Phase Function Density Deconvolution with Heteroscedastic Measurement Error of Unknown Type

Linh Nghiem * Cornelis J. Potgieter
Department of Statistical Science, Southern Methodist University

Abstract

The empirical phase function was recently proposed by Delaigle & Hall (2016) as a tool for nonparametric density deconvolution. Their estimator was developed for the case of additive, homoscedastic measurement error. The present paper considers how the empirical phase function can be used in the case of heteroscedastic measurement error. A weighted empirical phase function (WEPF) is proposed where the weights are used to adjust for heteroscedasticity of measurement error. The properties of the WEPF estimator are considered and simulation results show that the weighting can result in large decreases in MISE when estimating the phase function. The estimation of the weights from replicate observations is also discussed. Finally, the construction of a deconvolution density estimator using the WEPF is compared to the heteroscedastic data deconvolution estimator of Delaigle & Meister (2008). The WEPF estimator proves to be competitive, especially when considering that it does not require any knowledge of the distribution of measurement error.

1 Introduction

Assume that it is of interest to estimate the density function \( f_X(x) \) of a random variable \( X \), but that one is only able to observe contaminated versions of \( X \), say \( W = X + \varepsilon \), where \( \varepsilon \) represents measurement error. The problem of estimating the density function of \( X \) based on an observed sample \( W_1, W_2, \ldots, W_n \) with \( W_i = X_i + \varepsilon_i, i = 1, \ldots, n \) is known as density deconvolution. Here, the \( X_i \) are an iid sample from a distribution with density \( f_X \), and \( \varepsilon_i \) represents the measurement error of the \( i^{th} \) observation. The \( \varepsilon_i \) are both mutually independent and independent of the \( X_i \).

The work presented here builds on the seminal paper of Delaigle and Hall (2016), who developed methodology to estimate the density function \( f_X \) without parametric assumptions.

*Corresponding Author: lnghiem@mail.smu.edu
for the measurement error, and also without the need for replicate observations. Their method makes use of the phase function in estimating the true density $f_X$. The method is based on the minimal assumption on the measurement error: the measurement error terms $\varepsilon_i$ are only assumed to be symmetric around 0, to be independent of the $X_i$, to have equal variance, and to have a non-negative characteristic function. The present work investigates how the phase function approach can be used for density deconvolution when the measurement error is heteroscedastic in nature.

The nonparametric density deconvolution problem was first considered by [Carroll and Hall (1988)] and [Stefanski and Carroll (1990)], who assumed that the distribution of the measurement error was fully known. The development that followed in the literature mostly considered the case of known measurement error, and generally treated the measurement error as homoscedastic, see [Fan (1991a), Fan (1991b), Fan and Truong (1993), Hall and Qiu (2005), and Lee et al. (2010)]. The case of heteroscedastic measurement error was considered by [Fan (1992)] and [Delaigle and Meister (2008)]. The problem of the measurement error having an unknown distribution was considered by [Diggle and Hall (1993)] and [Neumann and Hössjer (1997)], who assume that samples of error data are available, and by [Delaigle et al. (2008)] who use replicate data to estimate the entire characteristic function of the measurement error. [McIntyre and Stefanski (2011)] considered the heteroscedastic case with replicate observations. Their work assumed the measurement errors all follow a normal distribution with unknown variances only. The phase function deconvolution approach developed by [Delaigle and Hall (2016)] is groundbreaking in that they estimate the density function $f_X$ with unknown measurement error and without the need for replicate data. [Kato and Sasaki (2016)] also considered measurement error with unknown distribution, and constructed uniform confidence bands for the density function.

The model considered in this paper assumes the observed data are of the form $W_i = X_i + \sigma_i\varepsilon_i$ where the $X_i$ are an iid sample from $f_X$, the measurement error terms $\varepsilon_i$ are independent and each $\varepsilon_i$ has a symmetric distribution with non-negative characteristic function and satisfies $E(\varepsilon_i) = 0$ and $\text{Var}(\varepsilon_i) = 1$. The $\sigma_i$ are non-negative constants and represent measurement error heteroscedasticity. Specifically, $\text{Var}(W_i) = \sigma^2_X + \sigma^2_i$ where $\sigma^2_X$ denotes the variance of $X$. Additionally, it is assumed that the random variable $X$ does not have a symmetric component. Specifically, there is no symmetric random variable $U$ for which $X$ can be decomposed as $X = X_0 + U$ for arbitrary random variable $X_0$. This indecomposability is required to make the distribution of $X$ identifiable using the phase function – see [Delaigle and Hall (2016)] for a detailed discussion on how reasonable this assumption is in practice.

