2, \sigma_{k2}^2), \quad (14)$$

$$\pi_\varepsilon \sim \mathrm{Stick}(\alpha_\varepsilon), \qquad (p_k, \tilde{\mu}_k, \sigma_{k1}^2, \sigma_{k2}^2) \sim \mathrm{Unif}(0,1)\,\mathrm{Normal}(0, \sigma_{\tilde{\mu}}^2)\,\mathrm{IG}(a_\varepsilon, b_\varepsilon)\,\mathrm{IG}(a_\varepsilon, b_\varepsilon). \quad (15)$$

2.5 Choice of Hyper-parameters and Posterior Calculations

Appendix B describes the choice of hyper-parameters, while Appendix C gives the details of posterior computations.

3 Model Diagnostics

In practical deconvolution problems, the basic structural assumptions on the measurement errors may be dictated by prominent features of the data extracted by simple diagnostic tools and expert knowledge of the data generating process. Conditional heteroscedasticity, in particular, is easy to identify from the scatterplot of S_W^2 on $\bar{W}$, where $\bar{W}$ and S_W^2 denote the subject specific sample mean and variance, respectively (Eckert, et al., 1997). The multiplicative structural assumption (2) on the measurement errors provides one particular way of accommodating conditional heteroscedasticity in the model. When at least 4 replicates are available for sufficiently many subjects, one can define the pairs (W_{ij1}, $C_{ij2j3j4}$) for all i and for all $j_1 \neq j_2 \neq j_3 \neq j_4$, where $C_{ij2j3j4} = \{(W_{ij2} - W_{ij3})/(W_{ij2} - W_{ij4})\}$. When (2) is true, $C_{j2j3j4} = \{(\varepsilon_{j2} - \varepsilon_{j3})/(\varepsilon_{j2} - \varepsilon_{j4})\}$ is independent of W_{j1}. Therefore, the absence of non-random patterns in the plots of W_{j1} against C_{j2j3j4} and nonsignificant p-values in nonparametric tests of association between W_{j1} and C_{j2j3j4} for various $j_1 \neq j_2 \neq j_3 \neq j_4$ may be taken as indications that (2) is valid or that the departures from (2) are not severe. For those cases with m ($\geq$ 4) replicates per subject, the total number of possible such tests is $m!/(m-4)! = L$, say, where, for any positive integer r, $r! = r \cdot (r-1) \ldots 2 \cdot 1$. The p-values of these tests can be combined using the truncated product method of Zaykin, et al. (2002). The test statistic of this combined left-sided test is given by $T(\varsigma) = \prod\nolimits_{\ell=1}^{L} p_\ell^{1(p_\ell < \varsigma)}$, where p_ℓ denotes the p-value of the ℓ^{th} test and ς is a prespecified truncation limit. If $\min_\ell\{p_\ell\} \geq \varsigma$, the p-value of the combined test is trivially 1. Otherwise, the bootstrap procedure described in Zaykin, et al. (2002) may be used to estimate it.

4 Simulation Experiments

4.1 Background

The mean integrated squared error (MISE) of estimation of f_X by $\hat{f}_X$ is defined as $MISE = \int E\{f_X(x) - \hat{f}_X(x)\}^2 dx$. A Markov chain Monte Carlo (MCMC) algorithm, implemented for drawing samples from the posterior to calculate estimates of f_X and other functions of secondary interest, is detailed in Appendix C. Based on B simulated data sets, a Monte Carlo estimate of MISE is given by $MISE_{est}=B^{-1}\sum_{b=1}^{B}\sum_{i=1}^{N}\{f_X(X_i^{\Delta})-\hat{f}_X^{(b)}(X_i^{\Delta})\}^2\Delta_i$, 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*.

The simulation experiments are designed to evaluate the MISE performance of the proposed models for a wide range of possibilities. The Bayesian deconvolution models proposed in this article all take semiparametric routes to model conditional heteroscedasticity assuming a multiplicative structural assumption on the measurement errors. Performance of the proposed models is first evaluated for 'semiparametric truth scenarios' when the truth conforms to the assumed multiplicative structure. Efficiency of the proposed models will also be illustrated for 'nonparametric truth' scenarios when the truth departs from the assumed multiplicative structure.

The reported estimated MISE are all based on *B* = 400 simulated data sets. For the proposed methods 5,000 MCMC iterations were run in each case with the initial 3,000 iterations discarded as burn-in. In our R code, with *n* = 500 subjects and m_i = 3 proxies for each subject, on an ordinary desktop, 5,000 MCMC iterations for models I, II and III required approximately 5 minutes, 10 minutes and 25 minutes, respectively. In comparison, the method of Staudenmayer, et al. (2008) and the nonparametric alternative described below in Section 4.3 took approximately 100 minutes.

4.2 Semiparametric Truth

This subsection presents the results of simulation experiments comparing our methods with the method of Staudenmayer, et al. (2008), referred to as the SRB method. The methods are compared over a factorial combination of three sample sizes (*n* = 250,500,1000), two densities for X $\{f_X^1(X)=0.5\,\mathrm{Normal}(X|0,0.75)+0.5\,\mathrm{Normal}(X|3,0.75)$ and $f_X^2(X)=0.8\,\mathrm{Normal}(X|0,0.75)+0.2\,\mathrm{Normal}(X|3,0.75)\}$, nine different types of distributions for the scaled errors (six light-tailed and three heavy-tailed, see Table 1 and Figure 1), and one variance function $v(X) = (1 + X/4)^2$. For each subject, m_i = 3 replicates were simulated. The MISE are presented in Table 2. Additional simulation results, where the true f_X is a normalized mixture of B-splines, are presented in the supplementary materials.

4.2.1 Results for Light-tailed Error Distributions—This section discusses MISE performances of the models for the 36 (3×2×6) cases where the scaled errors were light-tailed, distributions (a)–(f), see Table 1 and Figure 1. Results of the simulation experiments show that all three models proposed in this article significantly out-performed the SRB model in all 36 cases considered. When measurement errors are normally distributed, the reductions in MISE over the SRB method for all three models and for all six possible combination of sample sizes and true *X* distributions are more than 50%. This is particularly interesting, since the SRB method was originally proposed for normally distributed errors, even more so because our Model-II and Model-III relax the normality assumption on the measurement errors.

4.2.2 Results for Heavy-tailed Error Distributions—This section discusses MISE performances of the models for the 18 (3 × 2 × 3) cases where the distribution of scaled errors were heavy-tailed, distributions (g), (h) and (i), see Table 1 and Figure 1. Results for the error distribution (g) are summarized in Figure 2. The SRB model and Model-I assume normally distributed errors; Model-II assumes skew-normal errors whose tail behavior is similar to that of normal distributions. The results show the MISE performances of these three models to be very poor for heavy-tailed error distributions and the MISE increased with an increase in sample size due to the presence of an increasing number of outliers. Model-III, on the other hand, can accommodate heavy-tails in the error distributions and is, therefore, very robust to the presence of outliers. MISE patterns produced by Model-III for heavy-tailed errors were similar to that for light-tailed errors, and improvements in MISE over the other models were huge. For example, when the density for the scaled was (i), a mixture of Laplace densities with a very sharp peak at zero, for $n = 1000$, the improvements in MISEs over the SRB model were $54.03/0.94 \approx 57$ times for the 50–50 mixture of normals and $57.87/0.83 \approx 70$ times for the 80–20 mixture of normals.

In simpler settings, when the measurement errors are independent of the variable of interest and have a known density, Fan (1991a, 1991b, 1992) showed that the dificulty of a deconvolution problem depends directly on the shape (more specifically the smoothness) of the measurement error distribution. The results of our simulation experiments provide empirical evidence in favor of a similar conclusion in more complicated and realistic deconvolution scenarios, where the measurement errors show strong patterns of conditional heteroscedasticity, and illustrate the importance of modeling the shape of the error distribution when it is unknown.

4.3 Nonparametric Truth

This subsection is aimed at providing some empirical support to the claim made in Section 2.4.3, where it was argued that for deconvolution problems the proposed semiparametric route to model the distribution of conditionally heteroscedastic measurement errors will often be more efficient than possible nonparametric alternatives, even when the truth departs from the assumed multiplicative structural assumption (2) on the measurement errors. This is done by comparing our Model III with a method that also models the density of interest by a DPMM like ours but employs the formulation of Pelenis (2014) to model the density of the measurement errors. This possible nonparametric alternative was reviewed in Section 2.4.3 and will be referred to as the NPM method. Recall that by modeling the mixture probabilities as functions of X the NPM model allows all aspects of the distribution of errors to vary with X, not just the conditional variance. In theory, the NPM model is, therefore, more flexible than Model-III as it can also accommodate departures from (2). However, in practice, for reasons described in Section 2.4.3, Model-III will often be more efficient than the NPM model, as is shown here.

