technische universität
dortmund

Bachelor Thesis

Deconvolution Problems in a Psychometric Context

Robin Grugel

February 17, 2020

Supervisor:

Prof. Dr. Philipp Doebler
Chair for Statistical Methods in Social Sciences
Faculty of Statistics
TU Dortmund University

Contents

1. Motivation

Estimating distributions of random variables is a difficult task, especially if the relevant variable is not directly observable and therefore latent. This problem is easily understood using intelligence test as an example. While those tests aim to assess the 'real intelligence' of a person, many factors can lead to deviations from the true score, which are difficult to measure, such as plain having a bad day. Given the classical test theory model $X = \tau + \epsilon$, with X as the observable indicator of the latent random variable τ and measurement error ϵ, the classic kernel density estimation isn't able to take ϵ into account.

This convolution of random variables is typically approached using characteristic functions, as the additive operation in the domain of density functions converts to the multiplication in the domain of characteristic functions and leads to the well studied procedure of deconvolving density estimation.

The central issue, which is aimed in this thesis, is the estimation of the error variance by exploiting the property of psychometric test models in using Cronbach's α. The focus is on the implementation of necessary procedures, simulation and evaluation for a variety of parameters and the application on data of an intelligence test.

The methodology of classical test theory, characteristic functions and their computational inversion, deconvolving kernel density estimation and related bandwidth selection will be briefly explained, before approaching the key issues.

The deconvolving density estimation performs very well for both basic models in comparison to the classic kernel density estimation, although the approach does not take differing error variances of τ-equivalence into account. There is no difference between the expected quality of those models regarding the accuracy, which confirms that Cronbach's α is a reasonable choice of obtaining an estimation for reliability. The procedure, despite equal precision, produces more reliable estimations for the error variance in parallel variables models.

The example of deconvolving estimation with data of intelligence research illustrates the benefit of density estimation in pointing out characteristics of distinctively shaped distribution, for examples skewness, including any measurement errors.

The benefit of this deconvolution approach is clearly highlighted and the method should be considered more frequently for further research projects and practical applications in a psychometric context.

2. Methodology

2.1. Classical Test Theory

The classical test theory of psychometric tests concerns stochastic measurement models of the type

$$X = \tau + \epsilon, \tag{1}$$

where X is the observable random variable or indicator, τ a latent random variable or true score and ϵ the measurement error. X is the indicator for τ as the only possible source of information about the latent factor that can be used for scientific inference. τ can also be considered as the expected value of intraindividual distribution of X, $\mathrm{E}(X|p_U = u)$ with p_U as a function that maps to index u of the individual observation for the set U of the entire population as described by Steyer and Eid (2013, p. 103). The assumption that τ and ϵ are stochastically independent is reasonable considering the idea that an error appears in the observable variable, regardless of the specific property that should be measured. Therefore you could define

$$\mathrm{Cov}(\tau, \epsilon) = 0, \tag{2}$$

which leads to the formula for the variance

$$\mathrm{Var}(X) = \mathrm{Var}(\tau + \epsilon) = \mathrm{Var}(\tau) + \mathrm{Var}(\epsilon) + 2\underbrace{\mathrm{Cov}(\tau, \epsilon)}_{=0} = \mathrm{Var}(\tau) + \mathrm{Var}(\epsilon), \tag{3}$$

for $\mathrm{Var}(X) < \infty$, so the variance does exist. This formulation highlights that observed variance can be divided in inter individual variation $\mathrm{Var}(\tau)$ and intra individual variation $\mathrm{Var}(\epsilon)$. Additionally, a common premise is

$$\mathrm{E}(\epsilon) = 0, \tag{4}$$

which assumes that the occurring measurement error has the expected value of zero and doesn't show any systematic behaviour. In case that τ is seen as the expected value, the above equations aren't assumptions rather than implications of the basic definition, meaning they can't be proven or falsified. For the used definition of τ as a non-observable random variable, it is necessary to make further assumptions for multiple indicators, tests or items.
Raykov and Marcoulides (2010, p. 122-123) report it is relevant to assume independence of two indicators or tests $X_1 = \tau_1 + \epsilon_1$ and $X_2 = \tau_2 + \epsilon_2$ regarding their measurement error by stating

$$\mathrm{Cov}(\epsilon_1, \epsilon_2) = 0 \tag{5}$$

and regarding their true scores similar to equation 2 by

$$\mathrm{Cov}(\epsilon_2, \tau_1) = 0 \text{ and } \mathrm{Cov}(\epsilon_1, \tau_2) = 0. \tag{6}$$

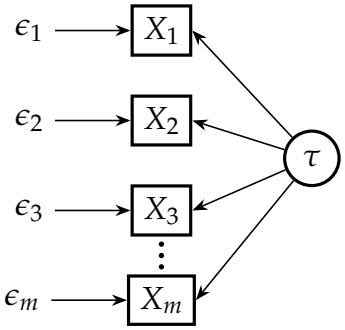

Figure 1: Path diagram of a typical measurement model of parallel or τ-equivalent variables

These assumptions are valid under the condition that there is no overlap in reference to the tested content. This should be especially considered by the researcher that designs and constructs measurement instruments. The basic goal in using the classical test theory model, is to make conclusions for the true score variable with help of indicators illustrated in figure 1.

2.1.1. Parallel Variables

The model of essentially parallel variables means that the variables $X_1 = \tau_1 + \epsilon_1$ and $X_2 = \tau_2 + \epsilon_2$ are independent in sense of equations 5 and 6, the variance of the errors are equal $\text{Var}(\epsilon_1) = \text{Var}(\epsilon_2)$ and the true score can be described by the equation

$$\tau_1 = \tau_2 + c \text{ , with } c \in \mathbb{R}. \tag{7}$$

If additionally $c = 0$ the variables are parallel instead of essentially parallel. This definition can be extended to a finite set of variables $X_1, ..., X_m$ for $m > 1$, so all variables are (essentially) parallel, if all variables are pairwise (essentially) parallel. The model of parallel variables means that all differences only are generated by the same measurement error, all instruments measure with the same precision. Essentially parallel variables have an additional scale parameter c as source of differences. In a specific application this model is more appropriate than the one of parallel variables. For example when performing tests at two distinct times, the learning effect can be taken into consideration.

2.1.2. τ-equivalent Variables

The model of essentially τ-equivalent variables generalizes the model of parallel variables in the sense of allowing different error variances for the variables. Therefore it is only necessary for $X_1 = \tau_1 + \epsilon_1$ and $X_2 = \tau_2 + \epsilon_2$ to be locally uncorrelated and the true score can be again described by the equation

$$\tau_1 = \tau_2 + c \text{ , with } c \in \mathbb{R}. \tag{8}$$

Here again for $c = 0$ the variables are τ-equivalent instead of essentially τ-equivalent. The generalization to more variables is analog to the formulation for parallel variables models. Unlike the parallel model the (essential) τ-equivalence takes different error variances into consideration, which seems useful because the assumption isn't necessarily valid. Therefore the restrictive assumptions of equally precise measurement tools can be relaxed and leads to more variety in designing measurement constructs.

2.1.3. Reliability

The reliability of a test X can be defined by the proportion of the variance of the true score τ to the variance of X by

$$\text{Rel}(X) = \frac{\text{Var}(\tau)}{\text{Var}(X)} = \frac{\text{Var}(\tau)}{\text{Var}(\tau) + \text{Var}(\epsilon)}, \tag{9}$$

with $\text{Var}(X) > 0$. It is easy to see, that for the reliability $\text{Rel}(X) \in [0, 1]$ holds true, because $\text{Var}(X) \geq \text{Var}(\tau)$. Practically the quotient gives information on how strongly the observed indicator is related to the true score. Keeping that in mind it makes sense to illustrate this idea in an alternative definition. Raykov and Marcoulides (2010, p. 138) express the reliability

$$\begin{aligned}
\text{Rel}(X) &= \frac{\text{Var}(\tau)}{\text{Var}(X)} = \frac{(\text{Var}(\tau))^2}{\text{Var}(X)\text{Var}(\tau)} && (10)\\
&= \frac{(\text{Cov}(\tau + \epsilon, \tau))^2}{\text{Var}(X)\text{Var}(\tau)} = \frac{(\text{Cov}(X, \tau))^2}{\text{Var}(X)\text{Var}(\tau)} && (11)\\
&= \left(\frac{\text{Cov}(X, \tau)}{\sqrt{\text{Var}(X)\text{Var}(\tau)}} \right)^2 && (12)\\
&= (\rho(X, \tau))^2 && (13)
\end{aligned}$$

as the squared correlation ρ between X and τ, which is also defined for finite and non zero variances. Additionally allowing $\text{Var}(\tau) = 0$ defining $\text{Rel}(X) = 0$, means no true variance, leads to the same exact reliability concept. The reliability in sense of internal consistency, using actual items, is mandatory for valid measurement concepts. A variety of reliability definitions for special test designs exists and is frequently used. Definition 9 leads directly to an alternative representation of the relationship between error variance and reliability

$$\text{Var}(\epsilon) = \text{Var}(X)(1 - \text{Rel}(X)) = \text{Var}(\tau) \left(\frac{1 - \text{Rel}(X)}{\text{Rel}(X)} \right), \tag{14}$$

which also can be seen as a function depending on the reliability and true score variance. Practically context the error variance is estimated by replacing $\text{Rel}(X)$ by an appropriate estimator $\widehat{\text{Rel}}(X)$. The relationship could also be used to obtain the error variance for a fixed reliability and true score variance in context of a

simulation study. In practice the reliability is obtained via different estimators, the most frequently used one is Cronbach's α. Cronbach (1951) and Guttman (1945) proposed the same estimator for the reliability for the sum $X = \sum_{i=1}^{m} X_i$ of m components

$$\alpha = \left(\frac{m}{m-1}\right)\left(1 - \frac{\sum_{i=1}^{m} \text{Var}(X_i)}{\text{Var}\left(\sum_{i=1}^{m} X_i\right)}\right). \tag{15}$$

Cronbach's α does not change, if X is replaced by $\overline{X}$. In practice $\widehat{\alpha}$ can be obtained by replacing the theoretical variance by the empirical version. For (essentially) τ-equivalent, and as a special case also for (essentially) parallel variables,

$$\alpha = \text{Rel}(X) \tag{16}$$

holds true. Only given the assumption of formula 5 α is a valid lower bound

$$\alpha \leq \text{Rel}(X). \tag{17}$$

Cronbach's α can still be useful in a context of single-factor models meaning τ-congeneric models pointed out by Raykov and Marcoulides (2017). In following considerations Cronbach's α will be used in equation 14 to estimate the variance of measurement errors.

2.2. Characteristic Functions

The characteristic function of a continuous univariate random variable X as a function $\varphi : \mathbb{R} \to \mathbb{C}$ is defined by Ushakov (1999, p. 1) as

$$\varphi_X(t) = \text{E}(\exp(iXt)) = \int_{-\infty}^{+\infty} \exp(ixt) f(x) dx \tag{18}$$

with $t \in \mathbb{R}$, $i = \sqrt{-1}$, and $f(x)$ the probability density function of X and for a discrete variable X as

$$\varphi_X(t) = \text{E}(\exp(iXt)) = \sum_{n=1}^{\infty} p_n \exp(iX_n t) \tag{19}$$

with $p_n = P(X = x_n)$ the corresponding probability measure for $n \in \mathbb{N}$. Therefore it exists a one-to-one relation between the probability density function and the characteristic function of a random variable. In mathematical context this is called Fourier transform and is extensively studied in the Fourier analysis. The characteristic function with the term $\exp(iXt)$ is oscillating and allows location shifts of length 2π and scaling. Euler's formula for the complex exponential function

$$\exp(iXt) = \cos(Xt) + i \cdot \sin(Xt) \tag{20}$$

reveals the periodicity of these oscillations, which is integrated or summed up in context of characteristic functions. Additionally Lukacs (1972) recognizes, that expected value exists for all bounded random variables like $\cos(Xt)$ and $\sin(Xt)$ are obviously bounded. For example Brémaud (2014) gives alternative definitions only differing in a factor 2π in the complex exponential function, which is valid considering the periodicity of it. As taught in basic probability courses and explained by Ushakov (1999, p. 3) the equation

$$\begin{aligned}\varphi_X(t) &= \mathrm{E}(\exp(iXt) = \mathrm{E}(\exp(i(\tau+\epsilon)t) && (21)\\ &= \mathrm{E}(\exp(i\tau t)\cdot\exp(i\epsilon t)) && (22)\\ &= \mathrm{E}(\exp(i\tau t))\cdot\mathrm{E}(\exp(i\epsilon t)) && (23)\\ &= \varphi_\tau(t)\cdot\varphi_\epsilon(t), && (24)\end{aligned}$$

is valid under the assumption of stochastic independence of τ and ϵ, which is the central reason for choosing estimation procedures using characteristic functions. Therefore it is possible to obtain the characteristic function $\varphi_\tau(t)$ through

$$\varphi_X(t) = \varphi_\tau(t)\cdot\varphi_\epsilon(t) \iff \varphi_\tau(t) = \frac{\varphi_X(t)}{\varphi_\epsilon(t)}, \tag{25}$$

which is only valid under the assumption, that $\varphi_\epsilon(t)$ does not vanish for any $t \in \mathbb{R}$. This property also holds true for the generalization to l stochastic independent random variables $X_1, ..., X_l$ in the sense that the sum transforms to the product of characteristic functions as follows:

$$\begin{aligned}\varphi_{\sum_{i=1}^{l} X_i}(t) &= \mathrm{E}\left(\exp\left(i\left(\sum_{i=1}^{l} X_i\right)t\right)\right) = \mathrm{E}\left(\exp\left(\sum_{i=1}^{l} iX_it\right)\right) && (26)\\ &= \mathrm{E}\left(\prod_{i=1}^{l}\exp(iX_it)\right) = \prod_{i=1}^{l}\mathrm{E}(\exp(iX_it)) && (27)\\ &= \prod_{i=1}^{l}\varphi_{X_i}(t). && (28)\end{aligned}$$

