Published in final edited form as:

J Am Stat Assoc. 2018 ; 113(521): 401–416. doi:10.1080/01621459.2016.1260467.

Bayesian Semiparametric Multivariate Density Deconvolution

Abhra Sarkar,

Department of Statistical Science, Duke University, Durham, NC 27708-0251, USA,
abhra.sarkar@duke.edu

Debdeep Pati,

Department of Statistics, Florida State University, Tallahassee, FL 32306-4330, USA,
debdeep@stat.fsu.edu

Antik Chakraborty,

Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX, 77843-3143
USA, antik@stat.tamu.edu

Bani K. Mallick, and

Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143,
USA, bmallick@stat.tamu.edu

Raymond J. Carroll

Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143,
USA, and School of Mathematical and Physical Sciences, University of Technology Sydney,
Broadway NSW 2007, Australia, carroll@stat.tamu.edu

Abstract

We consider the problem of multivariate density deconvolution when interest lies in estimating the distribution of a vector valued random variable $\mathbf{X}$ but precise measurements on $\mathbf{X}$ are not available, observations being contaminated by measurement errors $\mathbf{U}$. The existing sparse literature on the problem assumes the density of the measurement errors to be completely known. We propose robust Bayesian semiparametric multivariate deconvolution approaches when the measurement error density of $\mathbf{U}$ is not known but replicated proxies are available for at least some individuals. Additionally, we allow the variability of $\mathbf{U}$ to depend on the associated unobserved values of $\mathbf{X}$ through unknown relationships, which also automatically includes the case of multivariate multiplicative measurement errors. Basic properties of finite mixture models, multivariate normal kernels and exchangeable priors are exploited in novel ways to meet modeling and computational challenges. Theoretical results showing the flexibility of the proposed methods in capturing a wide variety of data generating processes are provided. We illustrate the efficiency of the proposed methods in recovering the density of $\mathbf{X}$ through simulation experiments. The methodology is applied to estimate the joint consumption pattern of different dietary components from contaminated 24 hour recalls. Supplementary Material presents substantive additional details.

Keywords

B-splines; Conditional heteroscedasticity; Latent factor analyzers; Measurement errors; Mixture models; Multivariate density deconvolution; Regularization; Shrinkage

1 Introduction

Many problems of practical importance require estimation of the density $f_{\mathbf{X}}$ of a vector valued random variable $\mathbf{X}$. Precise measurements on $\mathbf{X}$ may not, however, be available, observations being contaminated by measurement errors $\mathbf{U}$. Under the assumption of additive measurement errors, the observations are generated from a convolution of the density $f_{\mathbf{X}}$ of $\mathbf{X}$ and the density $f_{\mathbf{U}}$ of the measurement errors $\mathbf{U}$. The problem of estimating the density $f_{\mathbf{X}}$ from available contaminated measurements then becomes a problem of multivariate density deconvolution.

This article proposes novel Bayesian semiparametric density deconvolution approaches based on finite mixtures of latent factor analyzers for robust estimation of the density $f_{\mathbf{X}}$ when the measurement error density $f_{\mathbf{U}}$ is not known, but replicated proxies contaminated with measurement errors $\mathbf{U}$ are available for at least some individuals. The proposed deconvolution approaches are highly robust, not having to impose restrictive parametric assumptions on $f_{\mathbf{X}}$ or $f_{\mathbf{U}}$. Additionally, the variability of $\mathbf{U}$ is allowed to depend on the associated unobserved values of $\mathbf{X}$ through unknown relationships.

While the focus of the article will primarily be on additive measurement errors, importantly, the methodology for additive conditionally heteroscedastic measurement errors developed here also automatically encompasses the case of multivariate multiplicative measurement errors.

To the best of our knowledge, all existing multivariate deconvolution approaches assume that $\mathbf{U}$ is independent of $\mathbf{X}$ and that the error density $f_{\mathbf{U}}$ is completely known. Ours is thus the first paper that allows the density of the measurement errors to be unknown and free from parametric laws and additionally also accommodates conditional heteroscedasticity in the measurement errors.

The literature on the problem of univariate density deconvolution, in which context we denote the variable of interest by X and the measurement errors by U , is vast. Most of the early literature considered scenarios when the measurement error density f_U is completely known. Fourier inversion based deconvoluting kernel density estimators have been studied by Carroll and Hall (1988), Liu and Taylor (1989), Devroye (1989), Fan (1991a, 1991b, 1992) and Hesse (1999) among many others. For a review of these methods, the reader may be referred to Section 12.1 in Carroll, et al. (2006) and Section 10.2.3 in Buonaccorsi (2010). In reality f_U is rarely known. The problem of deconvolution when the errors are homoscedastic with an unknown density and replicated proxies are available for each subject has been addressed by Li and Vuong (1998). See also Diggle and Hall (1993), Neumann (1997), Carroll and Hall (2004) and the references therein. The assumptions of homoscedasticity of U and their independence from X are also often unrealistic. Flexible Bayesian density deconvolution approaches that allow U to be conditionally heteroscedastic have recently been developed in Staudenmayer, et al. (2008) and Sarkar, et al. (2014). Staudenmayer, et al. (2008) assumed the measurement errors to be normally distributed and used finite mixtures of B-splines to estimate f_X and a variance function that captured the conditional heteroscedasticity. Sarkar, et al. (2014) further relaxed the assumption of

normality of U employing flexible infinite mixtures of normal kernels induced by Dirichlet processes to estimate both f_X and f_U . Sieve based methods developed in Schennach (2004) and Hu and Schennach (2008) can also handle conditional heteroscedasticity.

In sharp contrast to the univariate case, the literature on multivariate density deconvolution is quite sparse. We can only mention Masry (1991), Youndjé and Wells (2008), Comte and Lacour (2013), Hazelton and Turlach (2009, 2010) and Bovy, et al. (2011). The first three considered deconvoluting kernel based approaches assuming the measurement errors U to be distributed independently from X according to a known probability law. Hazelton and Turlach (2009, 2011), working with the same assumptions on U , proposed weighted kernel based methods. Bovy, et al. (2011) modeled the density f_X using flexible mixtures of multivariate normal kernels, but they assumed f_U to be multivariate normal with known covariance matrices, independent from X . As in the case of univariate problems, the assumptions of a fully specified f_U , known covariance matrices, and independence from X are highly restrictive for most practical applications.

The focus of this article is on multivariate density deconvolution when f_U is not known but replicated proxies are available for at least some individuals. The proposed deconvolution approaches can additionally accommodate conditional heteroscedasticity in U . The problem is important, for instance, in nutritional epidemiology, where nutritionists are typically interested not just in the consumption behaviors of individual dietary components but also in their joint consumption patterns. The data are often available in the form of dietary recalls and are contaminated by measurement errors that show strong patterns of conditional heteroscedasticity.

As in Sarkar, et al. (2014), we use mixture models to estimate both f_X and f_U but the multivariate nature of the problem brings in new modeling challenges and computational obstacles that preclude straightforward extension of their univariate deconvolution approaches. Instead of using infinite mixtures induced by Dirichlet processes, we use finite mixtures of multivariate normal kernels with exchangeable Dirichlet priors on the mixture probabilities. The use of finite mixtures and exchangeable priors greatly reduces computational complexity while retaining essentially the same flexibility. Carefully constructed priors also allow automatic model selection and model averaging. To save space, detailed discussions on these important issues are moved to Section S.6 in the Supplementary Material.

We also exploit symmetric Dirichlet priors and properties of multivariate normal distributions and finite mixture models to develop a novel strategy that enables us to enforce a required zero mean restriction on the measurement errors. Our proposed technique, as opposed to the one adopted by Sarkar, et al. (2014), is particularly suitable for high dimensional applications and can be easily generalized to enforce moment restrictions on other types of finite mixture models.

It is well known that inverse Wishart priors, due to their dense parametrization, are not suitable for modeling covariance matrices in high dimensional applications. In deconvolution problems the issue is further complicated since X and U are both latent. This

results in numerically unstable estimates even for small and moderate dimensions, particularly when the true covariance matrices are sparse and the likelihood function is of complicated form. To reduce the effective number of parameters required to be estimated, we consider factor-analytic representation of the component specific covariance matrices with sparsity inducing shrinkage priors on the factor loading matrices.

Models for multivariate regression errors that assume normality but allow the covariance matrix to vary flexibly with associated precisely measured and possibly multivariate predictors have recently been developed in the literature (Hoff and Niu, 2012; Fox and Dunson, 2016, etc.). Unlike regression settings, exclusive relationships exist between different components of multivariate measurement errors $\mathbf{U}$ and different components of the associated multivariate latent ‘predictor’ $\mathbf{X}$ - the ℓ^{th} component U_ℓ of $\mathbf{U}$ contaminates only the ℓ^{th} component X_ℓ of $\mathbf{X}$ but not others. We thus deem covariance regression models that allow $\text{cov}(\mathbf{U}|\mathbf{X})$ to vary arbitrarily with all components of $\mathbf{X}$ to be inappropriate in multivariate measurement error settings. As discussed above, the assumption of multivariate normality is also particularly restrictive in measurement error problems. In this article, we develop a semiparametric approach that appropriately highlights the exclusive associations between U_ℓ and X_ℓ while allowing the distribution of $(\mathbf{U}|\mathbf{X})$ to depart from normality. Importantly, the model also arises naturally from multivariate multiplicative measurement error settings, automatically encompassing such cases. Diagnostic tools for checking model adequacy are also discussed.

The likelihood function for the conditional heteroscedastic model poses significant computational challenges. We overcome these obstacles by designing a novel two-stage procedure that exploits the unique properties of conditionally heteroscedastic multivariate measurement errors to our advantage. The procedure first estimates the variance functions characterizing $\text{var}(U_\ell|X_\ell)$ using reparametrized versions of the corresponding univariate submodels. The estimates obtained in the first stage are then plugged-in to estimate the remaining parameters in the second stage. Having two estimation stages, our deconvolution method for conditionally heteroscedastic measurement errors is not purely Bayesian. But they show good empirical performance and, with no other solution available in the existing literature, they provide at least workable starting points towards more sophisticated methodology.

