Accounting for Bandwidth Selection Variability in Estimating the Standard Errors of Kernel Equating

Kseniia Marcq

Master's Programme
Assessment, Measurement and Evaluation
120 credits

University of Oslo
Faculty of Educational Sciences

May 15, 2020

Popular Abstract

Standardized tests are commonly used in assessing individual performance, and their results greatly influence high stakes decisions. Because the tests are generally administered in alternate forms on multiple occasions, a statistical procedure known as equating is used to ensure that the results of those alternate forms are comparable across administrations. As any statistical procedure, equating is subject to sampling variability. The measure allowing to quantify this variability accurately is referred to as the standard error of equating. This paper focuses on one of the equating methods, kernel equating, and investigates the additional variability stemming from selecting a specific component within the procedure of kernel equating, and its effect on the standard error of kernel equating. We quantify this additional variability and account for it in modified formulas for the standard error of equating. The results of the present study suggest that the component does influence the standard error of equating, granted this influence is subtle. Hence, we are confident that implementing the methods introduced in this paper can improve the accuracy of the standard error of equating, and consequently, facilitate fairness and comparability of the equated results.

Acknowledgements

I would like to express my most profound appreciation for the help and support to the following persons who directly or indirectly have contributed to making this thesis a reality.

First and foremost, I would like to extend my deep and sincere gratitude to my supervisor, Dr Björn Andersson, for believing in me, his patience and ample time spent throughout this year. Without his continuous guidance, unwavering support, and immense knowledge, this thesis would not be possible.

My appreciation goes to the lecturers and staff at CEMO for their kindness and towering passion for what they do.

To my classmates, for their invaluable support and bright spirits which kept me going during stressful times.

Above all, I'm forever indebted to my family for their unparalleled love and encouragement in all my endeavours. To my parents, for teaching me to face challenges with strength and determination. To my husband, Erik, for his undying support and patience during these years.

Abstract

In standardized testing, equating is used to ensure fairness and comparability of the results across multiple test administrations. This paper focuses on one of the equipercentile observed-score equating methods, kernel equating. An essential step in kernel equating is obtaining the continuous approximations to the discrete score distributions by applying a kernel with a smoothing parameter, a bandwidth. When selecting a bandwidth, however, additional variability is introduced, which is currently not accounted for when calculating the standard errors of equating, consequently posing a threat to their accuracy. In this study, the asymptotic variance and standard error of the bandwidth parameter estimator are derived, and a modified method for calculating the standard error of equating, which accounts for bandwidth selection variability, is introduced. A simulation study is used to verify the derivations and confirm the accuracy of the modified method across several sample sizes as compared to the existing method and the Monte Carlo standard error estimates. The results show that the modified method of the standard error of equating calculation is relatively accurate under the considered conditions. Furthermore, the modified and the existing methods produce fairly similar results suggesting that the bandwidth variability impact on the standard error of equating is minimal.

Keywords: observed-score equating, kernel equating, bandwidth selection, bandwidth variability, standard errors of equating, delta method

Accounting for Bandwidth Selection Variability in Estimating the Standard Errors of Kernel Equating

Standardized testing is commonly used for assessing individual achievement, and its results greatly influence high-stakes decisions ranging from university admissions to various industry certifications. The framework of standardized testing generally requires the alternate test forms to be administered on multiple occasions. In consequence, the tests often differ in their difficulty from one administration to another, which poses a challenge with respect to comparability and fairness of their results. In order to address this challenge, a statistical procedure known as equating is employed with a paramount goal of adjusting the scores on the test forms so that they yield interchangeable results (Kolen & Brennan, 2014).

Observed-score equating is one of the fundamental methodologies used in test equating. Rooted in Classical Test Theory, it is concerned with establishing the equivalence of the observed scores on two test forms and incorporates both linear and equipercentile equating functions (von Davier, 2011). In this paper, we focus on one of the equipercentile observed-score equating methods, kernel equating, which was initially introduced by two ETS researchers, Holland and Thayer, in the late 1980s, followed by the work of von Davier, Holland and Thayer in 2004 (Holland & Thayer, 1989; von Davier, Holland, & Thayer, 2004).

The conceptual framework of kernel equating follows that of the equipercentile observed-score equating and posits a series of steps to obtain equated results: 1) pre-smoothing of the data to reduce sampling variability, 2) obtaining discrete score probability distributions, 3) obtaining continuous approximations to the discrete score distributions, 4) calculating of the equating function, and finally, 5) calculating of the standard error of equating (von Davier, 2011; von Davier et al., 2004). A feature distinguishing kernel equating from other equipercentile methods is that the continuous approximation of the score probability distributions is achieved through the use of a kernel with a bandwidth parameter. The bandwidth allows to make the density functions as smooth as possible while retaining the properties of the original distributions. Applying such a parameter, however, introduces additional variability. This variability is typically not accounted for when calculating the standard errors of equating, and therefore constitutes a threat to their accuracy (Holland, King, & Thayer, 1989; von Davier et al., 2004).

It should be self-evident that accurate estimation of the standard error of equating is integral to making correct inferences and comparisons. When estimated incorrectly, it can lead to unjustified certainty. This thesis addresses the issue of additional variability stemming from the bandwidth selection by modifying the existing method for calculating the standard error of equating (Holland et al., 1989).

We structure this paper as follows. In the subsequent subsections, we give a brief background to the kernel method of equating and expand on the issue of bandwidth variability and the standard error of equating calculation. Next, the asymptotic variance of the bandwidth parameter estimator is derived, which is then incorporated in a modified method for calculating the standard error of equating. This modified method is further verified and compared to the existing method using a simulation study. Lastly, the results are reported and discussed.

Data Collection Designs

Kernel equating, as a method within the framework of observed-score equating, consists of two fundamental components, the data collection design and the equating method (von Davier et al., 2004). Hence, before we focus on the equating itself, it is essential to review, if only briefly, the common approaches to collecting the data. There are several designs of data collection widely used in practice. Those can be roughly divided into two categories: designs which use examinees from a common population taking both test forms, and designs which use common items on both test forms (von Davier et al., 2004).

The first category of data collection designs includes the Equivalent-Groups (EG), the Single-Group (SG), and the Counterbalanced (CB) designs.

The Equivalent-Groups (EG) design. Two independent random samples are drawn from a common population, P, and one group takes test form X, while the other takes test form Y.

The Single-Group (SG) design. Both test forms X and Y are administered to a single sample from the population of test-takers, P. All examinees take both forms of the test.

The Counterbalanced (CB) design. Both test forms X and Y are administered to a single group of examinees drawn from a common population, P. The group is divided into subgroups so that one subgroup first takes test X followed by test Y, and the other - test Y followed by test X.

The second category of data collection designs uses common items instead of common examinees and is represented by the Non-Equivalent groups with Anchor Test (NEAT) design.

The Non-Equivalent groups with Anchor Test (NEAT) design. Test form X is administered to a sample from one population, P, while test form Y is administered to a sample from another population, Q. A number of anchor items, test form A, is administered to both samples from the populations P and Q.

The choice of an appropriate data collection design is subject to considerations of available sample size, time, and costs. The designs consecutively affect the equating procedure such that some designs, e.g. the Equivalent-Groups design, allow for a relatively straightforward comparison between the test forms while others are much more complex, e.g. the Non-Equivalent groups with Anchor Test design. A more detailed account of the considerations and procedures involved in various data collection designs can be found in von Davier et al. (2004).

Kernel Equating

Adopting the notation of von Davier et al. (2004), the two test forms are denoted as X and Y, and the target population is denoted as T. Let the possible observations on the test form X be x_j for $j = 1, \ldots, J$; and let the observations on the test form Y be y_k for $k = 1, \ldots, K$. Hence, the score probabilities can be defined as

$$r_j = \text{Prob}\{X = x_j|T\}, \tag{1}$$

and

$$s_k = \text{Prob}\{Y = y_k|T\}. \tag{2}$$

Further, an equipercentile equating function is defined in terms of cumulative distribution functions and is given by

$$F(x) = \text{Prob}(X \leq x|T), \tag{3}$$

and

$$G(y) = \text{Prob}(Y \leq y|T). \tag{4}$$

In the situation when the cumulative distribution functions are continuous, we can arrive to the equipercentile equating function of X to Y as follows

$$y = \text{Equi}_y(X) = G^{-1}(F(X)). \tag{5}$$

Strictly speaking, however, most score distributions are discrete, and their continuous approximation is required. Kernel equating addresses this problem by introducing a series of steps which can be applied to various data collection designs. Those steps include (1) pre-smoothing, (2) estimation of the score probabilities, (3) continuous approximation to the discrete score distributions, (4) equating, and (5) calculating the standard error of equating (von Davier et al., 2004). We now briefly review the first two steps and dedicate subsequent subsections to present the remaining steps in more detail as they pertain to the subject at hand.

(1) Pre-smoothing. In the first step, the score probabilities are estimated by fitting statistical models to the raw data until one is selected as an adequate fit. This can be achieved by fitting log-linear or item response theory, IRT, models. The methods are described in detail in Andersson and Wiberg (2017), and Holland and Thayer (1987), and are not repeated here.

(2) Estimation of the score probabilities. Having estimated the score distributions, the score probabilities can be obtained using a linear or non-linear transformation, the Design Function. The Design Function, DF, depends on the data collection design. For

instance, consider the Equivalent-Groups design and let $\boldsymbol{r}$ denote the column vector of the score probabilities of X given by $(r_1, ..., r_J)^t$ and $\boldsymbol{s}$ - the column vector of the score probabilities of Y given by $(s_1, ..., s_K)^t$. The Design function is then a simple identity function, i.e.

$$\begin{pmatrix} \boldsymbol{r} \\ \boldsymbol{s} \end{pmatrix} = \mathrm{DF}(\boldsymbol{r}, \boldsymbol{s}) = \begin{pmatrix} \boldsymbol{I_j} & 0 \\ 0 & \boldsymbol{I_k} \end{pmatrix} \begin{pmatrix} \boldsymbol{r} \\ \boldsymbol{s} \end{pmatrix}, \tag{6}$$

where $\boldsymbol{I_j}$ and $\boldsymbol{I_k}$ are $J_X \times J_X$ and $K_Y \times K_Y$ identity matrices. Design Functions for other data collection designs are given explicitly in von Davier et al. (2004).

Continuous Approximation and Equating

The third and essential step in kernel equating, distinguishing it from other equipercentile methods, is the continuous approximation to the discrete cumulative distribution functions, $\hat{F}(x)$ and $\hat{G}(y)$ to $\hat{F}_{h_X}(x)$ and $\hat{G}_{h_Y}(y)$, in order to compute the equating function (5). This is achieved by applying a kernel with a smoothing parameter, a bandwidth (von Davier et al., 2004). Following the notation of von Davier et al. (2004), let $\Phi(\cdot)$ denote the cumulative distribution function of the Gaussian distribution, and let h_X denote the bandwidth parameter. Then the Gaussian kernel smoothing of the distribution of X has a cumulative distribution function defined by

$$\hat{F}_{h_X}(x) = \sum_j \hat{r}_j \Phi(R_{jX}(x)), \tag{7}$$

where $R_{jX}(x)$ is given by

$$R_{jX}(x) = \frac{x - a_X x_j - (1 - a_X)\mu_X}{a_X h_X}, \tag{8}$$

and

$$a_X^2 = \frac{\sigma_X^2}{\sigma_X^2 + h_X^2}. \tag{9}$$

$\hat{G}_{h_Y}(y)$ is defined analogously.

It is evident from (7) - (9) that for the continuous approximation to be carried out bandwidth parameters, h_X and h_Y, have to be selected. The primary goal of introducing such parameters is to make the density functions as smooth as possible while retaining the properties of the original distributions. Various methods of selecting the bandwidth parameter have been suggested in previous research (Andersson & von Davier, 2014; von Davier et al., 2004). Of particular interest to this paper is a method described in von Davier et al. (2004) which selects the bandwidth parameter by minimizing the first part of the penalty function with respect to the bandwidth. The penalty function itself is based on the squared distances between the estimated proportions and the derivative of the continuous cumulative distribution function and is given by

$$\mathrm{PEN}_1(h_X) = \sum_j (\hat{r_j} - \hat{f}_{hX}(x_j))^2, \tag{10}$$

where $\hat{f}_{hX}(x_j)$ is a density function, found by differentiating (7) with respect to x, i.e.

$$f_{hX}(x) = \sum_j r_j \phi(R_{jX}(x)) \frac{1}{a_X h_X}, \tag{11}$$

and $R_{jX}(x)$ is given in (8). If the bandwidth is chosen correctly, the result of minimizing PEN_1 is a good approximation of the raw discrete distribution.

Once the continuous approximations are obtained, the equating function estimator for equating X to Y is given by

$$\hat{e}_Y(x) = e_Y(x; \hat{\boldsymbol{r}}, \hat{\boldsymbol{s}}) = G^{-1}_{h_Y}(F_{h_X}(x; \hat{\boldsymbol{r}}); \hat{\boldsymbol{s}}). \tag{12}$$

The equating function for equating Y to X is analogous and found by substitution.

Standard Error of Kernel Equating

The standard error of equating is the measure of random equating error or uncertainty which stems from the equating function being subject to estimation and thereby sampling variability. We largely base this subsection on the work of Holland et al. (1989) who derived the asymptotic standard error for the kernel method of equating using the standard delta method for computing large sample approximations to the sampling variances of functions of statistics. Adopting the notation of Holland et al. (1989), the standard error of equating for equating X to Y is defined as

$$\mathrm{SEE}_x(Y) = \sqrt{\mathrm{Var}(\hat{e}_Y(x))}. \tag{13}$$

The standard error of equating for equating Y to X is defined analogously.