This equation is the source for simple convolution formulas used in basic probability classes. Any characteristic function $\varphi_X(t)$ of a random variable X satisfies all of the following conditions analysed by Lukacs (1972), directly obtained from the definition 18

$$\begin{aligned}&(a)\ \varphi \text{ is uniformly continous in } \mathbb{R} && (29)\\ &(b)\ |\varphi(t)| \leq \varphi(0) = 1 \text{ for all } t \in \mathbb{R} && (30)\\ &(c)\ \varphi_{aX+b}(t) = \exp(itb)\varphi_X(at) \text{ for constants } a \text{ and } b && (31)\\ &(d)\ \overline{\varphi(-t)} = \varphi(t) \text{ with } \overline{\varphi(-t)} \text{ as the complex conjugate.} && (32)\end{aligned}$$

It now can easily be seen, that the representation of distributions in characteristic functions is useful for many purposes. It can be of use in obtaining distributions of weighted and scaled sums of random variables, even if those are not identically distributed. It should also be noted, that characteristic functions always exist, even if the variance as the second moment does not. This is the case for the Cauchy distribution. Focussing on the typical distributions used in parametric statistics, especially in a psychometric context, the first candidate is the normal or Gaussian distribution. The normally distributed random variable X has a symmetric density function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \text{ for } x \in \mathbb{R}, \tag{33}$$

with expected value $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}^+$. The corresponding characteristic function φ_X is defined by Lukacs (1970, p. 18) as

$$\varphi_X(t) = \exp\left(i\mu t - \frac{\sigma^2 t^2}{2}\right). \tag{34}$$

Transforming a standard normally distributed random variable X with

$$\widetilde{X} = e^{\mu+\sigma X} \tag{35}$$

leads to the log-normally distributed $\widetilde{X}$ (Henze (2017, p. 308)) with parameters $\mu \in \mathbb{R}$ and $\sigma^2 \in \mathbb{R}^+$, as in the common normal case, which implies $\log(\widetilde{X}) \sim N(\mu, \sigma^2)$. The Log-Normal distribution has a density function

$$f_{\widetilde{X}}(\widetilde{x}) = \frac{1}{\widetilde{x}\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{\log(\widetilde{x})-\mu}{\sigma}\right)^2\right) \text{ for } \widetilde{x} \in (0, \infty), \tag{36}$$

which is skewed to the right with the expected value $\mathrm{E}(\widetilde{X}) = \exp\left(\mu - \frac{\sigma^2}{2}\right)$ and variance

$$\mathrm{Var}(\widetilde{X}) = (\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2). \tag{37}$$

The associated characteristic function results as

$$\varphi_{\widetilde{X}}(t) = \int_0^\infty \exp(it\widetilde{x}) \frac{1}{\widetilde{x}\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{\log(\widetilde{x})-\mu}{\sigma}\right)^2\right) d\widetilde{x} \tag{38}$$

and cannot be displayed in closed form, but Holgate (1989) concerns the issue of alternative representations in detail. Another property of Fourier transforms of valid density functions can be illustrated using the characteristic functions of the Log-Normal and Normal distribution. For symmetric density functions, the imaginary

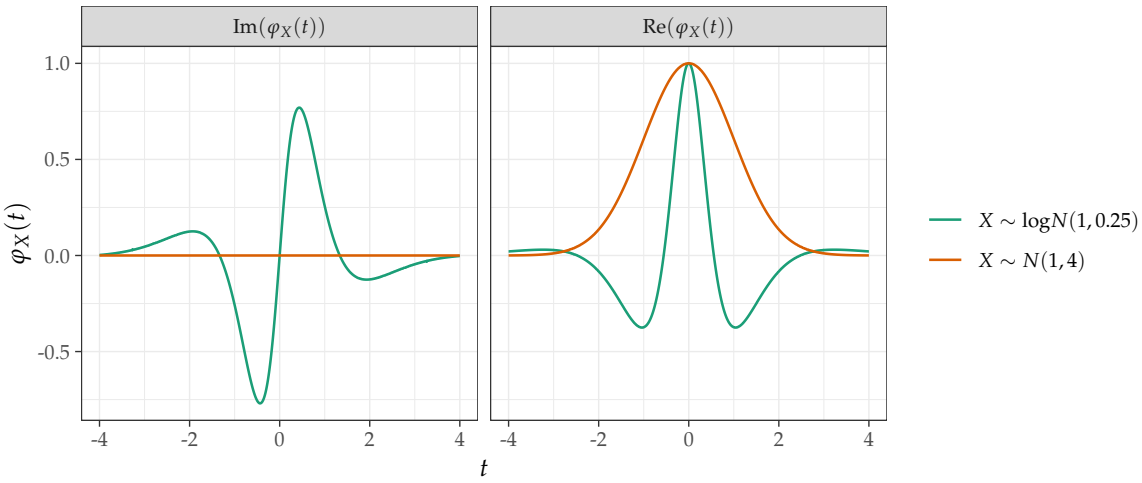

Figure 2: Example of the real and imaginary part of characteristic functions for $N(1,4)$- and $\log N(1,0.25)$-distributed random variable X

part of the corresponding characteristic function is always zero in comparison to asymmetric functions, which can be seen in figure 2.
Because of the one-to-one relationship between the probability density function and the characteristic function of a random variable in the sense of a Fourier transform it is possible to invert the characteristic function to a valid density. If a characteristic function $\varphi_X(t)$ is absolutely integrable, the belonging density function $f_X(x)$ can be defined by

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(ixt)\varphi_X(t)dt, \tag{39}$$

which is bounded and absolutely integrable as well. An important property concerning the inversion is the theorem of Parseval-Plancherel as defined in Ushakov (1999, p. 7), which states that for a characteristic function $\varphi_X(t)$ of an absolutely continuous distribution with density $f_X(x)$

$$|\varphi_X(t)|^2 \text{ is integrable} \iff f_X^2(x) \text{ is integrable} \tag{40}$$

holds true. It also implies

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |\varphi_X(t)|^2 dt = \int_{-\infty}^{\infty} f_X^2(x)dx, \tag{41}$$

which will be useful to derive formulas for the quality measure of density function estimators later.

For a detailed mathematical analysis of characteristic functions from a point of view of statistics and stochastic see Ushakov (1999) and Lukacs (1970).

2.3. Estimation Procedure

A simple estimation of the characteristic function φ_X for independently identically distributed random variables $X_1, ..., X_n$ is produced by the so-called empirical characteristic function

$$\widehat{\varphi}_X^{\text{naive}}(t) = \frac{1}{n} \sum_{j=1}^{n} \exp(iX_j t), \tag{42}$$

with $t \in \mathbb{R}$ and replacing with $x_1, ..., x_n$ in practice. This approach follows the common statistical practise of replacing the expectation under the assumption of underlying independently and identically distributed random variables $\exp(iX_1 t), ..., \exp(iX_n t)$ by the mean of the observed values. It already can be seen in figure 3, that the empirical characteristic function is not a good estimator for the true characteristic function because of different tail behaviour for large absolute values of t. In case of the desired probability density functions, respectively the characteristic functions, for $\varphi_\tau(t) = \frac{\varphi_X(t)}{\varphi_\epsilon(t)}$ under the assumption of a entirely known distribution of ϵ with characteristic function $\varphi_\epsilon(t)$ the estimator can be written as

$$\widehat{\varphi}_\tau(t) = \frac{\widehat{\varphi}_X(t)}{\varphi_\epsilon(t)} = \frac{1}{n} \sum_{j=1}^{n} \frac{\exp(iX_j t)}{\varphi_\epsilon(t)} \text{ for } \varphi_\epsilon(t) \neq 0 \ \forall t \in \mathbb{R}. \tag{43}$$

Meister (2009, p. 9-10) states, that $\widehat{\varphi}_\tau(t)$ is not integrable and has therefore no practical use for density estimation procedures, because no inverse Fourier transformat could be performed and therefore no density function could be obtained. An approach using possibilities of regularisation is necessary.

Choosing the classic kernel density estimation approach, the common estimator for a density function of a univariate random variable for identical independent distributed $X_1, ..., X_n$ is defined by Silverman (1990, p.45) as

$$\widehat{f}_X(x) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right). \tag{44}$$

The estimator represents a weighted sum of kernel functions K centred at every observed value only scaling with the parameter h. The bandwidth parameter $h \in \mathbb{R}^+$ ensures the smoothness of the estimator but also has to be determined separately. The function `density()` calculates the kernel density estimation in `R` with the commonly used optimal bandwidth

$$\widehat{h}_{\text{opt}} = 1.06 \sigma_X n^{-\frac{1}{5}} \tag{45}$$

and the gaussian kernel function

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right). \tag{46}$$

Silverman (1990) or Scott (1992) give further insights if an appropriate comparison is desired.

In the context of a deconvolution approach, the kernel function $K : \mathbb{R} \to \mathbb{R}$ must meet assumptions of integrability to lead to a valid characteristic function, respectively probability density function, but can conveniently be chosen from a set possible functions. The Fourier transform $\widehat{f}_X^{ft}(x)$ of $\widehat{f}_X(x)$ can then be carried out by

$$\widehat{\varphi}_X^{\text{kernel}}(t) = \widehat{f}_X^{ft}(t) \tag{47}$$

$$= \int_{-\infty}^{\infty} \exp(ixt) \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h}\right) dx \tag{48}$$

$$= \frac{1}{nh} \sum_{j=1}^{n} \int_{-\infty}^{\infty} \exp(ixt) K\left(\frac{x - X_j}{h}\right) dx \tag{49}$$

substitute $x = zh + X_j$ and $dx = hdz$

$$= \frac{1}{nh} \sum_{j=1}^{n} \int_{-\infty}^{\infty} \exp(i(zh + X_j)t) K(z) h dz \tag{50}$$

$$= \frac{1}{n} \sum_{j=1}^{n} \exp(iX_j t) \int_{-\infty}^{\infty} \exp(izht) K(z) dz \tag{51}$$

$$= \widehat{\varphi}_X^{\text{naive}}(t) \cdot \varphi_K(ht). \tag{52}$$

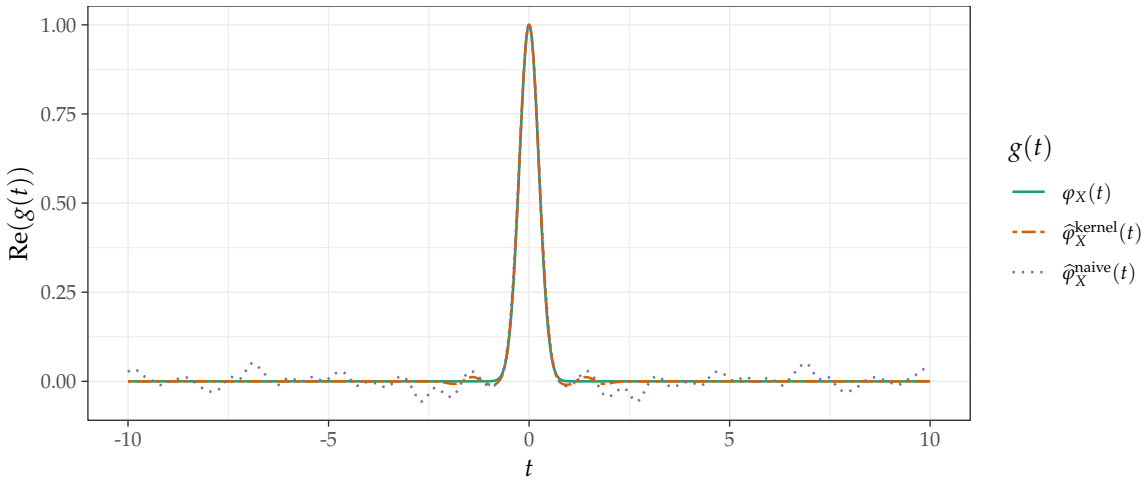

Figure 3: Real part of $\widehat{\varphi}_X^{\text{naive}}(t)$ and $\widehat{\varphi}_X^{\text{kernel}}(t)$ (with smoothing parameter $h = 0.35$) of the characteristic function for a sample of size $n = 1000$ from $X \sim N(1, 4)$ in comparison with true characteristic function from formula 34

φ_K is the Fourier transform of the chosen kernel function and $\widehat{\varphi}_X(t)$ the simple empirical characteristic function. The resulting estimator for the characteristic

function for τ can be written as

$$\widehat{\varphi}_\tau(t) = \frac{\widehat{\varphi}_X^{\text{naive}}(t) \cdot \varphi_K(ht)}{\varphi_\epsilon(t)}. \tag{53}$$

The kernel function has to be absolutely integrable or square-integrable. For this proof, it is sufficient to show that the Fourier transform of the kernel is bounded, which simplifies the choice for potential kernels in practical application. The actual proof can be examined in Stefanski and Carroll (1990) and further examined, regarding consistency concerns, with tools of the Fourier analysis provided by Meister (2009, p. 179-189). Choosing kernels among functions with bounded Fourier transforms leads to the deconvolving kernel estimator first introduced by Carroll and Hall (1988)

$$\widehat{f}_\tau(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-it\tau)\varphi_K(ht)\frac{1}{n}\sum_{j=1}^{n}\exp(iX_j t)\varphi_\epsilon(t)^{-1}dt. \tag{54}$$

Again the assumption of non vanishing and entirely known error density is necessary. $\widehat{f}_\tau(\omega)$ also takes complex values, which can simply be solved by only taking the real part. As shown in studies by Fan (1992) the choice of kernel function should additionally be made among functions whose Fourier transforms have a compact and symmetric support, also described by Delaigle and Gijbels (2004a). The simplest kernel $K^{\text{naive}}(x) = \frac{\sin(x)}{\pi x}$ with its Fourier transform $\varphi_{K^{\text{naive}}}(t) = \chi_{[-1,1]}(t)$ shows the integration bounds for estimator 54 with $t \in [-\frac{1}{h}, \frac{1}{h}]$, because following equation holds true:

$$\chi_{[-1,1]}(ht) = 1 \iff -1 \leq ht \leq 1 \iff -\frac{1}{h} \leq t \leq \frac{1}{h}. \tag{55}$$

The most frequently used kernel to avoid non-existence of the integral, leading to the density of the measurement error, is

$$K^{\text{imp}}(x) = \frac{48\cos x}{\pi x^4}\left(1 - \frac{15}{x^2}\right) - \frac{144\sin x}{\pi x^5}\left(2 - \frac{5}{x^2}\right), \tag{56}$$

which has the Fourier transform

$$\varphi_{K^{\text{imp}}}(t) = (1-t^2)^3\chi_{[-1,1]}(t). \tag{57}$$

In all further considerations and applications $K(t)$ will be seen as $K^{\text{imp}}(t)$ and therefore $\varphi_K(t)$ as $\varphi_{K^{\text{imp}}}(t)$.