This paper will consider estimation of the density function $f_X$ in this heteroscedastic setting. Note that the heteroscedasticity of the measurement error will require either that the constants $\sigma_i$ be known, or that there are replicate data so that the $\sigma_i$ can be estimated from the data. The paper is organized as follows. Section 2 considers estimation of the phase function and introduces a weighted empirical phase function (WEPF) which adjusts for heteroscedasticity in the data. A small simulation study compares two different weighting schemes. Section 3 shows how the WEPF can be inverted to estimate the density function $f_X$ and presents an approximation of the asymptotic mean integrated squared error for se-
lecting the bandwidth. The WEPF deconvolution estimator is compared to that of Delaigle and Meister (2008), who treat the heteroscedastic case with known measurement error distribution. Section 4 illustrates the method using data from the Framingham Heart Study and Section 5 has some concluding remarks.

2 Phase Function Estimation

2.1 The Weighted Empirical Phase Function (WEPF)

The phase function of a random variable $X$, denoted $\rho_X(t)$, is defined as the characteristic function of $X$ standardized by its norm,

$$\rho_X(t) = \frac{\phi_X(t)}{|\phi_X(t)|}$$

with $\phi_X(t)$ the characteristic function of $X$ and $|z| = (z\bar{z})^{1/2}$ denoting the norm function with $\bar{z}$ the complex conjugate of $z$. Let $W = X + \sigma \varepsilon$ with $\varepsilon$ symmetric about 0 and characteristic function $\phi_\varepsilon(t) \geq 0$ for all $t$. It is easy to verify that the random variables $W$ and $X$ have the same phase function, $\rho_W(t) = \rho_X(t)$. Delaigle and Hall (2016) use this relation and an empirical estimate of $\phi_W(t)$ in equation (1) to estimate the phase function, see their paper for details on implementation.

In the case of heteroscedastic errors, we propose to use a weighted empirical phase function (WEPF) to adjust for heteroscedasticity. Define function

$$\hat{\phi}_W(t|q) = \sum_{j=1}^{n} q_j \exp(itW_j)$$

where $q = \{q_1, \ldots, q_n\}$ denotes a set of non-negative constants that sum to 1. Function (2) is a weighted empirical characteristic function and noting random variable $W_i = X_i + \sigma_i \varepsilon_i$ has characteristic function $\phi_{W_i}(t) = \phi_X(t)\phi_{\varepsilon_i}(\sigma_i t)$, $i = 1, \ldots, n$, it follows that

$$\mathbb{E}[\hat{\phi}_W(t|q)] = \phi_X(t) \sum_{j=1}^{n} q_j \phi_{\varepsilon_j}(\sigma_j t).$$

The WEPF is defined as

$$\hat{\rho}_W(t|q) = \frac{\hat{\phi}_W(t|q)}{|\hat{\phi}_W(t|q)|} = \frac{\sum_j q_j \exp(itW_j)}{\left(\sum_j \sum_k q_j q_k \exp(it(W_j - W_k))\right)^{1/2}}.$$  \hspace{1cm} (3)

For $q_{eq} = \{1/n, \ldots, 1/n\}$, $\hat{\rho}_W(t|q_{eq})$ essentially reduces to the phase function proposed by Delaigle and Hall (2016). Use of weights choice $q_{eq}$ will be referred to as the empirical phase
function (EPF) estimator. Other choices of weights can serve as an adjustment for heteroscedasticity—observations with large measurement error variance can be down-weighted to have smaller contribution to the phase function estimate.

The WEPF given in (3) is an asymptotically unbiased estimator of the phase function of $X$. The asymptotic variance (as a function of $t$) of the WEPF follows from a standard application of the functional delta method and the asymptotic normality follows from the Lindeberg-Feller central limit theorem applied to $n^{1/2}[\hat{\phi}_W(t|q) - \phi_W(t)]$ provided, for example, $\max_j q_j \to 0$ as $n \to \infty$. The expression for the asymptotic variance is available in a longer version of the paper, but is only informative in showing that the asymptotic variance of $\hat{\rho}_W(t|q)$ depends on $\phi_{\epsilon_j}(t)$ $j = 1, \ldots, n$, the characteristic functions of the measurement error components. While one would ideally like to choose weights $q$ that minimize said asymptotic variance, this is unrealistic as the method considered here makes no parametric assumptions about the measurement error. A much simpler weighting scheme is proposed here, relying only on knowledge of the measurement error variances.