In the simulation experiments the true conditional distributions that generate the measurement errors are designed to be of the form $f_{U|X}(U|X)=\sum_{k=1}^{K}\pi_k(X)f_{cU}(U|\sigma_{Uk}^2,\boldsymbol{\theta}_{Uk})$, where each component density has mean zero, the k^{th} component has variance σ_{Uk}^2, and $\boldsymbol{\theta}_{Uk}$ denotes additional parameters. For the true and the fitted mixture probabilities we used the formulation of Chung and Dunson (2009) that allows easy posterior computation through data augmentation techniques. That is, we took $\pi_k(X)=V_k(X)\prod_{\ell=1}^{k-1}\{1-V_\ell(X)\}$ with $V_k(X)=\Phi(\alpha_k-\beta_k|X-X_k^*|)$ for k = 1; 2;…; (*K* − 1) and $\pi_K(X)=\{1-\sum_{k=1}^{K-1}\pi_k(X)\}$. The truth closely resembles the NPM model and clearly departs from the assumptions of Model III. The conditional variance is now given by $\mathrm{var}(U|X)=\sum_{k=1}^{K}\pi_k(X)\sigma_{UK}^2$. The two competing models are then compared over a factorial combination of three sample sizes (*n* = 250, 500, 1000), two densities for $X-f_X^1$ and f_X^2, as defined in Section 4.2, and three different choices for the component densities $f_{cU}-$(j) $\mathrm{Normal}(0,\sigma_{Uk}^2)$, $(\mathrm{k})\mathrm{SN}(\cdot|0,\sigma_{Uk}^2,\lambda_U)$ and (l) $SN(\cdot|0,\sigma_{Uk}^2,\lambda_{Uk})$. In each case, *K* = 8 and the parameters specifying the true mixture probabilities are set at α_k = 2, β_k = 1/2 for all *k* with X_k^* taking values in {−1.9, −1, 0, 1, 2.5, 4, 5.5} in that order. We chose the priors for α_k; β_k and X_k^* as in Chung and Dunson (2009). The component specific variance parameters σ_{Uk}^2 are set by minimizing the sum of squares of $g(X)=\{(1+X/4)^2-\sum_{k=1}^{K}\pi_k(X)\sigma_{Uk}^2\}$ on a grid. For the density (k) we set λ_U = 7. For the density (l) λ_{Uk} take values in {7, 3, 1, 0, −1, −3, −7}, with λ_{Uk} decreasing as *X* increases. For each subject, m_i = 3 replicates were simulated.

The estimated MISE are presented in Table 3. The results show that Model III vastly outperforms the NPM model in all 18 (3 × 2 × 3) cases even though the truth actually conforms to the NPM model closely. The reductions in MISE are particularly significant when the true density of interest is a 50–50 mixture of normals. The results further emphasize the need for flexible and efficient semiparametric deconvolution models such as the ones proposed in this article.

5 Application in Nutritional Epidemiology

5.1 Data Description and Model Validation

Dietary habits are known to be leading causes of many chronic diseases. Accurate estimation of the distributions of dietary intakes is important in nutritional epidemiologic surveillance and epidemiology. One large scale epidemiologic study conducted by the National Cancer Institute, the Eating at America's Table (EATS) study (Suber, et al., 2001), serves as the motivation for this paper. In this study *n* = 965 participants were interviewed m_i = 4 times over the course of a year and their 24 hour dietary recalls (W_{ij}'s) were recorded. The goal is to estimate the distribution of true daily intakes (X_i's).

Figure 3 shows diagnostic plots (as described in Section 3) for daily intakes of folate. Conditional heteroscedasticity of measurements errors is one salient feature of the data, clearly identifiable from the plot of subject-specific means versus subject-specific variances. We did not see any non-random pattern in the scatterplots of W_{j1} vs C_{j2j3j4} for various $j_1 \neq j_2 \neq j_3 \neq j_4$. A combined p-value of 1 given by nonparametric tests of association combined by the truncated product method of Zaykin, et al. (2002) with truncation limit as high as 0.50 is also strong evidence in favor of independence of W_{j1} and C_{j2j3j4} for all $j_1 \neq j_2 \neq j_3 \neq j_4$. By the arguments presented in Section 3, model (1)–(2) may therefore be assumed to be valid for reported daily intakes of folate. Data on many more dietary components were recorded in the EATS study. Due to space constraints, it is not possible to present diagnostic plots for other dietary components. However, it should be noted that the combined p-values for nonparametric tests of association between W_{j1} and C_{j2j3j4} for various $j_1 \neq j_2 \neq j_3 \neq j_4$ for *all* 25 dietary components, for which daily dietary intakes were recorded in the EATS study, are greater than 0.50 even for a truncation limit as high as 0.50, see Table S.1 of the supplementary materials.

5.2 Results for Daily Intakes of Folate

Estimates of the density of daily intakes of folate and other nuisance functions of secondary importance produced by different deconvolution models are summarized in Figure 4. When the density of scaled errors is allowed to be flexible, as in Model-III, the estimated density of daily folate intakes is visibly very different from the estimates when the measurement errors are assumed to be normally or skew-normally distributed, as in Model-I, Model-II or the SRB model, particularly in the interval of 3–6 mcg. Estimated 90% credible intervals for $f_X(3.7)$ for Model-I is (0.167, 0.283), for Model-II is (0.237, 0.375), and for Model-III is (0.092, 0.163). Since the credible interval for Model-III is disjoint from the credible intervals for the other models, the differences in the estimated densities at 3.7 may be considered to be significant.

Our analysis also showed that the measurement error distributions of *all* dietary components included in the EATS study deviate from normality and exhibit strong conditional heteroscedasticity. These findings emphasize the importance of flexible conditionally heteroscedastic error distribution models in nutritional epidemiologic studies.

6 Summary and Discussion

6.1 Summary

We have considered the problem of Bayesian density deconvolution in the presence of conditionally heteroscedastic measurement errors. Attending to the specific needs of deconvolution problems, three different approaches were considered for modeling the distribution of measurement errors. The first model made the conventional normality assumption about the measurement errors. The next two models allowed, with varying degrees of flexibility, the distribution of measurement errors to deviate from normality. In all these models conditional heteroscedasticity was also modeled nonparametrically. The proposed methodology, therefore, makes important contributions to the density deconvolution literature, allowing both the distribution of interest and the distribution of measurement errors to deviate from standard parametric laws, while at the same time accommodating conditional heteroscedasticity. Efficiency of the models in recovering the true density of interest was illustrated through simulation experiments, and in particular we showed that our method vastly dominates that of Staudenmayer, et al. (2008). Results of the simulation experiments suggested that all the models introduced in this article out-perform previously existing methods, even while relaxing some of the restrictive assumptions of previous approaches. Simulation experiments also showed that our Bayesian semiparametric deconvolution approaches proposed in this article will often be more efficient than possible nonparametric alternatives, even when the true data generating process deviates from the assumed semiparametric framework.

6.2 Data Transformation and Homoscedasticity

In our application area of nutrition, many researchers assume that W is unbiased for X in the original scale that the nutrient is measured, i.e., $E(WjX) = X$ as in our model, see Willett (1998), Spiegelman, et al. (1997, 2001, 2005) and Kipnis, et al. (2009). It is this original scale of X then that is of scientific interest in this instance. An alternative technique is a transform-retransform method: attempt to transform the W_{ij} data to make it additive and with homoscedastic measurement error, fit in the transformed scale, and then back-transform the density. For example, if $W_{ij}=X_i \exp(U_{ij}-\sigma_u^2/2)$ where $U_{ij}=\mathrm{Normal}(0,\sigma_u^2)$, then $\log(W_{ij})=\log(X_i)-\sigma_u^2/2+U_{ij}$, the classical homoscedastic deconvolution problem with target $X_*=\log(X)-\sigma_u^2/2$. One could then use any homoscedastic deconvolution method to estimate the density of $X*$, and then from that estimate the density of X. Our methods obviously apply to such a problem. We have used the kernel deconvolution R package “decon” (Wang and Wang, 2011), the only available set of programs, and compared it to our method both using transform-retransform with homoscedasticity and by working in the original scale, using Model III. In a variety of target distributions for X and a variety of sample sizes, our methods consistently have substantially lower MISE.

It is also the case though that transformations to a model such as $h(W) = h(X) + U$ with $U=\mathrm{Normal}(0,\sigma_u^2)$ do not satisfy the unbiasedness condition in the original scale. In the log-transformation case, there is a multiplicative bias, but in the cube-root case, $E(W)=E(X)+3\sigma_u^2 E(X^{1/3})$, a model that many in nutrition would find uncomfortable and, indeed, objectionable.