Concerning the error density, Fan (1991a) classified distributions into groups of ordinary smooth and supersmooth to investigate convergence under those specific conditions, preventing too excessive oscillations in the tails of the characteristic functions. Therefore the distribution of the random variable ϵ is called ordinary smooth of order β, if the matching characteristic function $\varphi_\epsilon(t)$ meet the condition

$$b_0|t|^{-\beta} \leq |\varphi_\epsilon(t)| \leq b_1|t|^{-\beta} \text{ for } t \to \infty, \tag{58}$$

with positive real constants b_0, b_1 and β. On the other hand, the distribution of ϵ is called supersmooth, if $\varphi_\epsilon(t)$ satisfies the condition

$$b_0|t|^{\beta_0}\exp\left(\frac{|t|^\beta}{\gamma}\right) \leq |\varphi_\epsilon(t)| \leq b_1|t|^{\beta_1}\exp\left(\frac{|t|^\beta}{\gamma}\right) \text{ for } t \to \infty, \tag{59}$$

with positive real constants b_0, b_1, β, γ and real constants β_0 and β_1. A typical supersmooth density is belonging to the family of normal distributions, which was proved by Fan (1991b) to have a very slow rate of convergence. Meister (2009, p. 33-35) tackles this problem from a detailed mathematical point of view using the mean integrated squared error (MISE). In further considerations, simulations and practical applications, the distribution of the random measurement error will be assumed as normal and therefore supersmooth. Especially in assessing the results of simulation, the slow convergence rate should be kept in mind.

2.4. Numerical Inversion Method

The task of calculating the integral of the continuous Fourier transform (CFT) isn't easily achievable, because no analytical integration can be carried out via computation. Therefore a numerical solution has to be found. There is a variety of methods in classical numerical integration procedures, that are typically used. These approaches can't be applied meaningful on their own to the case of highly oscillating functions like the CFT. Therefore it is reasonable to take this property into account by considering the discrete Fourier transform (DFT) as an approximation. A truncation to a compact support $[a, b]$ should be selected first. In case of obtained distribution density functions it is very important to choose the truncation wisely to get a valid approximation. Given a characteristic function $\varphi_X(t)$ of the random variable X, the belonging density $f_X(x)$ can be approximated in the following manner:

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp(-itx)\varphi_X(t)dt \tag{60}$$

$$\approx \frac{1}{2\pi}\int_{b}^{a} \exp(-itx)\varphi_X(t)dt \tag{61}$$

$$= \frac{1}{2\pi} I[\varphi_X(t), a, b](w). \tag{62}$$

Inverarity (2002) describes a possible method to evaluate Fourier integrals of the type $I[f(t), a, b](w)$ and their inverse for non symmetric integration limits (support $[a, b]$) and different ranges of the transformed variables. Because of this flexibility, this method is useful for transformations from probability space to frequency space back and forth for varying distributions. The range $[c, d]$ of the transformed variable should be especially considered because for common continuous distributions the probability vanishes nowhere, but the probability mass typically tends to 0 for absolutely increasing truncations $\lim_{c,d\to\infty}[-|c|, |d|]$ in the tails of unimodal density

functions. The methods described by Bailey and Swarztrauber (1991) will be applied on transforms in the sense of probability measures. The procedure called the fractional Fourier transform (FRFT) concerns the case, where the input for Fourier integrals or DFTs is only partially non-zero and is able to speed up the conventionally used FFT, which is explained in detail by Bailey and Swarztrauber (1994). It uses a generalized form of the DFT. The DFT of a sequence of m elements g_j will lead as described by Inverarity (2002) to the output

$$F_k(g_j, m) = \sum_{j=0}^{m-1} g_j \exp\left(-\frac{2\pi ijk}{m}\right), \tag{63}$$

for $k = 0, ..., m-1$. The inverse DFT is defined as

$$F_k^{-1}(g_j, m) = \frac{1}{m} \sum_{j=0}^{m-1} g_j \exp\left(\frac{2\pi ijk}{m}\right), \tag{64}$$

only differing in a scaling factor $\frac{1}{m}$ and a different sign in the exponential function. The integral

$$I[f(t), a, b](w) = \int_a^b f(t) \exp(-iwt) dt \tag{65}$$

can now be approximated for $k = 1, ..., m$ by computing

$$I[f(t), a, b](w_k) \approx \beta \exp(-i(aw_k + \delta k^2)) F_k^{-1}(F_p(y_j, 2m) F_p(z_j(\delta, m), 2m), 2m), \tag{66}$$

for the following quantities. The already defined integration support $[a, b]$ and limits for the obtain transform variable space c and d are used in

$$\beta = \frac{b-a}{m} \text{ and } \gamma = \frac{d-c}{m}, \tag{67}$$

as the step size of the frequency respectively transform variable space. In preparation for the evaluation of the Fourier integral in style of the midpoint rule, the factor

$$\delta = \frac{\beta\gamma}{2} \tag{68}$$

is calculated from the previous quantities. The next step is to initialize the values for the frequency and transform variable space with

$$t_j = a + \beta j \text{ and } w_k = c + \gamma k, \tag{69}$$

to determine the desired range of values to display. The sequences of y_i and $z_j(\delta, m)$ for $j = 0, ..., 2m-1$ are defined by

$$y_j = \begin{cases} f(t_j) \exp(-ij(\beta c + \delta j)) & \text{for } j \in [0, m-1] \\ 0 & \text{for } j \in [m, 2m-1] \end{cases} \tag{70}$$

and

$$z_j(\delta, m) = \begin{cases} \exp(i\delta j^2) & \text{for } j \in [0, m-1] \\ \exp(i\delta(j-2m)^2) & \text{for } j \in [m, 2m-1], \end{cases} \tag{71}$$

to ensure, that the desired length of periodicity $2m$ is achieved and no indices will be summed up repeatedly. The DFT and their inverse can be easily computed by efficient algorithms called fast Fourier transforms (FFT). Therefore it is useful to define the resolution m as 2^l for $l \in \mathbb{N}$, to benefit from fast computation. For observed values that are in no suitable length, the input data can be padded with zeros to be able to exploit the properties of the FFT. The proposed procedure can easily be implemented in R. For an approximation of the inverse Fourier integral, the factor $\frac{1}{2\pi}$ must be considered. Furthermore, in a practical context of density estimation, the determination of c and d have to be performed appropriately. In the implemented version the decision for the limits was made in favour of $c =: \min(x_1, ..., x_n) - \widehat{\sigma}_x$ and $d =: \max(x_1, ..., x_n) + \widehat{\sigma}_x$ for an observed sample $x = x_1, ..., x_n$, to also display probability mass, which was estimated because of possibly large values of the chosen bandwidth h.

2.5. Fast Fourier Transform

The DFT can be calculated easily, but has a high complexity in numerical evaluation of $O(m^2)$ floating-point operations for m sampled elements that are used, because of reoccurring evaluations. The fast Fourier transform is a method to calculate the DFT that reduces the complexity to $O(m\log(m))$ and relies on the basic principle of dividing the sum of m elements into two sums of length $\frac{m}{2}$ like

$$F_k(g_j, m) = \sum_{j=0}^{m-1} g_j \exp\left(-\frac{2\pi ijk}{m}\right) \tag{72}$$

$$= \sum_{j=0}^{m/2-1} g_{2j} \exp\left(-\frac{2\pi i(2j)k}{m}\right) + \sum_{j=0}^{m/2-1} g_{2j+1} \exp\left(-\frac{2\pi i(2j+1)k}{m}\right) \tag{73}$$

$$= \underbrace{\sum_{j=0}^{m/2-1} g_j^{\text{even}} \exp\left(-\frac{2\pi ijk}{m/2}\right)}_{\text{DFT of even indexed elements}} + \exp\left(-\frac{2\pi ik}{m}\right) \underbrace{\sum_{j=0}^{m/2-1} g_j^{\text{odd}} \exp\left(-\frac{2\pi ijk}{m/2}\right)}_{\text{DFT of odd indexed elements}} \tag{74}$$

$$= F_k(g_j^{\text{even}}, m/2) + \exp\left(-\frac{2\pi ik}{m}\right) F_k(g_j^{\text{odd}}, m/2), \tag{75}$$

for $k = 0, ..., m-1$ and dividing into sums of even- and odd-indexed elements. Further, the property of the complex exponential function is exploited to reduce operations of sums. The periodicity means that

$$\exp\left(-\frac{2\pi i(k+m/2)}{m}\right) = \exp\left(-\frac{2\pi ik}{m} - \pi i\right) \tag{76}$$

$$= \exp\left(-\frac{2\pi ik}{m}\right)\exp(-\pi i) \tag{77}$$

$$= \exp\left(-\frac{2\pi ik}{m}\right)(\underbrace{\cos(-\pi)}_{=-1} + i \cdot \underbrace{\sin(-\pi)}_{=0}) \tag{78}$$

$$= -\exp\left(-\frac{2\pi ik}{m}\right), \tag{79}$$

which can be seen by applying Euler's formula 20 which leads to the representation

$$F_k(g_j, m) = F_k(g_j^{\text{even}}, m/2) + \exp\left(-\frac{2\pi ik}{m}\right) F_k(g_j^{\text{odd}}, m/2) \tag{80}$$

$$F_{k+\frac{m}{2}}(g_j, m) = F_{k+\frac{m}{2}}(g_j^{\text{even}}, m/2) - \exp\left(-\frac{2\pi ik}{m}\right) F_{k+\frac{m}{2}}(g_j^{\text{odd}}, m/2) \tag{81}$$

and illustrates the idea of further splitting up the sums in DFTs, that consist of pairs of transforms only differing in the sign. Therefore the calculation can be split down to the deepest level division by evaluating only those two quantities. This method is often called radix-2-procedure or Cooley-Tukey-Algorithm described in detail for the first time by Cooley and Tukey (1965). Gauss (1886) used the method circa 150 years earlier for astronomical calculations without analysing numerical properties and advantages. In figure 4 the approach is commonly illustrated for an example with a sample of size $m = 8$. The exploitation of the relationship in equation 80 is visualized by the direction of the lines on the right side of the figure. Horizontal and upright diagonal lines imply a positive sign whereas horizontal and downright diagonal lines imply a negative sign in the addition operation with the complex exponential quantity.

This illustration could be extended for a greater m to the deepest level by further splitting up to $\frac{m}{2}$ DFTs which calculate the DFT from only two elements. It makes therefore perfectly sense, that this method only applies best for $m = 2^l$ for $l \in \mathbb{N}$. The exponential factor can be precomputed to further reduce computational expenses. In case of division into $\frac{m}{2}$ DFTs in practical implementations the input has to be put in the right order. This can be achieved by applying the method of bit reverse seen in figure 1, reordering the bits of the binary representation of the indexes of input sequences.

It is easy to use the fast Fourier transform in practice because a variety of implementations mostly in `C`, `C++` and `FORTRAN` exists, which can be used in statistical purposes using `R` and calling the regarding function. In `R`'s `base`-package the function `fft()` uses the mixed radix implementation of Singleton (1979), which is a modification of the Cooley-Tukey-Algorithm. A variety of algorithms and their details are described by Rao et al. (2010).