Note that $E(W_i) = E(X) = \mu$. As such, for weights $q$, the estimator $\hat{\mu}_q = \sum_{j=1}^n q_j W_j$ is an unbiased estimator of $\mu$. The weights

$$q_i^* = \sigma_{W_i}^{-2} \left[ \sum_{j=1}^n \sigma_{W_j}^{-2} \right]^{-1} = (\sigma_X^2 + \sigma_i^2)^{-1} \left[ \sum_{j=1}^n (\sigma_X^2 + \sigma_j^2)^{-1} \right]^{-1}$$

result in a minimum variance estimator of $\mu$. This does have a connection to the phase function, as $\rho_X'(0) = \mu$—see the supplemental material of Delaigle & Hall, 2016 for the connection between the phase function and the odd moments of the underlying distribution. Define $q_{opt} = \{q_1^*, \ldots, q_n^*\}$ denote the vector of mean-optimal weights and let $\text{WEPF}_{opt}$ denote the weighted empirical phase function estimator calculated using the mean-optimal weights. Both the performance of the EPF and the $\text{WEPF}_{opt}$ will be considered for estimating the phase function and density function.

2.2 Estimating the Variance Components

In practice, it is often the case that neither the measurement error variances $\sigma_1^2, \ldots, \sigma_n^2$ nor $\sigma_X^2$ are known. These quantities can be easily estimated from replicate observations. This section describes how to estimate the variance components for a heteroscedastic measurement error variance model.

Consider replicate observations, $W_{ij} = X_i + \tau_i \epsilon_{ij}$, $j = 1, \ldots, n_i$, $i = 1, \ldots, n$ with $\min_i n_i \geq 2$. Note that $W_{ij} - W_{ij'} = \tau_i (\epsilon_{ij} - \epsilon_{ij'})$ and thus $E [(W_{ij} - W_{ij'})^2] = 2\tau_i^2$ for $j \neq j'$. Define grand mean

$$\bar{W} = \frac{1}{n} \sum_{i=1}^n \left[ \frac{1}{n_i} \sum_{j=1}^{n_i} W_{ij} \right] = \frac{1}{n} \sum_{i=1}^n X_i + \frac{1}{n} \sum_{i=1}^n \left[ \frac{\tau_i}{n_i} \sum_{j=1}^{n_i} \epsilon_{ij} \right]$$
and note that \( E(\bar{W}) = \mu \) and
\[
\text{Var}(\bar{W}) = \frac{\sigma_X^2}{n} + \frac{1}{n^2} \sum_{i=1}^{n} \frac{\tau_i^2}{n_i}.
\]

It can also be shown that
\[
E \left( (W_{ij} - W)^2 \right) = \sigma_X^2 + \tau_i^2 + o(n^{-1}).
\] (5)

Subsequently, estimates for the variance components are given by
\[
\hat{\tau}_i^2 = \frac{1}{n_i(n_i - 1)} \sum_{j=1}^{n_i-1} \sum_{j'=j+1}^{n_i} (W_{ij} - W_{ij'})^2, \quad i = 1, \ldots, n,
\]
and, motivated by (5),
\[
\hat{\sigma}_X^2 = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{n_i} (W_{ij} - \bar{W})^2 - \frac{1}{n} \sum_{i=1}^{n} \hat{\tau}_i^2
\]
with \( N = \sum_{i} n_i \). The analysis then proceeds by defining individual-level averages \( W_i = \frac{1}{n_i-1} \sum_{j=1}^{n_i} W_{ij} \) and noting that \( W_i = X_i + \sigma_i \epsilon_i \) where \( \sigma_i = \tau_i/\sqrt{n_i} \) and \( \epsilon_i \) has a distribution symmetric about 0 with a non-negative characteristic function. The estimate of \( \sigma_i \) is given by \( \hat{\sigma}_i = \hat{\tau}_i/\sqrt{n_i} \).

2.3 Simulation Study

A small simulation study was conducted to compare the performance of the EPF and WEPF\(_{opt}\) estimators. The true \( X_i \) data were sampled from the following three distributions, (1) \( X \sim \chi^2(3)/\sqrt{6} \) (Scaled \( \chi^2_3 \)), (2) \( X \sim (0.5N(1, 1) + 0.5\chi^2(5))/\sqrt{0.5} \) (Mixture 1), and (3) \( X \sim (0.5N(5, 0.6^2) + 0.5N(2.5, 1))/\sqrt{2.2425} \) (Mixture 2). In the simulations, the measurement error was taken to be normal in all instances, with error variance structures considered outlined in Table 2.