Of course, other fields would be amenable to unbiasedness on a transformed scale, and hope that the measurement error is homoscedastic on that scale. Even in this problem, our methodology is novel and dominates other methods that have been proposed previously. Our methods apply to this problem, allowing flexible Bayesian semiparametric models for the density of X in the transformed scale, flexible Bayesian semiparametric models for the density of the measurement errors, and, if desired, at the same time build modeling robustness lest there be any remaining heteroscedasticity. We have experimented with this ideal case, and even here our methods substantially dominate those currently in the literature. It also must be remembered too that it is often not possible to transform to additivity with homoscedasticity: one example in the EATS data of Section 5, where this occurs with vitamin B for the Box-Cox family. Details are available from the first author.

6.3 Extensions

Application of the Bayesian semiparametric methodology, introduced in this article for modeling conditionally heteroscedastic errors with unknown distribution where the conditioning variable is not precisely measured, is not limited to deconvolution problems. An important extension of this work and the subject of an ongoing research project is an application of the proposed methodology to errors-in-variables regression problems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Carroll's research was supported by grant R37-CA05730 from the National Cancer Institute. Mallick's research was supported in part by National Science Foundation grant DMS0914951. Staudenmayer's work was supported in part by NIH grants CA121005 and R01-HL099557. The authors thank Jeff Hart, John P. Buonaccorsi and Susanne M. Schennach for their helpful suggestions. The authors also acknowledge the Texas A&M University Brazos HPC cluster that contributed to the research reported here. This publication is based in part on work supported by Award Number KUS-CI-016-04, made by King Abdullah University of Science and Technology (KAUST).

References

Azzalini A. A class of distributions which includes the Normal ones. Scandinavian Journal of Statistics. 1985; 12:171–178.

Carroll RJ, Hall P. Optimal rates of convergence for deconvolving a density. Journal of the American Statistical Association. 1988; 83:1184–1186.

Carroll RJ, Hall P. Low order approximations in deconvolution and regression with errors in variables. Journal of the Royal Statistical Society, Series B. 2004; 66:31–46.

Carroll RJ, Roeder K, Wasserman L. Flexible parametric measurement error models. Biometrics. 1999; 55:44–54. [PubMed: 11318178]

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

Chung Y, Dunson DB. Nonparametric Bayes conditional distribution modeling with variable selection. Journal of the American Statistical Association. 2009; 104:1646–1660. [PubMed: 23580793]

de Boor, C. A Practical Guide to Splines. New York: Springer; 2000.

Devroye L. Consistent deconvolution in density estimation. Canadian Journal of Statistics. 1989; 17:235–239.

Eckert RS, Carroll RJ, Wang N. Transformations to additivity in measurement error models. Biometrics. 1997; 53:262–272. [PubMed: 9147595]

Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. Statistical Science. 1996; 11:89–121.

Escobar MD, West M. Bayesian density estimation and inference using mixtures. Journal of the American Statistical Association. 1995; 90:577–588.

Fan J. On the optimal rates of convergence for nonparametric deconvolution problems. Annals of Statistics. 1991a; 19:1257–1272.

Fan J. Global behavior of deconvolution kernel estimators. Statistica Sinica. 1991b; 1:541–551.

Fan J. Deconvolution with supersmooth distributions. Canadian Journal of Statistics. 1992; 20:155–169.

Ferguson TF. A Bayesian analysis of some nonparametric problems. Annals of Statistics. 1973; 1:209–230.

Hesse CH. Data driven deconvolution. Journal of Nonparametric Statistics. 1998; 10:343–373.

Hu Y, Schennach SM. Instrumental variable treatment of nonclassical measurement error models. Econometrica. 2008; 76:195–216.

Kipnis V, Midthune D, Buckman DW, Dodd KW, Guenther PM, Krebs-Smith SM, Subar AF, Tooze JA, Carroll RJ, Freedman LS. Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. Biometrics. 2009; 65:1003–1010. [PubMed: 19302405]

Li T, Vuong Q. Nonparametric estimation of the measurement error model using multiple indicators. Journal of Multivariate Analysis. 1998; 65:139–165.

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

Liu MC, Taylor RL. A consistent nonparametric density estimator for the decon-volution problem. Canadian Journal of Statistics. 1989; 17:427–438.

Lo AY. On a class of Bayesian nonparametric estimates. I: Density estimates. Annals of Statistics. 1984; 12:351–357.

Neal RM. Markov chain sampling methods for Dirichlet process mixture models. Journal of Computational and Graphical Statistics. 2000; 9:249–265.

Pelenis J. Semiparametric Bayesian regression. Journal of Econometrics. 2014; 178:624–638.

Sethuraman J. A constructive definition of Dirichlet priors. Statistica Sinica. 1994; 4:639–650.

Spiegelman D, McDermott A, Rosner B. The regression calibration method for correcting measurement error bias in nutritional epidemiology. American Journal of Clinical Nutrition. 1997; 65 (supplement):1179S–1186S. [PubMed: 9094918]

Spiegelman D, Carroll RJ, Kipnis V. Efficient regression calibration for logistic regression in main study/internal validation study designs with an imperfect reference instrument. Statistics in Medicine. 2001; 20:139–160. [PubMed: 11135353]

Spiegelman D, Zhao B, Kim J. Correlated errors in biased surrogates: study designs and methods for measurement error correction. Statistics in Medicine. 2005; 24:1657–1682. [PubMed: 15736283]

Staudenmayer J, Ruppert D, Buonaccorsi JP. Density estimation in the presence of heteroscedastic measurement error. Journal of the American Statistical Association. 2008; 103:726–736.

Suber AF, Thompson FE, Kipnis V, Midthune D, Hurwitz P, McNutt S, McIntosh A, Rosenfeld S. Comparative validation of the block, Willet, and National Cancer Institute food frequency questionnaires. American Journal of Epidemiology. 1990; 154:1089–1099.

Wang X, Wang B. Deconvolution estimation in measurement error models: the R package decon. Journal of Statistical Software. 2011; 39:1–24. [PubMed: 21572908]

West, M.; Müller, P.; Escobar, MD. Hierarchical priors and mixture models, with application in regression and density estimation. In: Smith, AFM.; Freeman, P., editors. Aspects of uncertainty: a tribute to D. V. Lindley. New York: Wiley; 1994. p. 363-386.

Willett, W. Nutritional Epidemiology. 2. Oxford University Press; 1998.

Yau P, Kohn R. Estimation and variable selection in nonparametric heteroscedastic regression. Statistics and Computing. 2003; 13:191–208.

Zaykin DV, Zhivotovsky LA, Westfall PH, Weir BS. Truncated product method for combining p-values. Genetic Epidemiology. 2002; 22:170–185. [PubMed: 11788962]

Appendix A Model Identifiability

Hu and Schennach (2008) showed that models such as ours are identified under very weak conditions. They show that when four variables, (Y, W, Z, X), where X is the only unobserved variate, are continuously distributed, their joint distribution is identified under the following conditions; their conditions are even weaker, but these suffice for our case.

Conditions 1

1. $f_{Y|W,Z,X} = f_{Y|X}$. *2.* $f_{W|Z,X} = f_{W|X}$. *3.* $\mathbb{E}(W \mid X) = X$. 4. The set $\{Y\colon f_{Y|X}(Y \mid X_1) \neq f_{Y|X}(Y \mid X_2)\}$ has positive probability under the marginal of Y for all $X_1 \neq X_2$. 5. The marginal, joint and conditional densities of (Y, W, Z, X) are bounded.

They also have a highly technical assumption about injectivity of operators, which is satisfied if the distributions of W given X and Z given X are complete. This means, for example, that if $\int g(W) f_{W|X}(W \mid X) dW = 0$ for all X, then $g \equiv 0$. This is a weak assumption and we comment upon it no further.

When $m_i \geq 3$, identifiability of our model (1)–(2) is assured as it falls within the general framework of Hu and Schennach (2008). To see this, replace their Y_i by our W_{i1}, their W_i by our W_{i2}, their Z_i by our W_{i3} and their X_i by our X_i. Conditions 3.1–3.4 then follow from the fact that (ε_{i1}, ε_{i2}, ε_{i3}, X_i) have a continuous distribution and are mutually independent with $E(\varepsilon_{ij}) = 0$. Condition 3.5 follows assuming the variance function v is continuous.

We conjecture that model (1)–(2) is identifiable even with $m_i \geq 2$ under very weak assumptions. We have numerical evidence to support the claim.