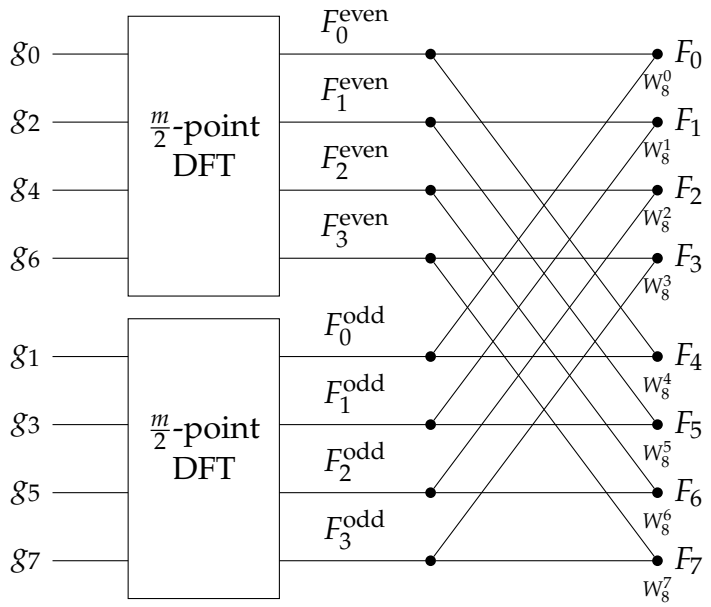

Figure 4: Example flow chart for a FFT using butterfly diagram for $m = 8$ elements; with abbreviations $F_k := F_k(g_j, m)$, $F_k^{\text{even}} := F_k(g_j^{\text{even}}, m/2)$, $F_k^{\text{odd}} := F_k(g_j^{\text{odd}}, m/2)$ and $W_m^k := \exp\left(-\frac{2\pi ik}{m}\right)$ for the factors

initial index	binary	reversed binary	new index
0	000	000	0
1	001	100	4
2	010	010	2
3	011	110	6
4	100	001	1
5	101	101	5
6	110	011	3
7	111	111	7

Table 1: Example for the bit reverse of input sequences indexes for $m = 8$ elements

2.6. Quality Measure and Adaptive Bandwidth Selection

In practice for finite data samples, the theoretical optimal bandwidth parameter based on an infinity large sample size might not lead to the best estimation of the distribution density function. In those cases, an adaptive bandwidth selection procedure is an approach for optimising the parameter by using the observed data for the optimization once, but also for the estimation itself.

A very common way to evaluate the quality of an estimated density function $\widehat{f}_\tau(\cdot, h)$ depending on the bandwidth h is the Mean Integrated Squared Error (MISE), which is defined as

$$\text{MISE}(\widehat{f}_\tau(\cdot;h)) = \text{E}\left(\int (\widehat{f}_\tau(\omega;h) - f_\tau(\omega))^2 d\omega\right) \tag{82}$$

$$= \int (\text{Bias}(\widehat{f}_\tau(\omega;h)))^2 d\omega + \int \text{Var}(\widehat{f}_\tau(\omega;h)) d\omega, \tag{83}$$

described in Delaigle and Gijbels (2004a), which provides a measure of nearness of the estimated to the targeted real density function. Using the deconvolving kernel density estimator in equation 54, including quantities concerning the error distribution of ϵ and selected kernel K and according characteristic function $\varphi_\tau(t)$ and $\varphi_\epsilon(t)$ respectively Fourier transform φ_K, Delaigle and Gijbels (2004a), using equation 41 write

$$\text{MISE}(\widehat{f}_\tau(\cdot;h)) = \frac{1}{2\pi n h} \int \frac{|\varphi_K(t)|^2}{|\varphi_\epsilon\left(\frac{t}{h}\right)|^2} dt \tag{84}$$

$$+ \frac{n-1}{2\pi n} \int |\varphi_\tau(t)|^2 |\varphi_K(ht)|^2 dt \tag{85}$$

$$+ R(f_\tau) - \frac{1}{\pi} \int |\varphi_\tau(t)|^2 \varphi_K(ht) dt, \tag{86}$$

where $R(g) = \int g^2(x)dx$, for any square integrable function g. As before the error density distribution function is assumed to be known exactly.

$\text{MISE}(\widehat{f}_\tau(\cdot,h))$ takes the minimum for the optimal bandwidth h_{opt}. The term f_τ and therefore $R(f_\tau)$ and $\varphi_\tau(t)$ aren't necessarily available in practical applications, therefore they have to be estimated. Estimations for $\text{MISE}\{\widehat{f}_\tau(\cdot;h)\}$ and the necessary functions can be obtained through different procedures. Fan (1991b) suggests the rule of thumb bandwidth

$$\widehat{h}_{\text{opt}} = \frac{\sqrt{2}\sigma_\epsilon}{\sqrt{\log(n)}}, \tag{87}$$

wherein practice σ_ϵ is replaced by $\widehat{\sigma}_\epsilon$. Meister (2009, p. 65-78) describes the cross-validated bandwidth selector and refers to Stefanski and Carroll (1990) and Hesse (1999) for detailed consistency properties. Delaigle and Gijbels (2002) describe the Plug-In bandwidth method and Delaigle and Gijbels (2004b) discuss two methods using bootstrap resampling. They compare all those selection procedures and come to realize, that the bootstrap and Plug-In selection methods outperform the other competitors. Because of that, the bootstrap selection procedure of the bandwidth without actually resampling observations will be explained and applied.

The optimal bandwidth h_{opt}, defined as the minimizer of the $\text{MISE}(\widehat{f}_\tau(\cdot;h))$ as mentioned before, has to be estimated by finding estimates for the unknown quantities related to f_τ. Therefore the resampling method of bootstrapping will be used to obtain needed quantities. For $\tau_1, ..., \tau_n$ the underlying density f_τ can be estimated by the deconvolving kernel estimator $\widehat{f}_\tau(\cdot;g)$ using the kernel K and the so called

pilot bandwidth g from the sample $X_1, ..., X_n$ including the measurement errors $\epsilon_1, ..., \epsilon_n$. $\widehat{f_\tau}(\cdot; g)$ represents a hypothetical real density function, to which the MISE as a measure of distance should be minimized. The pilot bandwidth g is not necessarily equal to the optimal bandwidth h_{opt} for $\widehat{f_\tau}(\cdot; h)$. By theoretically drawing a bootstrap sample $\tau_1^\star, ..., \tau_n^\star$ with replacement from the previously obtained $\widehat{f_\tau}(\cdot; g)$, this exact density can be estimated by $\widehat{f_\tau^\star}(\cdot; h)$ with bandwidth h. With the help of this quantity the original definition of the MISE can be modified in the following manner (involving $\mathrm{E}^\star$, $\mathrm{Var}^\star$ and $\mathrm{Bias}^\star$ depending on the observed data):

$$\mathrm{MISE}^\star(\widehat{f_\tau}(\cdot; h)) = \mathrm{E}^\star\left(\int (\widehat{f_\tau^\star}(\omega; h) - \widehat{f_\tau}(\omega; g))^2 d\omega\right) \tag{88}$$

$$= \int (\mathrm{Bias}^\star\{\widehat{f_\tau}(\omega; h)\})^2 d\omega + \int \mathrm{Var}^\star(\widehat{f_\tau^\star}(\omega; h)) d\omega, \tag{89}$$

and analogously split up as follows

$$\mathrm{MISE}^\star(\widehat{f_\tau}(\cdot; h)) = \frac{1}{2\pi n h} \int \frac{|\varphi_K(t)|^2}{|\varphi_\epsilon\left(\frac{t}{h}\right)|} dt \tag{90}$$

$$+ \frac{n-1}{2\pi n} \int |\widehat{\varphi}_\tau(t; g)|^2 |\varphi_K(ht)|^2 dt \tag{91}$$

$$+ R(\widehat{f_\tau}(\cdot; g)) - \frac{1}{\pi} \int |\widehat{\varphi}_\tau(t; g)|^2 \varphi_K(ht) dt. \tag{92}$$

$\widehat{\varphi}_\tau(t; g)$ as the Fourier transform of $\widehat{f_\tau}(\cdot; g)$ leads to an approximation of the MISE. For these bootstrap versions of the MISE, it is not necessary to draw actual samples rather than just calculate simple quantities based on the original observations. This is additionally simpler by considering the following definition of the MISE, which is equivalent to the previous one. Defining

$$\mathrm{MISE}_2^\star(\widehat{f_\tau}(\cdot; h)) = \frac{1}{2\pi n h} \int \frac{|\varphi_K(t)|^2}{|\varphi_\epsilon\left(\frac{t}{h}\right)|} dt \tag{93}$$

$$+ \frac{n-1}{2\pi n} \int |\widehat{\varphi}_\tau(t; g)|^2 |\varphi_K(ht)|^2 dt \tag{94}$$

$$- \frac{1}{\pi} \int |\widehat{\varphi}_\tau(t; g)|^2 \varphi_K(ht) dt, \tag{95}$$

reveals, that just $\widehat{\varphi}_\tau(\cdot, g)$ has to be estimated from the observed data, which can be easily achieved by using formula 53, only containing the empirical characteristic function $\widehat{\varphi}_\tau(t)$, which depends on the observed data.

The following question in regard to the choice of the so-called pilot bandwidth g, which allows to calculate the MISE in practice and thus enables an optimization procedure. The method chosen was proposed by Delaigle and Gijbels (2002) and is based on the asymptotic mean integrated squared error (AMISE) defined by

$$\text{AMISE}(\widehat{f}_\tau(\cdot;h)) = \frac{h^4}{4}\mu_2^2(K)R(f_\tau'') + \frac{1}{2\pi nh}\int \frac{|\varphi_K(t)|^2}{|\varphi_\epsilon\left(\frac{t}{h}\right)|^2}dt, \tag{96}$$

with $\mu_2(K) = \int z^2K(z)dz$. And using an estimation $R(\widehat{f}_\tau''(\cdot,g))$ leads to

$$\text{AMISE}^\star(\widehat{f}_\tau(\cdot,h)) = \frac{h^4}{4}\mu_2^2(K)R(\widehat{f}_\tau''(\cdot,g)) + \frac{1}{2\pi nh}\int \frac{|\varphi_K(t)|^2}{|\varphi_\epsilon(\frac{t}{h})|^2}dt, \tag{97}$$

which can easily been seen as this very quantity only contains g. Delaigle and Gijbels (2002) prove, that the MSE-optimal bandwidth g_r for $R(f_\tau^{(r)})$ with $r \in \mathbb{N}_0$ minimizes the absolute value of the asymptotic bias

$$\text{ABias}(R(\widehat{f}^{(r)}(\cdot;g_r))) = -g_r^2\mu_2(K)R(f_\tau^{(r+1)}) + \frac{1}{2\pi n g_r^{2r+1}}\int t^{2r}\frac{|\varphi_K(t)|^2}{|\varphi_\epsilon\left(\frac{t}{g_r}\right)|^2}dt \tag{98}$$

and recommends a two-stage procedure beginning with the estimation

$$\widehat{R(f_\tau^{(4)})} = \frac{8!\widehat{\sigma}_\tau^{-9}}{2^9 4!\sqrt{\pi}}, \tag{99}$$

which is based on the assumption of a parametric normal model for f_τ, where $\widehat{\sigma}_\tau^2$ is an adequate estimator for the variance of the true score variable. A formula to further calculate $R(\widehat{f}^{(r)}(\cdot,g_r))$ is described in Delaigle and Gijbels (2004b) as

$$R(\widehat{f}^{(r)}(\cdot;g_r)) = \frac{1}{2\pi}\int |\widehat{\varphi}_\tau(t,g_r)|^2 = \frac{1}{2\pi g_r^{2r+1}}\int t^{2r}\frac{|\widehat{\varphi}_X\left(\frac{t}{g_r}\right)|^2}{|\varphi_\epsilon\left(\frac{t}{g_r}\right)|^2}|\varphi_K(t)|^2dt, \tag{100}$$

as a part of the alternative Plug-In bandwidth selection method mentioned before. With all those quantities it is possible to obtain the optimal bandwidth described in algorithm 1.

Algorithm 1: Bootstrap bandwidth selection by Delaigle and Gijbels (2004a) including a two-stage procedure obtaining the pilot bandwidth g

Input : observed set x of contaminated data
deconvolution kernel density estimator $\widehat{f}_\tau$
kernel function K

Output: optimal bandwidth $\widehat{h}_{\text{opt}}$

Step 0: estimate $\widehat{R(f_\tau^{(4)})}$ for $R(f_\tau^{(4)})$ given by equation 99 using the estimation $\widehat{\sigma}_\tau^2 = \widehat{\sigma}_X^2 - \widehat{\sigma}_\epsilon^2$ for σ_τ^2

Step 1: Obtain g_3 by minimizing 98 and calculate $\widehat{R(f_\tau^{(3)})}$ using estimator 100

Step 2: Obtain g_2 by minimizing 98

Final: Use g_2 as pilot bandwidth g to minimize $\text{MISE}_2^\star(\widehat{f}_\tau(\cdot, h))$ given in 95 and obtain $\widehat{h}_{\text{opt}}$

The described two-stage bandwidth procedure can be replaced by choosing g according to the rule of thumb bandwidth estimator in formula 87. From a practical point of view, the integrations necessary to evaluate all used quantities are bounded by the compact support of the Fourier transform φ_K of the kernel function K. Therefore they can easily be performed in `R` using the `integrate()`-function.

3. Practical Application

All procedures and estimation techniques are implemented in `R` by R Core Team (2019). The visual illustrations and tables are created with the help of packages `ggplot2` and `gridExtra` by Wickham (2016), `RColorBrewer` by Neuwirth (2014), `xtable` by Dahl et al. (2019), `ggpubr` by Kassambara, 2019 and `tables` by Murdoch (2019). Simulations are carried out using `parallel` and `Rmpi` by Yu (2002) on the computing cluster of Faculty of Statistics at TU Dortmund University. The complete `R`-Code is available on Github[1] and OSF[2].

The central idea of the proposed approach is to estimate the error variance using Cronbach's α in equation 14 for an observed score $\overline{X}$ by assuming a classical test theory model. This leads to $\widehat{\sigma}^2_{\epsilon}$, which will be used to obtain the estimated density $\widehat{f}_\tau(\cdot)$ for the true score variable τ.