The article is organized as follows. Section 2 details the models. Model identifiability issues and implementation details, including the choice of hyper-parameters and Markov chain Monte Carlo (MCMC) algorithms to sample from the posterior, are discussed in the Supplementary Material. Section 4 discusses model identifiability issues. Section 5 presents theoretical results showing flexibility of the proposed models. Simulation studies comparing the proposed deconvolution methods to a naive method that ignores measurement errors are presented in Section 6. Section 7 presents an application of the proposed methodology in estimation of the joint consumption pattern of dietary intakes from contaminated 24 hour recalls in a nutritional epidemiologic study. Section 8 includes a discussion. An unnumbered section concludes the article with a description of the Supplementary Material.

2 Deconvolution Models

The goal is to estimate the unknown joint density of a p -dimensional multivariate random variable $\mathbf{X}$. There are $i = 1, \dots, n$ subjects. Precise measurements of $\mathbf{X}$ are not available. Instead, for $j = 1, \dots, m_i$ replicated proxies $\mathbf{W}_{ij}$ contaminated with measurement errors $\mathbf{U}_{ij}$ are available for each subject i . The replicates are assumed to be generated by the model

$$\mathbf{W}_{ij} = \mathbf{X}_i + \mathbf{U}_{ij}. \quad (1)$$

Given $\mathbf{X}_i$, $\mathbf{U}_{ij}$ are independently distributed with $E(\mathbf{U}_{ij}|\mathbf{X}_i) = \mathbf{0}$. The marginal density of $\mathbf{W}_{ij}$ is denoted by $f_{\mathbf{W}}$. The implied conditional distributions of $\mathbf{W}_{ij}$ and $\mathbf{U}_{ij}$, given $\mathbf{X}_i$, are denoted by $f_{\mathbf{W}|\mathbf{X}}$ and $f_{\mathbf{U}|\mathbf{X}}$, respectively.

2.1 Modeling the Density $f_{\mathbf{X}}$

In this article $f_{\mathbf{X}}$ is specified as a mixture of multivariate normal kernels

$$f_{\mathbf{X}}(\mathbf{X}) = \sum_{k=1}^{K_{\mathbf{X}}} \pi_{\mathbf{X},k} \text{MVN}_p(\mathbf{X} | \boldsymbol{\mu}_{\mathbf{X},k}, \boldsymbol{\Sigma}_{\mathbf{X},k}), \quad (2)$$

where $\text{MVN}_p(\cdot | \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a p -dimensional multivariate normal density with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. For the rest of this subsection, the subscript $\mathbf{X}$ is kept implicit to keep the notation clean.

We assign a finite Dirichlet prior to the mixture probability vector $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)^T$ as

$$\boldsymbol{\pi} \sim \text{Dir}(\alpha/K, \dots, \alpha/K). \quad (3)$$

Here $\text{Dir}(\alpha_1, \dots, \alpha_K)$ denotes a finite dimensional Dirichlet distribution on the K -dimensional unit simplex with concentration parameter $(\alpha_1, \dots, \alpha_K)$. Given K and the latent cluster membership indices, the prior is conjugate. The symmetry of the assumed Dirichlet prior helps in additional reduction of computational complexity by simplifying MCMC mixing issues. Provided K is sufficiently large, a carefully chosen α can impart the posterior with certain properties that simplify model selection and model averaging issues by influencing the posterior to concentrate in regions that favor empty redundant components, see Section S.1 and Section S.6 of the Supplementary Material. We assign conjugate multivariate normal priors to the component specific mean vectors $\boldsymbol{\mu}_k$, so that

$$\boldsymbol{\mu}_k \sim \text{MVN}_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0). \quad (4)$$

The conjugacy again helps in simplifying posterior calculations. Later on, we will employ similar mixture models for the density of the measurement errors, and this conjugacy, along with some basic properties of multivariate normal kernels, will also help us enforce the mean

zero restriction on the measurement errors. For the component specific covariance matrices Σ_k , we first consider conjugate inverse Wishart priors

$$\Sigma_k \sim \text{IW}_p(\nu_0, \Psi_0). \quad (5)$$

Here $\text{IW}_p(\nu, \Psi)$ denotes an inverse Wishart density on the space of $p \times p$ positive definite matrices with mean $\Psi/(\nu - p - 1)$. While the conjugacy of the inverse Wishart priors helps in simplifying posterior calculations, in complex high dimensional problems its dense parameterization may result in numerically unstable estimates, particularly when the covariance matrices are sparse. In a deconvolution problem the issue is compounded further by the nonavailability of the true $\mathbf{X}_i$'s. To reduce the effective number of parameters to be estimated, we consider a parsimonious factor-analytic representation of the component specific covariance matrices:

$$\Sigma_k = \Lambda_k \Lambda_k^T + \Omega, \quad (6)$$

where Λ_k are $p \times q_k$ factor loading matrices and Ω is a diagonal matrix with non-negative entries. In practical applications q_k will typically be much smaller than p , inducing parsimonious characterizations of the unknown covariance matrices Σ_k . Model (2) can be equivalently represented as

$$\Pr(C_i = k) = \pi_k, \quad (7)$$

$$(\mathbf{X}_i | C_i = k) = \boldsymbol{\mu}_k + \Lambda_k \boldsymbol{\eta}_i + \boldsymbol{\Delta}_i, \quad (8)$$

$$\boldsymbol{\eta}_i \sim \text{MVN}_p(\mathbf{0}, \mathbf{I}_p), \quad \boldsymbol{\Delta}_i \sim \text{MVN}_p(\mathbf{0}, \Omega), \quad (9)$$

where C_i are the mixture labels associated with $\mathbf{X}_i$, $\boldsymbol{\eta}_i$ are latent factors, and $\boldsymbol{\Delta}_i$ are errors with covariance $\Omega = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$.

The above characterization of Σ_k is not unique, since for any semi-orthogonal matrix $\mathbf{P}$ the loading matrix $\Lambda_k^1 = \Lambda_k \mathbf{P}$ also satisfies (6). Since interest lies primarily in estimating the density $f_{\mathbf{X}}$, identifiability of the latent factors is, however, not required. This also allows the loading matrices to have a-priori a potentially infinite number of columns. Sparsity inducing priors, that favor more shrinkage as the column index increases, can then be used to shrink the redundant columns towards zero. In this article, we do this by adapting the shrinkage priors proposed in Bhattacharya and Dunson (2011) that allow easy posterior computation.

Let $\Lambda_k = ((\lambda_{k,jh}))_{j=1, h=1}^{p, \infty}$, where j and h denote the row and the column indices, respectively. For $h = 1, \dots, \infty$, we assign priors as follows

$$\lambda_{k,jh} \sim \text{Normal}(0, \phi_{k,jh}^{-1} \tau_{k,h}^{-1}), \quad \phi_{k,jh} \sim \text{Ga}(\nu/2, \nu/2), \quad (10)$$

$$\tau_{k,h} \sim \prod_{\ell=1}^h \delta_{k,\ell}, \quad \delta_{k,\ell} \sim \text{Ga}(a_\ell, 1), \quad \sigma_j^2 \sim \text{Inv-Ga}(a_\sigma, b_\sigma). \quad (11)$$

Here $\text{Ga}(a, \beta)$ denotes a Gamma distribution with shape parameter a and rate parameter β and $\text{IG}(a, b)$ denotes an inverse-Gamma distribution with shape parameter a and scale parameter b . In the k^{th} component factor loading matrix Λ_k , the parameters $\{\phi_{k,jh}\}_{j=1}^p$ control the local shrinkage of the elements in the h^{th} column, whereas $\tau_{k,h}$ controls the global shrinkage. When $a_h > 1$ for $h = 2, \dots, \infty$, the sequence $\{\tau_{k,h}\}_{h=1}^\infty$ becomes stochastically increasing and thus favors more shrinkage as the column index h increases.

In addition to inducing adaptive sparsity and hence numerical stability, by favoring more shrinkage as the column index increases, the shrinkage priors play another important role in making the proposed factor analytic model highly robust to misspecification of the number of latent factors, allowing us to adopt simple strategies to determine the number of latent factors to be included in the model in practice. Details are deferred to Section S.1 in the Supplementary Material.

Throughout the rest of the paper, mixtures with inverse Wishart prior on the covariance matrices will be referred to as MIW models and mixtures of latent factor analyzers will be referred to as MLFA models.

For a review of finite mixture models and mixtures of latent factor analyzers, without moment restrictions or sparsity inducing priors and with applications in measurement error free scenarios, see Fokoué and Titterton (2003), Frühwirth-Schnatter (2006), Mengersen, et al. (2011) and the references therein. For other types of shrinkage priors, see Brown and Griffin (2010), Carvalho, et al. (2010), Bhattacharya, et al. (2014) etc.

2.2 Modeling the Density of the Measurement Errors

2.2.1 Independently Distributed Measurement Errors—In this section, we develop models for the measurement errors $\mathbf{U}$ assuming them to be independent from $\mathbf{X}$. That is, we assume $f_{\mathbf{U}|\mathbf{X}} = f_{\mathbf{U}}$ for all $\mathbf{X}$. This remains the most extensively researched deconvolution problem for both univariate and multivariate cases. The techniques developed in this section will also provide crucial building blocks for more realistic models in Section 2.2.2. The measurement errors and their density are now denoted by $\mathbf{e}_{ij}$ and $f_{\mathbf{e}}$, respectively, for reasons to become obvious shortly in Section 2.2.2.

As in Section 2.1, a mixture of multivariate normals can be used to model the density $f_{\mathbf{e}}$ but the model now has to satisfy a mean zero constraint. That is

$$f_{\mathbf{e}}(\mathbf{e}) = \sum_{k=1}^{K_{\mathbf{e}}} \pi_{\mathbf{e},k} \text{MVN}_p(\mathbf{e} | \boldsymbol{\mu}_{\mathbf{e},k}, \boldsymbol{\Sigma}_{\mathbf{e},k}), \quad (12)$$

$$\text{subject to } \sum_{k=1}^{K_{\mathbf{e}}} \pi_{\mathbf{e},k} \boldsymbol{\mu}_{\mathbf{e},k} = \mathbf{0}. \quad (13)$$

To get numerically stable estimates of the density of the errors, latent factor characterization of the covariance matrices with sparsity inducing shrinkage priors as in Section 2.1 may again be used. Details are curtailed to avoid unnecessary repetition and we only present the mechanism to enforce the zero mean restriction on the model. The subscript $\mathbf{e}$ is again dropped in favor of cleaner notation. In later sections, the subscripts $\mathbf{X}$ and $\mathbf{e}$ reappear to distinguish between the parameters associated with $f_{\mathbf{X}}$ and $f_{\mathbf{e}}$, when necessary.