Appendix B Choice of Hyper-Parameters

For the DPMM prior for f_X, the prior variance of each σ_k^2 is $\sigma_0^4/\{(\gamma_0-2)^2(\gamma_0-1)\}$, whereas the prior variance of each μ_k, given σ_k^2 is σ_k^2/ν_0. Small values of γ_0 and ν_0 imply large prior variance and hence non-informativeness. We chose $\gamma_0 = 3$ and $\nu_0 = 1/5$. The prior marginal mean and variance of X, obtained by integrating out all but the hyper-parameters, are given by μ_0 and $\sigma_0^2(1+1/\nu_0)/(\gamma_0-1)$ respectively. Taking an empirical Bayes type approach, we set $\mu_0 = \bar{\mathbf{W}}$ and $\sigma_0^2 = S_{\mathbf{W}}^2(\gamma_0-1)/(1+1/\nu_0)$, where $\bar{\mathbf{W}}$ is the mean of the subject-specific sample means $\bar{\mathbf{W}}_{1:n}$, and $S_{\mathbf{W}}^2$ is an estimate of the across subject variance from a one way random effects model. To ensure noninformativeness, hyper-parameters appearing in the prior for f_ε are chosen as $\sigma_\mu \tilde{=} 3$, $a_\varepsilon = 1$ and $b_\varepsilon = 1$. For real world applications, the values of A and B may not be known. We set $[A, B] = [\min(\bar{\mathbf{W}}_{1:n}) - 0.1\ \mathrm{range}(\bar{\mathbf{W}}_{1:n}), \max(\bar{\mathbf{W}}_{1:n}) + 0.1\ \mathrm{range}(\bar{\mathbf{W}}_{1:n})]$. The DP concentration parameters α_X and α_ε could have been assigned gamma hyper-priors (Escobar and West, 1995), but in this article we kept them fixed at $\alpha_X = 0.1$ and $\alpha_\varepsilon = 1$, respectively. The prior mean and standard deviation of λ were set at $\mu_{0\lambda} = 0$ and $\sigma_{0\lambda} = 4$. For modeling the variance functions v and $\tilde{v}$, quadratic (q=2) B-splines based are used. See the supplementary materials for detailed expressions. The B-splines are based on (2 × 2 + 10 + 1) = 15 knot points that divide the interval $[A, B]$ into $K = 10$ subintervals of equal length. We take $X_0 = t_5$. The identifiability restriction on the variance function for Model III now becomes $\{\exp(\xi_3) + \exp(\xi_4)\} = 2$. The inverse-gamma hyper-prior on the smoothing parameter σ_ξ^2 is non-informative if b_ξ is small relative to $\xi^{\mathrm{T}} P \xi$. We chose $a_\xi = b_\xi = 0.1$.

Appendix C Posterior Inference

Define cluster labels $\mathbf{C}_{1:n}$, where $C_i = k$ if X_i is associated with the k^{th} component of the DPMM. Similarly for Model-III, define cluster labels $\{Z_{ij}\}_{i,j=1}^{n,m_i}$, where $Z_{ij} = k$ if ε_{ij} comes from the k^{th} component of (14). Let $N = \sum_{i=1}^{n} m_i$ denote the total number of observations. With a slight abuse of notation, define $\mathbf{W}_{1:N} = \{W_{ij}\}_{i,j=1}^{n,m_i}$ and $\boldsymbol{Z}_{1:N} = \{Z_{ij}\}_{i,j=1}^{n,m_i}$. Then for Model-I, $f_{W|X}(W_{ij} \mid X_i, \xi)$ = Normal $\{W_{ij} \mid X_i, v(X_i, \xi)\}$; for Model-II, $f_{W|X}(W_{ij} \mid X_i, \xi, \lambda)$ = SN$\{W_{ij} \mid X_i, v(X_i, \xi), \lambda\}$; and for Model-III, given $Z_{ij} = k$,

$$\begin{aligned} & f_{W|X}(W_{ij}|X_i,\boldsymbol{\xi},p_k,\mu_{k1},\mu_{k2},\sigma^2_{k1},\sigma^2_{k2}) \\ &\quad = p_k\text{Normal}\{W_{ij}|X_i \\ &\quad + \tilde{v}(X_i,\boldsymbol{\xi})^{1/2}\mu_{k1}, \tilde{v}(X_i,\boldsymbol{\xi})\sigma^2_{k1}\} + (1-p_k)\text{Normal}\{W_{ij}|X_i \\ &\quad\quad + \tilde{v}(X_i,\boldsymbol{\xi})^{1/2}\mu_{k2}, \tilde{v}(X_i,\boldsymbol{\xi})\sigma^2_{k2}\} \end{aligned}$$

. In what follows ζ denotes a generic variable that collects all other parameters of a model, including $\mathbf{X}_{1:n}$, that are not explicitly mentioned.

It is possible to integrate out the random mixture probabilities from the prior and posterior full conditionals of the cluster labels. Classical algorithms for fitting DPMMs make use of this and work with the resulting Polya urn scheme. Neal (2000) provided an excellent review of this type of algorithm for both conjugate and non-conjugate cases. In this article, the parameters specific to DPMMs are updated using algorithms specific to those models and other parameters are updated using the Metropolis-Hastings algorithm. In what follows, the generic notation $q(\textit{current} \to \textit{proposed})$ denotes the proposal distributions of the Metropolis-Hastings steps proposing a move from the *current* value to the *proposed* value.

The starting values of the MCMC chain are determined as follows. Subject-specific sample means $\bar{\mathbf{W}}_{1:n}$ are used as starting values for $\mathbf{X}_{1:n}$. Each C_i is initialized at i with each X_i coming from its own cluster with mean $\mu_i = X_i$ and variance $\sigma_i^2=\sigma_0^2$. In addition, σ^2_ξ is initialized at 0.1. The initial value of ξ is obtained by maximizing $\ell(\xi \mid 0.1, \bar{\mathbf{W}}_{1:n})$ with respect to ξ, where $\ell(\boldsymbol{\xi}|\sigma^2_\xi, \mathbf{X}_{1:n})$ denotes the conditional log-posterior of ξ. The parameters of the distribution of scaled errors are initialized at values that correspond to the special standard normal case. For example, for Model-II, λ is initialized at zero. For Model-III, Z_{ij}'s are all initialized at 1 with $(p_1,\tilde{\mu}_1,\sigma^2_{11},\sigma^2_{12})=(0.5,0,1,1)$. The MCMC iterations comprise the following steps.

1. **Updating the parameters of the distribution of** X: Conditionally given $\mathbf{X}_{1:n}$, the parameters specifying the DPMM for f_X can be updated using a Gibbs sampler (Neal, 2000, Algorithm 2). The full conditional of C_i is given by

$$\begin{aligned} p(C_i=k, k\in \mathbf{C}_{-i}|\mathbf{X}_{1:n},\mathbf{C}_{-i},\boldsymbol{\zeta}) &= b\frac{n_{-i,k}}{n-1+\alpha_X}\text{Normal}(X_i|\mu_k,\sigma^2_k), \\ p(C_i\notin \mathbf{C}_{-i}|\mathbf{X}_{1:n},\mathbf{C}_{-i},\boldsymbol{\zeta}) &= b\frac{\alpha_X}{n-1+\alpha_X}\text{t}_{2\gamma_0}(t_i), \end{aligned}$$

where b denotes the appropriate normalizing constant; for each i, $\mathbf{C}_{-i} = \mathbf{C}_{1:n} - \{C_i\}$; $n_{-i,k} = \Sigma_{\{l:l\neq i\}} 1_{\{cl=k\}}$ is the number of c_l's that equal k in $\mathbf{C}_{-i}$; and $t_i=\gamma_0^{1/2}(X_i-\mu_0)/\{\sigma_0(1+1/\nu_0)^{1/2}\}$. t_m denotes the density of a t-distribution with m degrees of freedom.

For all $k \in \mathbf{C}_{1:n}$, we update (μ_k, σ^2_k) using the closed-form joint full conditional given by $\{(\mu_k,\sigma^2_k|\mathbf{X}_{1:n},\boldsymbol{\zeta})\}=\text{NIG}(\mu_{nk},\sigma^2_{nk}/\nu_{nk},\gamma_{nk},\sigma^2_{nk})$, where $n_k=\sum_{i=1}^n 1_{\{C_i=k\}}$ is the number of X_i's associated with the k^{th} cluster; $\nu_{nk} = (\nu_0 +$

n_k); $\gamma_{nk} = (\gamma_0 + n_k/2)$; $\mu_{nk} = (\nu_0\mu_0 + n_k\Sigma_{\{i:C_i=k\}} X_i)/(\nu_0 + n_k)$ and $\sigma_{nk}^2=\sigma_0^2+(\sum_{\{i:C_i=k\}} X_i^2+\nu_0\mu_0^2-\nu_{nk}\mu_{nk}^2)/2$.