3.1. Simulation Procedure

To evaluate the quality of the proposed estimation procedure in practice, it is useful to calculate the integrated squared error (ISE) defined as

$$\text{ISE}_j(\widehat{f}_\tau(\cdot; h_j)) = \int (\widehat{f}_\tau(\omega; h_j) - f_\tau(\omega))^2 d\omega, \tag{101}$$

for $j = 1, ..., k$ simulated datasets for τ and ϵ. The true MISE can therefore be estimated by calculating the mean of all k ISE terms like

$$\widehat{\text{MISE}}(\widehat{f}_\tau(\cdot)) = \frac{1}{k} \sum_{j=1}^{k} \widehat{\text{ISE}}_j(\widehat{f}_\tau(\cdot; h_j)). \tag{102}$$

The numerical integration will again be carried out using `integrate()`. The obtained estimate $\widehat{f}_\tau$ is discrete-valued by definition of the numerical inversion procedure with a specific resolution. To ensure comparability with other density estimation methods, the estimated set of function values will additionally be interpolated linearly using `R`'s `approxfun()`. This procedure does not change occurring values and especially for a very high resolution of the FFT, which leads to a very fine grid of function evaluations, the interpolation simplifies without interference with the precision. Proceeding similarly, the classical kernel density estimator $\widehat{f}_X(x)$ described in 44 will be used as a baseline comparison, when the measurement error is not considered.

To evaluate the quality of density estimation in practical implementations, it is relevant to use data, whose actual distribution is known. Therefore two underlying distributions for the true score variable τ were selected, that are used in psychometric

[1] `https://github.com/robingrugel/Deconvolution-Problems-in-a-Psychometric-Context`
[2] `https://osf.io/92ze4/`

context. A varying number of indicators for the different models and for several fixed values of reliability should be considered to achieve a good insight by a broad comparison. Given the m observed values for each indicator $X_1, ..., X_m$ according to the model of (essentially) parallel variables $X_i = \tau + c_i + \epsilon_i$ for $i = 1, ..., m$, the score results from calculating the mean $\overline{X} = \tau + \overline{c} + \overline{\epsilon}$. To obtain the characteristic function of τ and respectively the density function, formulas 25 and 31 can be applied, leading to

$$\varphi_{\overline{X}}(t) = \varphi_{\tau+\overline{c}+\overline{\epsilon}}(t) \tag{103}$$

$$= \varphi_{\tau}(t)\exp(i\overline{c}t)\varphi_{\overline{\epsilon}}(t) \tag{104}$$

$$\iff \varphi_{\tau}(t) = \frac{\varphi_{\overline{X}}(t)}{\exp(i\overline{c}t)\varphi_{\overline{\epsilon}}(t)}. \tag{105}$$

This shows the considerable difference to the proposed estimation procedure of the transform of shifting parameters $\overline{c}$. By assuming $\overline{c} = 0$, the estimation can still be done without any changes. This assumption could be valid for special measurement concepts, for example in context of achievement scores by designing a well balanced composition of items having task in varying difficulty. Alternatively $\overline{c}$ could be estimated using

$$\mathrm{E}(\overline{X}) = \mathrm{E}(\tau + \overline{c}) = \mathrm{E}(\tau) + \overline{c}, \tag{106}$$

assuming $\mathrm{E}(\tau) = 0$ and replacing $\mathrm{E}(\overline{X})$ by the mean of the observed scores of $\overline{x}_1, ..., \overline{x}_n$. The regular estimator could be expanded by the transform of estimated mean shifting parameters. The focus is on the behaviour of Cronbach's α as an estimator for the reliability, which will be generating the variance of the measurement error. As shifting parameters do not influence the variance at all they can be neglected in this approach. The error variance for $\overline{X}$ can be written as $\mathrm{Var}(\overline{X}) = \mathrm{Var}(\tau + \overline{c} + \overline{\epsilon}) = \mathrm{Var}(\tau) + \mathrm{Var}(\overline{\epsilon})$, which leads to

$$\mathrm{Var}(\overline{\epsilon}) = \mathrm{Var}(\tau)\left(\frac{1 - \mathrm{Rel}(\overline{X})}{\mathrm{Rel}(\overline{X})}\right) \tag{107}$$

$$\iff \sum_{i=1}^{m}\mathrm{Var}(\epsilon_i) = m^2\mathrm{Var}(\tau)\left(\frac{1 - \mathrm{Rel}(\overline{X})}{\mathrm{Rel}(\overline{X})}\right) =: \sigma^2_{\mathrm{sum}}, \tag{108}$$

using equation 14. In case of parallel variables σ^2_{sum} will be divided in m equal variances. For τ-equivalent variables the sum has to be divided into unequal parts. This division should differ clearly from the case of parallel variables. Therefore a absolutely pathological division is chosen. As a result it will be assumed, that a high proportion of the sum of error variances will be caused by one indicator. Therefore proportion p of σ^2_{sum} is assigned to $\sigma^2_{\epsilon_1}$ and further assigning p of the remaining variance to the next indicator $\sigma^2_{\epsilon_2}$ till $m - 1$, finally assigning the rest (means the difference) to $\sigma^2_{\epsilon_m}$. This can easily be done by defining weights for $w_i = p(1-p)^{i-1}$ for $i = 1, ..., m-1$ and $w_m = 1 - \sum_{i=1}^{m-1} w_i$ and calculating $\sigma^2_{\epsilon_i} = w_i\sigma^2_{\mathrm{sum}}$ for $i = 1, ..., m$.

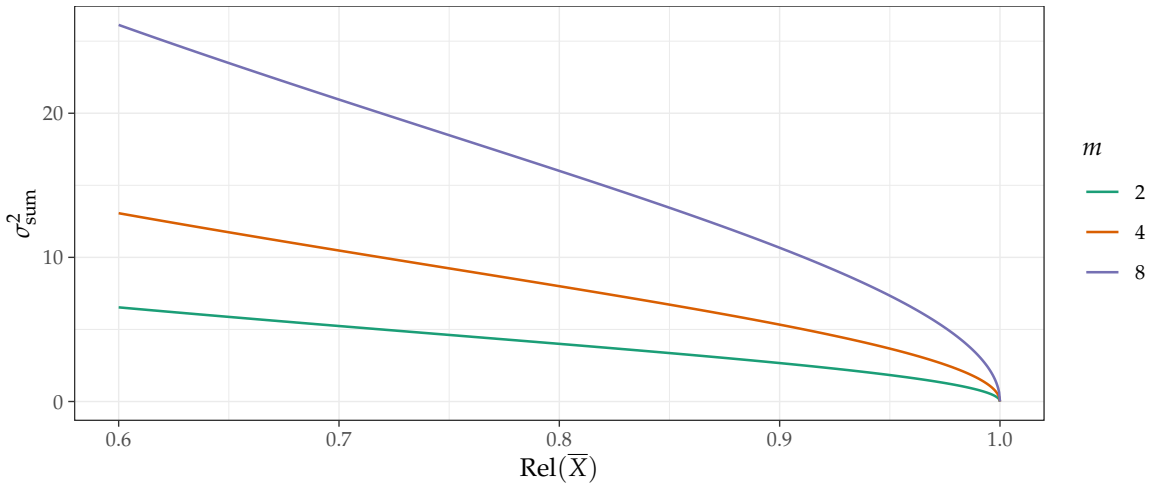

Figure 5: Plot of the sum of error variances σ^2_{sum} as a function of the reliability $\text{Rel}(\overline{X})$ for fixed true score variance $\text{Var}(\tau) = 4$ and m indicators

When plotting σ^2_{sum} depending on the true reliability $\text{Rel}(\overline{X})$ like in figure 5 it is clearly observable, why there are rule of thumbs, that strongly recommand values for the reliability above 0.9. The error variance decreases monotonously with increasing reliability. Especially when considering that measurement tools consist of a higher number of indicators the reliability is more important in obtaining valid inference. This can be recognized by the steeper negative gradient. The common reason for highly reliable measurement tools is the reduction of measurement errors, which is well illustrated in the behaviour of error variance in the area with high reliability.

The deconvolving kernel density estimator will be obtained for samples sizes $n \in \{250, 2500, 10000\}$ and $m \in \{2, 4, 8\}$ measurement tools of true reliability $\text{Rel}(\overline{X}) \in \{0.8, 0.9, 0.95, 0.99\}$. An corresponding error variance with fixed distribution of the true score $\tau \sim N(1, 4)$ and $\tau \sim \log N(1, 0.25)$ for the parallel and τ-equivalent CTT models will be used. The simulation described in algorithm 2 will be conducted for $k = 1000$. The split of the variance is chosen pathologically in form of 80% of variance being explained by one indicator. The estimated MISE is used for the comparison of precision for the different parameter combinations possible. It seems reasonable to assume that $k = 1000$ leads to a valid estimation for the MISE.

Algorithm 2: Algorithm for simulation study to compare quality measures of different parameter inputs

Input : sample size n
Number of variables m
Distribution of the true scroe variable τ
True reliability $\text{Rel}(\overline{X})$
Type of CTT model
Number of simulations k

Output: $\widehat{\text{MISE}}(\widehat{f}_\tau(\cdot)) = \frac{1}{k}\sum_{j=1}^{k}\widehat{\text{ISE}}_j(\widehat{f}_\tau(\cdot))$

for $j \in \{1, ..., k\}$ **do**
  Sample n values of τ
  Calculate σ^2_{sum}
  if *CTT model is τ-equivalent* **then**
    Divide σ^2_{sum} with m different weights as described into $\sigma^2_{\epsilon_1}, ..., \sigma^2_{\epsilon_m}$
  end
  else
    Divide σ^2_{sum} with m equal weights into m σ^2_ϵ
  end
  Convolve τ with ϵ_i into $x_i = \tau + \epsilon_i$ for $i = 1, ..., m$ by sampling n random numbers from $N(0, \sigma^2_{\epsilon_i})$
  Obtain estimated Cronbach's $\widehat{\alpha}$ for $x_1, ..., x_m$ and then estimate $\widehat{\sigma}^2_{\overline{\epsilon}}$
  Calculate $\overline{x}$
  Obtain density estimation $\widehat{f}_\tau(\cdot)$ with $\widehat{\sigma}^2_{\overline{\epsilon}}$
  Evaluate $\widehat{\text{ISE}}_j(\widehat{f}_\tau(\cdot))$
end

3.2. Simulation Results

Figure 6 illustrates the results of the simulation from table 4 and 5 in the appendix. The first and most obvious development in quality of the proposed estimator concerns the sample size in the way of approximating the true density function better with increasing sample size. This is the case for all distributions and assumed CTT models. It is noticeable that with increasing reliability the estimated MISE decreases, which is desired and expected. The MISE does not decrease in the same proportion for all parameter combinations, as it is substantially steeper for small sample sizes and becomes increasingly flatter with bigger samples. Another primary result is the fundamental difference between the estimations for the normal and log-normal distribution. The density estimator for the highly skewed log-normal distribution is clearly not as precise in approximation as for the normal distribution. Comparing the deconvolving density estimator to the classical kernel density estimator as a baseline confirms that the consideration of measurement errors is a useful concept. It is not surprising that for increasing reliability the kernel density estimation approximates the deconvolving estimation regarding the quality, because with increasing reliability the measurement error decreases and vanishes for $\text{Rel}(\overline{X}) = 1$ as seen in figure 5. The difference between the classic kernel and deconvolution approach is largest for low reliability and increases with the sample size. This observation again highlights the slow convergence behaviour for the supersmooth normal error distribution. A curious observation is that the baseline estimation performs superior for small sample sizes of the skewed log-normal distribution for highly reliable measurement tools. This behaviour vanishes as sample sizes increase, resulting in a more substantial difference, than for the normal distribution. Therefore a mathematical examination for highly skewed distributions would be an appropriate research project. The figure highlights especially that basically no difference in quality between the parallel and the τ-equivalent model is noticeable, which is the case although the chosen split of the variance is highly pathological. This can be explained by the well performing estimation of reliability by Cronbach's α. An additional reason is that the score calculated is the arithmetic mean which leads to $\widehat{\text{Var}}(\overline{X})$ being basically the same for both models, which explains that the estimated error variance does not differ in expectation. The number of indicators shows no impact on the estimated MISE at all.