The first two distributions were used in simulation studies by [Delaigle and Hall (2016)] and are right-skewed; the third distribution is bimodal. All three distributions were scaled to have unit variance. The measurement error \( \epsilon_i \) was sampled from a normal distribution with mean 0 and variance structure given as Case 1 in Table 2. The measurement error variances were assumed known. For each candidate distribution of \( X \), \( N = 1000 \) \( W \) samples were generated with sample sizes \( n = 250, 500, \) and 1000. For each simulated dataset, \( q_{opt} \) was calculated using equation (4) and the WEPF\(_{opt}\) estimator was computed using these weights. Additionally, the EPF estimator was calculated. As the quality of the empirical characteristic function decreases with increasing \( t \), the suggestion of [Delaigle and Hall (2016)] was followed and the estimated phase functions were only computed on the interval \([-t^*, t^*]\), where \( t^* \) is the smallest \( t > 0 \) such that \( |\hat{\phi}_W(t|q)| < n^{-1/4} \).
Two performance criteria are used for evaluating the two weighting schemes. Firstly, the pointwise mean square error (MSE) ratios of the two estimators are plotted in Figure 1. This was done not only for the phase function, but also for its real and imaginary components individually. Note that at \( t = 0 \), all the phase function estimates are equal to 1 and the MSE ratio is undefined at \( t = 0 \). Also, the pointwise MSE ratio is symmetric in \( t \) and is therefore only plotted for \( t > 0 \). Secondly, mean integrated square error (MISE) ratios of the two weighting schemes were computed and are reported in Tables 1 and 3.

![Figure 1: The ratio MSE\(_{eq}/MSE_{opt}\), comparing the EPF and WEPF\(_{opt}\) estimators, sample size \( n = 500 \) and normal measurement errors with variance structure given in Case 1 (Table 2) using 1000 samples. The solid, dashed, and dotted lines represent respectively the MSE ratios for the phase function estimate as well as its real and imaginary parts.](image)

Figure 1 shows the pointwise MSE ratio of the EPF and WEPF\(_{opt}\) estimators as a function of \( t \). A MSE ratio greater than 1 indicates that the WEPF\(_{opt}\) has smaller MSE than the EPF estimator at that \( t \). For all the three underlying distributions, the mean-optimal weights result in better performance than equal weights. The improvement is quite substantial in the case of the scaled \( \chi^2_3 \) distribution and less so for the two mixture distributions.

| Distribution         | \( n = 250 \) | \( n = 500 \) | \( n = 1000 \) |
|----------------------|---------------|---------------|---------------|
| \( X \sim \chi^2(3)/\sqrt{6} \) | 1.1348        | 1.3407        | 1.7135        |
| \( X \sim \text{Mixture 1} \)    | 1.0824        | 1.1436        | 1.1280        |
| \( X \sim \text{Mixture 2} \)    | 1.0119        | 1.0262        | 1.0004        |

Table 1: The MISE ratio \( \text{MISE}_{eq}/\text{MISE}_{opt} \) when estimating the phase function of \( X \) with normal measurement error and variance structure given in Case 1 of Table 2 based on 1000 samples.

Table 1 gives the MISE ratios \( \text{MISE}_{eq}/\text{MISE}_{opt} \) of the EPF and WEPF\(_{opt}\) estimators for sample sizes \( n = 250, 500 \) and 1000. An MISE ratio greater than 1 indicates better performance of the WEPF\(_{opt}\) estimator compared to the EPF estimator. For the scaled \( \chi^2_3 \)
distribution and Mixture 1, WEPF\textsubscript{opt} performs better than the EPF, while their performance is nearly identical for Mixture 2.

Next, the effect of different underlying measurement error variance structures on the total MISE ratio of the equally weighted and mean-optimal WEPF was examined. The sample size was fixed at \( n = 1000 \) and the three different measurement error variance structures considered are outlined in Table 2. The ratios \( \text{MSE}_{eq}/\text{MSE}_{opt} \) based on 1000 simulated datasets are reported in Table 3.