Without the mean restriction and under conjugate multivariate normal priors $\boldsymbol{\mu}_k \sim \text{MVN}_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, the posterior full conditional of $\boldsymbol{\mu}^{K \times 1} = (\boldsymbol{\mu}_1^T, \dots, \boldsymbol{\mu}_K^T)^T$ is given by

$$\text{MVN}_{Kp} \left\{ \begin{pmatrix} \boldsymbol{\mu}_1^0 \\ \boldsymbol{\mu}_2^0 \\ \vdots \\ \boldsymbol{\mu}_K^0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_1^0 & 0 & \dots & 0 \\ 0 & \boldsymbol{\Sigma}_2^0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \boldsymbol{\Sigma}_K^0 \end{pmatrix} \right\} \equiv \text{MVN}_{Kp}(\boldsymbol{\mu}^0, \boldsymbol{\Sigma}^0), \quad (14)$$

where $\mathbf{e}_{ij}$ and other conditioning variables are implicitly understood. Explicit expressions of $\boldsymbol{\mu}^0$ and $\boldsymbol{\Sigma}^0$ in terms of the conditioning variables can be found in Section S.1 in the Supplementary Material. The posterior full conditional of $\boldsymbol{\mu}$ under the mean restriction can then be obtained easily by further conditioning the distribution in (14) by

$\boldsymbol{\mu}_R = \sum_{k=1}^K \pi_k \boldsymbol{\mu}_k = \mathbf{0}$ and is given by

$$(\boldsymbol{\mu} | \boldsymbol{\mu}_R = \mathbf{0}) \sim \text{MVN}_{Kp} \{ \boldsymbol{\mu}^0 - \boldsymbol{\Sigma}_{1,R}^0 (\boldsymbol{\Sigma}_{R,R}^0)^{-1} \boldsymbol{\mu}_R^0, \boldsymbol{\Sigma}^0 - \boldsymbol{\Sigma}_{1,R}^0 (\boldsymbol{\Sigma}_{R,R}^0)^{-1} \boldsymbol{\Sigma}_{R,1}^0 \}, \quad (15)$$

where and

$$\boldsymbol{\mu}_R^0 = \sum_{k=1}^K \pi_k \boldsymbol{\mu}_k^0 = E(\boldsymbol{\mu}_R), \boldsymbol{\Sigma}_{k,K} = \pi_k \boldsymbol{\Sigma}_k^0 = \text{cov}(\boldsymbol{\mu}_k, \boldsymbol{\mu}_R), \boldsymbol{\Sigma}_{R,R}^0 = \boldsymbol{\Sigma}_{K+1,K+1} = \sum_{k=1}^K \pi_k^2 \boldsymbol{\Sigma}_k^0 = \text{cov}(\boldsymbol{\mu}_R)$$

and $\boldsymbol{\Sigma}_{R,1}^0 = (\boldsymbol{\Sigma}_{1,K+1}, \boldsymbol{\Sigma}_{2,K+1}, \dots, \boldsymbol{\Sigma}_{K,K+1})$. To sample from this singular density, we can

first sample from the non-singular distribution of $\{(\mu_1^T, \mu_2^T, \dots, \mu_{K-1}^T)^T | \mu_R = \mathbf{0}\}$, which can also be trivially obtained from (15), and then set $\mu_K = -\sum_{k=1}^{K-1} \pi_k \mu_k / \pi_K$.

2.2.2 Conditionally Heteroscedastic Measurement Errors—We now consider the case when the variances of the measurement errors depend on the associated unknown values of $\mathbf{X}$ through unknown relationships.

Interpreting the conditioning variables $\mathbf{X}$ broadly as predictors, one can loosely connect our problem of modeling conditionally heteroscedastic $\mathbf{U}$ to the problem of covariance regression (Hoff and Niu, 2012; Fox and Dunson, 2016, etc.), where the covariance of the multivariate regression errors are allowed to vary flexibly with precisely measured and possibly multivariate predictors. In such problems, the dimension of the regression errors is unrelated to the dimension of the predictors and different components of the regression errors are assumed to be equally influenced by different components of the predictors. In multivariate deconvolution problems, in contrast, the dimension of $\mathbf{U}_{ij}$ is exactly the same as the dimension of $\mathbf{X}_i$, the ℓ^{th} component $U_{ij\ell}$ being the measurement error associated exclusively with $X_{i\ell}$. See Figure 1. While different components of $\mathbf{U}_{ij}$ may be correlated, this exclusive association between $U_{ij\ell}$ and $X_{i\ell}$ implies that the dependence of $U_{ij\ell}$ on $\mathbf{X}_i$ should be explained primarily through $X_{i\ell}$. Figure 7, for instance, suggests strong conditional heteroscedasticity patterns and it is plausible to assume that this conditional variability in $U_{ij\ell}$ can be explained mostly through $X_{i\ell}$ only. It is interesting to note these contrasts between conditionally varying regression and measurement errors become particularly prominent in the multivariate set up. Additionally, the aforementioned covariance regression approaches all assume multivariate normality of the regression errors. As discussed in the introduction, such strong parametric assumptions on the error distribution are particularly restrictive in measurement error problems. Additional detailed discussions of these important issues and resulting modeling implications can be found in Section S.5 of the Supplementary Material. They preclude direct application of existing covariance regression approaches to multivariate deconvolution problems but warrant models that can highlight the aforementioned unique dependence relationships, accommodate distributional flexibility while enforcing the mean zero restriction, and produce computationally stable estimates even in the absence of precise information on the conditioning variable $\mathbf{X}$.

The semiparametric approach that we adopt in this article achieves distributional flexibility, enforces the mean zero restriction, accommodates the exclusive relationships between $U_{ij\ell}$ and $X_{i\ell}$ but ignores the weak dependencies of $U_{ij\ell}$ on $\{X_{im}\}_{m \neq \ell}$ depicted in Figure 1(b). Specifically, we let

$$(\mathbf{U}_{ij} | \mathbf{X}_i) = \mathbf{S}(\mathbf{X}_i) \mathbf{e}_{ij}, \quad (16)$$

where $\mathbf{S}(\mathbf{X}_i) = \text{diag}\{s_1(X_{i1}), s_2(X_{i2}), \dots, s_p(X_{ip})\}$ and $\mathbf{e}_{ij}$, henceforth referred to as the ‘scaled errors’, are distributed independently of $\mathbf{X}_i$. Model (16) implies that $\text{cov}(\mathbf{U}_{ij} | \mathbf{X}_i) = \mathbf{S}(\mathbf{X}_i) \text{cov}(\mathbf{e}_{ij}) \mathbf{S}(\mathbf{X}_i)$ and marginally $\text{var}(U_{ij\ell} | \mathbf{X}_i) = s_{\ell}^2(X_{i\ell}) \text{var}(e_{ij\ell})$, a function of $X_{i\ell}$ only. The

techniques developed in Section 2.2.1 can now be employed to model the density of $\mathbf{e}_{ij}$ allowing different components of $\mathbf{U}_{ij}$ to be correlated and their joint density to deviate from multivariate normality.

We model the variance functions S_{ℓ}^2 , denoted also by v_{ℓ} using positive mixtures of B-spline basis functions with smoothness inducing priors on the coefficients as in Staudenmayer, et al. (2008). For the ℓ^{th} component, partition an interval $[A_{\ell}, B_{\ell}]$ of interest into L_{ℓ} subintervals using knot points $A_{\ell} = t_{\ell,1} = \dots = t_{\ell,q+1} < t_{\ell,q+2} < t_{\ell,q+3} < \dots < t_{\ell,q+L_{\ell}} < t_{\ell,q+L_{\ell}+1} = \dots = t_{\ell,2q+L_{\ell}+1} = B_{\ell}$. A flexible model for the variance functions is given by

$$v_{\ell}(X_{i\ell}) = s_{\ell}^2(X_{i\ell}) = \sum_{j=1}^{J_{\ell}} b_{q,j,\ell}(X_{i\ell}) \exp(\xi_{j\ell}) = \mathbf{B}_{q,J_{\ell},\ell}(X_{i\ell}) \exp(\boldsymbol{\xi}_{\ell}), \quad (17)$$

$$(\boldsymbol{\xi}_{\ell} | J_{\ell}, \sigma_{\xi,\ell}^2) \propto (2\pi\sigma_{\xi,\ell}^2)^{-J_{\ell}/2} \exp \left\{ -\boldsymbol{\xi}_{\ell}^T P_{\ell} \boldsymbol{\xi}_{\ell} / (2\sigma_{\xi,\ell}^2) \right\}, \quad \sigma_{\xi,\ell}^2 \sim \text{Inv-Ga}(a_{\xi}, b_{\xi}). \quad (18)$$

Here $\{b_{q,j,\ell}\}_{j=1}^{J_{\ell}}$ denote $J_{\ell} = (q+L_{\ell})$ B-spline bases of degree q as defined in de Boor (2000), $\boldsymbol{\xi}_{\ell} = \{\xi_{1\ell}, \xi_{2\ell}, \dots, \xi_{J_{\ell}\ell}\}^T$; $\exp(\boldsymbol{\xi}_{\ell}) = \{\exp(\xi_{1\ell}), \exp(\xi_{2\ell}), \dots, \exp(\xi_{J_{\ell}\ell})\}^T$; and $P_{\ell} = D_{\ell}^T D_{\ell}$, where D_{ℓ} is a $J_{\ell} \times (J_{\ell} + 2)$ matrix such that $D_{\ell} \boldsymbol{\xi}_{\ell}$ computes the second differences in $\boldsymbol{\xi}_{\ell}$. The prior $P_0(\boldsymbol{\xi}_{\ell} | \sigma_{\xi,\ell}^2)$ induces smoothness in the coefficients because it penalizes $\sum_{j=1}^{J_{\ell}} (\Delta^2 \xi_{j\ell})^2 = \boldsymbol{\xi}_{\ell}^T P_{\ell} \boldsymbol{\xi}_{\ell}$, the sum of squares of the second order differences in $\boldsymbol{\xi}_{\ell}$ (Eilers and Marx, 1996). The parameters $\sigma_{\xi,\ell}^2$ play the role of smoothing parameter - the smaller the value of $\sigma_{\xi,\ell}^2$, the stronger the penalty and the smoother the variance function. The inverse-Gamma hyperpriors on $\sigma_{\xi,\ell}^2$ allow the data to have influence on the posterior smoothness and make the approach data adaptive.