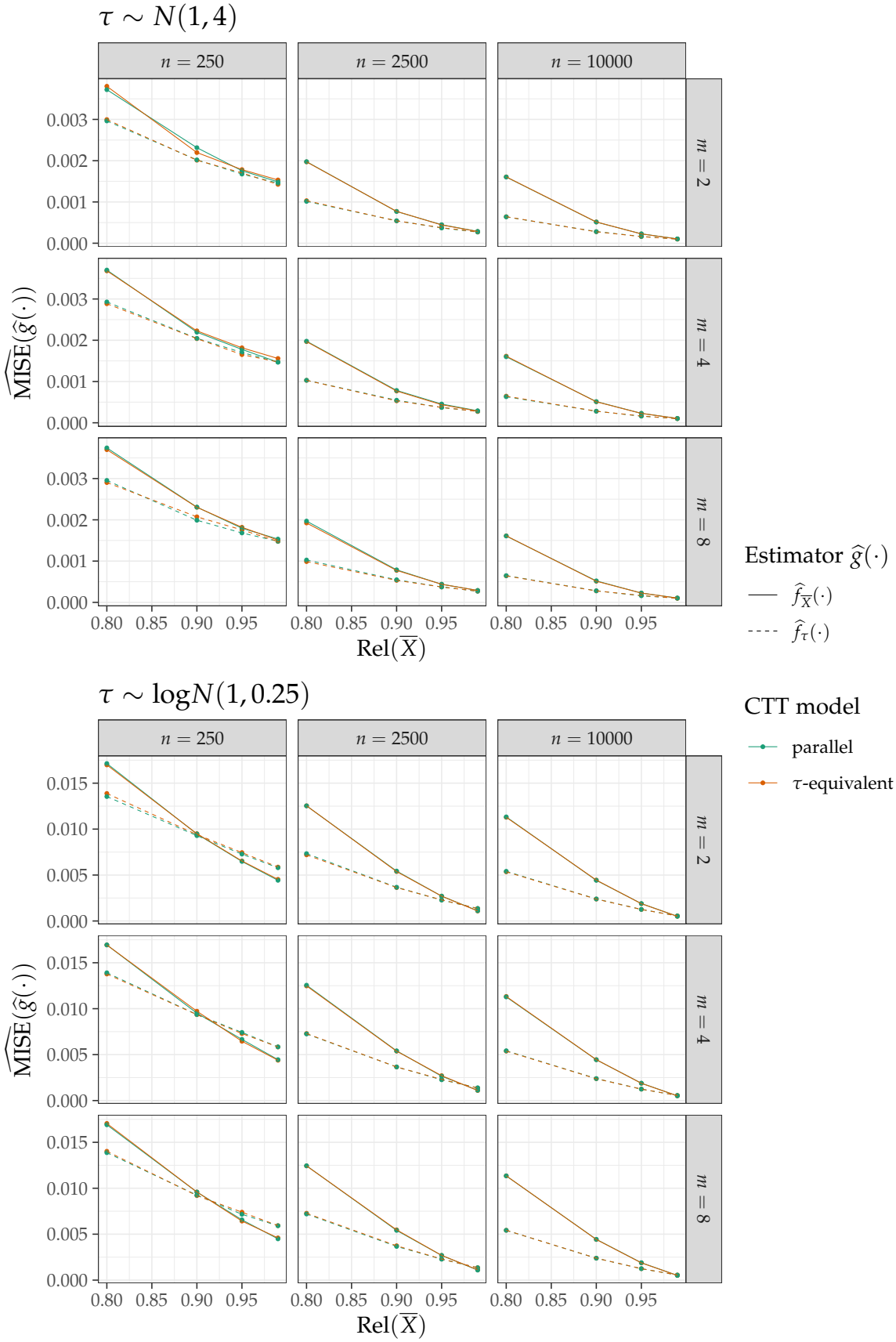

Figure 6: $\widehat{\text{MISE}}(\widehat{g}(\cdot))$ with $k = 1000$ randomly generated data samples τ convolved with $\epsilon_i \sim N(0, \sigma^2_{\epsilon_i})$ chosen depending on Rel($\overline{X}$) for n, m, $\widehat{g}(\cdot)$ and a specific CTT model

A very important different between parallel and τ-equivalent variables can still be pointed out by examining the variance of $\widehat{\sigma}^2_{\widehat{\epsilon}}$. A small simulation for both CTT models, the number of indicators and the reliability highlights the previously missed difference. Figure 7 displays the standard deviation of simulated error variances from table 6 in the appendix. A significant difference between the CTT models can be observed. The estimated error variance has less variation in the model of parallel variables. This difference decreases with increasing sample size, number of indicators and reliability. Consequently, the distinction between models and the number of components becomes less important for more reliable tools and an appropriate sample size.

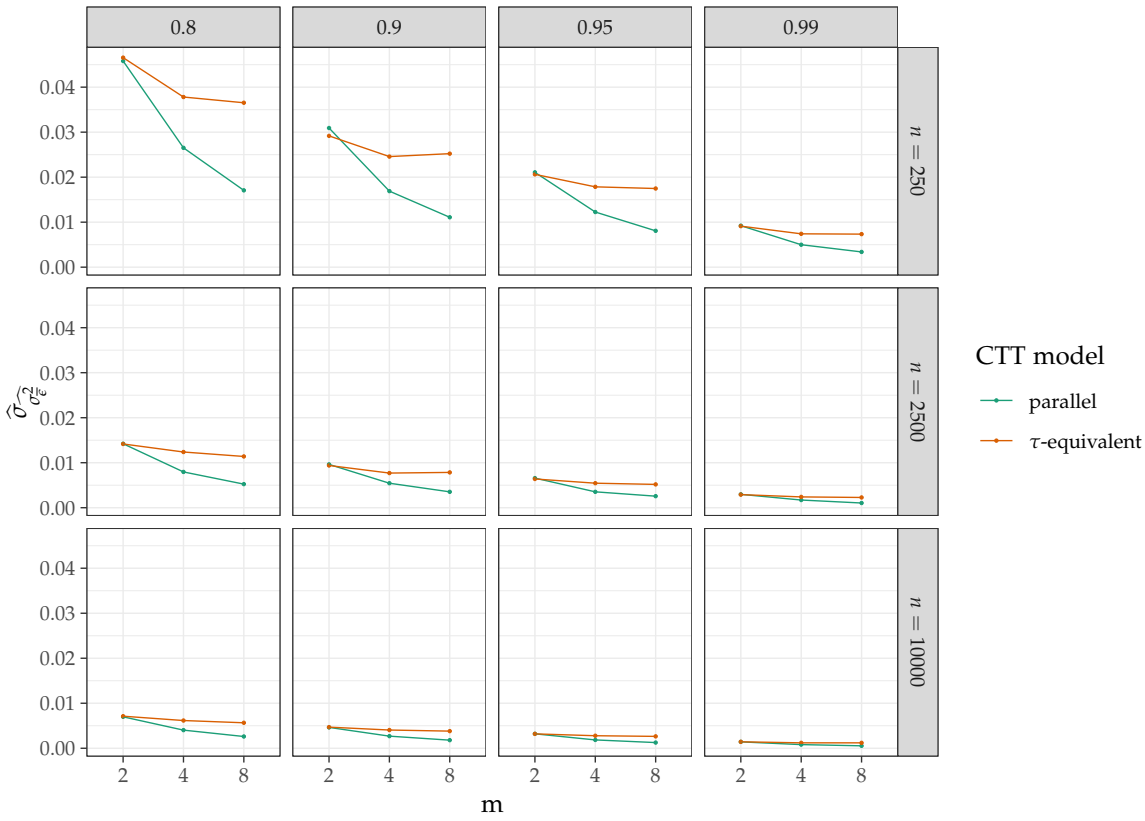

Figure 7: standard deviation $\widehat{\sigma}_{\widehat{\sigma^2_{\widehat{\epsilon}}}}$ with $k = 1000$ randomly generated data samples $\tau \sim N(1, 4)$ convolved with $\epsilon_i \sim N(0, \sigma^2_{\epsilon_i})$ chosen depending on $\text{Rel}(\overline{X})$ for n, m and a specific CTT model

In conclusion the findings confirm the benefit of considering measurement errors, especially for larger sample sizes. In the appendix, figures 10 and 11 illustrate examples for parameter combinations of the simulation and while the deconvolution estimation outperforms the classical kernel density estimator and improving with increasing reliability, it still does not necessarily approximate the true density perfectly. A curious observation is in particular, that the estimation is very accurate for the tails of the distribution and not fitting the body at the same time.

3.3. Real Data in Intelligence Research

The method of deconvolving kernel density estimation will now be applied to a real data set to illustrate possible usage. Therefore data of results in intelligence testing done in 1933-1936 and 2006 in Estonia, already analysed by O. Must and A. Must (2013), was selected. The original data was collected and later analysed by Tork (1940) in his dissertation titled "The Intelligence of Estonian Children". As described in O. Must and A. Must (2014) the National Intelligence Test (NIT) aimed to measure the intelligence of school children in grades 3 to 8 based on American Army Mental Tests and was translated into Estonian. The test consisted of two parallel forms encompassing 5 subtest, that measured in Arithmetical Reasoning ($A1$), Sentence Completion ($A2$), Concepts ($A3$), Same-Different (A), Symbol-Digit ($A5$), Computation ($B1$), Information ($B2$), Vocabulary ($B3$), Analogies ($B4$) and Comparisons ($B5$). O. Must and A. Must (2013) provided the original scoring algorithm which they used for descriptive and inference statistics, which also will be used in this demonstration. The same test was used again in 2006 to examine intelligence over time, more specifically the rise in intelligence quotients in progress of time, the so called Flynn effect. For more detailed scientific information see Steen (2009). In 1933-1936 $n_1 = 817$ students in the grades 4-6 and in 2006 $n_2 = 913$ students in the grades 6-8 participated as seen in table 2. It is interesting to see, that the membership to a grade clearly differs over time. The explanation for the differing grades is the change of the school system of Estonia.

grade	4	5	6	7	8
in 1933-1936	266	221	402	0	0
in 2006	0	0	243	343	327

Table 2: Number of students in grades depending on the time of survey

The composition of those samples of students shown in table 3 reveals, that age distribution is very similar.

age in years	10	11	12	13	14	15	16
in 1933-1936	4	55	173	261	241	123	32
in 2006	0	1	118	344	313	123	11

Table 3: Number of students in specific ages depending on the time of survey

The gender ratio was not balanced at both times with 1.206 (in 1933-1936) and 1.210 (in 2006) times more girls than boys. This imbalance is not important in comparing the two surveys, because the imbalance stays roughly the same by design. The chosen samples can be considered valid for comparison because only mild differences can be found.

In context of the deconvolving density estimation, it has to be assumed, that the available indicators are none essentially parallel or τ-equivalent variables. The assumption seems valid as the design of the measurement tool and the scoring

algorithm results in centering and scaling of the values of subscores. As the estimation is purely for illustration purposes, there will be no further testing of model assumptions. The already preprocessed subscores (see the scoring algorithm in O. Must and A. Must (2013)) described above will simply be aggregated to the mean, for both times of survey, as input for the deconvolving density estimation. The goal is to estimate the true scores $\tau_{1933/36}$ and τ_{2006}, in this context as a form of an intelligence score.

Figure 8 illustrates that the estimated density functions $\widehat{f}_{\tau_{1933/36}}(\cdot)$ and $\widehat{f}_{\tau_{2006}}(\cdot)$ clearly differ in shape. While the distribution of the intelligence score is relatively wide but nearly symmetric in 1933-1936, the distribution in 2006 shows a more narrow density. The flat top and tails of the distribution in the first period imply more diverse distributed intelligence scores through more frequently observed low and high values, but fewer intermediates. The probability mass of the density in 2006 is noticeably more concentrated while showing steeper tails and a slight tendency of being skewed to the left. This implies that one is less likely to observe extremely low or extremely high values. Possibly the change of schooling systems or education in general combined with a broader availability of schooling is the reason for this observation. In addition, the density of the more recent survey is clearly shifted towards higher values. It seems reasonable to expect higher values of intelligence nowadays than in the beginning of the 20th century. This confirms the central statement of the Flynn effect on the basis of this approach. It is worth to stress, that the interpretation of these estimations should be made with caution, because the model assumptions are not verified technically and a more thoroughly conducted descriptive analysis is necessary. In this application, the deconvolving density estimation is still a useful tool because not only the location of distributions but also the shape is taken into consideration. Especially the tail behaviour and skewness is displayed, while the measurement errors are considered, giving a deeper insight the distribution of latent variables.

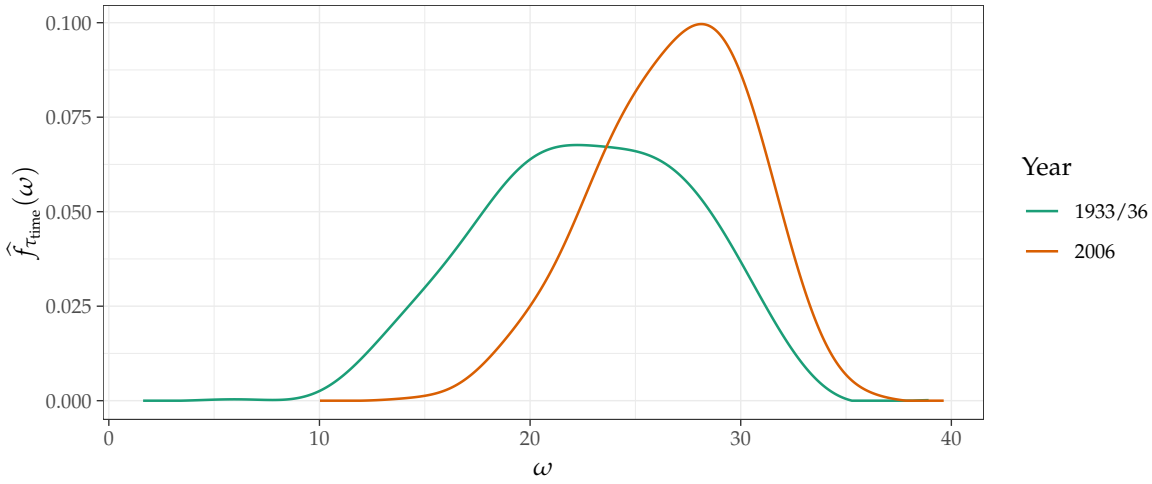

Figure 8: $\widehat{f}_{\tau_{\text{time}}}(\omega)$ for two periods times of survey 2006 and 1933-1936

4. Discussion and Conclusion

In this thesis, an approach is introduced which considers measurement errors of psychometric test theory models in a density estimation procedure. This procedure is based on estimating the error variance using Cronbach's α as an estimator for the reliability in psychometric context. The deconvolving density estimator is used to exploit the representation of measurement models in the domain of characteristic functions to take the error into account with help of the estimated variance.
A simulation study investigates the properties and benefits of the deconvolution approach, revealing a good performance compared to a classic kernel density estimation. The results show that performance increases with sample size and no difference in expected quality between the observed CTT models can be found. The parallel variables model however leads to am more reliable precision in estimating the error variance. This could be further investigated by a more elaborate simulation, which also reports intervals based on standard errors of the quality measure.
The application of the procedure illustrates its value by allowing distributions to have an unusual shape while taking measurement errors into account. One can be sure to obtain a valid density estimation and be able to compare them on a solid foundation. In processing the data concerning the Flynn effect no model assumptions were tested, this however would be necessary, if actual conclusions are supposed to be drawn on a scientific level.
The next step for future research approaches would be the focus on the shifting parameter c_i in essential CTT models. This would give the deconvolving density procedure higher value in practical applications.
The case of τ-equivalence, although performing equally well, is not taken into account explicitly and could be further studied as approached by Delaigle and Meister (2007) for different calculated scores from the indicators. The classical test theoretical models of parallel and τ-equivalent variables are sometimes too restrictive with assumptions that are hardly met. Therefore, it would be useful to also consider the model of (essentially) τ-congeneric variables $X_i = \lambda_i(\tau + c_i) + \epsilon_i$ for $i = 1, ..., m$ with specific factor loadings $\lambda_i > 0$ as illustrated in figure 9.