| Case   | Variance Structure                                                                 |
|--------|-------------------------------------------------------------------------------------|
| Case 1 | \( \sigma_i^2 = 0.25\sigma_X^2, i = 1, \ldots, n/2 \) and \( \sigma_i^2 = 0.75\sigma_X^2, i = n/2 + 1, \ldots, n \) |
| Case 2 | \( \sigma_i^2 = (0.25 + 0.5i/n)\sigma_X^2, i = 1, \ldots, n \)                     |
| Case 3 | \( \sigma_i^2 = (0.025 + 0.95i/n)\sigma_X^2, i = 1, \ldots, n \)                  |

Table 2: Three measurement error variance models used in simulations.

| Model          | Case 1 | Case 2 | Case 3 |
|----------------|--------|--------|--------|
| \( X \sim \chi^2(3)/\sqrt{6} \) | 1.7135 | 1.0599 | 1.4202 |
| \( X \sim \text{Mixture 1} \)     | 1.1280 | 0.9997 | 0.9938 |
| \( X \sim \text{Mixture 2} \)     | 1.0004 | 1.0045 | 1.0022 |

Table 3: The effect of error variance spacing on the MISE ratio \( \text{MSE}_{eq}/\text{MSE}_{opt} \) based on 1000 samples, sample size \( n = 1000 \).

Inspection of Table 3 reveals that WEPF\textsubscript{opt} estimator results in a lower MISE than the EPF estimator. This holds true for the scaled \( \chi^2 \) distribution, with the measurement error structure only affecting the size of the improvement. In the case of Mixture 1 and Mixture 2, there is virtually no difference between the WEPF\textsubscript{opt} and EPF estimators. In most cases here, the MISE ratio is nearly equal to 1. These simulation results suggest that, when the measurement error variances are known, there is no downside to using the WEPF\textsubscript{opt}, as the resulting estimator tends to perform no worse than the EPF estimator, but at times large gains in efficiency are possible. Weighting to adjust for heteroscedasticity can result in a greatly improved estimator of the phase function. In the next section, this is explored in the context of density deconvolution. The case where the measurement error variance needs to be estimated from replicate data will also be addressed.
3 Density Estimation

3.1 Constructing an Estimator of \( f_X \)

Outline here is a brief overview of how the method of [Delaigle and Hall (2016)] can be implemented using the WEPF to estimate the density function \( f_X \). Let \( \hat{\phi}_W(t|q) \) and \( \hat{\rho}_W(t|q) \) denote the weighted characteristic function and corresponding WEPF respectively. Let \( w(t) \) denote a non-negative weight function that decreases as \( |t| \) increases. Also let \( x_j, j = 1, \ldots, m \) denote arbitrary atoms with respective probability masses \( p_j \). Delaigle and Hall (2016) suggest sampling the \( x_j \) uniformly on the interval \([\min W_i, \max W_i]\) with \( m = 5\sqrt{n} \). The goal is then to find the set \( \{p_j\}_{j=1}^m \) that minimizes

\[
T(p) = \int_{-\infty}^{\infty} \left| \hat{\phi}_W(t|q) - |\hat{\phi}_W(t|q)|^{1/2} \sum_j p_j \exp(itx_j) \right|^2 w(t) dt \tag{6}
\]

under the constraint of also minimizing the variance function \( v(p) = \sum_{j=1}^m p_j x_j^2 - (\sum_{j=1}^m p_j x_j)^2 \). This non-convex optimization problem of finding the set \( \{\hat{p}_j\}_{j=1}^m \) can be solved using MATLAB. Details are given in [Delaigle and Hall (2016)]. The present numerical implementation differs only in that the estimated phase function is weighted to adjust for heteroscedasticity, but the optimization problem remains unchanged.

The function \( \sum_j p_j \exp(itx_j) \) in (6) is the characteristic function of a discrete distribution with probability mass \( p_j \) at the point \( x_j \) for \( j = 1, \ldots, m \). It serves as an approximation to the characteristic function of \( X \) and estimating \( \hat{f}_X(x) \) therefore requires some smoothing. To this end, let \( \phi(t|\hat{p}) = \sum_j \hat{p}_j \exp(itx_j) \) be the characteristic function with the \( \hat{p}_j \)'s the probability masses estimated by minimizing (6). The deconvolution density estimator based on the WEPF is

\[
\hat{f}_X(x) = \frac{1}{2\pi} \int \exp(-itx) \tilde{\phi}(t) K^R(\sigma^2) dt \tag{7}
\]

where

\[
\tilde{\phi}(t) = \begin{cases} \phi(t|\hat{p}), & \text{for } t \leq t^* \\ r(t), & \text{for } t > t^* \end{cases}
\]

