# On the Use of Nonparametric ICC Estimation Techniques For Checking Parametric Model Fit

March 27, 2004

Young-Sun Lee
Teachers College, Columbia University

James A.Wollack
University of Wisconsin – Madison

Jeffrey Douglas
University of Illinois – Urbana Champaign

Paper presented at the annual conference of the American Educational Research Association, San Diego, CA.

Keywords: Nonparametric IRT and model fit, Kernel Smoothing, Isotonic Regression, ICC Estimation

# On the Use of Nonparametric ICC Estimation Techniques For Checking Parametric Model Fit

## Abstract

This study had two purposes, 1) to investigate the performance of three nonparametric ICC estimation procedures relative to a two parameter logistic (2PL) model using marginal maximum likelihood estimation (MMLE), both visually and numerically, and 2) to develop a statistical test for assessing the model fit of the 2PL by comparison with the nonparametric ICC estimation procedures. A simulation study was conducted to investigate these issues. Results from root integrated squared error (RISE) and mean absolute deviation (MAD) confirmed that the 2PL MMLE and smoothed isotonic regression estimation are comparatively good for ICC estimation when the items fit an underlying 2PL model. The smoothed isotonic regression estimation procedure employed with an appropriate kernel function, however, provided the best fit for non-model fitting items. In particular, smoothed isotonic regression yielded the smallest RISE and MAD values, while also satisfying the assumption of monotonicity. As the number of items and the sample size increased, the differences among the nonparametric ICC estimation procedures became less pronounced. The Type I probabilities of the statistical test for model fit were very close to those expected for all sample sizes and test lengths. Power to detect items not fitting the 2PL was best for the smoothed isotonic regression method, but was very good for all three nonparametric ICC methods in all conditions studied.

## Purpose

Many educational testing programs rely on item and ability calibration algorithms that are rooted in item response theory (IRT). IRT includes a family of models which, if appropriate, provides substantial information about item and examinee performance. However, often overlooked is the fact that IRT is premised on some rather strong assumptions, namely unidimensionality and local independence (LI). In addition, an assumption inherent in nonparametric IRT (NIRT) and often made in parametric IRT (PIRT) is that of monotonicity. Therefore, for situations where monotonicity is assumed in PIRT, the only difference between PIRT and NIRT pertains to the relationship between the probability of correct response, $P(\theta)$, and examinee ability, $\theta$. In PIRT, this relationship*,* given by the item characteristic curve (ICC; item response function (IRF)), is assumed to be of a pre-specified form that is either logistic or normal ogive. In practice, however, IRT assumptions are often violated – a result that can cause poor estimation of item parameters and examinees' ability.

Model-based PIRT models are desirable when the model fits the data. Recently, however, it has increasingly come to be recognized that ICCs cannot always be modeled well within the PIRT models such as the two- or three-parameter logistic or normal ogive models (Douglas, 1997, 1999; Ramsay, 1991, 1995). Also, it has become essential to check the appropriateness of the modeling of PIRT. As a result, many researchers have begun to explore the use of NIRT models (Mokken 1971; Mokken & Lewis, 1982; Sijtsma & Molenaar, 1987) for estimating ICCs without restricting them to assume any particular parametric form (Ramsay, 1988, 1991; Ramsay & Abrahamowicz, 1989; Ramsay & Winsberg, 1991). Accordingly, many attempts have been made to test the goodness-of-fit of PIRT models (Kingston & Dorans, 1985; Stone, 2000; Fischer & Molenaar, 1995). Stone

(2000) has studied a goodness-of-fit test statistic based on the $\chi^2$ distribution in PIRT using a Monte Carlo resampling procedure.

To date, the use of nonparametric regression approaches in checking the fit of PIRT models has not been fully investigated. Douglas and Cohen (2001) used kernel smoothing to investigate the fit of PIRT models by comparing them to models fitted under nonparametric assumptions. It was shown that a nonparametric ICC estimation technique was able to capture the irregularity of parametric ICC with sufficient flexibility. However, because kernel smoothing relies on local averaging, this estimation technique may not result in a monotonically increasing ICC, thereby violating a common IRT assumption. One method for fitting nonparametric ICCs which satisfies the monotonicity assumption is through isotonic regression (Barlow, et al., 1972; Lee, 2002). However, the quality of fit of ICCs based on isotonic regression estimates has not been evaluated.