Since $s_{\ell}^2(X_{i\ell}) \text{var}(\varepsilon_{ij\ell}) = \{s_{\ell}^2(X_{i\ell})c\} \{\text{var}(\varepsilon_{ij\ell})/c\}$ for any $c > 0$, the variance functions s_{ℓ}^2 's can not be uniquely determined without additional restrictions on $\text{var}(\varepsilon_{ij\ell})$. Separate identifiability of $\mathbf{S}$ and $\mathbf{f}_{\ell}$ is, however, not required for inference on $\mathbf{f}_{\mathbf{X}}$ or to assess the conditional variability in $\mathbf{U}_{ij\ell}$. The latter, for instance, may simply be obtained as $\text{var}(U_{ij\ell} | X_i) = s_{\ell}^2(X_{i\ell}) \text{var}(\varepsilon_{ij\ell})$. We thus avoid additional identifiability restrictions that would further compound modeling challenges. Adjustments made to the estimates of s_{ℓ}^2 and $\mathbf{f}_{\ell}$ to enable comparisons with the corresponding true values in simulation experiments are discussed in Section S.3 in the Supplementary Material.

2.2.3 Multiplicative Measurement Errors—In this section we consider the case of multivariate multiplicative measurement errors. The replicates are now assumed to be generated by the model

$$\mathbf{W}_{ij} = \mathbf{X}_i \circ \tilde{\mathbf{U}}_{ij}, \quad (19)$$

where $\circ$ denotes element wise product and the errors $\tilde{\mathbf{U}}_{ij}$ are distributed independently of $\mathbf{X}_i$ with $E(\tilde{\mathbf{U}}_{ij}) = \mathbf{1}$. Importantly, model (19) can be reformulated to arrive at model (16) as

$$\mathbf{W}_{ij} = \mathbf{X}_i \circ \tilde{\mathbf{U}}_{ij} = \mathbf{X}_i + \mathbf{U}_{ij}, \quad \text{with} \quad \mathbf{U}_{ij} = \mathbf{X}_i \circ (\tilde{\mathbf{U}}_{ij} - \mathbf{1}) = \mathbf{S}(\mathbf{X}_i) \mathbf{e}_{ij}. \quad (20)$$

with $E(\mathbf{U}_{ij}|\mathbf{X}_i) = \mathbf{X}_i \circ E(\tilde{\mathbf{U}}_{ij} - \mathbf{1}) = \mathbf{0}$, $\mathbf{S}(\mathbf{X}_i) = \text{diag}\{s_1(X_{i1}), \dots, s_p(X_{ip})\}$ with $s_\ell(X_{i\ell}) = X_{i\ell}$ and $\mathbf{e}_{ij} = (\tilde{\mathbf{U}}_{ij} - \mathbf{1})$ are independent of $\mathbf{X}_i$ with $E(\mathbf{e}_{ij}) = \mathbf{0}$. This observation precludes the need for separate methodology to be developed for the problem of multivariate density deconvolution in the presence of multiplicative measurement errors and further emphasizes the importance of the additive conditionally heteroscedastic measurement error model (16) developed in Section 2.2.2.

3 Posterior Inference

Inference is based on samples drawn from the posterior using MCMC algorithms. A Gibbs sampler for the independent error case discussed in Section 2.2.1 is presented in Section S.2 of the Supplementary Material. For the conditionally heteroscedastic case discussed in Section 2.2.2, the full conditionals of the parameters characterizing the variance functions do not have closed form expressions. MCMC algorithms where we tried to integrate Metropolis-Hastings (MH) steps within the Gibbs sampler to generate samples from the full posterior were numerically unstable and failed to converge sufficiently quickly. To address this challenge, we designed a novel two-stage procedure. For each k , we first estimate the functions $s_\ell(X_{i\ell})$ by fitting the univariate deconvolution models $W_{ij\ell} = X_{i\ell} + s_\ell(X_{i\ell})e_{ij\ell}$. High precision estimates of the variance functions $s_\ell^2(X_{i\ell})$ can be obtained using the univariate deconvolution models. See Figure 2 in the main article and Figure S.7 in the Supplementary Material for illustrations. Parameters characterizing other components of the full model are then sampled using a Gibbs sampler keeping the estimates of the variance functions fixed. Additional details are deferred to Sections S.3 and S.4 of the Supplementary Material.

4 Model Identifiability

This section presents a discussion of model identifiability issues. The density of interest $f_{\mathbf{X}}$ is identifiable under mild technical assumptions. In the case of independently distributed measurement errors considered in Section 2.2.1 of the main paper, appealing to Li and Vuong (1998), the densities $f_{\mathbf{X}}$ and $f_{\mathbf{e}}$ are identifiable provided $m_i \geq 2$ replicates are available for some individuals, and the characteristics functions $\phi_{\mathbf{X}}(\mathbf{t}) = E\{\exp(\mathbf{t}^T \mathbf{X})\}$ and $\phi_{\mathbf{e}}(\mathbf{t}) = E\{\exp(\mathbf{t}^T \mathbf{e})\}$ are non-vanishing everywhere.

In the case of conditionally heteroscedastic measurement errors considered in Section 2.2.2 of the main paper, appealing to Hu and Schennach (2004), the densities $f_{\mathbf{X}}$ and $f_{\mathbf{U}|\mathbf{X}}$ are identifiable provided $m_i \geq 3$ replicates are available for some individuals, the joint,

conditional and marginal densities of $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mathbf{X}$ are all bounded, and the density $f_{\mathbf{X}|\mathbf{W}}$ is bounded complete in the sense that the unique solution to $\int f_{\mathbf{X}|\mathbf{W}}(\mathbf{X})g(\mathbf{X})d\mathbf{X} = 0$ for all $\mathbf{W}$ and for all bounded $g(\mathbf{X})$ is $g(\mathbf{X}) = 0$ for all $\mathbf{X}$. The following lemma provides a sufficient condition for the density $f_{\mathbf{X}|\mathbf{W}}$ to be bounded complete.

Lemma 1

$f_{\mathbf{X}|\mathbf{W}}$ is bounded complete if $E\{\exp(\mathbf{u}^T \mathbf{X}|\mathbf{W})\}$ is non-vanishing everywhere for all $\mathbf{W}$.

Proof—By Theorem 10C of Goldberg (1961), since $E\{\exp(\mathbf{u}^T \mathbf{X}|\mathbf{W})\}$ is non-vanishing everywhere for all $\mathbf{W}$, the closed linear span of $f_{\mathbf{X}|\mathbf{W}}(\cdot)$ is $L_1(\mathbb{R})$. By Hahn-Banach Theorem, the dual space of $L_1(\mathbb{R})$ is $L_\infty(\mathbb{R})$ and there is an isometric isomorphism from $L_\infty(\mathbb{R})$ to $L_1(\mathbb{R})$ given by $g \mapsto \Phi_g$ where $\Phi_g(f_{\mathbf{X}|\mathbf{W}}) = \int f_{\mathbf{X}|\mathbf{W}}(\mathbf{X})g(\mathbf{X})d\mathbf{X}$ for all $\mathbf{W}$. Since the closed linear span of $f_{\mathbf{X}|\mathbf{W}}(\cdot)$ for all $\mathbf{W}$ is $L_1(\mathbb{R})$, $\int f_{\mathbf{X}|\mathbf{W}}(\mathbf{X})g(\mathbf{X})d\mathbf{X} = 0$ for all $\mathbf{W}$ implies that the mapping Φ_g is identically 0. By the isometric isomorphism above, it follows that g should be identically 0.

Different types of completeness of densities are often used as key identifying conditions in measurement error problems. See, for example, d'Haultfoeuille (2011) and Carroll, et al. (2010). Here, we have provided a general sufficient condition for bounded completeness to hold true and a novel proof using functional analysis techniques. Loosely speaking, if the density $f_{\mathbf{X}|\mathbf{W}}(\mathbf{X})$ varies with $\mathbf{X}$, its characteristic function does not vanish. Without sufficient variability of the density of $\mathbf{X}|\mathbf{W}$, observations on $\mathbf{W}$ do not have enough information to recover the density of $\mathbf{X}$.

Model parameters specifying the components $f_{\mathbf{X}}, f_{\mathbf{e}}, \sigma$ etc. are not separately identifiable. For inference on identifiable functional model components, identifiability of individual parameters is, however, not required. Indeed, the mixture models and the associated priors were so chosen that the mixture components remain unidentifiable. This helps simplify MCMC mixing issues. See Section S.6 of the Supplementary Material.

5 Model Flexibility

This section presents a theoretical study of the flexibility of the proposed models. Proofs of the results are presented in the Supplementary Material. We focus on the deconvolution models for conditionally heteroscedastic measurement errors, the case of independently distributed errors following as a special case. First we show that componentwise our models for the density $f_{\mathbf{X}}$ of $\mathbf{X}$, the density $f_{\mathbf{e}}$ of the scaled errors $\mathbf{e}$, and the variance functions v are all highly flexible. Building on these results, we then show that our proposed deconvolution models can accommodate a large class of data generating processes.

Let the generic notation Π denote a prior on some class of random functions. Also let $\mathcal{T}$ denote the target class of functions to be modeled by Π . The support of Π throws light on the flexibility of Π . For Π to be a flexible prior, one would expect that $\mathcal{T}$ or a large subset of $\mathcal{T}$ would be contained in the support of Π .

For investigating the flexibility of priors for density functions, a relevant concept is that of Kullback-Leibler (KL) support. The KL divergence between two densities f_0 and f , denoted by $d_{KL}(f_0, f)$, is defined as $d_{KL}(f_0, f) = \int f_0(Z) \log \{f_0(Z)/f(Z)\} dZ$. Let Π_f denote a prior assigned to a random density f . A density f_0 is said to belong to the KL support of Π_f if $\Pi_f\{f: d_{KL}(f_0, f) < \delta\} > 0 \forall \delta > 0$. The class of densities in the KL support of Π_f is denoted by $KL(\Pi_f)$.

Let $\mathcal{A}$ be the class of target densities to be modeled by the prior Π_f . Let $\mathcal{S}$ denote the support of $\mathcal{A}$ and $\tilde{\mathcal{A}} \subseteq \mathcal{A}$ denote the class of densities that satisfy the following fairly minimal set of regularity conditions. Since $\tilde{\mathcal{A}}$ is a large subclass of $\mathcal{A}$, its inclusion in the KL support of Π_f would establish the flexibility of Π_f .