with \( t^* \) being the smallest \( t > 0 \) such that \( |\hat{\phi}_W(t|q)| < n^{-1/4} \). Here, \( K^R(t) \) denotes the Fourier transform of a deconvolution kernel function and \( r(t) \) denotes a ridging function. The ridging function ensures that the estimator is well-behaved outside the range \([-t^*, t^*]\). The proposed choice of ridging function is \( r(t) = \hat{\phi}_W(t|\hat{p})/\hat{\rho}_L(t) \), with \( \hat{\rho}_L(t) \) the characteristic function of a Laplace distribution with variance equal to an estimator of \( \sigma^2 = \sum_j q_j \sigma_j^2 \), the weighted sum of the measurement error variances. In application here, the common choice \( K^R(t) = (1 - t^2)^3 \) for \( |t| \leq 1 \) is used. As in [Delaigle and Hall (2016)], the weight function is chosen to be \( w(t) = \omega(t)|\sum_j p_j \exp(itx_j)| \) with \( \omega(t) \) the Epanechnikov kernel rescaled to the interval \([-t^*, t^*]\). This choice of weight function avoids numerical difficulties that can arise when dividing by very small numbers.
3.2 Bandwidth Selection

One can show that the phase function deconvolution estimator that accounts for heteroscedasticity as proposed in this paper is an approximation of the estimator

\[ \hat{f}_n(x) = \frac{1}{2\pi} \int \exp(-i t x) K^\text{ht}(ht) \frac{\hat{\phi}_W(t|q)}{\sum_j q_j \phi_{\epsilon_j}(\sigma_j t)} \, dt \]

with \( \hat{\phi}_W(t|q) \) defined in (2). Note that (8) is an estimator that one could compute if the measurement error distribution were known, but that it is different from the heteroscedastic estimator proposed by Delaigle and Meister (2008). Evaluation the integrated squared error (ISE) of (8), \( \text{ISE} = \int [\hat{f}_n(x) - f_X(x)]^2 \, dx \), and evaluating \( E[\hat{\phi}_W(t|q)] \) and \( E[|\hat{\phi}_W(t|q)|^2] \), it follows that the mean integrated squared error (MISE) is given by

\[ \text{MISE} = \frac{1}{2\pi} \int |\phi_X(t)|^2 [K^\text{ht}(ht) - 1]^2 \, dt + \frac{1}{2\pi} \int [K^\text{ht}(ht)]^2 \frac{\sum_j q_j^2}{\left( \sum_j q_j \phi_{\epsilon_j}(\sigma_j t) \right)^2} \, dt \]

\[- \frac{1}{2\pi} \int |\phi_X(t)|^2 [K^\text{ht}(ht)]^2 \frac{\sum_j q_j^2 \phi_{\epsilon_j}^2(\sigma_j t)}{\left( \sum_j q_j \phi_{\epsilon_j}(\sigma_j t) \right)^2} \, dt. \]

Using an argument similar to that of Delaigle and Meister (2008) when evaluating the asymptotic MISE (AMISE) of their heteroscedastic estimator, one can show that the last term of (9) is negligible, giving

\[ \text{AMISE} = \frac{1}{2\pi} \int |\phi_X(t)|^2 [K^\text{ht}(ht) - 1]^2 \, dt + \frac{1}{2\pi} \int [K^\text{ht}(ht)]^2 \frac{\sum_j q_j^2}{\left( \sum_j q_j \phi_{\epsilon_j}(\sigma_j t) \right)^2} \, dt \]

In the present application, both \( \phi_X(t) \) and \( \phi_{\epsilon_j}(t) \), \( j = 1, \ldots, n \) are unknown. However, note that \( |\phi_X(t)|^2 = \phi_X(t) \phi_X(-t) \) is the characteristic function of the random variable \( X - X' \), where \( X, X' \) are iid \( f_X \). Regardless of the shape of \( f_X \), the random variable \( X - X' \) is symmetric about 0 and has variance \( 2\sigma_X^2 \). This suggests replacing \( |\phi_X(t)|^2 \) with the characteristic function of a symmetric distribution with mean 0 and variance \( 2\sigma_X^2 \). Appropriate choices might be the normal distribution, i.e. substituting \( \exp(-\hat{\sigma}_X^2 t^2) \) for \( |\phi_X(t)|^2 \), or the Laplace distribution, i.e. substituting \((1 + \hat{\sigma}_X^2 t^2)^{-1}\). Additionally, one can use appropriate approximations for \( \phi_{\epsilon_j}(\sigma_j t) \). For example, the Laplace choice is a reasonable one, see Meister (2006) and Delaigle (2008). One can therefore substitute \((1 + 0.5\hat{\sigma}_j^2 t^2)^{-1}\) for \( \phi_{\epsilon_j}(\sigma_j t) \). This Normal-Laplace substitution gives approximate AMISE function