Treating the bandwidth parameters h_X and h_Y as constants, Holland et al. (1989) assert that all the uncertainty in the equating function comes from the estimation of the score probabilities $\boldsymbol{r}$ and $\boldsymbol{s}$. Hence, the variance of the equating function, and in turn the standard error of equating, reflect the data collection design, the choice of the pre-smoothing technique used in the estimation of the population score probabilities, and the equating function itself.

Reiterating the notation used previously, let $\boldsymbol{r}$ and $\boldsymbol{s}$ define the vectors of the pre-smoothed score distributions. The calculation of the standard error of equating per Holland et al. (1989) then requires two components: a vector of derivatives of the equating function e_Y with respect to $\boldsymbol{r}$ and $\boldsymbol{s}$, and the asymptotic covariance matrix $\Sigma_{(\hat{\boldsymbol{r}},\hat{\boldsymbol{s}})}$. Using the delta method, the variance of the equating function $\hat{e}_Y$ can then be expressed as

$$\mathrm{Var}(\hat{e}_Y(x;\hat{\boldsymbol{r}},\hat{\boldsymbol{s}})) = [\partial e_Y]'\Sigma_{(\hat{\boldsymbol{r}},\hat{\boldsymbol{s}})}\partial e_Y, \tag{14}$$

where $\Sigma_{(\hat{\boldsymbol{r}},\hat{\boldsymbol{s}})}$ is the covariance matrix of the independently estimated score probabilities given by

$$\Sigma_{(\hat{\boldsymbol{r}},\hat{\boldsymbol{s}})} = \begin{bmatrix} \Sigma_{\hat{\boldsymbol{r}}} & 0 \\ 0 & \Sigma_{\hat{\boldsymbol{s}}} \end{bmatrix}. \tag{15}$$

The matrix $\Sigma_{(\hat{\boldsymbol{r}},\hat{\boldsymbol{s}})}$ is a matrix with dimensions of $(J_X + K_Y) \times (J_X + K_Y)$ where J_X is the dimension of $\boldsymbol{r}$, and K_Y is the dimension of $\boldsymbol{s}$ (von Davier et al., 2004). The calculation of the $\Sigma_{(\hat{\boldsymbol{r}},\hat{\boldsymbol{s}})}$ for different equating methods is beyond the scope of this paper and can be found in Andersson and Wiberg (2017), Holland et al. (1989) and von Davier et al. (2004).

The second component, ∂e_Y, is a vector of first order derivatives of the equating function e_Y with respect to the estimated score probabilities $\boldsymbol{r}$ and $\boldsymbol{s}$, i.e.

$$\partial_{e_Y} = \left[\frac{\partial e_Y}{\partial \boldsymbol{r}}, \frac{\partial e_Y}{\partial \boldsymbol{s}}\right]. \tag{16}$$

Recalling (12), the derivatives needed to compute the ∂_{e_Y} are defined in Holland et al.

(1989) as

$$\frac{\partial e_Y}{\partial r_j} = \frac{1}{G'}\frac{\partial F(x, \boldsymbol{r})}{\partial r_j}, \tag{17}$$

$$\frac{\partial e_Y}{\partial s_k} = -\frac{1}{G'}\frac{\partial G(e_Y(x); \boldsymbol{s})}{\partial s_k}, \tag{18}$$

where $\frac{\partial e_Y}{\partial \boldsymbol{r}}$ is a row vector with dimensions $1 \times J_X$, $\frac{\partial e_Y}{\partial \boldsymbol{s}}$ is a row vector with dimensions $1 \times K_Y$, G' is the density evaluated at $e_Y(x)$, i.e.

$$G' = \frac{\partial G(e_Y(x); \boldsymbol{s})}{\partial y}, \tag{19}$$

and

$$\frac{\partial F(x; \boldsymbol{r})}{\partial r_j} = \Phi(R_{jX}(x; \boldsymbol{r})) - M_{jX}(x; \boldsymbol{r})\frac{\partial F(x; \boldsymbol{s})}{\partial x}, \tag{20}$$

where $\frac{\partial F(x;\boldsymbol{s})}{\partial x}$ is given in (11), $R_{jX}(x; \boldsymbol{r})$ - in (8), and

$$M_{jX}(x; \boldsymbol{r}) = \frac{1}{2}(x - \mu_X)(1 - a_X^2)z_{jX}^2 + (1 - a_X)x_j, \tag{21}$$

where z_{jX} is defined as

$$z_{jX} = \frac{x_j - \mu_X}{\sigma_X}. \tag{22}$$

The derivatives of e_X are analogous to those above and can be computed by substitution.

At this point, it is important to emphasize that the method for the standard error of equating calculation described above treats the bandwidth parameters h_X and h_Y as fixed and not as functions of $\boldsymbol{r}$ and $\boldsymbol{s}$. Hence, the additional variability introduced by the bandwidth selection is currently not accounted for in the estimation of the standard errors of equating, and consequently poses a challenge with respect to their accuracy (Holland et al., 1989; von Davier et al., 2004).

Therefore, the objective of this thesis is to introduce a modified method of calculating the standard error of equating which accounts for the additional variability stemming from the bandwidth selection and compare it to the current method of calculating the standard error of equating (Holland et al., 1989), and the Monte Carlo standard error across several sample sizes.

In order to narrow the aspects of the subject, we formulate our research questions as follows:

1. Using Monte Carlo standard error as a criterion, what is the effect of the sample size on the accuracy of the modified method of calculating the asymptotic standard error of equating?

2. Using Monte Carlo standard error as a criterion, how do the modified and the existing (Holland et al., 1989) methods of calculating the asymptotic standard error of equating compare with respect to their accuracy?

Method

This section is structured as follows. First, the bandwidth parameter estimator variance and standard error are derived. Next, we introduce a modified method for calculation of the analytical standard error of equating, which accounts for bandwidth variability. Finally, the simulation setup designed to illustrate the use of the modified method and compare it to the current method of the standard error of equating calculation (Holland et al., 1989) as well as the Monte Carlo standard error, is outlined.

Asymptotic Variance and Standard Error of the Bandwidth Parameter Estimator

Before deriving the asymptotic variance of the bandwidth parameter estimator, we see it appropriate to restate the multivariate delta method (Rao, 1973). Adopting the notation of Rao (1973), let the $(k \times 1)$-dimensional random vector $\sqrt{n}(T_{kn} - \theta_k)$ converge to a multivariate normal distribution with zero mean and covariance Σ. Let g denote a vector-valued function with components g_1,..., g_q, such that all the entries of g are differentiable, then $\sqrt{n}(g(T_{kn}) - g(\theta_k))$ converges to a multivariate normal distribution with zero mean and covariance of $G\Sigma G'$, i.e.

$$\sqrt{n}(g(T_{kn}) - g(\theta_k)) \xrightarrow{d} N(0, G\Sigma G'), \quad (23)$$

where G is the Jacobian matrix of partial derivatives of g with respect to θ_k.

The bandwidth parameter estimator, however, is not defined explicitly but rather in terms of other asymptotically normal variables. Therefore, we use a generalization of the delta method presented by Benichou and Gail (1989) which facilitates computing the asymptotic variance of the implicitly defined bandwidth parameter estimator.

Following the notation of von Davier et al. (2004), let h_X denote the bandwidth parameter selected to minimize PEN_1 defined by (10), and $\boldsymbol{r}$ – the vector of estimated score probabilities. Consider further that PEN_1 is a continuously differentiable function of the estimated score probabilities $\boldsymbol{r}$ in h_X, and the function is minimized so that $\frac{\partial \text{PEN}_1}{\partial h'_X} = 0$. Applying the implicit function theorem (Rao, 1973), we can then define h_X as a function of $\boldsymbol{r}$ such that $h_X = g_{h_X}(\boldsymbol{r})$, and compute the partial derivatives of $g_{h_X}(\boldsymbol{r})$ with respect to $\boldsymbol{r}$ as

$$\frac{\partial g_{h_X}(\boldsymbol{r})'}{\partial \boldsymbol{r}} = -\left(\frac{\partial^2 \text{PEN}_1}{\partial h'_X}\right)^{-1} \frac{\partial^2 \text{PEN}_1}{\partial h'_X \partial \boldsymbol{r'}}, \tag{24}$$

where $\frac{\partial^2 \text{PEN}_1}{\partial h'_X}$ is a scalar second order partial derivative of PEN_1 with respect to h_X, and $\frac{\partial^2 \text{PEN}_1}{\partial h'_X \partial \boldsymbol{r'}}$ is a $1 \times J_X$ vector of second order partial derivatives of PEN_1 with respect to $\boldsymbol{r}$. The $\frac{\partial^2 \text{PEN}_1}{\partial h'_X}$ and $\frac{\partial^2 \text{PEN}_1}{\partial h'_X \partial \boldsymbol{r'}}$ derivatives are unequivocally calculated using the chain rule and implicit differentiation. The equations, however, are lengthy, and we summarize them in Appendix III.

Let $\Sigma_{\hat{\boldsymbol{r}}}$ denote the asymptotic covariance matrix of the estimated score probabilities $\boldsymbol{r}$ with dimensions $J_X \times J_X$ where J_X is the dimension of $\boldsymbol{r}$. By applying the delta method for implicit functions (Benichou & Gail, 1989), we can define the asymptotic variance of the bandwidth parameter estimator h_X as

$$\text{Var}(\hat{h_X}) = \frac{\partial g_{h_X}(\boldsymbol{r})'}{\partial \boldsymbol{r}} \Sigma_{\hat{\boldsymbol{r}}} \left[\frac{\partial g_{h_X}(\boldsymbol{r})'}{\partial \boldsymbol{r}}\right]', \tag{25}$$

and its standard error as

$$\text{SE}(\hat{h_X}) = \sqrt{\text{Var}(\hat{h_X})}. \tag{26}$$

The variance and the standard error of $\hat{h}_Y$ are analogous to those given for $\hat{h}_X$ and can be computed by substituting X by Y and $\boldsymbol{r}$ by $\boldsymbol{s}$.

Standard Error of Equating Accounting for Bandwidth Variability

As mentioned previously, the existing method of calculating the standard error of equating fixes the bandwidth parameters h_X and h_Y to be constant, and in doing so ignores the variance they attribute to the equating functions $e_Y(x)$ and $e_X(y)$ (Holland et al., 1989). We introduce a modification to this existing method by expanding the formula for the standard error of equating calculation to account for the additional variance stemming from the bandwidth selection.

Treating h_X as a function of the estimated score probabilities $\boldsymbol{r}$ (refer to the previous subsection), we redefine (14) as

$$\mathrm{Var}(\hat{e}_Y(x;\hat{\boldsymbol{r}},\hat{\boldsymbol{s}},\hat{h})) = \frac{\partial e_Y(x;\hat{\boldsymbol{r}},\hat{\boldsymbol{s}},\hat{h})}{\partial(\hat{\boldsymbol{r}}',\hat{\boldsymbol{s}}',\hat{h}')}\Sigma_{(\hat{\boldsymbol{r}},\hat{h}_X,\hat{\boldsymbol{s}},\hat{h}_Y)}\left[\frac{\partial e_Y(x;\hat{\boldsymbol{r}},\hat{\boldsymbol{s}},\hat{h})}{\partial(\hat{\boldsymbol{r}}',\hat{\boldsymbol{s}}',\hat{h}')}\right]', \tag{27}$$

where $\Sigma_{(\hat{\boldsymbol{r}},\hat{h}_X,\hat{\boldsymbol{s}},\hat{h}_Y)}$ is a $(((J_X+1)+(J_X+1))\times((K_Y+1)+(K_Y+1)))$-matrix of the estimated score probabilities $\boldsymbol{r}$ and $\boldsymbol{s}$, and the bandwidth parameters h_X and h_Y; and $\frac{\partial e_Y(x;\hat{\boldsymbol{r}},\hat{\boldsymbol{s}},\hat{h})}{\partial(\boldsymbol{r}',\boldsymbol{s}',h')}$ is a $(J_X \times (J_X+K_Y+1+1))$-matrix of derivatives of the equating function $\hat{e}_Y(x;\hat{\boldsymbol{r}},\hat{\boldsymbol{s}},\hat{h})$ with respect to the estimated score probabilities $\boldsymbol{r}$, $\boldsymbol{s}$, and the two bandwidth parameters, h_X and h_Y.

Recall that the derivatives of $e_Y(x)$ with respect to $\boldsymbol{r}$, $\boldsymbol{s}$ are given in (17) and (18). We, therefore, define the additional derivatives of $e_Y(x)$ with respect to the bandwidths h_X and h_Y by

$$\frac{\partial e_Y}{\partial h_X} = \frac{1}{G'}\sum_j \hat{r}_j \frac{\partial \Phi(R_{jX}(x))}{\partial h_X}, \tag{28}$$

where

$$\frac{\partial \Phi(R_{jX}(x))}{\partial h_X} = \sum_j \hat{r}_j \phi(R_{jX}(x))\frac{\partial R_{jX}(x)}{\partial h_X}, \tag{29}$$

and

$$\frac{\partial e_Y}{\partial h_Y} = -\frac{1}{G'}\sum_k \hat{r}_k \frac{\partial \Phi(R_{kY}(y))}{\partial h_Y}, \tag{30}$$

where

$$\frac{\partial \Phi(R_{kY}(y))}{\partial h_Y} = \sum_k \hat{r}_k \phi(R_{kY}(y))\frac{\partial R_{kY}(y)}{\partial h_Y}, \tag{31}$$

and G' is defined in (19), $R_{jX}(x)$, $R_{kY}(y))$ - in (8).