Conditions 1

1. f_0 is continuous on $\mathcal{S}$ except on a set of measure zero.
2. The second order moments of f_0 are finite.
3. For some $r > 0$ and for all $\mathbf{z} \in \mathcal{S}$, there exist hypercubes $C_r(\mathbf{z})$ with side length r and $\mathbf{z} \in C_r(\mathbf{z})$ such that

$$\int f_0(\mathbf{z}) \log \left\{ \frac{f_0(\mathbf{z})}{\inf_{\mathbf{t} \in C_r(\mathbf{z})} f_0(\mathbf{t})} \right\} d\mathbf{z} < \infty.$$

Let $\Pi_{\mathbf{X}}$ be a generic notation for both the MIW and the MLFA prior on $f_{\mathbf{X}}$ defined in Section 2.1. Similarly, let $\Pi_{\mathbf{e}}$ be a generic notation for both the MIW and the MLFA prior on $f_{\mathbf{e}}$ defined in Section 2.2. When the measurement errors are distributed independently of $\mathbf{X}$, the support of $f_{\mathbf{X}}$, say $\mathcal{X}$, may be taken to be any subset of $\mathbb{R}^p$. For conditionally heteroscedastic measurement errors, the variance functions $s_{\mathcal{E}}^2(\cdot)$ that capture the conditional variability are modeled by mixtures of B-splines defined on closed intervals $[A_k, B_k]$. In this case, the support of $f_{\mathbf{X}}$ is assumed to be the closed hypercube $\mathcal{X} = [A_1, B_1] \times \cdots \times [A_p, B_p]$. Let $\mathcal{A}_{\mathbf{X}}$ denote the set of all densities on $\mathcal{X}$, the target class of densities to be modeled by $\Pi_{\mathbf{X}}$ and $\tilde{\mathcal{A}}_{\mathbf{X}} \subseteq \mathcal{A}_{\mathbf{X}}$ denote the class of densities $f_{0\mathbf{X}}$ that satisfy Conditions 1. Similarly, let $\mathcal{A}_{\mathbf{e}}$ denote the set of all densities on $\mathbb{R}^p$ that have mean zero and $\tilde{\mathcal{A}}_{\mathbf{e}} \subseteq \mathcal{A}_{\mathbf{e}}$ denote the class of densities $f_{0\mathbf{e}}$ that satisfy Conditions 1. The following Lemma establishes the flexibility of the models for $f_{\mathbf{X}}$ and $f_{\mathbf{e}}$.

Lemma 2

1. $\tilde{\mathcal{A}}_{\mathbf{X}} \subseteq KL(\Pi_{\mathbf{X}})$ 2. $\tilde{\mathcal{A}} \subseteq KL(\Pi_{\mathbf{e}})$.

For investigating the flexibility of models for general classes of functions, a relevant concept is that of sup norm support. The sup norm distance between two functions g_0 and g , denoted by $\|g_0 - g\|_{\infty}$, is defined as $\|g_0 - g\|_{\infty} = \sup_Z |g_0(Z) - g(Z)|$. Let Π_g denote a prior assigned to a random function g . A function g_0 is said to belong to the sup norm support of Π_g if $\Pi_g\{g: \|g_0 - g\|_{\infty} < \delta\} > 0 \forall \delta > 0$. The class of functions in the sup norm support of Π_g is denoted by $SN(\Pi_g)$.

Let Π_V denote the prior on the variance functions based on mixtures of B-spline basis functions defined in Section 2.2.2. For notational convenience we consider the case of a univariate variance function supported on $[A, B]$. Extension to the multivariate case with variance functions supported on $\mathcal{X}$ is technically trivial. Let $\mathcal{C}_+[A, B]$ denote the set of continuous functions from $[A, B]$ to $\mathbb{R}^+$. Also, for $\alpha \in (q+1)$, let $\mathcal{C}_+^\alpha[A, B] \subseteq \mathcal{C}_+[A, B]$ denote the set of functions that are α_0 times continuously differentiable, and for all $v_0 \in \mathcal{C}_+^\alpha[A, B]$, $\|v_0\|_\alpha < \infty$, where α_0 is largest integer less than or equals to α and the seminorm is defined by $\|v_0\|_\alpha = \sup_{X, X' \in [A, B], X \neq X'} \{ |v_0^{(\alpha_0)}(X) - v_0^{(\alpha_0)}(X')| / |X - X'|^{\alpha - \alpha_0} \}$.

The local support properties of B-splines make the models for the variance functions very flexible as is indicated by the following lemma.

Lemma 3

$$\mathcal{C}_+^\alpha[A, B] \subseteq \mathcal{C}_+[A, B] \subseteq SN(\Pi_V).$$

Although technically the sup norm distance between linear combinations of B-splines and any continuous function can be made arbitrarily small by increasing the number of knots, for obvious reasons the actual bounds for the sup norm distance may not be very sharp if the function to be modeled is wiggly. However, for most applications of practical importance, the true variance function may be assumed to be smooth, that is, to belong to some $\mathcal{C}_+^\alpha[A, B]$ with $\alpha \geq 1$. Therefore, for practical reasons, it is only important that the smaller Hölder class of functions $\mathcal{C}_+^\alpha[A, B]$ belongs to the sup norm support of Π_V . As shown in Section S. 7.2 of the Supplementary Material, the bounds for sup norm distance in this case will also be much sharper.

Since the models for the variance functions v_l and the models for the density of the scaled errors f_e are separately very flexible, under model (16) on the measurement errors, the implied conditional and joint densities are also expected to be very flexible. This is investigated in the next lemma. For a given $\mathbf{X}$, let $\Pi_{U|\mathbf{X}}$ denote the prior for $f_{U|\mathbf{X}}$ induced by Π_e and Π_V under model (16). Define $\tilde{\mathcal{F}}_{U|\mathbf{X}} = \{f_{0U|\mathbf{X}} : f_{0U|\mathbf{X}}(\mathbf{U}) = \prod_{k=1}^p s_{0k}^{-1}(X_k) f_{0e}\{\mathbf{S}_0^{-1}(\mathbf{X})\mathbf{U}\}, s_{0k}^2 \in \mathcal{C}_+[A_k, B_k] \text{ for } k = 1, \dots, p, f_{0e} \in \tilde{\mathcal{F}}_e\}$. Also let $\Pi_{U|V}$ denote the prior for the unknown conditional density of $\mathbf{U}$ induced by Π_e and Π_V under model (16). Define $\tilde{\mathcal{F}}_{U|V} = \{f_{0U|V} : \text{for any given } \mathbf{X} \in \mathcal{X}, f_{0U|V} = f_{0U|\mathbf{X}} \in \tilde{\mathcal{F}}_{U|\mathbf{X}}\}$. Finally, let $\Pi_{\mathbf{X}, \mathbf{U}}$ denote the prior for the joint density of $(\mathbf{X}, \mathbf{U})$ induced by $\Pi_{\mathbf{X}}$, Π_e and Π_V under model (16). Define $\tilde{\mathcal{F}}_{\mathbf{X}, \mathbf{U}} = \{f_{0, \mathbf{X}, \mathbf{U}} : f_{0, \mathbf{X}, \mathbf{U}}(\mathbf{X}, \mathbf{U}) = f_{0, \mathbf{X}}(\mathbf{X}) f_{0U|\mathbf{X}}(\mathbf{U}|\mathbf{X}), \text{ where } f_{0\mathbf{X}} \in \tilde{\mathcal{F}}_{\mathbf{X}} \text{ and } f_{0U|\mathbf{X}} \in \tilde{\mathcal{F}}_{U|\mathbf{X}} \text{ for all } \mathbf{X} \in \mathcal{X}\}$.

Lemma 4

1. $\tilde{\mathcal{F}}_{U|\mathbf{X}} \subseteq KL(\Pi_{U|\mathbf{X}})$ for any given $\mathbf{X} \in \mathcal{X}$.
2. For any $f_{0U|V} \in \tilde{\mathcal{F}}_{U|V}$, $\Pi_{U|V}\{\sup_{\mathbf{X} \in \mathcal{X}} d_{KL}(f_{0U|\mathbf{X}}, f_{U|\mathbf{X}}) < \delta\} > 0$ for all $\delta > 0$.
3. $\tilde{\mathcal{F}}_{\mathbf{X}, \mathbf{U}} \subseteq KL(\Pi_{\mathbf{X}, \mathbf{U}})$.

The flexibility of the implied model for the marginal density $f_{\mathbf{W}}$ is the subject of our final result. Since the only observed quantities are $\mathbf{W}_{ij}$, the support of the induced prior on $f_{\mathbf{W}}$ tells us about the types of likelihood functions the model can approximate.

Let $\Pi_{\mathbf{W}}$ denote the prior for the density of $\mathbf{W}$ induced by $\Pi_{\mathbf{X}}$, $\Pi_{\mathbf{e}}$ and $\Pi_{\mathbf{V}}$ under model (16). Also let $\tilde{\mathcal{F}}_{\mathbf{W}} = \{f_{0\mathbf{W}} : f_{0\mathbf{W}}(\mathbf{W}) = \int f_{0\mathbf{X}}(\mathbf{X})f_{0\mathbf{U}|\mathbf{X}}(\mathbf{W}-\mathbf{X})d\mathbf{X}, f_{0\mathbf{X}} \in \tilde{\mathcal{F}}_{\mathbf{X}}, f_{0\mathbf{U}|\cdot} \in \tilde{\mathcal{F}}_{\mathbf{U}|\cdot}\}$, the class of densities $f_{0\mathbf{W}}$ that can be obtained as the convolution of two densities $f_{0\mathbf{X}}$ and $f_{0\mathbf{U}|\cdot}$, where $f_{0\mathbf{X}} \in \tilde{\mathcal{F}}_{\mathbf{X}}$ and $f_{0\mathbf{U}|\cdot} \in \tilde{\mathcal{F}}_{\mathbf{U}|\cdot}$.

Since the supports of $\Pi_{\mathbf{X}}$ and $\Pi_{\mathbf{U}|\mathbf{X}}$ are large, it is expected that the support of $\Pi_{\mathbf{W}}$ will also be large. However, because convolution is involved, investigation of KL support of $\Pi_{\mathbf{W}}$ is a difficult problem. A weaker but relevant concept is that of L_1 support. The L_1 distance between two densities f_0 and f , denoted by $\|f_0 - f\|_1$, is defined as $\|f_0 - f\|_1 = \int |f_0(Z) - f(Z)|dZ$. A density f_0 is said to belong to the L_1 support of Π_f if $\Pi_f(f : \|f_0 - f\|_1 < \delta) > 0 \forall \delta > 0$. The class of densities in the L_1 support of Π_f is denoted by $L_1(\Pi_f)$. The following theorem shows that the L_1 support of $\Pi_{\mathbf{W}}$ is large.