\[ \hat{\text{A}}(h) = \frac{1}{2\pi} \int \exp(-\hat{\sigma}_X^2 t^2) [K^\text{ht}(ht) - 1]^2 \, dt + \frac{1}{2\pi} \int [K^\text{ht}(ht)]^2 \frac{\sum_j q_j^2}{\left( \sum_j q_j (1 + 0.5\hat{\sigma}_j^2 t^2)^{-1} \right)^2} \, dt \]

(10)

and the value of \( h \) that minimizes the above function can then be used to evaluate the density deconvolution estimator in equation (7).
3.3 Simulation Study

Simulation studies were done to evaluate the performance of the equal-weighted and mean-optimal weighted phase function deconvolution density estimators. Additionally, as it is already established in the literature, the Delaigle and Meister (2008) estimator for heteroscedastic data was also calculated. The three candidate distributions for $X$ as described in Section 2.3 were considered. Both normal and Laplace distributions were considered for the measurement error, each in conjunction with the three measurement error variance models outlined in Table 2 being considered. In all cases the sample size was taken to be $n = 500$. Due to the computational cost of evaluating the phase function deconvolution estimators, a total of 500 samples were generated for each combination of $X$-distribution and variance model. For the phase-function estimators, the approximate AMISE bandwidth minimizing \( (10) \) was computed. The bandwidth of the Delaigle-Meister estimator was a two-stage plug-in bandwidth as suggested in their paper. For all the three deconvolution estimators, the integrated squared errors (ISE) was computed for each sample.

The simulation results are compiled in Table 4 with three columns corresponding to the setting where the measurement error variances are assumed known, while the last four columns correspond to the case with $J = 2$ replicates per observation and estimating the variance components as outlined in Section 2.2. The simulation with replicate observations contains results for the Delaigle-Meister estimator both using the estimated variances and treating the variances as known. Note that the simulations with replicate observations use the individual-level average data $W_i = (W_{i1} + W_{i2})/2$ to compute the deconvolution estimators and are therefore not directly comparable to the simulation without replication and measurement error variances assumed known. Due to the presence of outliers in the ISE calculations, the median as well as the first and third quartiles of $10 \times ISE$ are reported.

Inspection of the no-replicate simulation results (columns 4 through 6) reveals that the Delaigle-Meister (D&M) estimator tends to have the smallest median ISE, although there are a few instances in which the phase function estimators outperform the D&M estimator, notably for Mixture 2 and Laplace measurement error. It is also clear that calculating the mean-optimal weights is very advantageous in this setting, with the mean-optimally weighted estimator having smaller median ISE than the equally weighted estimator in all but one instance. Overall, one can conclude that the WEPF estimator performs very well and compares favorably to the D&M estimator, which benefits from specification of the measurement error distribution.

Inspection of simulation results with replicates present (columns 7 through 10) is very insightful. Note that the measurement error variances here are estimated based on only $J = 2$ replicates for each observation. As such, one might not expect good performance. However, the two phase function estimators perform very favorable when compared to the D&M estimator with known measurement error variances. The mean-optimally weighted estimator generally performs better than the equally weighted estimators in terms of median ISE, although there are two exceptions. It is interesting that weights estimated based on
only two replicates give such good performance. Also revealing is that the WEPF estimator performs significantly better than the D&M estimator with estimated variances, with the median ISE of the mean-optimally weighted estimator often reflecting more than a 50% reduction in median ISE when compared to the D&M counterpart.

4 Analysis of Framingham Data

In this section, the EPF and WEPF\textsubscript{opt} density deconvolution estimators are illustrated using a classical dataset in the deconvolution literature, a subset of the Framingham Heart Study. The data consists of several variables related to coronary heart disease for \( n = 1615 \) patients. For each patient, two measurements of long-term systolic blood pressure (SBP) were collected at each of two examinations. As per Carroll et al. (2006), let \( M_{ij} \) be the average of the two measurements at exam \( j \) for \( j = 1, 2 \), and let \( W_{ij} = \log(M_{ij} - 50) \). The \( W_{ij} \) are assumed to be related to true long-term SBP, \( X_i \) according to \( W_{ij} = Y_i + \sigma_i \varepsilon_{ij} \) with \( Y_i = \log(X_i - 50) \). Density deconvolution is therefore used to estimate the density on the \( Y \)-scale, \( \hat{f}_Y(y) \), after which it follows that \( \hat{f}_X(x) = (x - 50)^{-1}\hat{f}_Y[\log(x - 50)], \ x > 50 \).