The second component in (27), the matrix $\Sigma_{(\hat{\boldsymbol{r}},\hat{h}_X,\hat{\boldsymbol{s}},\hat{h}_Y)}$, incorporates $\frac{\partial g_{h_X}(\boldsymbol{r})'}{\partial \boldsymbol{r}}$ defined in (24), $\frac{\partial g_{h_Y}(\boldsymbol{s})'}{\partial \boldsymbol{s}}$ calculated analogously, and $\Sigma_{(\hat{\boldsymbol{r}},\hat{\boldsymbol{s}})}$ given in (15) and is computed as follows

$$\Sigma_{(\hat{\boldsymbol{r}},\hat{h}_X,\hat{\boldsymbol{s}},\hat{h}_Y)} = \left[\frac{\partial g_{h_X}(\boldsymbol{r})'}{\partial \boldsymbol{r}}, \frac{\partial g_{h_Y}(\boldsymbol{s})'}{\partial \boldsymbol{s}}\right] \Sigma_{(\hat{\boldsymbol{r}},\hat{\boldsymbol{s}})} \left[\frac{\partial g_{h_X}(\boldsymbol{r})'}{\partial \boldsymbol{r}}, \frac{\partial g_{h_Y}(\boldsymbol{s})'}{\partial \boldsymbol{s}}\right]', \tag{32}$$

where $\left[\frac{\partial g_{h_X}(\boldsymbol{r})'}{\partial \boldsymbol{r}}, \frac{\partial g_{h_Y}(\boldsymbol{s})'}{\partial \boldsymbol{s}}\right]$ is a $(((J_X+1)+(K_Y+1)) \times (J_X+K_Y))$ zero identity matrix with ones on the diagonal and $\frac{\partial g_{h_X}(\boldsymbol{r})'}{\partial \boldsymbol{r}}$ placed in 1 in (J_X+1) and $\frac{\partial g_{h_Y}(\boldsymbol{s})'}{\partial \boldsymbol{s}}$ placed in 1 in (K_Y+1).

Lastly, considering (27) whose main diagonal elements are the corresponding variances of the equating function, we can define the standard error of equating which accounts for bandwidth variability as

$$\text{SEE}_Y(x) = \sqrt{\text{Var}(\hat{e}_Y(x; \hat{\boldsymbol{r}}, \hat{\boldsymbol{s}}, \hat{h}))}. \tag{33}$$

Simulation Design

Adapting the example of 20-item dichotomously scored parallel tests given in von Davier et al. (2004), data for two test forms X and Y were simulated using the 2 parameter logistic, 2PL, models within the framework of the item response theory, IRT (de Ayala, 2009). The discrimination parameters for both test forms were selected from the $U(1,2)$-distribution and the difficulty parameters – from the $N(0,1)$-distribution. These distributions were considered to mimic realistic item parameters used in standardized testing (National Center for Education Statistics, 2004).

The Equivalent-Groups design was used in which two independent random samples of individuals are drawn from a single common population, and groups take either of the test forms X and Y (von Davier et al., 2004). Dictated by the design, no differences in the latent distributions were enforced between groups, and the ability distributions followed $N(0,1)$. The Equivalent-Groups design was considered because of its simplicity. Relative to other data collection designs, it provided an opportunity for direct comparison of the results on the test forms X and Y without additional considerations or assumptions.

In order to systematically verify accuracy of the modified method of calculating the standard error of equating as well as to explore how well it performs in a variety of sample sizes ranging from small to relatively large, sample sizes 250, 1000, 4000, and 16000 were considered. The study was conducted using version 3.6.2 of R software environment (R Core Team, 2019) primarily employing packages *kequate* (Andersson, Bränberg, & Wiberg, 2013), *mirt* (Chalmers, 2012) and *numDeriv* (Gilbert & Varadhan, 2019). All the analyses were based on 10000 replications.

The study followed the recommended kernel equating procedure (von Davier et al., 2004), albeit with a few adjustments to verify the derivations presented in previous subsections. For each generated data set per sample size, a number of steps were carried out.

(1) Pre-smoothing. The package *mirt* (Chalmers, 2012) was used to pre-smooth the irregularities of the raw data by estimating IRT 2 parameter logistic, 2PL, models to obtain item parameter estimates. The expectation-maximization (EM) algorithm was used for estimation.

(2) Estimation of score probabilities. Under the Equivalent-Groups design, the score probabilities $\hat{r}_j$ and $\hat{s}_k$ were estimated based on the item parameter estimates and the assumed distribution of the latent variable (von Davier et al., 2004).

(3) Continuous approximation. Using the package *kequate* (Andersson et al., 2013) continuous approximations to the discrete distributions were obtained by applying a Gaussian kernel with an optimal bandwidth parameter (previous studies suggest different kernels provide similar equating results; Lee & von Davier, 2008). Optimal bandwidth parameters $\hat{h}_X$ and $\hat{h}_Y$ were obtained by minimizing the first part of the penalty function, PEN_1 (von Davier et al., 2004). When optimizing the penalty function, default tolerance of 1.50e-08 was used.

The analytical derivations for the bandwidth parameter estimator variance were paramount to the study. Hence, upon obtaining the optimal bandwidths, the average variance and standard errors of the bandwidth parameters were computed following the equations introduced in this section, and their accuracy was assessed using Monte Carlo standard error as a criterion. When calculating the asymptotic variance of the bandwidth parameter estimator, the bandwidth parameters h_X and h_Y were replaced with the estimated parameters $\hat{h}_X$ and $\hat{h}_Y$, and the asymptotic covariance matrices of the estimated score probabilities $\Sigma_{\hat{\boldsymbol{r}}}$ and $\Sigma_{\hat{\boldsymbol{s}}}$ were extracted from the *kequate* package output (Andersson et al., 2013).

(4) Equating. Upon obtaining continuous cumulative distribution functions, an equipercentile equating function was applied to equate the test forms X and Y.

(5) Calculating of the standard error of equating. The average analytical standard errors of equating were computed using the existing method for the standard error of equating calculation not accounting for the bandwidth variability (Holland et al., 1989), and the modified method of the standard error of equating calculation accounting for the bandwidth variability (refer to the previous subsection). The Monte Carlo standard errors (MCSE) were used as a criterion for comparing the accuracy of the modified and the existing methods of the standard error of equating calculation.

The analytical derivations used in computing the bandwidth variance and standard errors, as well as the standard error of equating, were verified numerically using the R package *numDeriv* (Gilbert & Varadhan, 2019). The R syntax code is available for review in the supplementary material (Appendix II).

Results

The study largely depended on the accuracy of the asymptotic variance and standard error of the bandwidth parameter estimator derivations presented in the Method section and Appendix III. The results of the simulation for the standard errors of the bandwidth parameter estimators h_X and h_Y given in Table 1 confirmed that the derivations were correct, and the asymptotic standard errors of the bandwidth parameter estimator (ASE) were accurate as witnessed by comparison to the Monte Carlo standard error estimates (MCSE). As can be expected for asymptotic variance approximation (Ferguson, 1996), the differences between the asymptotic standard errors and the Monte Carlo standard errors were lager in smaller sample sizes.

Table 1

ASE and MCSE for the bandwidth parameters h_X and h_Y

N	h_X parameter		h_Y parameter	
	ASE	MCSE	ASE	MCSE
250	0.0099	0.0097 (0.0000)	0.0139	0.0136 (0.0000)
1000	0.0052	0.0052 (0.0000)	0.0076	0.0077 (0.0000)
4000	0.0026	0.0026 (0.0000)	0.0039	0.0040 (0.0000)
16000	0.0013	0.0013 (0.0000)	0.0020	0.0020 (0.0000)

Note. N = sample size; ASE = asymptotic standard error; MCSE = Monte Carlo standard error.

Subsequently incorporating the bandwidth selection variability into computing the modified standard errors of equating, Table 2 and Table 3 present the standard errors of equating calculated using the modified method introduced in this paper, the existing method (Holland et al., 1989), and the Monte Carlo simulation estimates. Additionally, the standard errors of the Monte Carlo standard error estimates are provided in parenthesis of Tables 1, 2 and 3, and given a large number of replications those are sufficiently low.

When compared to the Monte Carlo standard error estimates, the modified asymptotic standard errors of equating which take bandwidth variability into account (ASEE_{mod}) were fairly accurate for all sample sizes. Furthermore, the modified asymptotic standard errors of equating in most cases appeared to be nearly identical to those not accounting for bandwidth variability (ASEE), suggesting that the bandwidth selection influence on the standard errors of equating was tenuous.

Supporting the previous finding were estimates of absolute aggregate differences between the standard errors of equating for two pairs, ASEE - MCSE and ASEE_{mod} - MCSE. With the smallest sample size (N = 250), the differences were 0.0054 and 0.0056 for pairs ASEE - MCSE and ASEE_{mod} - MCSE, respectively. The differences subsided

Table 2

ASEE, modified ASEE accounting for bandwidth variability, and MCSE

	N = 250			N = 1000		
Score	ASEE_{mod}	ASEE	MCSE	ASEE_{mod}	ASEE	MCSE
0	0.408	0.412	0.422 (0.004)	0.208	0.210	0.212 (0.002)
1	0.537	0.540	0.548 (0.005)	0.274	0.275	0.277 (0.003)
2	0.593	0.595	0.600 (0.006)	0.300	0.301	0.302 (0.003)
3	0.601	0.601	0.605 (0.006)	0.302	0.302	0.303 (0.003)
4	0.582	0.581	0.584 (0.006)	0.292	0.291	0.291 (0.003)
5	0.554	0.550	0.553 (0.005)	0.277	0.275	0.275 (0.003)
6	0.525	0.519	0.522 (0.005)	0.263	0.259	0.260 (0.003)
7	0.501	0.492	0.495 (0.005)	0.251	0.246	0.247 (0.002)
8	0.481	0.471	0.474 (0.005)	0.241	0.235	0.236 (0.002)
9	0.466	0.455	0.459 (0.004)	0.233	0.227	0.228 (0.002)
10	0.455	0.444	0.448 (0.004)	0.228	0.222	0.223 (0.002)
11	0.448	0.437	0.442 (0.004)	0.224	0.219	0.219 (0.002)
12	0.443	0.433	0.439 (0.004)	0.221	0.216	0.217 (0.002)
13	0.441	0.432	0.438 (0.004)	0.220	0.216	0.216 (0.002)
14	0.439	0.432	0.439 (0.004)	0.220	0.216	0.217 (0.002)
15	0.436	0.431	0.438 (0.004)	0.219	0.216	0.217 (0.002)
16	0.428	0.425	0.433 (0.004)	0.215	0.213	0.214 (0.002)
17	0.410	0.408	0.416 (0.004)	0.206	0.206	0.206 (0.002)
18	0.375	0.375	0.383 (0.004)	0.189	0.189	0.190 (0.002)
19	0.320	0.322	0.328 (0.003)	0.161	0.162	0.162 (0.002)
20	0.244	0.248	0.250 (0.002)	0.124	0.126	0.125 (0.001)

Note. N = sample size; ASEE_{mod} = modified asymptotic standard error of equating accounting for bandwidth variability; ASEE = asymptotic standard error of equating; MCSE = Monte Carlo standard error.

Table 3

ASEE, modified ASEE accounting for bandwidth variability, and MCSE

	N = 4000			N = 16000		
Score	$ASEE_{mod}$	ASEE	MCSE	$ASEE_{mod}$	ASEE	MCSE
0	0.105	0.105	0.104 (0.001)	0.052	0.053	0.051 (0.000)
1	0.138	0.138	0.137 (0.001)	0.069	0.069	0.068 (0.001)
2	0.150	0.149	0.149 (0.001)	0.075	0.075	0.074 (0.001)
3	0.151	0.150	0.149 (0.001)	0.076	0.075	0.075 (0.001)
4	0.146	0.144	0.144 (0.001)	0.073	0.073	0.072 (0.001)
5	0.139	0.137	0.136 (0.001)	0.069	0.069	0.068 (0.001)
6	0.132	0.129	0.128 (0.001)	0.066	0.065	0.064 (0.001)
7	0.126	0.122	0.122 (0.001)	0.063	0.061	0.061 (0.001)
8	0.120	0.117	0.116 (0.001)	0.060	0.059	0.058 (0.000)
9	0.117	0.113	0.113 (0.001)	0.058	0.057	0.056 (0.000)
10	0.114	0.111	0.110 (0.001)	0.057	0.056	0.055 (0.000)
11	0.112	0.109	0.109 (0.001)	0.056	0.055	0.054 (0.000)
12	0.111	0.108	0.107 (0.001)	0.055	0.054	0.054 (0.000)
13	0.111	0.108	0.107 (0.001)	0.055	0.054	0.054 (0.000)
14	0.111	0.108	0.108 (0.001)	0.055	0.054	0.054 (0.000)
15	0.109	0.108	0.108 (0.001)	0.055	0.054	0.054 (0.000)
16	0.108	0.107	0.106 (0.001)	0.054	0.053	0.053 (0.000)
17	0.103	0.103	0.102 (0.001)	0.052	0.051	0.052 (0.000)
18	0.095	0.094	0.094 (0.001)	0.047	0.047	0.047 (0.000)
19	0.081	0.080	0.080 (0.001)	0.040	0.041	0.041 (0.000)
20	0.063	0.062	0.061 (0.001)	0.031	0.032	0.031 (0.000)

Note. N = sample size; $ASEE_{mod}$ = modified asymptotic standard error of equating accounting for bandwidth variability; ASEE = asymptotic standard error of equating; MCSE = Monte Carlo standard error.

as the sample size grew with ASEE - MCSE = 0.0008 and ASEE_{mod} - MCSE = 0.0028, and ASEE - MCSE = 0.0006 and ASEE_{mod} - MCSE = 0.0017 for sample sizes 1000 and 4000, respectively. Lastly, the differences for the largest sample size (N = 16000) were notably lower with ASEE - MCSE = 0.0004 and ASEE_{mod} - MCSE = 0.0009.