2. **Updating $\mathbf{X}_{1:n}$:** Because the X_i's are conditionally independent, the full conditional of X_i is given by $p(X_i|\mathbf{W}_{1:N},\boldsymbol{\zeta}) \propto \hat{f}_X(X_i|\boldsymbol{\zeta}) \times \prod_{j=1}^{m_i} f_{W|X}(W_{ij}|X_i,\boldsymbol{\zeta})$. We use a Metropolis-Hastings sampler to update the X_i's with proposal $q(X_i \to X_{i,new})=TN(X_{i,new}|X_i,\sigma_X^2,[A,B])$, where σ_X = (the range of $\bar{\mathbf{W}}_{1:n}$)/6 and TN($\cdot$ | m, s^2, [ℓ,u]) denotes a truncated normal distribution with location m and scale s restricted to the interval [ℓ, u].

3. **Updating the parameters of the distribution of scaled errors:** For Model-II and Model-III, the parameters involved in the distribution of scaled errors have to be updated.

   For Model-II, the distribution of scaled error is SN(0, 1, λ), involving only the parameter λ. The full conditional of λ is given by $p(\lambda|\mathbf{W}_{1:N},\boldsymbol{\zeta}) \propto p_0(\lambda) \times \prod_{i=1}^{n}\prod_{j=1}^{m_i} f_{W|X}(W_{ij}|\lambda,\boldsymbol{\zeta})$. We use Metropolis-Hastings sampler to update λ with random walk proposal $q(\lambda \to \lambda_{new})=\mathrm{Normal}(\lambda_{new}|\lambda,\sigma_\lambda^2)$.

   For Model-III, we use Metropolis-Hastings samplers to update the latent parameters $\mathbf{Z}_{1:N}$ as well as the component specific parameters (p_k, $\tilde{\mu}_k$, $\sigma_{k1}^2,\sigma_{k2}^2$)'s (Neal, 2000, Algorithm 5). We propose a new value of Z_{ij}, say $Z_{ij,new}$, according to its marginalized conditional prior

$$p(Z_{ij}=k, k \in \boldsymbol{Z}_{-ij}|\boldsymbol{Z}_{-ij})=N_{-ij,k}/(N-1+\alpha_\varepsilon),$$
$$p(Z_{ij} \notin \boldsymbol{Z}_{-ij}|\boldsymbol{Z}_{-ij})=\alpha_\varepsilon/(N-1+\alpha_\varepsilon),$$

   where, for each (i, j) pair, $\mathbf{Z}_{-ij} = \mathbf{Z}_{1:N} - \{Z_{ij}\}$; $N_{-ij,k} = \Sigma_{[rs:rs\neq ij]} 1_{\{Z_{rs}=k\}}$, the number of Z_{rs}'s in $\mathbf{Z}_{-ij}$ that equal k. If $Z_{ij,new} \not\in \mathbf{Z}_{-ij}$, we draw ($pZ_{ij,new}$, $\mu\tilde{Z}_{ij,new}$, $\sigma_{Z_{ij,new}1}^2,\sigma_{Z_{ij,new}2}^2$) from the prior p_0(p, $\tilde{\mu}$, σ_1^2,σ_2^2. We update Z_{ij} to its proposed value with probability

$$\min\left\{1, \frac{f_{W|X}(W_{ij}|p_{Z_{ij,new}},\tilde{\mu}_{Z_{ij,new}},\sigma_{Z_{ij,new}1}^2,\sigma_{Z_{ij,new}2}^2,\boldsymbol{\zeta})}{f_{W|X}(W_{ij}|p_{Z_{ij}},\tilde{\mu}_{Z_{ij}},\sigma_{Z_{ij}1}^2,\sigma_{Z_{ij}2}^2,\boldsymbol{\zeta})}\right\}.$$

For all $k \in \mathbf{Z}_{1:N}$, we propose a new value for (p_k, $\tilde{\mu}_k$, $\sigma^2_{k1}, \sigma^2_{k2}$) with the proposal $q\{\boldsymbol{\theta}_k=(p_k, \tilde{\mu}_k, \sigma^2_{k1}, \sigma^2_{k2}) \rightarrow (p_{k,new}, \tilde{\mu}_{k,new}, \sigma^2_{k1,new}, \sigma^2_{k2,new}) =\boldsymbol{\theta}_{k,new}\}=\mathrm{TN}(p_{k,new}|p_k, \sigma^2_p, [0,1]) \times \mathrm{Normal}(\tilde{\mu}_{k,new}|\tilde{\mu}_k, \sigma^2_{\tilde{\mu}}) \times \mathrm{TN}(\sigma^2_{k1,new}|\sigma^2_{k1}, \sigma^2_\sigma, [\max\{0, \sigma^2_{k1}-1\}, \sigma^2_{k1}+1]) \times \mathrm{TN}(\sigma^2_{k2,new}|\sigma^2_{k2}, \sigma^2_\sigma, [\max\{0, \sigma^2_{k2} -1\}, \sigma^2_{k2}+1])$. We update θ_k to the proposed value $\theta_{k,new}$ with probability

$$\min\left\{1, \frac{q(\boldsymbol{\theta}_{k,new} \rightarrow \boldsymbol{\theta}_k)}{q(\boldsymbol{\theta}_k \rightarrow \boldsymbol{\theta}_{k,new})} \frac{\prod_{\{ij:z_{ij}=k\}} f_{W|X}(W_{ij}|\boldsymbol{\theta}_{k,new}, \boldsymbol{\zeta})\, p_0(\boldsymbol{\theta}_{k,new})}{\prod_{\{ij:z_{ij}=k\}} f_{W|X}(W_{ij}|\boldsymbol{\theta}_k, \boldsymbol{\zeta})\, p_0(\boldsymbol{\theta}_k)}\right\}.$$

4. **Updating the parameters of the variance function:** The full conditional for ξ is given by $p(\boldsymbol{\xi}|\mathbf{W}_{1:N}, \boldsymbol{\zeta}) \propto p_0(\boldsymbol{\xi}) \times \prod_{i=1}^{n}\prod_{j=1}^{m_i} f_{W|X}(W_{ij}|\boldsymbol{\xi}, \boldsymbol{\zeta})$. We use Metropolis-Hastings sampler to update ξ with random walk proposal $q(\xi \rightarrow \xi_{new})$ = MVN(ξ_{new} | ξ, Σ_ξ). For Model III, the identifiability restriction is imposed by replacing $\xi_{new,3}$ = log{2 − exp($\xi_{new,4}$)}.

   Finally, we update the hyper-parameter σ^2_ξ using its closed-form full conditional $(\sigma^2_\xi|\boldsymbol{\xi}, \boldsymbol{\zeta})=\mathrm{IG}\{a_\xi+(J+2)/2, b_\xi+\boldsymbol{\xi}^{'} P\boldsymbol{\xi}/2\}$.

The covariance matrix Σ_ξ of the proposal distribution for ξ is taken to be the inverse of the negative Hessian matrix of $l(\xi \mid 0.1, \bar{\mathbf{W}}_{1:n})$ evaluated at the chosen initial value of ξ. See Appendix D for more details. Other variance parameters appearing in the proposal distributions are tuned to get good acceptance rates for the Metropolis-Hastings samplers, the values σ_λ = 1, σ_p = 0.01 and σ_σ = 0.1 working well in the examples considered. In simulation experiments, 5,000 MCMC iterations with the initial 3,000 discarded as burn-in produced very stable estimates of the density and the variance function.