For the SBP data, the EPF and WEPF\textsubscript{opt} were estimated, the latter with mean-optimal weights \( q_{opt} \) using variance components estimated as described in Section 2.2. For both the EPF and WEPF\textsubscript{opt}, deconvolution bandwidths were estimated using (10). These two estimators are shown in Figure 2 together with the Delaigle & Meister (2008) estimator using the same estimated variances and Laplace measurement error. (The D&M estimator was also calculated for normal measurement error and was nearly identical.) A naive kernel estimator of the data is also shown for comparative purposes. It can be seen that the WEPF\textsubscript{opt} and EPF deconvolution density estimators are similar, except in the left tail where the EPF estimator has a spurious bump. The two estimators based on phase functions suggest that the distribution of \( X \) may be multi-modal, while the D&M estimator is unimodal and positive skew.
Table 4: Density estimation for $n = 500$ with no replicates and with $J = 2$ replicates for each observation. The median, as well as first and third quartiles, $[Q_1, Q_3]$, of $10 \times$ ISE of density estimators under 500 simulations.
Discussion

This paper presents a method for phase density deconvolution with heteroscedastic measurement error of unknown type and builds on the work of Delaigle and Hall (2016) who considered the homoscedastic case. Two estimators are proposed, one using equally weighted observations and the other using mean-optimal weights to adjust for heteroscedasticity of the measurement error. A method based on approximating the AMISE is proposed for bandwidth selection in both instances. In the simulation settings considered, the WEPF$_{opt}$ estimator generally performed better than the EPF estimator, although there were instances where their performance was comparable. The simulation results suggest that mean-optimal weighting of observations will not have a detrimental effect on estimating the density function, and big gains are sometimes possible. When the measurement error variances are known, the method of Delaigle and Meister (2008) will outperform both phase function estimators, although the latter are still competitive in this setting. When there are only 2 replicates per individual from which to estimate the measurement error variances, the phase function methods performed substantially better than the Delaigle & Meister estimator. This suggests that the phase function methods have some inherent robustness against variance estimate deviation from the true values, and that the phase function density estimators can generally do the same as Delaigle and Meister (2008) estimator with much less assumption.
on measurement error.

References

Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83, 1184–1186.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: a modern perspective*. CRC press.

Delaigle, A. (2008). An alternative view of the deconvolution problem. *Statistica Sinica*, pages 1025–1045.

Delaigle, A. and Hall, P. (2016). Methodology for non-parametric deconvolution when the error distribution is unknown. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78, 231–252.

Delaigle, A., Hall, P., and Meister, A. (2008). On deconvolution with repeated measurements. *The Annals of Statistics*, pages 665–685.

Delaigle, A. and Meister, A. (2008). Density estimation with heteroscedastic error. *Bernoulli*, pages 562–579.

Diggle, P. J. and Hall, P. (1993). A fourier approach to nonparametric deconvolution of a density estimate. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 523–531.

Fan, J. (1991a). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 97–110.

Fan, J. (1991b). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, pages 1257–1272.

Fan, J. (1992). Deconvolution with supersmooth distributions. *Canadian Journal of Statistics*, 20, 155–169.

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

Hall, P. and Qiu, P. (2005). Discrete-transform approach to deconvolution problems. *Biometrika*, pages 135–148.

Kato, K. and Sasaki, Y. (2016). Uniform confidence bands in deconvolution with unknown error distribution. *arXiv preprint arXiv:1608.02251*.

Lee, M., Shen, H., Burch, C., and Marron, J. (2010). Direct deconvolution density estimation of a mixture distribution motivated by mutation effects distribution. *Journal of Nonparametric Statistics*, 22, 1–22.
McIntyre, J. and Stefanski, L. A. (2011). Density estimation with replicate heteroscedastic measurements. *Annals of the Institute of Statistical Mathematics*, 63, 81–99.

Meister, A. (2006). Density estimation with normal measurement error with unknown variance. *Statistica Sinica*, pages 195–211.

Neumann, M. H. and Hössjer, O. (1997). On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, 7, 307–330.

Stefanski, L. A. and Carroll, R. J. (1990). Deconvolving kernel density estimators. *Statistics*, 21, 169–184.