Discussion

As a special case of the equipercentile observed-score equating, the kernel method of equating relies on the equipercentile equating function in which number-correct scores are transformed into percentile rank scores from test form X to the scale of test form Y, and the scores from the two test forms with the same percentile rank are considered to be equivalent (von Davier et al., 2004). However, in order to obtain those equivalent scores, the continuous approximation to the discrete score distributions is necessary. To satisfy this requirement kernel equating uses a Gaussian kernel with a smoothing parameter, a bandwidth. The selected bandwidth then determines the characteristics of the continuous approximations of the raw discrete distributions (von Davier et al., 2004).

The most commonly used in practice method for bandwidth selection is minimizing a penalty function with respect to the bandwidth parameter (von Davier et al., 2004). The bandwidth, in turn, is influenced by the estimated score probabilities and therefore is subject to variability. This variability, however, is not currently accounted for when calculating the standard error of equating (Holland et al., 1989), challenging its accuracy and, ultimately, the fairness of the equated results.

The present study explored the issue of the additional variability stemming from the bandwidth selection and its impact on the standard error of equating. Building on the existing methodology of Holland et al. (1989) and von Davier et al. (2004), we derived the asymptotic variance of the bandwidth parameter estimator using the delta method for implicit functions (Benichou & Gail, 1989) and incorporated those derivations to expand the existing formulas for calculating the standard errors of equating (Holland et al., 1989), therefore introducing a modified method which accounted for bandwidth selection variability. A simulation study with eight data sets generated for a wide range of sample sizes was used to illustrate the results of the modified method as compared to the current method of the standard error of equating calculation (Holland et al., 1989) and the Monte Carlo standard error.

The results offered several observations which could be valuable to the testing industry. Firstly, the modified method for the standard error of equating calculation was fairly accurate and close to the Monte Carlo standard error estimates for all sample sizes, including relatively low ones (i.e., N = 250), suggesting that the method is, in fact, suitable for practical use. Secondly, using the Monte Carlo standard error as a criterion, the results of the study indicate that overall the existing (Holland et al., 1989) and the modified methods for the standard error of equating calculation produced similar results at times favouring the existing method, suggesting that the bandwidth selection impact on the standard error of equating was minimal.

It is important to note that in this study, we derived the modified asymptotic standard error of equating for two parallel test forms of 20 items in the setting of the Equivalent-Groups data collection design. It can be the case that the bandwidth selection influence on the standard error of equating is greater for other data collection and equating designs, as well as for tests with a larger or smaller number of items. It would, therefore, be beneficial for future theoretical and empirical studies to focus on determining the bandwidth selection impact on the standard error of equating in those additional equating scenarios.

As a final note, we believe that it is theoretically more sound to use a method which successfully accounts for all sources of variability, however negligible those may be. Introducing the modifications to the formulas for the standard error of equating calculation akin to those explored in this study can improve the accuracy of the standard errors of equating, and consequently, facilitate fairness and comparability of the equated results.

References

Andersson, B., Bränberg, K., & Wiberg, M. (2013). Performing the Kernel Method of Test Equating with the Package **kequate**. *Journal of Statistical Software*, *55*(6), 1–25. doi: 10.18637/jss.v055.i06

Andersson, B., & von Davier, A. A. (2014). Improving the Bandwidth Selection in Kernel Equating. *Journal of Educational Measurement*, *51*, 223–238. doi: 10.1111/jedm.12044

Andersson, B., & Wiberg, M. (2017). Item Response Theory Observed-Score Kernel Equating. *Psychometrika*, *82*, 48–66. doi: 10.1007/s11336-016-9528-7

Benichou, J., & Gail, M. H. (1989). A Delta Method for Implicitly Defined Random Variables. *The American Statistician*, *43*, 41–44. doi: 10.1080/00031305.1989.10475608

Chalmers, R. P. (2012). **mirt** : A Multidimensional Item Response Theory Package for the *R* Environment. *Journal of Statistical Software*, *48*(6), 1–29. doi: 10.18637/jss.v048.i06

de Ayala, R. J. (2009). *The Theory and Practice of Item Response Theory.* New York, NY: Guilford Press.

Ferguson, T. S. (1996). *A Course in Large Sample Theory.* New York, NY: Routledge.

Gilbert, P., & Varadhan, R. (2019). numderiv: Accurate Numerical Derivatives [Computer software manual]. Retrieved from `http://optimizer.r-forge.r-project.org/` (R package version 2016.8-1.1)

Holland, P. W., King, B. F., & Thayer, D. T. (1989). The Standard Error of Equating for the Kernel Method of Equating Score Distributions. *ETS Research Report Series*, *1989*(1), i–52. doi: 10.1002/j.2330-8516.1989.tb00332.x

Holland, P. W., & Thayer, D. T. (1987). Notes on the Use of Log-Linear Models for Fitting Discrete Probability Distributions. *ETS Research Report Series*, *1987*(2), i–40. doi: 10.1002/j.2330-8516.1987.tb00235.x

Holland, P. W., & Thayer, D. T. (1989). The Kernel Method of Equating Score Distributions. *ETS Research Report Series*, *1989*(1), i–45. doi:

10.1002/j.2330-8516.1989.tb00333.x

Kolen, M. J., & Brennan, R. L. (2014). *Test Equating, Scaling, and Linking* (3rd ed.). New York, NY: Springer New York.

Lee, Y.-H., & von Davier, A. A. (2008). Comparing Alternative Kernels for the Kernel Method of Test Equating: Gaussian, Logistic, and Uniform Kernels. *ETS Research Report Series*, *2008*(1), i–26. doi: 10.1002/j.2333-8504.2008.tb02098.x

National Center for Education Statistics. (2004). *NAEP technical documentation - Mathematics Assessment IRT Parameters.* Retrieved 2020-04-01, from `https://nces.ed.gov/nationsreportcard/tdw/analysis/scaling`$_irt_math.aspx$ (Library Catalog: nces.ed.gov Publisher: National Center for Education Statistics)

R Core Team. (2019). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rao, C. R. (1973). *Linear Statistical Inference and its Applications* (2nd ed.). New York, NY: Wiley.

von Davier, A. A. (2011). *Statistical Models for Test Equating, Scaling, and Linking.* New York, NY: Springer New York. doi: 10.1007/978-0-387-98138-3

von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). *The Kernel Method of Test Equating.* New York, NY: Springer-Verlag New York.

Appendix I

GDPR Documentation

The present study did not require obtaining any GDPR (General Data Protection Regulation) documentation as no personal data were collected or processed. Therefore, we present a mock NSD (Norwegian Centre for Research Data) application form.

NOTIFICATION FORM (ENGLISH TRANSLATION) – NSD

- Personal data
- Types of data
- Project information
- Responsibility for data processing
- Sample and criteria
- Third persons
- Documentation
- Other approvals
- Processing
- Information security
- Duration of project
- Additional information

Personal data

Which personal data will be processed?

N/A. No personal data will be collected or processed.

Personal data are any data about an identified or identifiable natural person (data subject). Pseudonymized data are also considered personal data. "Pseudonymization" means processing collected data in way that the data can no longer be linked to individual persons, without the use of additional information. This usually involves removing identifiable information such as name, national ID number, contact details etc. from the collected data and giving each data subject a code/number. A scrambling key is the file/list of names and codes that makes it possible to identify individuals in the collected data. The scrambling key should be stored separately form the rest of the data. NB: processing pseudonymized data is still considered processing personal data, even if you do not have access to the scrambling key, and even if the scrambling key is being stored by an external party, such as SSB, the National registry etc.

Types of data

Name. First name and surname. N/A

National ID number or other personal identification number 11-digit personal identifier, D number, or other national identification number. N/A

Date of birth. N/A

Address or telephone number. N/A

Email address, IP address or other online identifiers. An email address is a unique address that is assigned to the user of an electronic mail service. An IP address is a unique address that is assigned to a device (e.g. a computer) in a computer network like the Internet. Dynamic IP addresses may also be considered personal data in certain cases. Cookies are an example of an online identifier. NB! If you are going use an online survey, and the service provider (data processor) will have access to email addresses or IP addresses, you must indicate this here. N/A

Photographs or video recordings of persons. Photographs and video recordings of faces are usually considered to be personal data. N/A

Audio recordings of persons. Audio recordings where personal data are recorded and/or where there exists a scrambling key that links the audio recordings to individual persons on the recordings. The voice of the person speaking may be considered personal data in combination with other background information. N/A

GPS data or other geolocation data. Data which indicate the geographical location of a person. N/A

Demographic data that can identify a natural person. E.g. a combination of information such as municipality of residence, workplace, position, age, gender etc. N/A

Genetic data. Personal data relating to the inherited or acquired genetic characteristics of a natural person, which give unique information about the physiology or health of that person. N/A

Biometric data. E.g. fingerprint, handprint, facial form, retina and iris scan, voice recognition, DNA. N/A

Other data that can identify a natural person. N/A

Will special categories of personal data or personal data relating to criminal convictions and offences be processed? N/A

Racial or ethnic origin. This includes belonging to an ethnic group, population, cultural sphere or society that has common characteristics. For example, information that a person is Sami is not considered to say anything about race, but it says something about ethnicity. N/A

Political opinions. That a person is a member of a political party and/or what a person voted in an election, including political opinions and beliefs. However, this does not include information that a person is a conservative, radical or labor party supporter. N/A

Religious beliefs. That a person is a member of a religious organization/congregation. This does not include information that a person has a subscription to a religious newspaper. N/A

Philosophical beliefs. That a person is a member of a philosophical association, or that a person believes that knowledge is acquired through logical speculation and observation. N/A

Trade Union Membership. That a person is a member of a trade union that organizes employees within the same industry/subject area, e.g. LO, NTL, NAR etc. N/A

Health data. Personal data concerning a natural person's physical or mental health, including use of healthcare services. N/A

Sex life or sexual orientation. A person's sexual orientation (homosexual, lesbian, bisexual etc.) and/or sexual behavior (e.g. that a personal has been unfaithful, indecent exposure, offensive gestures/language) N/A

Criminal convictions and offences. Personal data concerning convictions and offences or related to security measures. N/A

Project Information

Title

Accounting for Bandwidth Selection Variability in Estimating Standard Errors of Kernel Equating

Project description

Give a description of the project's scientific purpose/research question

The objective of this project is to introduce a modified method of calculating the standard error of equating, which accounts for the additional variability introduced by the bandwidth selection, and compare it to the current method of calculating the standard error of equating, and the Monte Carlo standard error across several sample sizes.

We formulate our research questions as follows:

1. How do the current and modified methods of calculating the standard error of equating compare with respect to the accuracy of the standard error of equating?

2. Comparing between the two methods, what is the effect of the sample size with respect to the accuracy of the standard error of equating?

The project focuses on analytical derivations which are verified using a simulation study with artificially generated data.

Subject area

- Social sciences
- Statistics
- Educational Measurement

Will the collected personal data be used for other purposes, in addition to the purpose of this project? N/A

Personal data should only be processed for specified, explicit and legitimate purposes. This means that each purpose for processing personal data must be identified and described clearly and accurately. In order for a purpose to be considered legitimate, it must also be in accordance with ethical and legal norms.

Explain why it is necessary to process personal data. N/A

Explain why the personal data are adequate, relevant and limited to what is necessary for the purposes for which they are being processed. This includes limiting the amount of collected data to that which is necessary to realize the purposes of data collection. N/A

External funding

- The Research Council of Norway (Norges forskningsråd - NFR) N/A
- Public authorities. E.g. research commissioned by a ministry N/A
- Other. E.g. funding from a pharmaceutical company or from private actors N/A

Type of project

- Research Project and PhD thesis
- Student project, Master's thesis
- Student project, Bachelor's thesis
- Other student projects

Responsibility for data processing

Neither the student nor the supervisor will handle personal data.

Data controller N/A

The institution responsible for the processing of personal data. The data controller determines the purposes for which, and the manner in which, personal data are processed.

Project leader (research assistant/ supervisor or research fellow/ PhD candidate) Kseniia Marcq - Master Student, UiO

Björn Andersson – supervisor, Associate Professor, CEMO, UiO, bjorn.andersson@cemo.uio.no

Will the responsibility for processing personal data be shared with other institutions (joint data controllers)? N/A

If two or more institutions together decide the purposes for which personal data are processed, they are joint data controllers.