The posterior estimate of f_X is given by the unconditional predictive density $f_X(\cdot \mid \mathbf{W}_{1:N})$. A Monte Carlo estimate of $f_X(\cdot \mid \mathbf{W}_{1:N})$, based on M samples from the posterior, is given by

$$\hat{f}_X(X|\mathbf{W}_{1:N})=M^{-1}\sum_{m=1}^{M}\Big[\sum_{k=1}^{k^{(m)}}\{n_k^{(m)}/(\alpha_X+n)\}\,\mathrm{Normal}(X|\mu_k^{(m)}, \sigma_k^{(m)2}) +\{\alpha_X/(\alpha_X+n)\}\,\mathrm{t}_{2\gamma_0}(t_X),$$

where $t_X=t(X)=\gamma_0^{1/2}(X-\mu_0)/\{\sigma_0(1+1/\nu_0)^{1/2}\}$, $(\mu_k^{(m)}, \sigma_k^{(m)2})$ is the sampled value of (μ_k, σ^2_k) in the m^{th} sample, $n_k^{(m)}$ is the number of X_i's associated with the k^{th} cluster, and $k^{(m)}$ is the total number of active clusters. With $(p_k^{(m)}, \tilde{\mu}_k^{(m)}, \sigma_{k1}^{(m)2}, \sigma_{k2}^{(m)2})$, $N_k^{(m)}$ and $k_\varepsilon^{(m)}$ defined in a similar fashion, the posterior Monte Carlo estimate of f_ε for Model-III is

$$\hat{f}_{\varepsilon}(\varepsilon|\mathbf{W}_{1:N})=M^{-1}\sum_{m=1}^{M}\Big[\sum_{k=1}^{k_{\varepsilon}^{(m)}}\{N_{k}^{(m)}/(\alpha_{\varepsilon}+N)\}\,f_{c\varepsilon}(\varepsilon|p_{k}^{(m)},\tilde{\mu}_{k}^{(m)},\sigma_{k1}^{(m)2},\sigma_{k2}^{(m)2})+\{\alpha_{\varepsilon}/(\alpha_{\varepsilon}+N)\}\int f_{c\varepsilon}(\varepsilon|p,\tilde{\mu},\sigma_{k1}^{2},\sigma_{k2}^{2})dp_{0}(p,\tilde{\mu},\sigma$$

The integral above can not be exactly evaluated. Monte Carlo approximation may be used. If $N \gg \alpha_{\varepsilon}$, the term may simply be neglected. For Model II, f_{ε} can be estimated by $\hat{f}_{\varepsilon}(\varepsilon|\mathbf{W}_{1:N})=\sum_{m=1}^{M}\mathrm{SN}(\varepsilon|0,1,\lambda^{(m)})/M$. For Models I and II, an estimate of the variance function v can similarly be obtained as $\hat{v}(X|\mathbf{W}_{1:N})=\sum_{m=1}^{M}v(X|\boldsymbol{\xi}^{(m)})/M$. An estimate of the restricted variance function $\tilde{v}$ for Model III can be obtained using a similar formula. For Model III, $\hat{v}$ and a scaled version of $\hat{f}_{\varepsilon}$, scaled to have unit variance, can be obtained using the estimate of $\tilde{v}(X_0)$.

Appendix D Initial Values and Proposals for ξ

The conditional posterior log-likelihood of ξ for Model-I is given by

$$\ell(\boldsymbol{\xi}|\sigma_{\xi}^{2},\mathbf{X}_{1:n})=-\frac{1}{2\sigma_{\xi}^{2}}\boldsymbol{\xi}^{\mathrm{T}}P\boldsymbol{\xi}-\sum_{i=1}^{n}\left\{\frac{m_i}{2}\log v(X_i,\boldsymbol{\xi})+\sum_{j=1}^{m_i}\frac{1}{2v(X_i,\boldsymbol{\xi})}(W_{ij}-X_i)^2\right\}.$$

The initial values for the M-H sampler for ξ is obtained as $\xi^{(0)}$ = arg max $\ell(\xi \mid 0.1, \bar{\mathbf{W}}_{1:n})$. Numerical optimization is performed using the optim routine in R with the analytical gradient supplied.

The covariance matrix of the random walk proposal for ξ is taken to be the inverse of the negative of the matrix of second partial derivatives of $\ell(\xi \mid 0.1, \bar{\mathbf{W}}_{1:n})$ evaluated at $\xi^{(0)}$. Expressions for the gradient and the second derivatives are given below.

$$\frac{\partial\ell(\boldsymbol{\xi}|\sigma_{\xi}^{2},\mathbf{X}_{1:n})}{\partial\xi_k}=-\frac{(P\boldsymbol{\xi})_k}{\sigma_{\xi}^{2}}-\sum_{i=1}^{n}\left\{m_i-\sum_{j=1}^{m_i}\frac{(W_{ij}-X_i)^2}{v(X_i,\boldsymbol{\xi})}\right\}\frac{b_{2,k}(X_i)\exp(\xi_k)}{2v(X_i,\boldsymbol{\xi})},$$

$$\frac{\partial^2\ell(\boldsymbol{\xi}|\sigma_{\xi}^{2},\mathbf{X}_{1:n})}{\partial\xi_k^2}=-\frac{(P)_{kk}}{\sigma_{\xi}^{2}}-\sum_{i=1}^{n}\left\{\sum_{j=1}^{m_i}\frac{(W_{ij}-X_i)^2}{v(X_i,\boldsymbol{\xi})}-\frac{m_i}{2}\right\}\frac{b_{2,k}(X_i)^2}{v(X_i,\boldsymbol{\xi})^2}\exp(2\xi_k)-\sum_{i=1}^{n}\left\{m_i-\sum_{j=1}^{m_i}\frac{(W_{ij}-X_i)^2}{v(X_i,\boldsymbol{\xi})}\right\}\frac{b_{2,k}(X_i)\exp(\xi_k)}{2v(X_i,\boldsymbol{\xi})},$$

$$\frac{\partial^2\ell(\boldsymbol{\xi}|\sigma_{\xi}^{2},\mathbf{X}_{1:n})}{\partial\xi_k\partial\xi_{k'}}=-\frac{(P)_{kk'}}{\sigma_{\xi}^{2}}-\sum_{i=1}^{n}\left\{\sum_{j=1}^{m_i}\frac{(W_{ij}-X_i)^2}{v(X_i,\boldsymbol{\xi})}-\frac{m_i}{2}\right\}\frac{b_{2,k}(X_i)b_{2,k'}(X_i)}{v(X_i,\boldsymbol{\xi})^2}\exp(\xi_k+\xi_{k'}).$$

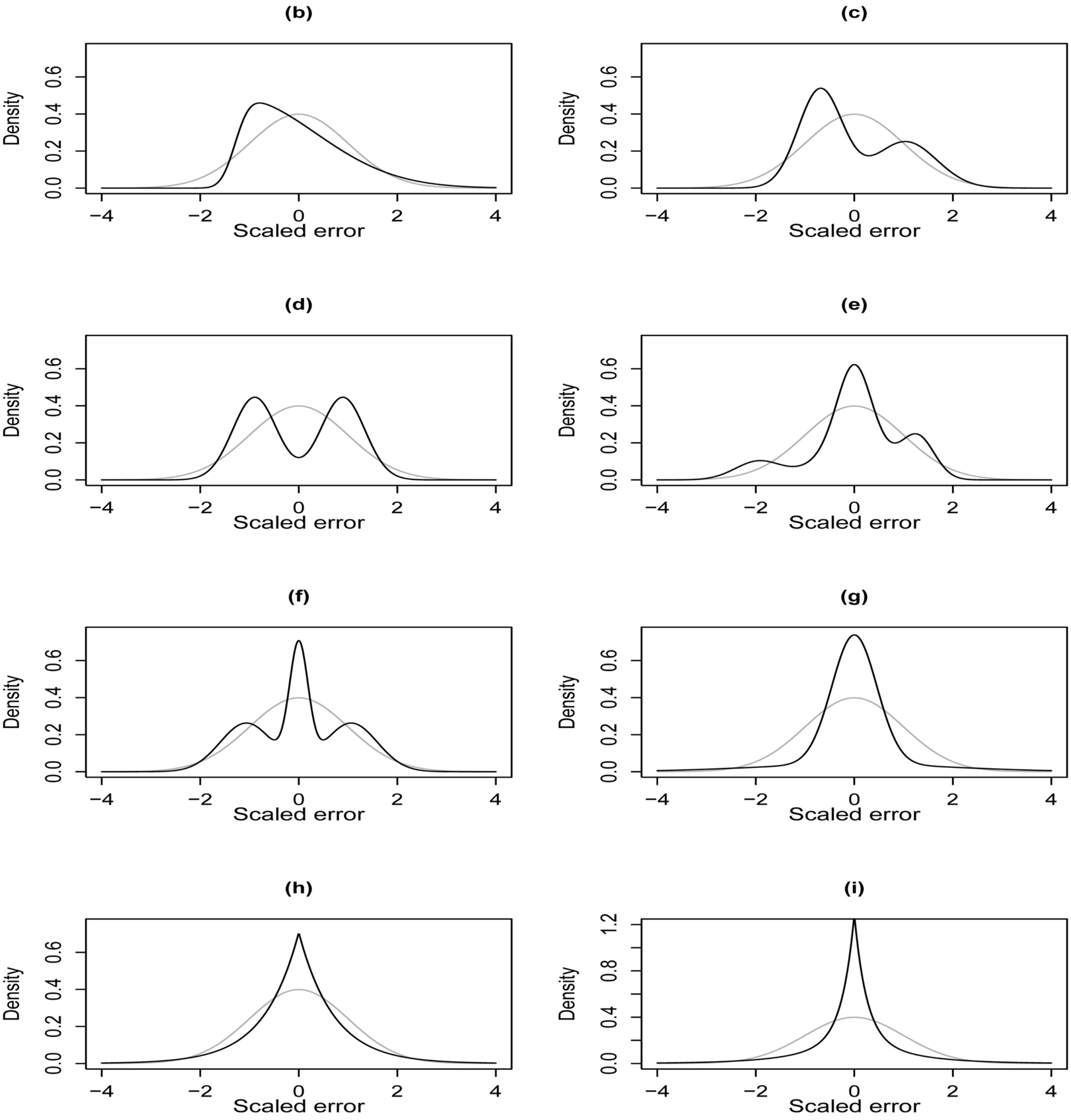

Figure 1.