Theorem 1

$$\tilde{\mathcal{F}}_{\mathbf{W}} \subseteq L_1(\Pi_{\mathbf{W}}).$$

The proofs of these results are deferred to Section S.7 of the Supplementary Material. The proofs require that the number of mixture components K be allowed to vary over $\mathbb{N}$, the set of all positive integers, through priors, denoted by the generic notation $P_0(K)$, that assign positive probability to all $K \in \mathbb{N}$. Posterior computation for such methods will be computationally intensive, specially in a complicated multivariate set up like ours. In our implementation, we thus keep the number of mixture components fixed at finite values.

6 Simulation Experiments

The mean integrated squared error (MISE) of estimation of $f_{\mathbf{X}}$ by $\hat{f}_{\mathbf{X}}$ is defined as $MISE = E_{f_{\mathbf{X}}} \int \{f_{\mathbf{X}}(\mathbf{X}) - \hat{f}_{\mathbf{X}}(\mathbf{X})\}^2 d\mathbf{X}$. Based on B simulated data sets, a Monte Carlo estimate of MISE is given by $MISE_{est} = B^{-1} \sum_{b=1}^B \sum_{m=1}^M \{f_{\mathbf{X}}(\mathbf{X}_{b,m}) - \hat{f}_{\mathbf{X}}^{(b)}(\mathbf{X}_{b,m})\}^2 / p_0(\mathbf{X}_{b,m})$, where $\{\mathbf{X}_{b,m}\}_{b=1, m=1}^{B, M}$ are random samples from the density p_0 . We designed simulation

experiments to evaluate the MISE performance of the proposed models for a wide range of possibilities. The MISEs we report here are all based on 100 simulated data sets and $M = 10^6$ samples generated from each of the two densities (a) $p_0 = f_{\mathbf{X}}$, the true density of $\mathbf{X}$, and (b) p_0 that is uniform on the hypercube with edges $\min_k \{\mu_{\mathbf{X},k} - 31_p\}$ and $\max_k \{\mu_{\mathbf{X},k} + 31_p\}$. With carefully chosen initial values and proposal densities for the MH steps, we were able to achieve quick convergence for the MCMC samplers. The use of exchangeable Dirichlet priors helped simplify mixing issues (Geweke, 2007). See Section S.6.2 in the Supplementary Material for additional discussions. We programmed our methods in R. In each case, we ran 3000 MCMC iterations and discarded the initial 1000 iterations as burn-in. The post burn-in samples were thinned by a thinning interval of length 5. For the univariate samplers, 1000 MCMC iterations with a burn-in of 500 sufficed to produce stable estimates of the variance functions. In our experiments with much larger iteration numbers and burn-

ins, the MISE performances remained practically the same. This being the first article that tries to solve the problem of multivariate density deconvolution when the measurement error density is unknown, the proposed MIW and MLFA models have no competitors. We thus compared our models with a naive Bayesian method that ignores measurement errors and treats the subject specific means as precisely measured observations instead, modeling $f_{\mathbf{X}}$ by a finite mixture of multivariate normals as in (2) with inverse Wishart priors on the component specific covariance matrices.

We considered two choices for the sample size $n = 500, 1000$. For each subject, we simulated $m_i = 3$ replicates. The true density of $\mathbf{X}$ was chosen to be

$f_{\mathbf{X}}(\mathbf{X}) = \sum_{k=1}^{K_{\mathbf{X}}} \pi_{\mathbf{X},k} \text{MVN}_p(\mathbf{X} | \boldsymbol{\mu}_{\mathbf{X},k}, \boldsymbol{\Sigma}_{\mathbf{X},k})$ with $p = 4$, $K_{\mathbf{X}} = 3$, $\boldsymbol{\pi}_{\mathbf{X}} = (0.25, 0.50, 0.25)^T$, $\boldsymbol{\mu}_{\mathbf{X},1} = (0.8, 6, 4, 5)^T$, $\boldsymbol{\mu}_{\mathbf{X},2} = (2.5, 4, 5, 6)^T$ and $\boldsymbol{\mu}_{\mathbf{X},3} = (6, 4, 2, 4)^T$. For the density of the measurement errors $f_{\mathbf{e}}$ we considered two choices, namely

1. $f_{\mathbf{e}}^{(1)}(\mathbf{e}) = \text{MVN}_p(\mathbf{e} | \mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{e}})$, and
2. $f_{\mathbf{e}}^{(2)}(\mathbf{e}) = \sum_{k=1}^{K_{\mathbf{e}}} \pi_{\mathbf{e},k} \text{MVN}_p(\mathbf{e} | \boldsymbol{\mu}_{\mathbf{e},k}, \boldsymbol{\Sigma}_{\mathbf{e},k})$ with $K_{\mathbf{e}} = 3$, $\boldsymbol{\pi}_{\mathbf{e}} = (0.2, 0.6, 0.2)^T$, $\boldsymbol{\mu}_{\mathbf{e},1} = (-0.3, 0, 0.3, 0)^T$, $\boldsymbol{\mu}_{\mathbf{e},2} = (-0.5, 0.4, 0.5, 0)^T$ and $\boldsymbol{\mu}_{\mathbf{e},3} = -(\boldsymbol{\pi}_{\mathbf{e},1} \boldsymbol{\mu}_{\mathbf{e},1} + \boldsymbol{\pi}_{\mathbf{e},2} \boldsymbol{\mu}_{\mathbf{e},2}) / \boldsymbol{\pi}_{\mathbf{e},3}$.

For the component specific covariance matrices, we set $\boldsymbol{\Sigma}_{\mathbf{X},k} = \mathbf{D}_{\mathbf{X}} \boldsymbol{\Sigma}_{\mathbf{X},0} \mathbf{D}_{\mathbf{X}}$ for each k , where $\mathbf{D}_{\mathbf{X}} = \text{diag}(0.75^{1/2}, \dots, 0.75^{1/2})$. Similarly, $\boldsymbol{\Sigma}_{\mathbf{e},k} = \mathbf{D}_{\mathbf{e}} \boldsymbol{\Sigma}_{\mathbf{e},0} \mathbf{D}_{\mathbf{e}}$ for each k , where $\mathbf{D}_{\mathbf{e}} = \text{diag}(0.3^{1/2}, \dots, 0.3^{1/2})$. For each pair of $f_{\mathbf{X}}$ and $f_{\mathbf{e}}$, we considered four types of covariance structures for $\boldsymbol{\Sigma}_{\mathbf{X},0} = \{(\sigma_{ij}^{\mathbf{X},0})\}$, and $\boldsymbol{\Sigma}_{\mathbf{e},0} = \{(\sigma_{ij}^{\mathbf{e},0})\}$, namely

1. Identity (I): $\boldsymbol{\Sigma}_{\mathbf{X},0} = \boldsymbol{\Sigma}_{\mathbf{e},0} = \mathbf{I}_p$
2. Latent Factor (LF): $\boldsymbol{\Sigma}_{\mathbf{X},0} = \boldsymbol{\Lambda}_{\mathbf{X}} \boldsymbol{\Lambda}_{\mathbf{X}} + \boldsymbol{\Omega}_{\mathbf{X}}$, with $\boldsymbol{\Lambda}_{\mathbf{X}} = (0.7, \dots, 0.7)^T$ and $\boldsymbol{\Omega}_{\mathbf{X}} = \text{diag}(0.51, \dots, 0.51)$, and $\boldsymbol{\Sigma}_{\mathbf{e},0} = \boldsymbol{\Lambda}_{\mathbf{e}} \boldsymbol{\Lambda}_{\mathbf{e}} + \boldsymbol{\Omega}_{\mathbf{e}}$, with $\boldsymbol{\Lambda}_{\mathbf{e}} = (0.5, \dots, 0.5)^T$ and $\boldsymbol{\Omega}_{\mathbf{e}} = \text{diag}(0.75, \dots, 0.75)$,
3. Autoregressive (AR): $\sigma_{ij}^{\mathbf{X},0} = 0.7^{|i-j|}$ and $\sigma_{ij}^{\mathbf{e},0} = 0.5^{|i-j|}$ for each (i, j) , and
4. Exponential (EXP): $\sigma_{ij}^{\mathbf{X},0} = \exp(-0.5|i-j|)$ and $\sigma_{ij}^{\mathbf{e},0} = \exp(-0.9|i-j|)$ for each (i, j) .

The parameters were chosen to produce a wide variety of one and two dimensional marginal densities, see Figure 4 and also Figure 6. Scale adjustments by multiplication with $\mathbf{D}_{\mathbf{X}}$ and $\mathbf{D}_{\mathbf{e}}$ were done so that the simulated values of each component of $\mathbf{X}$ fall essentially in the range $(-2, 6)$ and the simulated values of all components of $\mathbf{e}$ fall essentially in the range $(-3, 3)$. For conditionally heteroscedastic measurement errors, we set the true variance functions at $s_{\ell}^2(X) = (1 + X/4)^2$ for each component ℓ . A total of 16 ($2 \times 1 \times 2 \times 4$) cases were thus considered for both independent and conditionally heteroscedastic measurement errors.

We first discuss the results of the simulation experiments when the measurement errors $\mathbf{U}$ were independent of $\mathbf{X}$. The estimated MISEs are presented in Table 1. When the true $f_{\mathbf{e}}$ was

a single component multivariate normal, the MLFA model produced the lowest MISE when the true covariance matrices were diagonal. In all other cases the MIW model produced the best results. When the true $f_{\mathbf{e}}$ was a mixture of multivariate normals, the model complexity increases and the performance of the MIW model started to deteriorate. In this case, the MLFA model dominated the MIW model when the true covariance matrices were either diagonal or had a latent factor characterization.