Joint data controllers N/A

Institution

Institution not found in the list

Institution

Country

Postal address

Email address

Telephone number

Sample and criteria

Whose personal data will be processed?

You must describe each group of people whose personal data you will be processing. Add and describe each sample individually. N/A. No personal data will be collected or processed.

Sample 1 Describe the sample N/A

Recruitment or selection of the sample N/A

Describe how the sample will be recruited and how initial contact with the sample will be made. For example, whether you will make initial contact during fieldwork or via your own network, or whether a school, hospital or organization will contact its pupils, patients or members on your behalf. If the sample will not be recruited but will be selected from a registry or an administrative system etc., describe how the selection will be carried out and what the selection criteria will be.

Age N/A

Will you include adults (18 y.o. +) who do not have the capacity to consent?

i.e. the person has reduced capacity or lacks capacity to consent. For example, the person may have mental/cognitive impairment, significant physical/emotional ailments, or may be unconscious, conditions which make it difficult or impossible for the person to gain sufficient understanding in order to give valid consent. The central aspect is whether the person is capable of understanding the purpose of the processing/project in question, and of understanding potential positive and negative consequences (immediate and long-term).

Types of personal data - sample 1 N/A

Name N/A

National ID number or other personal identification number N/A

Date of birth N/A

Address or telephone number N/A

Email address, IP address or other online identifier N/A

Photographs or video recordings of persons N/A

Audio recordings of persons N/A

GPS data or other geolocation data N/A

Demographic data that can identify a natural person N/A

Genetic data N/A

Biometric data N/A

Other data that can identify a natural person N/A

Methods /data sources - sample 1. N/A

Select and/or describe the method(s) for collecting personal data and/or the source(s) of data N/A

Personal interview N/A

Group interview Online survey Paper-based survey N/A

Participant observation - Non-participant observation N/A

Field experiment / field intervention N/A

Web-based experiment N/A

Tests for pedagogical research / psychological tests N/A

Medical examination and/or physical tests N/A

Human biological material N/A

Social media – open forum N/A

Social media – closed forum N/A

Discussion board/forum for online newspapers/online debates N/A

Big data N/A

Medical records N/A

Biobank N/A

Data from another research project N/A

Other N/A

Statistics Norway - SSB N/A

Criminal records (Det sentrale straffe- og politiopplysningsregisteret, SSP) N/A

Medical Birth Registry of Norway (Medisinsk fødselsregister, MFR) N/A

Norwegian Registry of Pregnancy Termination (Register over svangerskapsavbrudd) N/A

Norwegian Cardiovascular Disease Registry (Hjerte- og karregisteret) N/A

Norwegian Cause of Death Registry (Dødsarsaksregisteret, DÅR) N/A

Norwegian Prescription Database - NorPD (Reseptregisteret) N/A

Norwegian Immunisation Registry (Nasjonalt vaksinasjonsregister, SYSVAK) Norwegian Surveillance System for Communicable Diseases (Meldesystem for smittsomme sykdommer, MSIS) N/A

Norwegian Surveillance System for use of antibiotics and healthcare related infections (Norsk overvåkingssystem for antibiotikabruk og helsetjenesteassosierte infeksjoner, NOIS) N/A

Norwegian Surveillance System for Antimicrobial Drug Resistance (Norsk overvåkingssystem for antibiotikaresistens hos mikrober, NORM) Norwegian Surveillance System for Virus Resistance (Norwegian Surveillance System for Virus

Resistance, RAVN) N/A Norwegian Patient Registry (Norsk pasientregister, NPR) IPLOS-registeret Kommunalt pasient- og brukerregister (KPR) N/A Cancer registry of Norway (Kreftregisteret) N/A

Genetic Mass Survey of Newborns (Genetisk masseundersøkelse av nyfødte) N/A

Reseptformidleren N/A

Forsvarets helseregister N/A

Helsearkivregisteret N/A

Helseundersøkelsen i Nord Trøndelag (HUNT) N/A

Tromsø-undersøkelsen N/A

SAMINOR N/A

Den norske mor og barn undersøkelsen (MoBa) N/A

Nasjonalt register for langtids mekanisk ventilasjon N/A

Nasjonalt kvalitetsregister for barnekreft N/A

Norsk Kvalitetsregister Øre-Nese-Hals –Tonsilleregisteret N/A

Norsk vaskulittregister & biobank (NorVas) N/A

Norsk Parkinsonregister & biobank N/A

Norsk karkirurgisk register (NORKAR) N/A

Norsk hjertinfarkregister N/A

Gastronet N/A

Norsk register for analinkontinens N/A

Nasjonalt barnehofteregister N/A

Norsk kvalitetsregister for artrittsykdommer (NorArtritt) N/A

Norsk nakke- og ryggregister N/A

Nasjonalt korsbåndregister N/A

Nasjonalt register for leddproteser N/A

NorKog N/A

Norsk MS-register og biobank N/A

Nasjonalt register for KOLS N/A

Nasjonalt kvalitetsregister for lymfom og lymfoide leukemier N/A

Nasjonalt kvalitetsregister for lungekreft N/A

Nasjonalt kvalitetsregister for føflekkreft N/A

Nasjonalt kvalitetsregister for brystkreft N/A

Nasjonalt kvalitetsregister for prostatakreft N/A

Nasjonalt kvalitetsregister for tykk- og endetarmskreft N/A

Nasjonalt register for ablasjonsbehandling og elektrofysiologi i Norge (ABLA NOR) N/A

Norsk register for invasiv kardiologi (NORIC) N/A

Norsk hjertesviktregister N/A

Norsk pacemaker- og ICD- register N/A

Nasjonalt kvalitetsregister for gynekologisk kreft N/A

Norsk register for gastrokirurgi (NoRGast) N/A

Nasjonalt kvalitetsregister for behandling av spiseforstyrrelser (NorSpis) N/A

Information - sample 1

Will you inform the sample about processing their personal data? N/A

How? N/A

Written information (on paper or electronically)

Oral information

See what you must give inform about and preferably use our template for the information letter.

Information should be given in writing or electronically. Only in special cases is it applicable to give oral information, if a participant asks for this. See what you must give information about.

Upload information letter N/A

Upload copy of oral information N/A

Explain why the sample will not be informed about the processing of their personal data. N/A. No personal data will be collected or processed. + Add sample

Third persons

Will you be processing personal data about third persons? This includes data about persons who are not included in the sample/are not participating in the project; information provided by a data subject that relates to another identified or identifiable natural person. Examples of this are when a data subject is asked about their mother's and father's education or country of origin, or when pupils are asked about their teacher's teaching methods. N/A. No personal data will be collected or processed.

Describe the third persons N/A

Types of personal data about third persons N/A

Name N/A

National ID number or other personal identification number N/A

Date of birth N/A

Address or telephone number N/A

Email address, IP address or other online identifiers N/A

Photographs or video recordings of persons N/A

Demographic data that can identify a natural person N/A

Genetic data N/A

Biometric data N/A

Other data that can identify a natural person N/A

Which sample will provide information about third persons? N/A

Will third persons consent to the processing of their personal data? N/A

Will third persons receive information about the processing of their personal data? N/A

Explain why third persons will not be informed. N/A

Documentation

Total number of data subjects in the project (Data subjects: persons whose personal data you will be processing)

- N/A
- 1-99
- 100-999

- 1000-4999
- 5000-9999
- 10.000-49.999
- 50.000-100.000
- 100.000+

How can data subjects get access to their personal data or how they can have their personal data corrected or deleted? N/A

Rights of data subjects (participants) include the right to access one's own personal data and to receive a copy of one's data if asked for. A data subject can request that their personal data are corrected if they feel that the information is wrong or lacking, and the data subject can withdraw consent and request that their personal data are deleted. Give a short description of the procedure for how a data subject can get access to their personal data, and how they can have their personal data corrected or deleted.

Other approvals

Will you obtain any of the following approvals or permits for the project? N/A

Indicate if you will obtain any of the following approvals or permits in order carry out the project.

No approvals or permits are required as no personal data will be collected or used.

- Ethical approval from The Regional Committees for Medical and Health Research Ethics (REC).
- Confidentiality permit (exemption from the duty of confidentiality) from the Regional Committees for Medical and Health Research Ethics (REC)
- Approval from own management for internal quality-assurance and evaluation of health services (intern kvalitetssikring) (The Health Personnel Act § 26)

- Confidentiality permit (exemption from the duty of confidentiality) from the Norwegian Directorate of Health, for quality-assurance and evaluation of health services (kvalitetssikring) (The Health Personnel Act § 29b)
- Biobank
- Confidentiality permit (exemption from the duty of confidentiality) from Statistics Norway (SSB). Statistics Norway has the authority to grant a confidentiality permit for the data that they manage, e.g. data about population, education, employment and social security.
- Approval from The Norwegian Medicines Agency (Statens legemiddelverk, SLV). E.g. for a clinical drugs trial
- Confidentiality permit (exemption from the duty of confidentiality) from a department or directorate
- Other approval. E.g. from a Data Protection Officer

Processing

Where will the personal data be processed? In the framework of this project, data will be simulated, meaning no personal data will be collected, used, stored or processed.

“Processing” includes any collecting, registering, storing, collating, transferring etc. of data. You must indicate all processing of personal data that will take place in the project.

- Computer belonging to the institution responsible for the project N/A
- Computer owned/operated by the data controller. For example, processing data in a private or communal user area on the institution’s server. N/A
- Mobile device belonging to the data controller. Mobile device owned/operated by the data controller. A mobile device can be a laptop, camera, mobile phone etc. N/A

- Physically isolated computer belonging to the data controller. Not connected to other computers or to a network, neither internally nor externally. N/A

- External service or network. Such as providers of cloud storage, online surveys or data storage (such as TSD). Use of an external service or server requires that a data processor agreement is made between the data controller and the external party. N/A

- Private device. Data collection or storage on private devices such as your own computer or mobile phone etc. is not recommended and must be clarified with the institution responsible for the project. N/A

Who will be processing/have access to the collected personal data? N/A

- Project leader N/A

- Student (student project) N/A

- Internal co-workers. Employees of the data controller. N/A

- External co-workers/collaborators inside the EU/EEA. Employees of other institutions that have formalized cooperation with the data controller, or employees of other institutions that are joint data controllers. N/A

- Data processor. An external person or entity that processes personal data on behalf of the data controller, such as an online survey provider, cloud storage provider, translator or transcriber. There must be a data processor agreement or other legal agreement between the data controller and the external party. N/A

- Others with access to the personal data. N/A

Which others will have access to the collected personal data? N/A

Will the collected personal data be made available to a third party or international organisation outside the EEA? This includes when personal data are sent to and stored in a country outside the EEA, or when persons outside this area are given access to personal data stored within the EEA. This means that you cannot use a service provider or outsourced supplier outside the EEA, unless there is a valid basis for the transfer of personal data. Yes No N/A

Give the name of the institution/organisation N/A Give the country of the institution/organisation N/A On what basis will the collected personal data be transferred? N/A

Personal data can be transferred on the basis of an adequate level of protection (art. 45) or on the basis of appropriate safeguards (art. 46). Personal data can also be transferred on the basis of the exception for special situations, but only if the transfer is not repeated, concerns only a limited number of data subjects, is necessary for the purposes of compelling legitimate interests pursued by the data controller (which are not overridden by the interests or rights and freedoms of the data subject), and if the data controller has assessed all the circumstances surrounding the data transfer and has provided suitable safeguards with regard to the protection of personal data (art. 49).

Information Security

No personal data will be used in the project. Therefore, identification and/or security issues are irrelevant for this project.

Will directly identifiable personal data be stored separately from the rest of the collected data (in a scrambling key)?

It is common practice to remove directly identifiable data (name, national ID number, contact details etc.) from the collected data and give each data subject a code/number. A scrambling key is the file/list of names and codes that makes it possible to directly identify data subjects in the collected data. It should be stored separately from the rest of the collected data. In practice, this means that the scrambling key cannot be stored in the same network as the rest of the data, unless the scrambling key is encrypted. Yes No N/A

Explain why directly identifiable personal data will be stored together with the rest of the collected data. N/A

For reasons of information security we recommend the use of a scrambling key in most projects, especially in projects where special categories of personal data (previously “sensitive” personal data) or personal data relating to criminal convictions and offences will be processed.

Which technical and practical measures will be used to secure the personal data?

- Personal data will be anonymized as soon as no longer needed. N/A

  Anonymization involves processing the data in such a way that no individual persons can be identified in the data that you’re left with, i.e. the data can no longer be linked to individual persons in any way.

  Anonymization usually involves: *deleting directly identifiable personal data (including scrambling key/list of names) *deleting or rewriting indirectly identifiable personal data (e.g. deleting or categorizing variables such as age, place of residence, school etc.) *deleting or editing audio recordings, photographs and video recordings.

- Personal data will be transferred in encrypted form. N/A

  Encryption is a mathematical method for ensuring confidentiality in that information cannot be read by unauthorized persons. For example, using an encrypted VPN tunnel or equivalent measure for external login to work-place network.

- Personal data will be stored in encrypted form. N/A

  Encryption is a mathematical method for ensuring confidentiality in that information cannot be read by unauthorized persons. For example, the encryption of a hard drive to ensure the confidentiality of data when the computer is turned off.

- Record of changes. N/A

  Changes in the collected data are recorded/documented with the time of the change and information about the person who made that change.