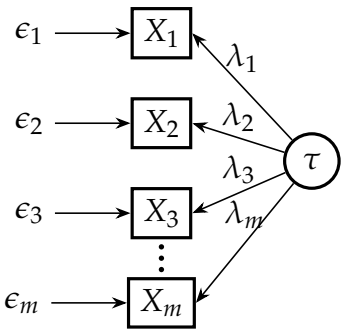

Figure 9: Path diagram of a typical measurement model of τ-congeneric variables

Therefore a score $\overline{X}$ for m τ-congeneric variables would lead to the characteristic function

$$\varphi_{\tau\bar{\lambda}}(t) = \frac{\varphi_{\overline{X}}(t)}{\exp(i\bar{\lambda}\bar{c}t)\varphi_{\bar{\varepsilon}}(t)}, \tag{109}$$

which does not represent the desired true score τ directly. It is worth to address this task after fully understanding the case of more restrictive models. The estimation of error variance using Cronbach's α would still be valid for τ-congeneric models in special cases of factor loadings as evaluated by Raykov (1997).
It would be additionally interesting to investigate estimation behaviour, if the assumed relationship 5 is not valid, errors are correlated and Cronbach's α can only be seen as a lower bound of the reliability.
In conclusion, the deconvolution approach in a psychometric context is a valuable tool, which requires a deeper evaluation in research projects.

A. Additional Figures and Tables

			parallel		τ-equivalent	
		$\text{Rel}(\overline{X})$	$\widehat{f}_{\tau}(\cdot)$	$\widehat{f}_{\overline{X}}(\cdot)$	$\widehat{f}_{\tau}(\cdot)$	$\widehat{f}_{\overline{X}}(\cdot)$
$n = 250$	$m = 2$	0.8	0.00296	0.00372	0.00300	0.00380
		0.9	0.00202	0.00232	0.00201	0.00220
		0.95	0.00168	0.00176	0.00170	0.00178
		0.99	0.00146	0.00149	0.00143	0.00153
	$m = 4$	0.8	0.00293	0.00370	0.00289	0.00368
		0.9	0.00205	0.00220	0.00204	0.00223
		0.95	0.00170	0.00178	0.00165	0.00182
		0.99	0.00147	0.00147	0.00147	0.00156
	$m = 8$	0.8	0.00295	0.00374	0.00290	0.00370
		0.9	0.00199	0.00231	0.00207	0.00230
		0.95	0.00168	0.00180	0.00176	0.00182
		0.99	0.00148	0.00153	0.00148	0.00151
$n = 2500$	$m = 2$	0.8	0.00101	0.00198	0.00103	0.00197
		0.9	0.00054	0.00077	0.00054	0.00077
		0.95	0.00037	0.00045	0.00038	0.00044
		0.99	0.00027	0.00029	0.00027	0.00028
	$m = 4$	0.8	0.00103	0.00198	0.00103	0.00197
		0.9	0.00055	0.00078	0.00053	0.00077
		0.95	0.00037	0.00045	0.00037	0.00044
		0.99	0.00028	0.00029	0.00027	0.00028
	$m = 8$	0.8	0.00103	0.00197	0.00099	0.00192
		0.9	0.00055	0.00079	0.00053	0.00077
		0.95	0.00037	0.00044	0.00037	0.00044
		0.99	0.00027	0.00029	0.00027	0.00029
$n = 10000$	$m = 2$	0.8	0.00064	0.00160	0.00064	0.00161
		0.9	0.00028	0.00051	0.00028	0.00052
		0.95	0.00016	0.00023	0.00016	0.00023
		0.99	0.00010	0.00011	0.00010	0.00010
	$m = 4$	0.8	0.00063	0.00160	0.00065	0.00161
		0.9	0.00028	0.00051	0.00028	0.00051
		0.95	0.00016	0.00023	0.00016	0.00023
		0.99	0.00010	0.00011	0.00010	0.00011
	$m = 8$	0.8	0.00065	0.00161	0.00064	0.00161
		0.9	0.00028	0.00052	0.00028	0.00051
		0.95	0.00016	0.00023	0.00017	0.00023
		0.99	0.00010	0.00010	0.00010	0.00010

Table 4: $\widehat{\text{MISE}}(\widehat{g}(\cdot))$ with $k = 1000$ randomly generated data samples $\tau \sim N(1,4)$ convolved with $\epsilon_i \sim N(0, \sigma^2_{\epsilon_i})$ chosen depending on $\text{Rel}(\overline{X})$ for n, m, $\widehat{g}(\cdot)$ and a specific CTT model

			parallel		τ-equivalent	
		$\text{Rel}(\overline{X})$	$\widehat{f}_\tau(\cdot)$	$\widehat{f}_{\overline{X}}(\cdot)$	$\widehat{f}_\tau(\cdot)$	$\widehat{f}_{\overline{X}}(\cdot)$
$n = 250$	$m = 2$	0.8	0.01355	0.01714	0.01388	0.01699
		0.9	0.00929	0.00947	0.00940	0.00950
		0.95	0.00728	0.00648	0.00744	0.00653
		0.99	0.00580	0.00443	0.00587	0.00453
	$m = 4$	0.8	0.01391	0.01694	0.01378	0.01694
		0.9	0.00936	0.00951	0.00935	0.00972
		0.95	0.00742	0.00666	0.00729	0.00645
		0.99	0.00584	0.00445	0.00583	0.00439
	$m = 8$	0.8	0.01387	0.01691	0.01402	0.01705
		0.9	0.00925	0.00959	0.00920	0.00959
		0.95	0.00716	0.00656	0.00740	0.00643
		0.99	0.00591	0.00449	0.00594	0.00459
$n = 2500$	$m = 2$	0.8	0.00734	0.01253	0.00720	0.01254
		0.9	0.00367	0.00543	0.00364	0.00538
		0.95	0.00227	0.00271	0.00230	0.00270
		0.99	0.00137	0.00112	0.00136	0.00109
	$m = 4$	0.8	0.00725	0.01256	0.00729	0.01248
		0.9	0.00366	0.00541	0.00365	0.00538
		0.95	0.00226	0.00265	0.00230	0.00271
		0.99	0.00139	0.00112	0.00136	0.00109
	$m = 8$	0.8	0.00720	0.01244	0.00728	0.01243
		0.9	0.00367	0.00541	0.00374	0.00547
		0.95	0.00228	0.00269	0.00230	0.00269
		0.99	0.00135	0.00109	0.00135	0.00112
$n = 10000$	$m = 2$	0.8	0.00541	0.01133	0.00535	0.01129
		0.9	0.00240	0.00443	0.00239	0.00446
		0.95	0.00126	0.00190	0.00125	0.00188
		0.99	0.00055	0.00051	0.00055	0.00051
	$m = 4$	0.8	0.00542	0.01128	0.00538	0.01133
		0.9	0.00239	0.00444	0.00238	0.00445
		0.95	0.00124	0.00189	0.00124	0.00188
		0.99	0.00056	0.00051	0.00055	0.00051
	$m = 8$	0.8	0.00542	0.01134	0.00542	0.01135
		0.9	0.00238	0.00444	0.00239	0.00442
		0.95	0.00125	0.00189	0.00125	0.00188
		0.99	0.00055	0.00051	0.00056	0.00051

Table 5: $\widehat{\text{MISE}}(\widehat{g}(\cdot))$ with $k = 1000$ randomly generated data samples $\tau \sim \log N(1, 0.25)$ convolved with $\epsilon_i \sim N(0, \sigma^2_{\epsilon_i})$ chosen depending on $\text{Rel}(\overline{X})$ for n, m, $\widehat{g}(\cdot)$ and a specific CTT model

		$\text{Rel}(\overline{X})$	parallel	τ-equivalent
$n = 250$	$m = 2$	0.8	0.04580	0.04656
		0.9	0.03092	0.02917
		0.95	0.02107	0.02063
		0.99	0.00922	0.00912
	$m = 4$	0.8	0.02652	0.03781
		0.9	0.01690	0.02458
		0.95	0.01225	0.01785
		0.99	0.00499	0.00741
	$m = 8$	0.8	0.01707	0.03653
		0.9	0.01108	0.02522
		0.95	0.00808	0.01748
		0.99	0.00339	0.00734
$n = 2500$	$m = 2$	0.8	0.01421	0.01416
		0.9	0.00963	0.00938
		0.95	0.00658	0.00638
		0.99	0.00297	0.00291
	$m = 4$	0.8	0.00795	0.01238
		0.9	0.00544	0.00770
		0.95	0.00353	0.00545
		0.99	0.00171	0.00240
	$m = 8$	0.8	0.00524	0.01139
		0.9	0.00353	0.00784
		0.95	0.00256	0.00518
		0.99	0.00105	0.00228
$n = 10000$	$m = 2$	0.8	0.00699	0.00712
		0.9	0.00463	0.00469
		0.95	0.00321	0.00319
		0.99	0.00142	0.00143
	$m = 4$	0.8	0.00403	0.00616
		0.9	0.00269	0.00405
		0.95	0.00185	0.00279
		0.99	0.00081	0.00120
	$m = 8$	0.8	0.00261	0.00566
		0.9	0.00180	0.00381
		0.95	0.00127	0.00264
		0.99	0.00055	0.00120

Table 6: $\widehat{\sigma}_{\widehat{\sigma^2_{\bar{\epsilon}}}}$ with $k = 1000$ randomly generated data samples $\tau \sim N(1, 4)$ convolved with $\epsilon_i \sim N(0, \sigma^2_{\epsilon_i})$ chosen depending on $\text{Rel}(\overline{X})$ for n, m and a specific CTT model

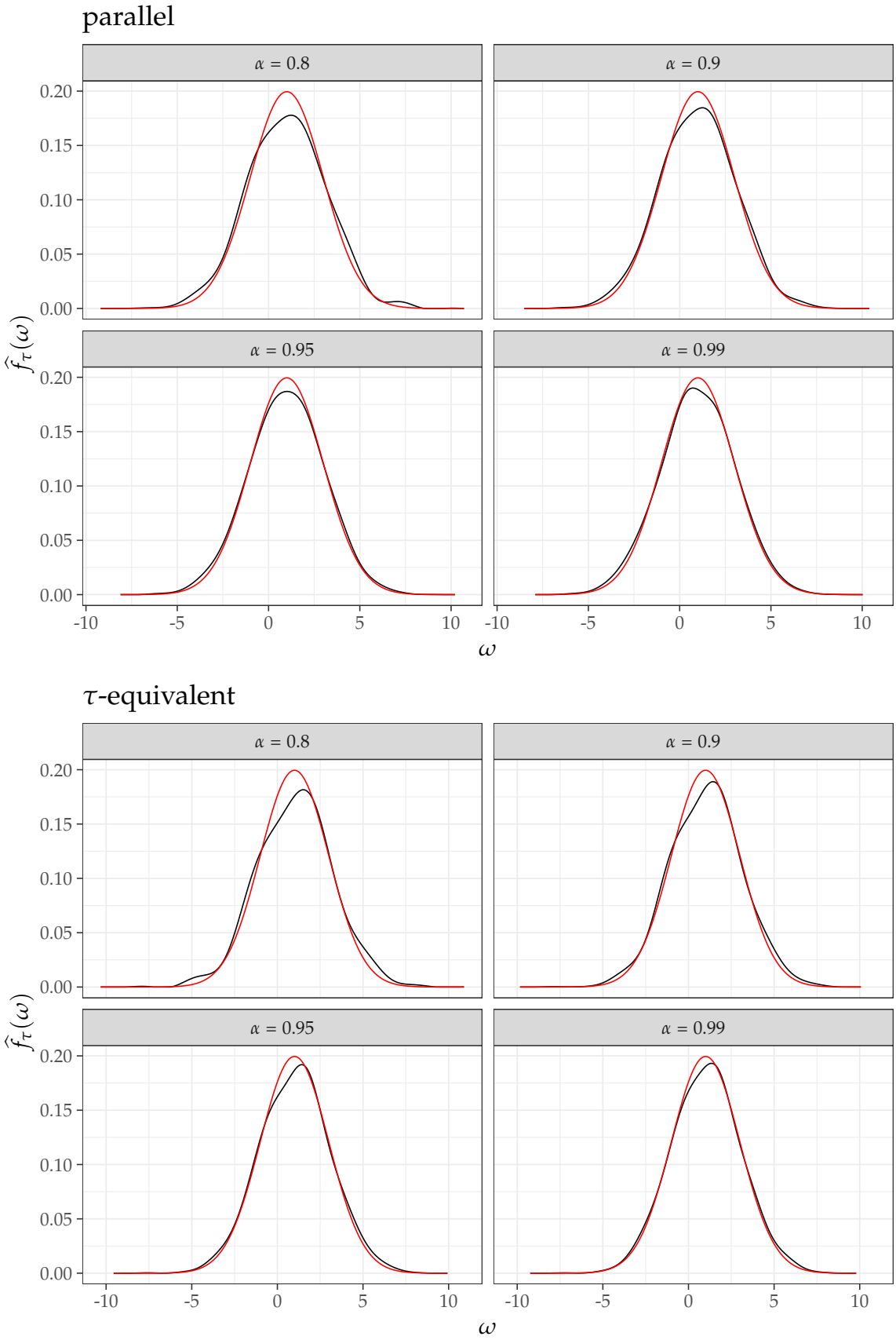

Figure 10: Example for estimations of $f_\tau(\omega)$ for $\tau \sim N(1,4)$ on basis of a mean score consisting of $m = 8$ subscores with sample size $n = 2500$ for true reliability $\text{Rel}(X) \in \{0.8, 0.9, 0.95, 0.99\}$ compared to the true density function in red