The distributions used to generate the scaled errors in the simulation experiment, superimposed over a standard normal density. The difierent choices cover a wide range of possibilities - (a) standard normal (not shown separately), (b) asymmetric skew-normal, (c) asymmetric bimodal, (d) symmetric bimodal, (e) asymmetric trimodal, (f) symmetric trimodal, (g) symmetric heavy-tailed, (h) symmetric heavy-tailed with a sharp peak at zero and (i) symmetric heavy-tailed with even a sharper peak at zero. The last six cases demonstrate the flexibility of mixtures of moment restricted two-component normals in capturing widely varying shapes.

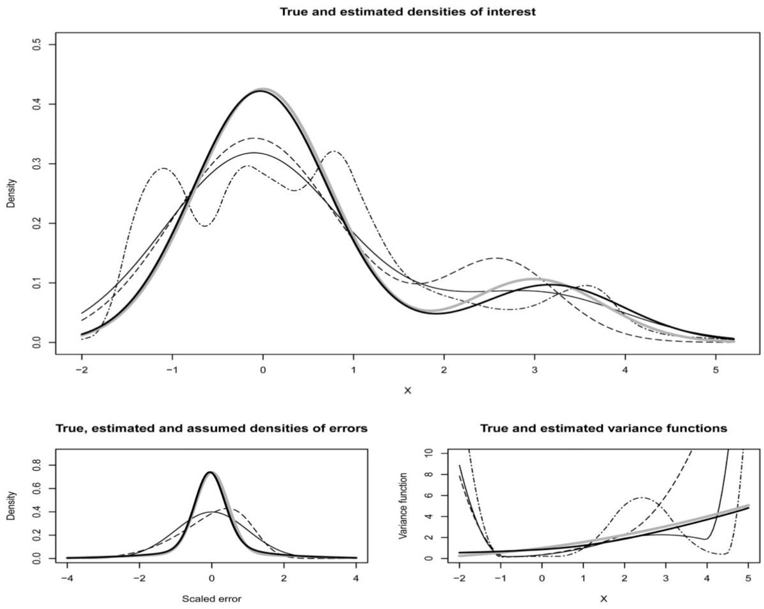

Figure 2.

Results for heavy-tailed error distribution (g) with sample size n=1000 corresponding to 25th percentile MISE. The top panel shows the estimated densities under different models. The bottom left panel shows estimated densities of scaled errors under Model-II (dashed line) and Model-III (solid bold line) superimposed over a standard Normal density (solid line). The bottom right panel shows estimated variance functions under different models. For the top panel and the bottom right panel, the solid thin line is for Model-I; the dashed line is for Model-II; the solid bold line is for Model-III; and the dot-dashed line is for the Model of Staudenmayer, et al. (2008). In all three panels the bold gray lines represent the truth.

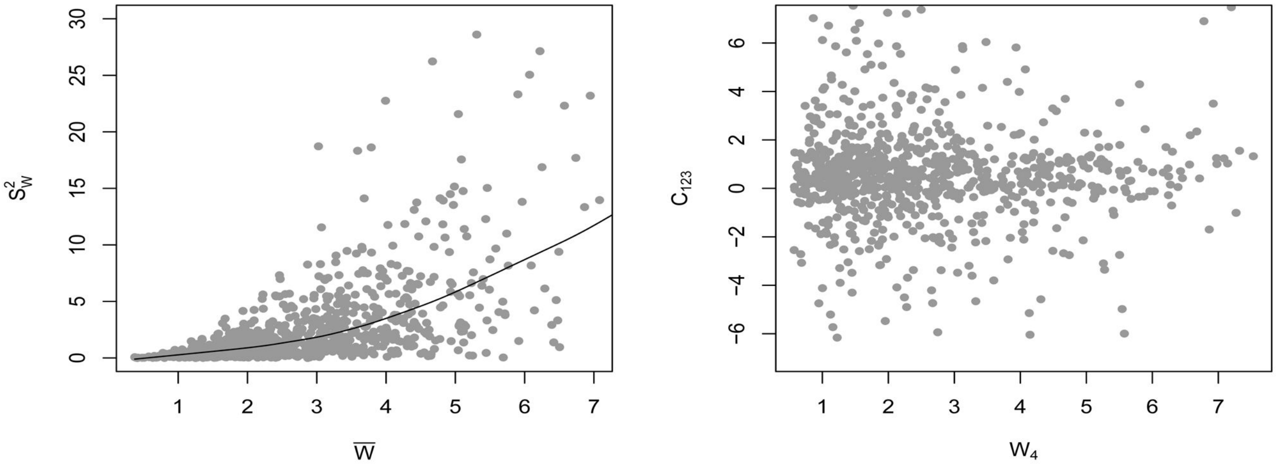

Figure 3.

Diagnostic plots for reported daily intakes of folate. The left panel shows the plot of $\bar{W}$ vs S_W^2 with a simple lowess fit superimposed. The right panel shows the plot of W_4 vs C_{123}.

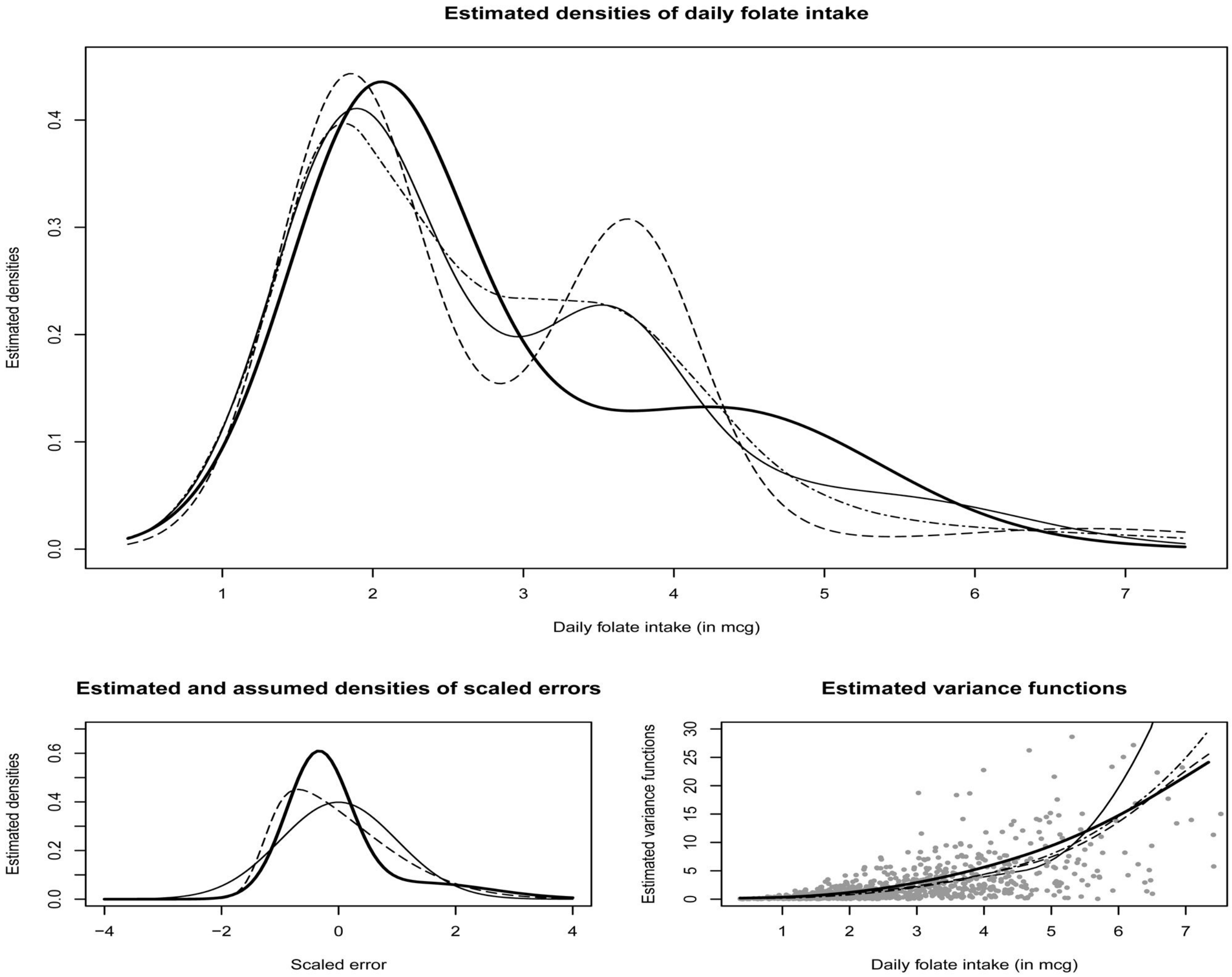

Figure 4.