- Multi-factor authentication. N/A

  A method of access control where a user is granted access after presenting two or more separate pieces of evidence to prove their identity (e.g. password + code sent by text message)

- Restricted access. N/A

  Blocking or restricting access to the collected data for unauthorized persons

- Access log. N/A

  An access log shows who has accessed the collected data and when

- Other security measures. N/A

  For example, locking away documents, automatic screen lock after a short time for mobile devices, partitioning of hard drive, checksum/integrity check etc.

Duration of project

Project period

Will personal data be stored beyond the end of project period? Personal data should not be further processed a way that is inconsistent with the initial purpose(s) for which the data were collected. Anonymous/anonymized data may be stored indefinitely, so long as nothing else has been agreed to by the data subjects.

- No, all collected data will be deleted
- No, the collected data will be stored in anonymous form. Stored in a form where the data can no longer be linked to individual persons in any way
- Yes, collected personal data will be stored until
- Yes, collected personal data will be stored indefinitely.
- Other No personal data will be processed. Questions of storing personal sensitive data are irrelevant for this project.

For what purpose(s) will the collected personal data be stored?

- Research
- Other No personal data will be processed. Questions of storing personal sensitive data are irrelevant for this project.

Where will the collected personal data be stored?

- At the institution responsible for the project (data controller)
- Other No personal data will be processed. Questions of storing personal sensitive data are irrelevant for this project.

Additional information

Will the data subjects be identifiable (directly or indirectly) in the thesis/publications for the project? If personal data are to be published, there should be a scientific purpose for this. Data is usually published in anonymous form. Yes No N/A

Explain why N/A

Additional information N/A

Here you can provide information that may have significance for our assessment of the project, including more detailed information about points covered in the form and information that is not covered by points in the form.

Other attachments N/A e.g. interview guide, questionnaire, information letter and consent form etc.

Appendix II

Data Management and Analysis Code

The documented syntax code for the simulation study is presented to allow for reproducibility of the findings. The study was conducted using version 3.6.2 of R software environment (R Core Team, 2019). The analyses were carried out for sample sizes 250, 1000, 4000, 16000, and were based on 10000 replications.

```
# load required packages
library(kequate)
library(mirt)
library(numDeriv)
library(matrixStats)
# FUNCTIONS
# data-generating function
irtresponse3pl <- function(alpha, delta, chi, theta){
  J <- length(alpha)
  N <- length(theta)
  res <- matrix(0, nrow = N, ncol = J)
  for(j in 1:J){
    res[, j] <- runif(N) < (chi[j] + (1 - chi[j]) /
                               (1 + exp(-alpha[j] * (theta - delta[j])))
   )
  }
  return(res)
}
# PEN1 function; in text - (10)
# adapted from kequate package  (Andersson et al., 2013)
PEN1 <- function(h, r, xx, var, mean){
  xx <- as.vector(xx)
  h <- as.vector(h)
  f <- numeric(length(xx))
  mean <- as.vector(mean)
```

```
  a <- as.vector(sqrt(var/(var+h^2)))
  for(j in 1:length(xx)){
    ff <- 0
    ff <- sum(r * dnorm((xx[j]-a*xx-(1-a)*mean)/(a*h)) / (a*h))
    f[j] <- ff
  }
  pen1 <- sum((r-f)^2)
  return(pen1)
}
# second partial derivative of PEN1 with respect to h;
# in text - (40)-(45)
d2PEN1dh2.function <- function(h, r, xx, mean, var){
  f <- numeric(length(xx))
  mean <- as.vector(mean)
  dfdh <- numeric(length(xx))
  d2fdh2 <- numeric(length(xx))
  a <- as.vector(sqrt(var/(var+h^2)))
  dadh <- -(h*(sqrt(var)))/((h^2 + var)^(3/2))
  d1divahdh <- -a^(-2)*dadh*(1/h)- (1/h^2)*(1/a)
  d2adh2 <- -sqrt(var)*((1/(h^2 + var)^(3/2)) - ((3*h^2)/(h^2 + var)^(5/2)))
  d21divahdh2 <- 1/(h*a^2)*(2/a*dadh^2 - d2adh2 + dadh* 1/h) +
    + (dadh*(1/(a^2*h^2)) + 2/(a*h^3))
  for (j in 1:length(xx)){
    Rx <- (xx[j]-a*xx-(1-a)*mean)/(a*h)
    dRdh <- ((mean-xx)*dadh)*(1/(a*h)) + (xx[j] - a*xx - (1-a)*mean)*d1divahdh
    d2Rdh2 <- (mean-xx)*(d2adh2*(1/(a*h)) + dadh*d1divahdh) + ((mean - xx)*dadh*d1divahdh + (xx[j]-a*xx-(1-a)*mean)*d21divahdh2)
    dphiRxdh <- -dnorm(Rx)*dRdh*Rx
    d2phiRxdh2 <- -dphiRxdh*dRdh*Rx-dnorm(Rx)*d2Rdh2*Rx-dnorm(Rx)*dRdh^2
```

```
  ff<-0
  ff <- sum(r*dnorm(Rx)/(a*h))
  f[j] <- ff
  df <- 0
  df <- sum(r*(dnorm(Rx)*d1divahdh+dphiRxdh*(1/(a*h))))
  dfdh[j] <- df
  d2f <- 0
  d2f <- sum(r*((d2phiRxdh2*1/(a*h) + dphiRxdh*d1divahdh) + (dphiRxdh
   *d1divahdh + dnorm(Rx)*d21divahdh2)))
  d2fdh2[j] <- d2f
 }
 d2PEN1dh2 <- -2*sum(d2fdh2*(r-f)-(dfdh^2))
 return(d2PEN1dh2)
}
# second partial derivative of PEN1 with respect to r_i;
# in text - (46)-(57)
d2PEN1dhdr.function <- function(h, r, xx){
 mean <- sum(r * xx)
 var <- sum(r * (xx - mean)^2)
 xx <- as.vector(xx)
 a <- as.vector(sqrt(var/(var+h^2)))
 dadh <- -(h*(sqrt(var)))/((h^2 + var)^(3/2))
 d1divahdh <- -a^(-2)*dadh*(1/h)-(1/h^2)*(1/a)
 f <- numeric(length(xx))
 dfdh <- numeric(length(xx))
 dfdr <- matrix(0, length(xx), ncol = length(r))
 d2fdhdr <- matrix(0, length(xx), ncol = length(r))
 d2PEN1dhdr <- numeric(length(r))
 for(j in 1:length(xx)){
  Rx <- (xx[j]-a*xx-(1-a)*mean)/(a*h)
  dRdh <- ((mean-xx)*dadh)*(1/(a*h)) + (xx[j] - a*xx - (1-a)*mean)*
   d1divahdh
```

```
    dphiRxdh <- -dnorm(Rx)*dRdh*Rx
    d1divadr <- - 1/2*a*(h^2/var)*((xx - mean)^2/var)
    dRdr <- (-1/(a*h))*((1/2)*(xx[j]-mean)*(1-a^2)*((xx-mean)^2/var)
   +(1-a)*xx)
    ff <-0
    ff <- sum(r*dnorm(Rx)/(a*h))
    f[j] <- ff
    df <- 0
    df <- sum(r*(dnorm(Rx)*d1divahdh+dphiRxdh*(1/(a*h))))
    dfdh[j] <- df
    dfdr[j, ] <- (1/h)*(dnorm(Rx)/a - dRdr * sum(r * dnorm(Rx) *  Rx)
   *(1/a) + d1divadr * sum(r * dnorm(Rx)))
    d2Rdhdr <- (1/(a*h^2) - a/var)*((1/2)*(xx[j]-mean)*(1-a^2)*((xx-
   mean)^2/var)+(1-a)*xx) + (-1/(a*h))*((-(xx[j]-mean)*((xx - mean)^2 /
    var)*a*dadh)-xx*dadh)
    P1 <- dphiRxdh*(1/a) + ((a*h)/var)*dnorm(Rx)
    P2 <- d2Rdhdr*sum(r*dnorm(Rx)*Rx)*(1/a) + dRdr*sum(r*dphiRxdh*Rx +
   r*dnorm(Rx)*dRdh)*(1/a) + dRdr*sum(r*dnorm(Rx)*Rx)*((a*h)/var)
    P3<- (-((xx-mean)^2/var)/(2*var))*(dadh*h^2 + 2*a*h)*sum(r*dnorm(Rx
   )) + d1divadr*sum(r*dphiRxdh)
    P <- P1 - P2 + P3
    d2fdhdr[j,] <- (-1/h^2)*(dnorm(Rx)/a - dRdr * sum(r * dnorm(Rx) *
   Rx)*(1/a) + d1divadr * sum(r * dnorm(Rx))) + (1/h)* P
  }
  for(i in 1:length(xx)){
    dr_i_j.vector <- numeric(length(xx))
    dr_i_j.vector[i] <- 1.0
    d2PEN1dhdr[i] <- -2 * sum(d2fdhdr[ ,i] * (r - f) + dfdh * (dr_i_j.
   vector - dfdr[ ,i]))
  }
  return(d2PEN1dhdr)
}
```

```
# first partial derivative of the equating function eY with respect to
    h_X
# in text - (28)-(29)
deYdh_X.function <- function(h, r, xx, var, mean, G.prime){
  mean <- as.vector(mean)
  xx <- as.vector(xx)
  h <- as.vector(h)
  a <- as.vector(sqrt(var/(var+h^2)))
  dadh <- -(h*(sqrt(var)))/((h^2 + var)^(3/2))
  d1divahdh <- -a^(-2)*dadh*(1/h)- (1/h^2)*(1/a)
  dFdh <- numeric(length(xx))
  for(j in 1:length(xx)){
    Rx <- (xx[j]-a*xx-(1-a)*mean)/(a*h)
    dRdh <- dadh*(mean-xx)*1/(a*h) + (xx[j]-a*xx-(1-a)*mean)*d1divahdh
    dF <- 0
    dF <- sum(r*dnorm(Rx)*dRdh)
    dFdh[j] <- dF
  }
  deYdhx <- (1/G.prime)*dFdh
  return(deYdhx)
}
# first partial derivative of the equating function eY with respect to
    h_Y
# in text - (30)-(31)
deYdh_Y.function <- function(h, r, xx, var, mean, G.prime, eqscore){
  mean <- as.vector(mean)
  xx <- as.vector(xx)
  h <- as.vector(h)
  a <- as.vector(sqrt(var/(var+h^2)))
  dadh <- -(h*(sqrt(var)))/((h^2 + var)^(3/2))
  d1divahdh <- -a^(-2)*dadh*(1/h)- (1/h^2)*(1/a)
  dGdh <- numeric(length(xx))
```

```
  for(j in 1:length(eqscore)){
    dRdh <- dadh*(mean-xx)*1/(a*h) + (xx[j]-a*xx-(1-a)*mean) *
   d1divahdh
    dG<-0
    dG <-sum(r*(dnorm((eqscore[j]-a*xx-(1-a)*mean)/(a*h)))*dRdh)
    dGdh[j] <- dG
  }
  deYdhx <- -(1/G.prime)*dGdh
  return(deYdhx)
}
# score probabilities function
stats <- function(model){
  freq <- fscores(model, method = "EAPsum", full.scores = FALSE)[,c(1,
   4, 5)]
  prop <- freq$expected/sum(freq$expected)
}
# G.prime function; the density of G evaluated at eY(x)
# in text - (19)
# adapted from kequate package (Andersson et al., 2013)
G.prime.function <-function(r, h, var, mean, eqscore, xx){
  h <- as.vector(h)
  a <- as.vector(sqrt(var/(var+h^2)))
  xx <- as.vector(xx)
  mean <- as.vector(mean)
  f <- numeric(length(eqscore))
  for(i in 1:length(eqscore)){
    ff <- 0
    ff <- sum(r*dnorm((eqscore[i]-a*xx-(1-a)*mean)/(a*h))/(a*h) )
    f[i]<-ff
  }
  return(f)
}
```

```
# first derivative of F(x) with respect to r
# in text - (20)-(22)
# adapted from kequate package (Andersson et al., 2013)
dFdr.function <-function(r, h, var, mean, F.prime, eqscore, xx){

  xx <- as.vector(xx)
  h <- as.vector(h)
  mean <- as.vector(mean)
  var <- as.vector(var)
  a <- as.vector(sqrt(var/(var+h^2)))
  dFdrest<-matrix(0, length(eqscore), ncol=length(xx))
  for(j in 1:length(eqscore)){
    Rx <- (eqscore[j]-a*xx-(1-a)*mean)/(a*h)
    Mx <- (1/2)*(eqscore[j]-mean)*(1-a^2)*(((xx-mean)/sqrt(var))^2)+(1-a)*xx
    dFdrest[j, ] <- pnorm(Rx) - Mx * F.prime[j]
  }
  return(dFdrest)
}

# SIMULATION SETUP
# item parameters for test forms X & Y
set.seed(1234567)
apar_X <- runif(20, 1, 2)
bpar_X <- rnorm(20)
cpar_X <- rep(0, 20)
set.seed(260688)
apar_Y <- runif(20, 1, 2)
bpar_Y <- rnorm(20)
cpar_Y <- rep(0, 20)
# sample sizes
n_sample_sizes <- c(250, 1000, 4000, 16000)
```

```
# n of iterations per sample size
n_iter <- 10000
# complete datasets with all iterations for all sample sizes
h_data <- matrix(ncol = 5, nrow = 0)
colnames(h_data) <- c("N", "h_x", "ASEhx", "h_y", "ASEhy")
ASEE_data <- matrix(ncol = 22, nrow = 0)
colnames(ASEE_data) <- c("n", c(1:21))
eqYx_data <- matrix(ncol = 22, nrow = 0)
colnames(eqYx_data) <- c("N", c(1:21))
ASEE_mod_data <- matrix(ncol = 22, nrow = 0)
colnames(ASEE_mod_data) <- c("N", c(1:21))
# temporary matrices to store iterations
h_mat <- matrix(ncol = 5, nrow = n_iter)
colnames(h_mat) <- c("N", "h_x", "ASEhx", "h_y", "ASEhy")
ASEE_mat <- matrix(ncol = 22, nrow = n_iter)
colnames(ASEE_mat) <- c("N", c(1:21))
eqYx_mat <- matrix(ncol = 22, nrow = n_iter)
colnames(eqYx_mat) <- c("N", c(1:21))
ASEE_mod_mat <- matrix(ncol = 22, nrow = n_iter)
colnames(ASEE_mod_mat) <- c("N", c(1:21))
# matrix to store summary of all iterations for a sample size
h_summary <- matrix(ncol = 5, nrow = length(n_sample_sizes))
colnames(h_summary) <- c("N", "mean_ASEhx", "MCSEhx", "mean_ASEhy", "MCSEhy")
ASEE_summary <- matrix(ncol = 22, nrow = length(n_sample_sizes))
colnames(ASEE_summary) <- c("N", c(1:21))
eqYx_summary <- matrix(ncol = 22, nrow = length(n_sample_sizes))
colnames(eqYx_summary) <- c("N", c(1:21))
ASEE_mod_summary <- matrix(ncol = 22, nrow = length(n_sample_sizes))
colnames(ASEE_mod_summary) <- c("N", c(1:21))

idx <- 1
```

```r
for (n in 1:length(n_sample_sizes)){
  for (iter in seq(1, n_iter)){
    seed <- seq(262626, 262626 + n_iter-1, 1)
    set.seed(seed[iter])
    print(n_sample_sizes[iter])
    # original data P and Q
    dat_og_P <- data.frame(irtresponse3pl(apar_X, bpar_X, cpar_X, rnorm(n_sample_sizes[n])))
    dat_og_Q <- data.frame(irtresponse3pl(apar_Y, bpar_Y, cpar_Y, rnorm(n_sample_sizes[n])))
    # pre-smooth raw data with mirt and extract score probability distributions
    mod_P <- mirt(dat_og_P[,1:20], 1, rep("2PL", 20), SE = TRUE)
    prop_R <- stats(mod_P)
    mod_Q <- mirt(dat_og_Q[,1:20], 1, rep("2PL", 20), SE = TRUE)
    prop_S <- stats(mod_Q)
    # minimize PEN1 with respect to h, extract minimum h_X and h_Y
    # tol = .Machine$double.eps^0.5 to match tolerance used in irtose function of kequate package (Andersson et al., 2013)
    h.x <- optimize(PEN1, c(0, 10), tol = .Machine$double.eps^0.5, r = prop_R, x = 0:20, var =  sum(prop_R * ((0:20) - sum(prop_R * (0:20)))^2), mean = sum(prop_R * (0:20)))$minimum
    h.y <- optimize(PEN1, c(0, 10), tol = .Machine$double.eps^0.5, r = prop_S, x = 0:20, var =  sum(prop_S * ((0:20) - sum(prop_S * (0:20)))^2), mean = sum(prop_S * (0:20)))$minimum
    # equate X and Y, and extract the ASEE without accounting for h, and the eqYx
    # (Andersson et al., 2013)
    sim2plan <- irtose("EG", mod_P, mod_Q, 0:20, 0:20)
    ASEE_eYx_no_h <- sim2plan@equating$SEEYx
    eqYx <- sim2plan@equating$eqYx
    # calculate asymptotic variance and SE of h_X and h_Y
```

```
d2PEN1dhx2 <- d2PEN1dh2.function(h = h.x, r = prop_R, xx = 0:20, var =  sum(prop_R * ((0:20) - sum(prop_R * (0:20)))^2), mean = sum(prop_R * (0:20)))
d2PEN1dhy2 <- d2PEN1dh2.function(h = h.y, r = prop_S, xx = 0:20, var =  sum(prop_S * ((0:20) - sum(prop_S * (0:20)))^2), mean = sum(prop_S * (0:20)))
d2PEN1dhdr <- d2PEN1dhdr.function(h = h.x, r = prop_R, xx = 0:20)
d2PEN1dhds <- d2PEN1dhdr.function(h = h.y, r = prop_S, xx = 0:20)
dhdr <- (-(d2PEN1dhx2)^-1) %*% d2PEN1dhdr
dhds <- (-(d2PEN1dhy2)^-1) %*% d2PEN1dhds
ASE_hx <- sqrt(dhdr %*% sim2plan@scores$covrs[1:21, 1:21] %*% t(dhdr))
ASE_hy <- sqrt(dhds %*% sim2plan@scores$covrs[22:42, 22:42] %*% t(dhds))
#  in order to calulate (27) (in text), we need 1) deYdr; 2)deYds; 3)deYdhx; 4)deYdhy.
F.primeY <- G.prime.function(r = prop_R, h = h.x, mean = sum(prop_R * (0:20)), var = sum(prop_R * ((0:20) - sum(prop_R * (0:20)))^2), eqscore = 0:20, xx = 0:20)
G.primeY <- G.prime.function(r = prop_S, h = h.y, mean = sum(prop_S * (0:20)), var = sum(prop_S * ((0:20) - sum(prop_S * (0:20)))^2), eqscore = eqYx, xx = 0:20)
dFdreY <- dFdr.function(prop_R, h.x, var = sum(prop_R * ((0:20) - sum(prop_R * (0:20)))^2), mean = sum(prop_R * (0:20)), F.primeY, eqscore = 0:20 , xx = 0:20)
dGdseY <- dFdr.function(prop_S, h.y, var = sum(prop_S * ((0:20) - sum(prop_S * (0:20)))^2), mean = sum(prop_S * (0:20)), G.primeY, eqscore = eqYx, xx = 0:20)
deYdr <- (1/G.primeY) * dFdreY # (von Davier et al., 2004)
deYds <- (-(1/G.primeY)) * dGdseY # (von Davier et al., 2004)
# first partial derivatives of eY with respect to h_x and h_y
# in text - (28)-(31)
```

```
    deYdhx <- deYdh_X.function(h.x, r = prop_R, xx = 0:20, var = sum(prop_R * ((0:20) - sum(prop_R * (0:20)))^2), mean = sum(prop_R * (0:20)), G.primeY)
    deYdhy <- deYdh_Y.function(h.y, r = prop_S, xx = 0:20, var = sum(prop_S * ((0:20) - sum(prop_S * (0:20)))^2), mean = sum(prop_S * (0:20)), G.primeY, eqscore = eqYx)
    # matrix with first derivatives of eY with respect to r, s, h_x and h_y
    # in text - 1st and 3rd components of (27)
    JeY <- matrix(0, nrow= 21, ncol=44)
    JeY[1:21, 1:21] <- deYdr
    JeY[1:21, 23:43] <- deYds
    JeY[, 22] <- deYdhx
    JeY[, 44] <- deYdhy
    # matrix RhxShy; in text - (32)
    drhxshy <- matrix(0, nrow = 44, ncol = 42)
    drhxshy[1:21, 1:21] <- diag(21)
    drhxshy[22, 1:21] <- dhdr
    drhxshy[23:43, 22:42] <- diag(21)
    drhxshy[44, 22:42] <- dhds
    covRhSh <- drhxshy %*% sim2plan@scores$covrs %*% t(drhxshy)
    # calculate variance of eYx and ASEE; in text - (27)
    ASEE_mod_eYx <- sqrt(diag(abs(JeY%*%covRhSh%*%t(JeY))))
    # record in the temporary matrix
    h_mat[iter,] <- c(n_sample_sizes[n], h.x, ASE_hx, h.y, ASE_hy)
    eqYx_mat[iter, 1:22] <- c(n_sample_sizes[n], eqYx)
    ASEE_mat[iter, 1:22] <- c(n_sample_sizes[n], ASEE_eYx_no_h)
    ASEE_mod_mat[iter, 1:22] <- c(n_sample_sizes[n], ASEE_mod_eYx)
  }
  h_data <- rbind(h_data, h_mat)
  eqYx_data <- rbind(eqYx_data, eqYx_mat)
  SEE_mod_data <- rbind(ASEE_mod_data, ASEE_mod_mat)
```

```
  ASEE_data <- rbind(ASEE_data, ASEE_mat)
  # results summaries for (1) ASEhx, (2) ASEE - NOT accounting for h, (3) ASEE_mod - accounting for h, (4) eqYx - MCSE
  h_summary[idx, ] <- c(n_sample_sizes[n], mean(h_mat[,3]), sqrt(var(h_mat[,2])), mean(h_mat[,5]), sqrt(var(h_mat[,4])))
  eqYx_summary[idx, ] <- c(n_sample_sizes[n], sqrt(colVars(eqYx_mat[,2:22])))
  ASEE_summary[idx, ] <- c(n_sample_sizes[n], colMeans(ASEE_mat[,2:22]))
  ASEE_mod_summary[idx, ] <- c(n_sample_sizes[n], colMeans(ASEE_mod_mat[,2:22]))
  idx <- idx + 1
}
```