The estimated MISEs for the cases when $\mathbf{U}$ were conditionally heteroscedastic are presented in Table 2. Models that accommodate conditional heteroscedasticity are significantly more complex compared to models that assume independence of the measurement errors from $\mathbf{X}$. The numerically more stable MLFA model thus out-performed the MIW model in all 32 cases. The improvements were particularly significant when the true covariance matrices were sparse and the number of subjects was small ($n = 500$). The true and estimated univariate and bivariate marginals of $f_{\mathbf{X}}$ produced by the MIW and the MLFA methods when the true density of the scaled errors was a mixture of multivariate normals ($f_{\mathbf{e}}^{(2)}$) and the component specific covariance matrices were diagonal (I) are summarized in Figure 3 and Figure 4, respectively. The true and estimated univariate and bivariate marginals for the density of the scaled errors $f_{\mathbf{e}}$ for this case produced by the two methods are summarized in Figure 5 and Figure 6, respectively. The true and the estimated variance functions produced by the univariate submodels are summarized in Figure 2. Comparisons between Figure 3 and Figure 4 illustrate the limitations of the MIW models in capturing high dimensional sparse covariance matrices and the improvements that can be achieved by the MLFA models. The estimates of $f_{\mathbf{e}}$ produced by the two methods are in better agreement. This may be attributed to the fact that many more residuals are available for estimating $f_{\mathbf{e}}$ than there are $\mathbf{X}_i$'s to estimate $f_{\mathbf{X}}$. Figure 2 in the main paper and Figures S.7 and S.16 in the Supplementary Material show that the univariate submodels can recover the true variance functions well. Additional figures when the true covariance matrices had auto-regressive structure (AR) are presented in the Supplementary Material. In this case the true covariance matrices were not sparse. The MLFA method still vastly dominated the MIW method when the sample size was small ($n = 500$). When the sample size was large ($n = 1000$) the two methods produced comparable results.

The proposed deconvolution methods, in particular the MLFA method, are highly scalable. In small scale simulations, not reported here, we tried $p = 6, 8$ and 10 and observed good empirical performance. We have focused here on $p = 4$ dimensional problems since with $p = 4$ the numbers of univariate and bivariate marginals, $p = 4$ and $\binom{p}{2} = 6$, remain manageable and the results are conveniently graphically summarized.

Additional small scale simulations for a variety of other distributions with similar MISE patterns are presented in the Supplementary Material.

7 Example

Dietary habits are known to be leading causes of many chronic diseases. Accurate estimation of the distributions of dietary intakes is thus important in nutritional epidemiologic

surveillance and epidemiology. Nutritionists are typically interested not just in the consumption patterns of individual dietary components but also in their joint consumption patterns. By the very nature of the problem, $\mathbf{X}$, the average long term daily intakes of the dietary components, can never be directly observed. Data are thus typically collected from a representative sample of the population in the form of dietary recalls, the subjects participating in the study remembering and reporting the type and amount of food they had consumed in the past 24 hours. The problem of estimating the joint consumption pattern of the dietary components from the contaminated 24-hour recalls then becomes a problem of multivariate density deconvolution.

A large scale epidemiologic study conducted by the National Cancer Institute, the Eating at America's Table (EATS) study (Subar, 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 ($\mathbf{W}_{ij}$'s) were recorded. The goal is to estimate the joint consumption patterns of the true daily intakes ($\mathbf{X}_i$'s).

To illustrate our methodology, we consider the problem of estimating the joint consumption pattern of four dietary components, namely (a) carbohydrate, (b) fiber, (c) protein and (d) a mineral potassium. Figure 7 shows the plots of subject-specific means versus subject-specific variances for daily intakes of the dietary components with the estimates of the variance functions produced by univariate submodels superimposed over them. As is clearly identifiable from this plot, conditional heteroscedasticity is a very prominent feature of the measurements errors contaminating the 24 hour recalls. The estimated univariate and bivariate marginal densities of average long term daily intakes of the dietary components produced by the MIW method and the MLFA method are summarized in Figure 8. The estimated univariate and bivariate marginal densities for the scaled errors are summarized in Figure 9. The estimated marginals of $\mathbf{X}$ produced by the two methods look quite different, while the estimated marginals of $\mathbf{e}$ are in close agreement. The estimated univariate and bivariate marginal densities of the long term intakes of the dietary components produced by the MIW model look irregular and unstable, whereas the estimates produced by the MLFA model look relatively more regular and stable. In experiments not reported here, we observed that the estimates produced by the MIW method were sensitive to the choice of the number of mixture components, but the estimates produced by the MLFA model were quite robust. The trace plots and the frequency distributions of the of the numbers of nonempty mixture components are summarized in Figures S.14 and S.15 in the Supplementary Material and provide some idea about the relative stability of the two methods. These observations are similar to that made in Section 6 for conditionally heteroscedastic measurement errors and sparse covariance matrices.

We next comment only on the estimates produced by the MLFA method assuming them to be closer to the truth. The estimates show that the long term daily intakes of the four dietary components are strongly correlated. The shapes of the bivariate consumption patterns suggest deviations from normality. Similarly, the shapes of the bivariate marginals for the scaled errors suggest that the measurement errors in the reported 24 hour recalls are positively correlated and deviate from normality. People who consume more are expected to do so for most dietary components. Strong correlations between the intakes of the dietary

components are thus somewhat expected. The correlations among different components of the measurement errors suggest that people usually have a tendency to either over-report or under-report the daily intakes. These findings illustrate the importance of robust but numerically stable multivariate deconvolution methods in nutritional epidemiologic studies.

Additional discussions on potentially far-reaching impact of our work on nutritional epidemiology studies are deferred to Section S.10 in the Supplementary Material.

8 Discussion

We considered the problem of multivariate density deconvolution when the measurement error density is not known but replicated proxies are available for some individuals. We used flexible finite mixtures of multivariate normal kernels with symmetric Dirichlet priors on the mixture probabilities to model both the density of interest and the density of the measurement errors. We proposed a novel technique to make the model for the density of the errors satisfy a zero mean restriction. We showed that the dense parametrization of inverse Wishart priors are not suitable for modeling covariance matrices in the presence of measurement errors. We proposed a numerically more stable approach based on latent factor characterization of the covariance matrices with sparsity inducing priors on the factor loading matrices. We built models for conditionally heteroscedastic additive measurement errors that also automatically accommodate multivariate multiplicative measurement errors.

The methodological contributions of this article are not limited to deconvolution problems. Mixtures of latent factor analyzers with sparsity inducing priors on the factor loading matrices can be used in other high dimensional applications including ordinary density estimation. The techniques proposed in Section 2.2.1 to enforce the mean zero moment restriction on the measurement errors can be readily used to model multivariate regression errors that are distributed independently of the predictors. The technique can also be adapted to relax the strong assumption of multivariate normality made by Hoff and Niu (2012) and Fox and Dunson (2016) in covariance regression problems.

As explained in Sections 2.2.2 and 2.2.3 in the main paper and also in Section S.5 in the Supplementary Material, the structural separability assumption (16) arises naturally in both additive and multiplicative multivariate measurement error settings. It would still be interesting, in future work, to consider more general covariance models that allow $\text{var}(U_{ij}|\mathbf{X})$ to be explained primarily by X_{ij} as in the current approach, but would allow the residual variability to be explained by the remaining components $\{X_{im}\}_{m \neq j}$ of $\mathbf{X}$. The current MCMC based implementation of the proposed methodology is computationally intensive. We are pursuing the development of faster algorithms for approximate posterior inference as the subject of a separate manuscript.

The question of consistency of Bayesian procedures is intimately related to the flexibility of the priors. For instance, in ordinary density estimation problems inclusion of the true density in the KL support of the prior is a sufficient condition to ensure weak consistency via the Schwartz theorem. In density deconvolution problems such a condition is not sufficient but is still required. The results from Section 5 thus provide crucial first steps in that direction.

We have not pursued the question of consistency of the proposed deconvolution methods any further in this article. It remains an important direction for future research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Pati's research was supported by Award N00014-14-1-0186 from the Office of Naval Research. Carroll and Mallick's research was supported by grant U01-CA057030 and by grant R01-CA194391, both from the National Cancer Institute. We acknowledge the Texas A&M University Brazos HPC cluster that contributed to the research reported here. We also thank the associate editor and the referees for their careful reviews.

References

- Bhattacharya A, Dunson DB. Sparse Bayesian infinite factor models. *Biometrika*. 2011; 98:291–306. [PubMed: 23049129]
- Bhattacharya A, Pati D, Pillai N, Dunson DB. Bayesian shrinkage. 2014 Unpublished manuscript.
- Brown PJ, Griffin JE. Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*. 2010; 5:171–188.
- Bovy J, Hogg DW, Rowies ST. Extreme deconvolution: inferring complete distribution functions from noisy, heterogeneous and incomplete observations. *Annals of Applied Statistics*. 2011; 5:1657–1677.
- Buonaccorsi JP. *Measurement Error: Models, Methods and Applications*. New York: Chapman and Hall/CRC; 2010.
- 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, Ruppert D, Stefanski LA, Crainiceanu CM. *Measurement Error in Nonlinear Models*. 2nd. Boca Raton: Chapman and Hall/CRC Press; 2006.
- Carvalho MC, Polson NG, Scott JG. The horseshoe estimator for sparse signals. *Biometrika*. 2010; 97:465–480.
- Comte F, Lacour C. Anisotropic adaptive density deconvolution. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*. 2013; 49:569–609.
- Devroye L. Consistent deconvolution in density estimation. *Canadian Journal of Statistics*. 1989; 17:235–239.
- Diggle PJ, Hall P. A Fourier approach to nonparametric deconvolution of a density estimate. *Journal of the Royal Statistical Society, Series B*. 1993; 55:523–531.
- Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Statistical Science*. 1996; 11:89–121.
- 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.
- Fokoué E, Titterton DM. Mixtures of factor analyzers. Bayesian estimation and inference by stochastic simulation. *Machine Learning*. 2003; 50:73–94.
- Fox EB, Dunson D. Bayesian nonparametric covariance regression. 2016 To appear in *Journal of Machine Learning Research*.
- Frühwirth-Schnatter S. *Finite Mixture and Markov Switching Models*. New York: Springer; 2006.