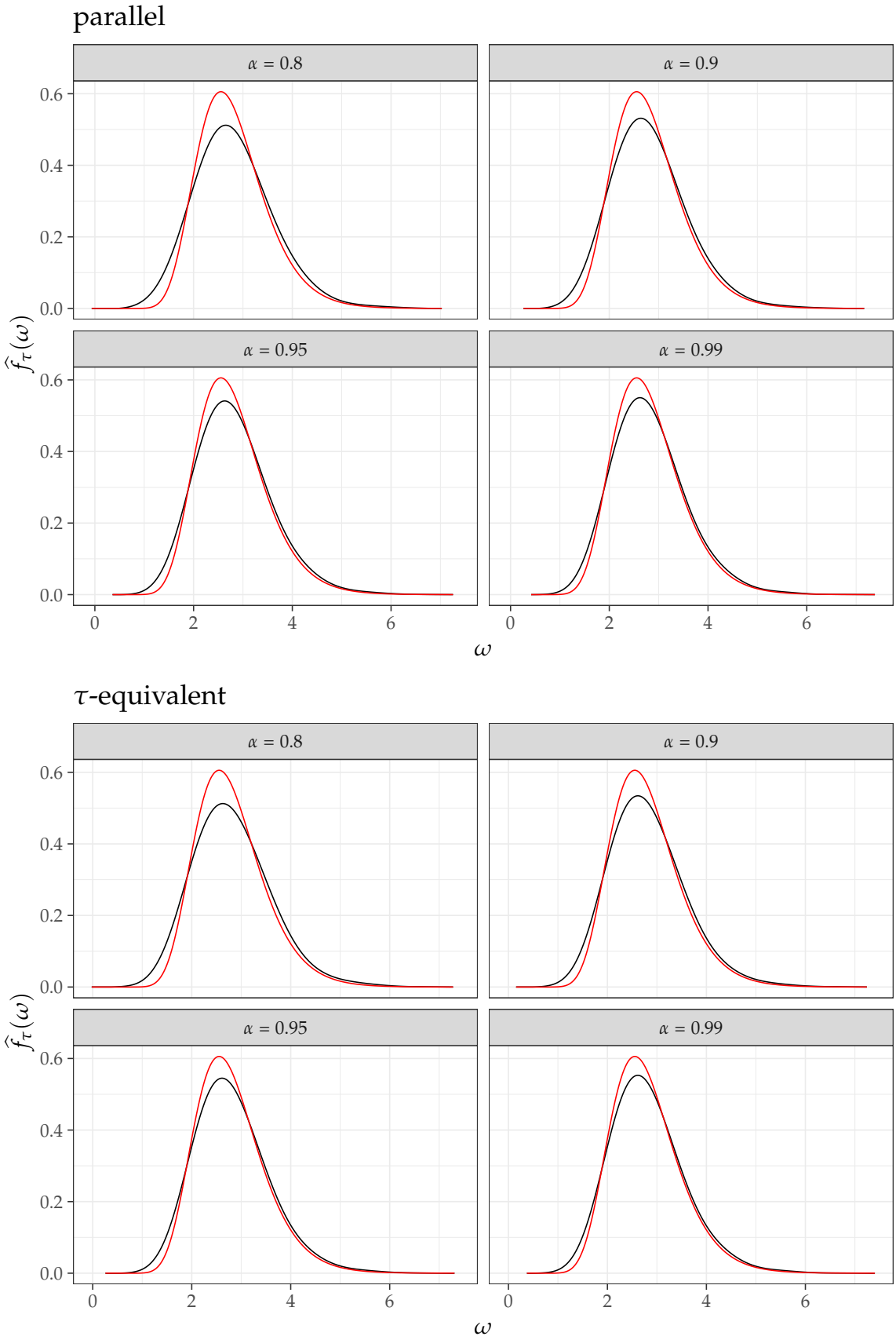

Figure 11: Example for estimations for $f_\tau(\omega)$ of $\tau \sim \log N(1, 0.25)$ on basis of a mean score consisting of $m = 8$ subscores with sample size $n = 2500$ for true reliability $\text{Rel}(X) \in \{0.8, 0.9, 0.95, 0.99\}$ compared to the true density function in red

References

Bailey, David H and Paul N Swarztrauber (1991). "The fractional Fourier transform and applications". In: *SIAM review* 33(3), pp. 389–404.

Bailey, David H and Paul N Swarztrauber (1994). "A fast method for the numerical evaluation of continuous Fourier and Laplace transforms". In: *SIAM Journal on Scientific Computing* 15(5), pp. 1105–1110.

Brémaud, Pierre (2014). "Fourier analysis of stochastic processes". In: *Fourier Analysis and Stochastic Processes*. Springer, pp. 119–179.

Carroll, Raymond J and Peter Hall (1988). "Optimal rates of convergence for deconvolving a density". In: *Journal of the American Statistical Association* 83(404), pp. 1184–1186.

Cooley, James W and John W Tukey (1965). "An algorithm for the machine calculation of complex Fourier series". In: *Mathematics of computation* 19(90), pp. 297–301.

Cronbach, Lee J (1951). "Coefficient alpha and the internal structure of tests". In: *psychometrika* 16(3), pp. 297–334.

Dahl, David B., David Scott, Charles Roosen, Arni Magnusson, and Jonathan Swinton (2019). *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4. URL: `https://CRAN.R-project.org/package=xtable`.

Delaigle, Aurore and Irène Gijbels (2002). "Estimation of integrated squared density derivatives from a contaminated sample". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(4), pp. 869–886.

Delaigle, Aurore and Irène Gijbels (2004a). "Bootstrap bandwidth selection in kernel density estimation from a contaminated sample". In: *Annals of the Institute of Statistical Mathematics* 56(1), pp. 19–47.

Delaigle, Aurore and Irène Gijbels (2004b). "Practical bandwidth selection in deconvolution kernel density estimation". In: *Computational statistics & data analysis* 45(2), pp. 249–267.

Delaigle, Aurore and Alexander Meister (2007). "Nonparametric regression estimation in the heteroscedastic errors-in-variables problem". In: *Journal of the American Statistical Association* 102(480), pp. 1416–1426.

Fan, Jianqing (1992). "Deconvolution with supersmooth distributions". In: *Canadian Journal of Statistics* 20(2), pp. 155–169.

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

Fan, Jianqing (1991b). "On the optimal rates of convergence for nonparametric deconvolution problems". In: *The Annals of Statistics*, pp. 1257–1272.

Gauss, CF (1886). "Theoria interpolationis methodo nova tractata Werke band 3, 265–327". In: *Göttingen: Königliche Gesellschaft der Wissenschaften*.

Guttman, Louis (1945). "A basis for analyzing test-retest reliability". In: *Psychometrika* 10(4), pp. 255–282.

Henze, Norbert (2017). *Stochastik für Einsteiger*. Vol. 1. Springer.

Hesse, Christian H (1999). "Data-driven deconvolution". In: *Journal of Nonparametric Statistics* 10(4), pp. 343–373.

Holgate, P (1989). "The lognormal characteristic function". In: *Communications in Statistics-Theory and Methods* 18(12), pp. 4539–4548.

Inverarity, Gordon W (2002). "Fast computation of multidimensional fourier integrals". In: *SIAM Journal on Scientific Computing* 24(2), pp. 645–651.

Kassambara, Alboukadel (2019). *ggpubr: 'ggplot2' Based Publication Ready Plots*. R package version 0.2.4. URL: https://CRAN.R-project.org/package=ggpubr.

Lukacs, Eugene (1970). *Characteristic functions*. Griffin.

Lukacs, Eugene (1972). "A survey of the theory of characteristic functions". In: *Advances in Applied Probability* 4(1), pp. 1–37.

Meister, Alexander (2009). *Deconvolution Problems in Nonparametric Statistics*. Springer. ISBN: 978-3-540-87556-7.

Murdoch, Duncan (2019). *tables: Formula-Driven Table Generation*. R package version 0.8.8. URL: https://CRAN.R-project.org/package=tables.

Must, Olev and Aasa Must (2013). "Changes in test-taking patterns over time". In: *Intelligence* 41(6), pp. 780–790.

Must, Olev and Aasa Must (2014). "Data from "Changes in test-taking patterns over time" concerning the Flynn Effect in Estonia". In: *Journal of Open Psychology Data* 2(1).

Neuwirth, Erich (2014). *RColorBrewer: ColorBrewer Palettes*. R package version 1.1-2. URL: https://CRAN.R-project.org/package=RColorBrewer.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Rao, Kamisetty Ramamohan, Do Nyeon Kim, and Jae Jeong Hwang (2010). *Fast Fourier transform-algorithms and applications*. Springer Science & Business Media.

Raykov, Tenko (1997). "Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components". In: *Multivariate behavioral research* 32(4), pp. 329–353.

Raykov, Tenko and George A Marcoulides (2010). *Introduction to psychometric theory*. Routledge.

Raykov, Tenko and George A Marcoulides (2017). "Thanks coefficient alpha, we still need you!" In: *Educational and Psychological Measurement* 79(1), pp. 200–210.

Scott, David W (1992). *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons.

Silverman, Bernard W (1990). *Density estimation for statistics and data analysis*. Routledge.

Singleton, RC (1979). *Mixed Radix Fast Fourier Transforms, in Programs for Digital Signal Processing*.

Steen, R Grant (2009). *Human intelligence and medical illness: Assessing the Flynn effect*. Springer Science & Business Media.

Stefanski, Leonard A and Raymond J Carroll (1990). "Deconvolving kernel density estimators". In: *Statistics* 21(2), pp. 169–184.

Steyer, Rolf and Michael Eid (2013). *Messen und Testen*. Springer-Verlag.

Tork, Juhan (1940). *Eesti laste intelligents: pedagoogiline, psühholoogiline ja sotsioloogiline uurimus*. Koolivara.

Ushakov, Nikolai G (1999). *Selected topics in characteristic functions*. Walter de Gruyter.

Wickham, Hadley (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN: 978-3-319-24277-4. URL: https://ggplot2.tidyverse.org.

Yu, Hao (2002). "Rmpi: Parallel Statistical Computing in R". In: *R News* 2(2), pp. 10–14. URL: https://cran.r-project.org/doc/Rnews/Rnews_2002-2.pdf.

Eidesstattliche Versicherung
(Affidavit)

Grugel, Robin

Name, Vorname
(Last name, first name)

194384

Matrikelnr.
(Enrollment number)

Ich versichere hiermit an Eides statt, dass ich die vorliegende Bachelorarbeit/~~Masterarbeit*~~ mit dem folgenden Titel selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

I declare in lieu of oath that I have completed the present Bachelor's/~~Master's*~~ thesis with the following title independently and without any unauthorized assistance. I have not used any other sources or aids than the ones listed and have documented quotations and paraphrases as such. The thesis in its current or similar version has not been submitted to an auditing institution.

Titel der Bachelor-/~~Masterarbeit*~~:
(Title of the Bachelor's/ ~~Master's*~~ thesis):

Deconvolution Problems in a Psychometric Context

*Nichtzutreffendes bitte streichen
(Please choose the appropriate)

Dortmund,

Ort, Datum
(Place, date)

Unterschrift
(Signature)

Belehrung:
Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden.
Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden. (§ 63 Abs. 5 Hochschulgesetz - HG -).

Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft.

Die Technische Universität Dortmund wird gfls. elektronische Vergleichswerkzeuge (wie z.B. die Software „turnitin") zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:
Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offense can be punished with a fine of up to €50,000.00. The competent administrative authority for the pursuit and prosecution of offenses of this type is the chancellor of TU Dortmund University. In the case of multiple or other serious attempts at deception, the examinee can also be unenrolled, section 63, subsection 5 of the North Rhine-Westphalia Higher Education Act (*Hochschulgesetz*).

The submission of a false affidavit will be punished with a prison sentence of up to three years or a fine.

As may be necessary, TU Dortmund will make use of electronic plagiarism-prevention tools (e.g. the "turnitin" service) in order to monitor violations during the examination procedures.

I have taken note of the above official notification:**

Dortmund,

Ort, Datum
(Place, date)

Unterschrift
(Signature)

**Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the Bachelor's/ ~~Master's~~ thesis is the official and legally binding version.