Appendix III

Computation of the Penalty Function Derivatives

In order to compute (24), two partial derivatives of the $\text{PEN}_1(h_X)$ function need to be defined, the second-order direct partial derivative of $\text{PEN}_1(h_X)$ with respect to h_X and the second-order cross partial derivative of $\text{PEN}_1(h_X)$ with respect to $\boldsymbol{r}$. Clearly, the computation of these derivatives subsequently requires defining the first partial derivative of $\text{PEN}_1(h_X)$ with respect to h_X.

Recalling (8) - (11), we define the first partial derivative of $\text{PEN}_1(h_X)$ with respect to h_X as

$$\begin{aligned}\frac{\partial \text{PEN}_1}{\partial h_X} &= \frac{\partial}{\partial h_X}[(\hat{r_j} - \hat{f}_{hX}(x_j)^2] \\ &= -2\sum_j \left((\hat{r_j} - \hat{f}_{hX}(x_j)) \frac{\partial \hat{f}_{hX}(x_j)}{\partial h_X} \right). \end{aligned} \tag{34}$$

We then need to calculate $\frac{\partial \hat{f}_{hX}(x_j)}{\partial h_X}$ as

$$\begin{aligned}\frac{\partial \hat{f}_{hX}(x_j)}{\partial h_X} &= \frac{\partial}{\partial h_X}\left[\sum_j \hat{r_j}\phi(R_{jX}(x))\frac{1}{a_X h_X}\right] \\ &= \sum_j \hat{r_j}\left(\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}\frac{1}{a_X h_X} + \phi(R_{jX}(x))\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right]\right), \end{aligned} \tag{35}$$

where $\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}$ and $\frac{\partial R_{jX}(x)}{\partial h_X}$ are defined as

$$\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X} = -\phi(R_{jX}(x))\frac{\partial R_{jX}(x)}{\partial h_X}R_{jX}(x), \tag{36}$$

$$\begin{aligned}\frac{\partial R_{jX}(x)}{\partial h_X} &= \frac{\partial}{\partial h_X}\left[\frac{x - a_X x_j - (1-a_X)\mu_X}{a_X h_X}\right] \\ &= \frac{\partial a_X}{\partial h_X}(\mu_X - x_j)\frac{1}{a_X h_X} + (x - a_X x_j - (1-a_X)\mu_X)\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right]. \end{aligned} \tag{37}$$

The remaining components needed for computing the first partial derivative with respect to h_X are then $\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right]$ and $\frac{\partial a_X}{\partial h_X}$. Thus we calculate

$$\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right] = -a_X^{-2}\frac{\partial a_X}{\partial h_X}\frac{1}{h_X} - \frac{1}{a_X h_X^2}, \tag{38}$$

$$\frac{\partial a_X}{\partial h_X} = \frac{\partial}{\partial h_X}\left[\frac{\sigma_X}{\sqrt{\sigma_X^2 + h_X^2}}\right] = -\frac{\sigma_X}{2(h_X^2+\sigma_X^2)^{\frac{3}{2}}}\frac{\partial h_X^2}{\partial h_X} + \frac{\partial \sigma_X^2}{\partial h_X}$$
$$= -\frac{\sigma_X h_X}{(h_X^2+\sigma_X^2)^{\frac{3}{2}}}. \tag{39}$$