- Geweke J. Interpretation and inference in mixture models: Simple MCMC works. *Computational Statistics & Data Analysis*. 2007; 51:3529–3550.
- Hazelton ML, Turlach BA. Nonparametric density deconvolution by weighted kernel estimators. *Statistics and Computing*. 2009; 19:217–228.
- Hazelton ML, Turlach BA. Semiparametric density deconvolution. *Scandinavian Journal of Statistics*. 2010; 37:91–108.
- Hesse CH. Data driven deconvolution. *Journal of Nonparametric Statistics*. 1999; 10:343–373.
- Hoff PD, Niu X. A covariance regression model. *Statistica Sinica*. 2012; 22:729–753.
- Hu Y, Schennach S. Instrumental Variable Treatment of Nonclassical Measurement Error Models. *Econometrica*. 2008; 76:195–216.
- Li T, Vuong Q. Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis*. 1998; 65:139–165.
- Liu MC, Taylor RL. A consistent nonparametric density estimator for the deconvolution problem. *Canadian Journal of Statistics*. 1989; 17:427–438.
- Masry E. Multivariate probability density deconvolution for stationary random processes. *IEEE Transactions on Information Theory*. 1991; 37:1105–1115.
- Mengersen KL, Robert CP, Titterton DM, editors. *Mixtures - Estimation and Applications*. Chichester: John Wiley; 2011.
- Neumann MH. On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*. 1997; 7:307–330.
- Sarkar A, Mallick BK, Staudenmayer J, Pati D, Carroll RJ. Bayesian semiparametric density deconvolution in the presence of conditionally heteroscedastic measurement errors. *Journal of Computational and Graphical Statistics*. 2014; 23:1101–1125. [PubMed: 25378893]
- Schennach S. Nonparametric regression in the presence of measurement error. *Econometric Theory*. 2004; 20:1046–1093.
- 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.
- Subar 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*. 2001; 154:1089–1099. [PubMed: 11744511]
- Youndjé E, Wells MT. Optimal bandwidth selection for multivariate kernel deconvolution. *TEST*. 2008; 17:138–162.

Figure 1.

Dependency structures in trivariate deconvolution problems with (a) independently distributed and (b) conditionally varying measurement errors. (c) Dependency structure in a trivariate regression problem with response $\mathbf{Y}$, regression errors $\mathbf{U}$ and bivariate predictor $\mathbf{X}$. The filled rectangular regions focus on the relationship between the (potentially conditionally varying) errors $\mathbf{U}$ and the (corresponding conditioning) variable $\mathbf{X}$. The unfilled and the shaded nodes signify latent and observable variables, respectively. The directed and the undirected edges represent one and two-way relationships, respectively. The solid black and the dashed gray edges in panel (b) signify strong and weak dependencies, respectively.

Figure 2.

Results for conditional variability $\text{var}(U|X) = s^2(X)\text{var}(\epsilon)$ produced by the univariate density deconvolution method for each component of $\mathbf{X}$ for the conditionally heteroscedastic error distribution $f_{\epsilon}^{(2)}$ with sample size $n = 1000$, $m_i = 3$ replicates for each subject and identity matrix (I) for the component specific covariance matrices. The results correspond to the data set that produced the median of the estimated integrated squared errors (ISE) out of a total of 100 simulated data sets for the MLFA (mixtures of latent factor analyzers) method. For each component of $\mathbf{X}$, the true variance function is $s^2(X) = (1+X/4)^2$. See Section 2.2.2 and Section S.3 in the Supplementary Material for additional details. In each panel, the true (lighter shaded green lines) and the estimated (darker shaded blue lines) variance functions are superimposed over a plot of subject specific sample means vs subject specific sample variances. The figure is in color in the electronic version of this article.

Figure 3.

Results for $f_{\mathbf{X}}$ produced by the MIW (mixtures with inverse Wishart priors) method for the conditionally heteroscedastic error distribution $f_{\epsilon}^{(2)}$ with sample size $n = 1000$, $m_j = 3$ replicates for each subject and identity matrix (I) for the component specific covariance matrices. The results correspond to the data set that produced the median of the estimated integrated squared errors (ISE) out of a total of 100 simulated data sets. See Section 6 for additional details. The upper triangular panels show the contour plots of the true two dimensional marginal densities. The lower triangular diagonally opposite panels show the

corresponding estimates. The numbers i, j at the bottom right corners of the off-diagonal panels show that the marginal densities f_{X_i}, X_j are plotted in those panels. The diagonal panels show the true (lighter shaded green lines) and the estimated (darker shaded blue lines) one dimensional marginals. The figure is in color in the electronic version of this article.

Figure 4.

Results for the $f_{\mathbf{X}}$ produced by the MLFA (mixtures of latent factor analyzers) method for the conditionally heteroscedastic error distribution $f_{\epsilon}^{(2)}$ with sample size $n = 1000$, $m_i = 3$ replicates for each subject and identity matrix (I) for the component specific covariance matrices. The results correspond to the data set that produced the median of the estimated integrated squared errors (ISE) out of a total of 100 simulated data sets. See Section 6 for additional details. The upper triangular panels show the contour plots of the true two dimensional marginal densities. The lower triangular diagonally opposite panels show the

corresponding estimates. The numbers i, j at the bottom right corners of the off-diagonal panels show that the marginal densities f_{X_i, X_j} are plotted in those panels. The diagonal panels show the true (lighter shaded green lines) and the estimated (darker shaded blue lines) one dimensional marginals. The figure is in color in the electronic version of this article.

Figure 5.

Results for the density of the scaled measurement errors $f_{\mathbf{e}}$ produced by the MIW (mixtures with inverse Wishart priors) method for the conditionally heteroscedastic error distribution $f_{\mathbf{e}}^{(2)}$ with sample size $n = 1000$, $m_i = 3$ replicates for each subject and identity matrix (I) for the component specific covariance matrices. The results correspond to the data set that produced the median of the estimated integrated squared errors (ISE) out of a total of 100 simulated data sets. See Section 6 for additional details. The upper triangular panels show the contour plots of the true two dimensional marginal densities. The lower triangular

diagonally opposite panels show the corresponding estimates. The numbers i, j at the bottom right corners of the off-diagonal panels show that the marginal densities $f_{e_i|e_j}$ are plotted in those panels. The diagonal panels show the true (lighter shaded green lines) and the estimated (darker shaded blue lines) one dimensional marginals. The figure is in color in the electronic version of this article.

Figure 6.

Results for the density of the scaled measurement errors f_{ϵ} produced by the MLFA (mixtures of latent factor analyzers) method for the conditionally heteroscedastic error distribution $f_{\epsilon}^{(2)}$ with sample size $n = 1000$, $m_i = 3$ replicates for each subject and identity matrix (I) for the component specific covariance matrices. The results correspond to the data set that produced the median of the estimated integrated squared errors (ISE) out of a total of 100 simulated data sets. See Section 6 for additional details. The upper triangular panels show the contour plots of the true two dimensional marginal densities. The lower triangular diagonally

opposite panels show the corresponding estimates. The numbers i, j at the bottom right corners of the off-diagonal panels show that the marginal densities $f_{e_i^* e_j}$ are plotted in those panels. The diagonal panels show the true (lighter shaded green lines) and the estimated (darker shaded blue lines) one dimensional marginals. The figure is in color in the electronic version of this article.

Figure 7.

Estimated variance functions $\text{var}(U|X) = s^2(X)\text{var}(e)$ produced by the univariate density deconvolution method for each component of $\mathbf{X}$ for the EATS data set with sample size $n = 965$, $m_i = 4$ replicates for each subject. See Section 7 for additional details. The figure is in color in the electronic version of this article.

Figure 8.

Results for the EATS data set for the $f_{\mathbf{X}}$. The off-diagonal panels show the contour plots of two-dimensional marginals estimated by the MIW method (upper triangular panels) and the MLFA method (lower triangular panels). The numbers i, j at the bottom right corners of the off-diagonal panels show that the marginal densities f_{X_i, X_j} are plotted in those panels. The diagonal panels show the one dimensional marginal densities estimated by the MIW method (darker shaded blue lines) and the MLFA method (lighter shaded green lines). The figure is in color in the electronic version of this article.

Figure 9.

Results for the EATS data set for the density of the scaled errors $f_{\mathbf{e}}$. The off-diagonal panels show the contour plots of two-dimensional marginals estimated by the MIW method (upper triangular panels) and the MLFA method (lower triangular panels). The numbers i, j at the bottom right corners of the off-diagonal panels show that the marginal densities f_{e_i, e_j} are plotted in those panels. The diagonal panels show the one dimensional marginal densities

estimated by the MIW method (darker shaded blue lines) and the MLFA method (lighter shaded green lines). The figure is in color in the electronic version of this article.

Mean integrated squared error (MISE) performance of MLFA (mixtures of latent factor analyzers) and MIW (mixtures with inverse Wishart priors) density deconvolution models described in Section 2 of this article for **homoscedastic** errors compared with a naive method that ignores measurement errors for different measurement error distributions. The minimum value in each row is highlighted.

Table 1

True Error Distribution	Covariance Structure	Sample Size	MISE $\times 10^4$		
			MLFA	MIW	Naive
(a) Multivariate Normal	I	500	1.24	3.05	8.01
		1000	0.59	1.33	6.58
	LF	500	6.88	6.33	33.41
		1000	5.15	3.10	32.42
	AR	500	11.91	5.51	27.17
		1000	9.82	2.78	26.01
	EXP	500	7.15	4.40	17.82
		1000	5.46	2.19	17.40
(b) Mixture of Multivariate Normal	I	500	1.28	3.24	5.97
		1000	0.64	1.37	4.99
	LF	500	7.28	7.51	31.62
		1000	4.17	4.34	31.48
	AR	500	10.43	6.66	30.74
		1000	7.75	4.35	28.90
	EXP	500	7.16	5.18	17.85
		1000	4.87	2.66	17.26

Mean integrated squared error (MISE) performance of MLFA (mixtures of latent factor analyzers) and MIW (mixtures with inverse Wishart priors) density deconvolution models described in Section 2 of this article for **conditionally heteroscedastic** errors compared with a naive method that ignores measurement errors for different measurement error distributions. The minimum value in each row is highlighted.

Table 2

True Error Distribution	Covariance Structure	Sample Size	MISE $\times 10^4$		
			MLFA	MIW	Naive
(a) Multivariate Normal	I	500	2.53	19.08	10.64
		1000	1.15	9.43	9.14
	LF	500	11.46	34.21	21.33
		1000	5.78	15.98	20.75
	AR	500	17.11	30.83	36.44
		1000	10.77	12.46	36.37
	EXP	500	11.63	26.99	24.28
		1000	6.67	10.56	23.36
(b) Mixture of Multivariate Normal	I	500	2.79	22.17	20.16
		1000	1.38	10.55	19.39
	LF	500	13.39	35.67	43.43
		1000	7.50	20.86	43.28
	AR	500	18.27	35.70	75.26
		1000	12.06	16.64	77.55
	EXP	500	12.11	34.50	48.76
		1000	7.59	13.74	50.02