Results for data on daily folate intakes from EATS example. The top panel shows the estimated densities of daily folate intake under different models. The bottom left panel shows estimated densities of scaled errors under Model-II (dashed line) and Model-III (solid bold line) superimposed over a standard Normal density (solid line). The bottom right panel shows estimated variance functions under different models. The gray dots represent subject-specific sample means (x-axis) and variances (y-axis). For the top panel and the bottom right panel, the solid thin line is for Model-I; the dashed line is for Model-II; the solid bold line is for Model-III; and the dot-dashed line is for the Model of Staudenmayer, et al. (2008).

Table 1

The distributions used to generate the scaled errors in the simulation experiment. Let MRTCN(K, π_ε, $\mathbf{p}$, $\tilde{\boldsymbol{\mu}}$ $\boldsymbol{\sigma}_1^2, \boldsymbol{\sigma}_2^2$) denote a K component mixture of moment restricted two-component normals: $\sum_{k=1}^{K} \pi_{\varepsilon k} f_{c\varepsilon}(\cdot|p_k, \tilde{\mu}_k, \sigma_{k1}^2, \sigma_{k2}^2)$. Then SMRTN denotes a scaled version of MRTCN, scaled to have variance one. Laplace(μ, b) denotes a Laplace distribution with location μ and scale b. SMLaplace(K, π_ε, **0, b**) denotes a K component mixture of Laplace densities: $\sum_{k=1}^{K} \pi_{\varepsilon k} \mathrm{Laplace}(0, b_k)$, scaled to have variance one. With μ_k denoting the k^{th} order central moments of the scaled errors, the skewness and excess kurtosis of the distribution of scaled errors are measured by the coeficients $\gamma_1 = \mu_3$ and $\gamma_2 = \mu_4 - 3$, respectively. The densities (a)–(f) are light-tailed, whereas the densities (g)–(i) are heavy-tailed. The shapes of these distributions are illustrated in Figure 1.

Distribution of scaled errors	Skewness (γ_1)	Excess Kurtosis (γ_2)
(a) Normal(0,1)	0	0
(b) Skew-normal(0,1,7)	0.917	0.779
(c) SMRTCN(1,1,0.4,2,2,1)	0.499	−0.966
(d) SMRTCN(1,1,0.5,2,1,1)	0	−1.760
(e) SMRTCN{2,(0.3,0.7),(0.6,0.5),(5,0),(1,4),(2,1)}	−0.567	−1.714
(f) SMRTCN{2,(0.3,0.7),(0.6,0.5),(0,4),(0.5,4),(0.5,4)}	0	−1.152
(g) SMRTCN{2,(0.8,0.2),(0.5,0.5),(0,0),(0.25,5),(0.25,5)}	0	7.524
(h) Laplace(0,$2^{-1/2}$)	0	3
(i) SMLaplace{2,(0.5,0.5),(0,0),(1,4)}	0	7.671

Table 2

Mean integrated squared error (MISE) performance of density deconvolution models described in Section 2 of this article (Models I, II and III) compared with the model of Staudenmayer, et al. (2008) (Model SRB) for different scaled error distributions. The true variance function was $v(X) = (1 + X/4)^2$. See Section 4.2 for additional details. The minimum value in each row is highlighted.

True Error Distribution	True *X* Distribution	Sample Size	MISE × 1000			
			SRB	Model1	Model2	Model3
(a)	50–50 mixture of normals	250	10.15	**5.31**	5.61	5.55
		500	6.64	**3.15**	3.16	3.34
		1000	4.50	**1.96**	2.08	2.21
	80–20 mixture of normals	250	9.60	**4.41**	4.47	4.52
		500	5.30	**2.34**	2.39	2.62
		1000	4.39	**1.31**	1.37	1.39
(b)	50–50 mixture of normals	250	11.79	7.80	**4.41**	4.55
		500	11.85	5.79	**3.11**	3.33
		1000	8.66	4.58	**1.91**	2.21
	80–20 mixture of normals	250	10.74	6.97	**4.52**	4.54
		500	7.94	4.17	**2.27**	2.60
		1000	6.16	3.08	**1.26**	1.39
(c)	50–50 mixture of normals	250	12.61	8.74	5.31	**4.60**
		500	9.27	4.91	3.57	**3.39**
		1000	9.15	4.13	2.53	**1.91**
	80–20 mixture of normals	250	9.27	6.46	4.65	**4.03**
		500	6.67	3.18	2.77	**2.37**
		1000	5.04	2.26	1.40	**1.26**
(d)	50–50 mixture of normals	250	10.10	7.71	9.94	**4.40**
		500	6.54	4.26	7.01	**2.70**
		1000	6.02	3.41	5.58	**1.40**
	80–20 mixture of normals	250	8.18	5.32	5.92	**3.43**
		500	4.45	2.67	4.30	**2.21**
		1000	4.40	1.74	3.31	**1.60**
(e)	50–50 mixture of normals	250	10.03	6.01	5.92	**4.03**
		500	9.38	3.87	3.57	**2.99**
		1000	8.39	2.42	2.25	**1.75**
	80–20 mixture of normals	250	7.82	3.97	4.44	**3.38**
		500	7.62	3.00	2.40	**2.01**
		1000	6.82	1.74	1.45	**1.17**
(f)	50–50 mixture of normals	250	9.35	5.82	6.52	**5.37**
		500	7.18	3.47	3.67	**3.62**
		1000	4.63	2.46	2.62	**2.10**

True Error Distribution	True *X* Distribution	Sample Size	MISE × 1000			
			SRB	Model1	Model2	Model3
	80–20 mixture of normals	250	9.17	4.75	4.80	**4.10**
		500	7.35	2.58	2.65	**2.52**
		1000	3.86	1.53	1.60	**1.45**
(g)	50–50 mixture of normals	250	15.68	11.78	10.38	**3.30**
		500	23.27	15.57	14.85	**2.07**
		1000	49.77	18.91	21.00	**1.12**
	80–20 mixture of normals	250	20.05	8.18	15.99	**3.10**
		500	36.46	10.83	17.23	**1.63**
		1000	48.70	18.53	17.77	**0.92**
(h)	50–50 mixture of normals	250	11.29	6.62	7.01	**5.18**
		500	15.07	8.07	7.24	**3.29**
		1000	18.79	12.04	8.41	**1.99**
	80–20 mixture of normals	250	11.34	7.18	7.05	**2.91**
		500	13.23	7.43	7.53	**1.67**
		1000	22.03	8.64	7.56	**1.03**
(i)	50–50 mixture of normals	250	19.34	7.69	9.90	**3.10**
		500	28.79	17.32	11.02	**2.14**
		1000	54.03	26.78	11.64	**0.94**
	80–20 mixture of normals	250	29.81	16.45	14.76	**2.74**
		500	48.41	20.94	14.99	**1.60**
		1000	57.87	23.80	16.59	**0.83**

Table 3

Mean integrated squared error (MISE) performance of Models III compared with the NPM model for different measurement error distributions. See Section 4.3 for additional details. The minimum value in each row is highlighted.

True Error Distribution	True *X* Distribution	Sample Size	MISE × 1000	
			NPM	Model3
(j)	50–50 mixture of normals	250 500 1000	29.25 23.83 20.11	**5.25** **3.61** **2.45**
	80–20 mixture of normals	250 500 1000	8.09 6.71 7.34	**4.62** **3.12** **2.05**
(k)	50–50 mixture of normals	250 500 1000	23.18 20.45 20.37	**4.81** **3.18** **2.13**
	80–20 mixture of normals	250 500 1000	11.62 8.26 8.01	**4.42** **2.77** **1.43**
(l)	50–50 mixture of normals	250 500 1000	21.69 17.72 16.43	**5.65** **3.86** **2.67**
	80–20 mixture of normals	250 500 1000	5.67 3.67 3.37	**4.71** **2.98** **2.01**