Using (34) - (39), we can then compute the second partial derivative of $\mathrm{PEN}_1(h_X)$ with respect to h_X as

$$\frac{\partial^2 \mathrm{PEN}_1}{\partial h_X^2} = \frac{\partial}{\partial h_X}\left[-2\sum_j\left((\hat{r_j} - \hat{f}_{hX}(x_j))\frac{\partial \hat{f}_{hX}(x_j)}{\partial h_X}\right)\right]$$
$$= -2\sum_j\left(\frac{\partial^2 \hat{f}_{hX}(x_j)}{\partial h_X^2}(\hat{r_j} - \hat{f}_{hX}(x_j)) - \left[\frac{\partial \hat{f}_{hX}(x_j)}{\partial h_X}\right]^2\right), \tag{40}$$

where $\frac{\partial \hat{f}_{hX}(x_j)}{\partial h_X}$ is defined in (35) and $\frac{\partial^2 \hat{f}_{hX}(x_j)}{\partial h^2}$ is given by

$$\frac{\partial^2 \hat{f}_{hX}(x_j)}{\partial h_X^2} = \frac{\partial}{\partial h_X}\left[\sum_j \hat{r_j}\left(\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}\frac{1}{a_X h_X} + \phi(R_{jX}(x))\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right]\right)\right]$$
$$= \sum_j \hat{r_j}\left(\frac{\partial^2[\phi(R_{jX}(x))]}{\partial h^2}\frac{1}{a_X h_X} + \frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right]\right) +$$
$$+\sum_j \hat{r_j}\left(\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right] + \phi(R_{jX}(x))\frac{\partial^2}{\partial h_X^2}\left[\frac{1}{a_X h_X}\right]\right). \tag{41}$$

Recall that $\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}$ is given in (36) and $\frac{\partial R_{jX}(x)}{\partial h_X}$ - in (37). Hence, we define $\frac{\partial^2[\phi(R_{jX}(x))]}{\partial h_X^2}$ and $\frac{\partial^2 R_{jX}(x)}{\partial h_X^2}$ as

$$\frac{\partial^2[\phi(R_{jX}(x))]}{\partial h_X^2} = \frac{\partial}{\partial h_X}\left[-\phi(R_{jX}(x))\frac{\partial R_{jX}(x)}{\partial h_X}R_{jX}(x)\right]$$
$$= -\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}\frac{\partial R_{jX}(x)}{\partial h_X}R_{jX}(x) - \phi(R_{jX}(x))\frac{\partial^2 R_{jX}(x)}{\partial h_X^2}R_{jX}(x) -$$
$$-\phi(R_{jX}(x))\left[\frac{\partial[R_{jX}(x)]}{\partial h_X})\right]^2, \tag{42}$$

$$\frac{\partial^2 R_{jX}(x)}{\partial h_X^2} = \frac{\partial}{\partial h_X}\left[\frac{\partial a_X}{\partial h_X}(\mu_X - x_j)\frac{1}{a_X h_X} + (x - a_X x_j - (1-a_X)\mu_X)\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right]\right]$$
$$= (\mu_X - x_j)\left(\frac{\partial^2 a_X}{\partial h_X^2}\frac{1}{a_X h_X} + \frac{\partial a_X}{\partial h_X}\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right]\right) +$$
$$+(\mu_X - x_j)\frac{\partial a_X}{\partial h_X}\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right] + (x - a_X x_j - (1-a_X)\mu_X)\frac{\partial^2}{\partial h_X^2}\left[\frac{1}{a_X h_X}\right]. \tag{43}$$

Consider further that $\frac{\partial}{\partial h_X}\left[\frac{1}{a_X h_X}\right]$ is defined in (38), $\frac{\partial a_X}{\partial h}$ - in (39). We can then observe that $\frac{\partial^2}{\partial h_X^2}\left[\frac{1}{a_X h_X}\right]$ can be computed as

$$\frac{\partial^2}{\partial h_X^2}\left[\frac{1}{a_X h_X}\right] = \frac{1}{h_X a_X^2}\left(\frac{2}{a_X}\left[\frac{\partial a_X}{\partial h_X}\right]^2 - \frac{\partial^2 a_X}{\partial h_X^2} + \frac{\partial a_X}{\partial h_X}\frac{1}{h_X}\right) + \\ + \left(\frac{\partial a_X}{\partial h_X}\frac{1}{a_X^2 h_X^2} + \frac{2}{a_X h_X^3}\right), \tag{44}$$

and

$$\frac{\partial^2 a_X}{\partial h_X^2} = \frac{\partial}{\partial h_X}\left[-\frac{\sigma_X h_X}{(h_X^2+\sigma_X^2)^{\frac{3}{2}}}\right] = \frac{-\sigma_X}{(h_X^2+\sigma_X^2)^{\frac{3}{2}}} - \frac{3h_X^2}{(h_X^2+\sigma_X^2)^{\frac{5}{2}}}. \tag{45}$$

Lastly, we can compute the second partial derivative of $\text{PEN}_1(h_X)$ with respect to $\boldsymbol{r}$ as follows

$$\frac{\partial^2 \text{PEN}_1}{\partial h_X \partial r_i} = \frac{\partial}{\partial r_i}\left[-2\sum_j\left((\hat{r_j} - \hat{f}_{h_X}(x_j))\frac{\partial \hat{f}_{h_X}(x_j)}{\partial h_X}\right)\right] \\ = -2\sum_j\left[\left(\frac{\partial r_j}{\partial h_X \partial r_i} - \frac{\partial \hat{f}_{h_X}(x_j)}{\partial r_i}\right)\frac{\partial \hat{f}_{h_X}(x_j)}{\partial h_X} + (\hat{r_j} - \hat{f}_{h_X}(x_j))\frac{\partial^2 \hat{f}_{h_X}(x_j)}{\partial h_X \partial r_i}\right], \tag{46}$$

where $\frac{\partial r_j}{\partial h_X \partial r_i} = 1$ if $i = j$, and $\frac{\partial r_j}{\partial h_X \partial r_i} = 0$ if $i \neq j$. Note that $\frac{\partial \hat{f}_{h_X}(x_j)}{\partial h_X}$ is given in (35). Then, the components needed for computing (46) are $\frac{\partial \hat{f}_{h_X}(x_j)}{\partial r_i}$ and $\frac{\partial^2 \hat{f}_{h_X}(x_j)}{\partial h_X \partial r_i}$. We define $\frac{\partial \hat{f}_{h_X}(x_j)}{\partial r_i}$ as

$$\frac{\partial \hat{f}_{h_X}(x_j)}{\partial r_i} = \frac{\partial}{\partial r_i}\left[\sum_j r_j \phi(R_{jX}(x))\frac{1}{a_X h_X}\right] \\ = \frac{1}{h_X}\left[\phi(R_{jX}(x))\frac{1}{a_X} - \frac{\partial R_{jX}(x)}{\partial r_i}\sum_j\left(r_j \phi(R_{jX}(x))R_{jX}(x)\right)\frac{1}{a_X}\right] + \\ + \frac{1}{h_X}\left[\frac{\partial}{\partial r_i}\left[\frac{1}{a_X}\right]\sum_j (r_j \phi(R_{jX}(x)))\right], \tag{47}$$

where $\frac{\partial R_{jX}}{\partial r_i}$ and $\frac{\partial}{\partial r}\left[\frac{1}{a_X}\right]$ are given in Holland et al. (1989) as

$$\frac{\partial}{\partial r_i}\left[\frac{1}{a_X}\right] = -\frac{1}{2}a_X \frac{h_X^2}{\sigma_X^2}\frac{x_i^2 - \mu_X^2}{\sigma_X^2}, \tag{48}$$

and

$$\frac{\partial R_{jX}}{\partial r_i} = \left(-\frac{1}{a_X h_X}\right)\left[\frac{1}{2}(x - \mu_X)(1 - a_X^2)\left(\frac{x_i^2 - \mu_X^2}{\sigma_X^2}\right) + (1 - a_X)x_i\right]. \tag{49}$$

We further define $\frac{\partial^2 \hat{f}_{hX}(x_j)}{\partial h_X \partial r_i}$ as

$$\frac{\partial^2 \hat{f}_{hX}(x_j)}{\partial h_X \partial r} = \frac{\partial}{\partial h_X}\left[\frac{\partial \hat{f}_{hX}(x_j)}{\partial r_i}\right]. \tag{50}$$

Given (47) is a lengthy expression, we further simplify the notation such that

$$\begin{aligned}\frac{\partial}{\partial h_X}\left[\frac{\partial \hat{f}_{hX}(x_j)}{\partial r_i}\right] &= \frac{\partial}{\partial h_X}[\frac{1}{h_X}] \times P + \frac{1}{h_X} \times \frac{\partial P}{\partial h_X} \\ &= -\frac{1}{h_X^2} \times P + \frac{1}{h_X} \times \frac{\partial P}{\partial h_X},\end{aligned} \tag{51}$$

where

$$\begin{aligned}P = \frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}\frac{1}{a_X} - \frac{\partial R_{jX}(x)}{\partial r_i}\sum_j (r_j \phi(R_{jX}(x))R_{jX}(x))\frac{1}{a_X} + \\ + \frac{\partial}{\partial r_i}\left[\frac{1}{a_X}\right]\sum_j (r_j \phi(R_{jX}(x))).\end{aligned} \tag{52}$$

Noting the three components in (52), $\frac{\partial P}{\partial h_X}$ can then be presented as follows

$$\frac{\partial P}{\partial h_X} = \frac{\partial P1}{\partial h_X} - \frac{\partial P2}{\partial h_X} + \frac{\partial P3}{\partial h_X}. \tag{53}$$

$\frac{\partial P1}{\partial h_X}$ is given by

$$\begin{aligned}\frac{\partial P1}{\partial h_X} &= \frac{\partial}{\partial h_X}\left[\phi(R_{jX}(x))\frac{1}{a_X}\right] \\ &= \frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}\frac{1}{a_X} + \phi(R_{jX}(x))\frac{a_X h_X}{\sigma_X^2},\end{aligned} \tag{54}$$

where $\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}$ is given in (36). $\frac{\partial P2}{\partial h_X}$ is defined as

$$\begin{aligned}\frac{\partial P2}{\partial h_X} &= \frac{\partial}{\partial h_X}\left[\frac{\partial R_{jX}(x)}{\partial r_i}\sum_j (r_j \phi(R_{jX}(x))R_{jX}(x))\frac{1}{a_X}\right] \\ &= \frac{\partial^2 R_{jX}(x)}{\partial h_X \partial r_i}\sum_j (r_j \phi(R_{jX}(x))R_{jX}(x))\frac{1}{a_X} + \\ &+ \frac{\partial R_{jX}(x)}{\partial r_i}\sum_j \left(r_j \frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}R_{jX}(x) + r_j \phi(R_{jX}(x))\frac{\partial R_{jX}(x)}{\partial h_X}\right)\frac{1}{a_X} + \\ &+ \frac{\partial R_{jX}(x)}{\partial r_i}\sum_j (r_j \phi(R_{jX}(x))R_{jX}(x))\frac{a_X h_X}{\sigma_X^2},\end{aligned} \tag{55}$$

where $\frac{\partial R_{jX}(x)}{\partial r_i}$ is defined in (49), $\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}$ - in (36), and $\frac{\partial^2 R_{jX}(x)}{\partial h_X \partial r_i}$ is given by

$$\begin{aligned}\frac{\partial^2 R_{jX}(x)}{\partial h_X \partial r_i} &= \frac{\partial}{\partial h_X}\left[(-\frac{1}{a_X h_X})(\frac{1}{2}(x-\mu_X)(1-a_X^2)\left(\frac{x^2-\mu_X^2}{\sigma_X^2}\right) + (1-a_X)x)\right] \\ &= \left(\frac{1}{a_X h_X^2} - \frac{a_X}{\sigma_X^2}\right)\left(\frac{1}{2}(x-\mu_X)(1-a_X^2)\left(\frac{x_i^2-\mu_X^2}{\sigma_X^2}\right) + (1-a_X)x\right) \\ &+ \left(-\frac{1}{a_X h_X}\right)\left(\left(-(x-\mu_X)\left(\frac{x^2-\mu_X^2}{\sigma_X^2}\right)a_X\frac{\partial a_X}{\partial h_X}\right) + x\frac{\partial a_X}{\partial h_X}\right).\end{aligned} \tag{56}$$

It remains to calculate $\frac{\partial P3}{\partial h_X}$ as follows

$$\begin{aligned}\frac{\partial P3}{\partial h_X} &= \frac{\partial}{\partial h_X}\left[\frac{\partial}{\partial r_i}\left[\frac{1}{a_X}\right]\sum_j (r_j\phi(R_{jX}(x)))\right] \\ &= \frac{\partial^2}{\partial h_X \partial r_i}\left[\frac{1}{a}\right]\sum_j (r_j\phi(R_{jX}(x))) + \frac{\partial}{\partial r_i}\left[\frac{1}{a_X}\right]\frac{\partial}{\partial h_X}\left[\sum_j (r_j\phi(R_{jX}(x)))\right] \\ &= -\left(\left(\frac{x_i^2-\mu_X^2}{\sigma_X^2}\right)\frac{1}{2\sigma_X^2}\right)\left(\frac{\partial a_X}{\partial h_X}h_X^2 + 2a_X h_X\right)\sum_j (r_j\phi(R_{jX}(x))) + \\ &+ \frac{\partial}{\partial r_i}\left[\frac{1}{a_X}\right]\sum_j\left(r_j\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}\right), \end{aligned} \tag{57}$$

where $\frac{\partial}{\partial r_i}\left[\frac{1}{a_X}\right]$ is given in (48), $\frac{\partial a_X}{\partial h_X}$ - in (39), and $\frac{\partial[\phi(R_{jX}(x))]}{\partial h_X}$ - in (36).

The partial derivatives of the $\mathrm{PEN}_1(h_Y)$ with respect to h_Y, $\frac{\partial \mathrm{PEN}_1}{\partial h_Y}$, $\frac{\partial^2 \mathrm{PEN}_1}{\partial h_Y^2}$ and $\frac{\partial^2 \mathrm{PEN}_1}{\partial h_Y \partial s_i}$, are computed analogously.