In this paper, we are concerned with the use of nonparametric ICC estimation techniques to assess the fit of PIRT models, via graphical inspection as well as through numerical measures. We checked the structural departure from PIRT models by comparing the fit of the two-parameter logistic model (2PL) with three different nonparametric ICC estimation procedures: isotonic regression, smoothed isotonic regression, and kernel smoothing ICC estimates.

## Nonparametric ICC Estimations

A nonparametric approach to estimating ICCs requires no mathematical model. In addition, no assumptions of any particular parametric forms of the ICCs are made in the estimation. Nonparametric approaches have been found to be more flexible in analyzing unknown curves and providing a better fit between dependent and independent variables than parametric procedures. In this paper, three nonparametric regression techniques are studied: kernel smoothing, isotonic regression, and smoothed isotonic regression.

The most commonly used approach for nonparametric ICC estimation is kernel smoothing, made popular because of its ease of use and computational convenience (Eubank, 1988). Kernel smoothing is based on local averaging which uses the mean of the response variables near a certain point as a representative point and produces reasonable approximations to the regression curve. In the context of IRT, Ramsay (1991) described kernel smoothing approaches in estimating dichotomous item response functions. The kernel estimate of the ICC, $\hat{P}_{\text{kernel},i}(\theta)$, is given below:

$$\hat{P}_{\text{kernel},i}(\theta) = \frac{\sum_{j=1}^{N} K\left(\frac{\theta - \theta_j}{h}\right) Y_{ij}}{\sum_{j=1}^{N} K\left(\frac{\theta - \theta_j}{h}\right)},$$

where $Y_{ij}$ is the response to item *i* for examinee *j* (*j = 1, …, N*); the kernel function $K(x)$, which is a nonnegative, continuous, and symmetric function, determines the shape of the distribution of kernel weights; and the bandwidth or smoothing parameter, *h*, indicates the size of the weights, controlling the amount of smoothing of the estimated regression function. The three common choices of kernel functions are the uniform $(K(x) = I[-1 \leq x \leq 1])$, quadratic $(K(x) = 1 - x^2 \text{ for } |x| \leq 1)$ , and Gaussian $(K(x) = \exp(-x^2/2))$.

Since the examinee's $\theta_j$ is unobservable and cannot be measured directly, it needs to be replaced with a reasonable estimator to determine how much weight is assigned to examinees *j*'s response to item *i*, $Y_{ik}$. An effective substitute for $\theta_j$ can be obtained by transforming each examinee's number correct score once examinees are ranked into the percentiles of a given latent trait distribution, *F* (Ramsay, 1991). For instance, when the standard normal distribution for the latent trait distribution, *F*, and the examinee's total score at the 90th percentile were used, then $\theta_j$ would be set to 1.282 which is equal to the 90th percentile of the standard normal distribution.

An important parameter, the smoothing parameter, *h*, is chosen by the user depending on the balance desired between bias and the variance of estimation, which are components of the mean square error (MSE) of the estimator in ICC estimation (Ramsay, 2000; Härdle 1990). The larger the bandwidth, the larger the bias and the smaller the sampling variance of the estimate. Ordinarily, the bottom line in choosing the bandwidth is to minimize the MSE of the estimator. In practice, it is recommended for researchers to select a bandwidth that produces a reasonable smoothness of the estimated function based on the empirical data-driven approach. The rule of thumb to choose bandwidth is recommended by letting *h* be proportional to $n^{-1/5}$ which is set as a default value in the program *TESTGRAF* (Ramsay, 2000).

Under the constraint of monotonicity of ICCs, Lee (2002) proposed the use of isotonic regression-based estimates in ICC estimation, which were motivated from Barlow et al. (1972) and Robertson et al. (1988). Isotonic regression is a least squares method for data fitting under order restrictions. To say that the estimated function is isotonic implies that the functions chosen to fit the data are non-decreasing functions of the independent variable. The isotonic regression for ICC estimation is defined as follows: Let the examinees' abilities follow the simple order, $\theta_1 \le \theta_2 \le ... \le \theta_k$. Next, let $P(\theta)$ be any given function of $\theta$, and let *F* be the collection of all isotonic functions of $\theta$. Then, $P^*(\theta)$ is the isotonic regression ICC estimate of $\theta$ if and only if $P^*(\theta)$ is isotonic and minimizes $\sum[P(\theta) - \hat{P}(\theta)]^2$, where $\hat{P}(\theta)$ is a member of *F*. Hence, $P^*(\theta)$ holds the constraint $P^*(\theta_1) \leq P^*(\theta_2) \leq \ldots \leq P^*(\theta_k)$.

The most widely used solution for computing isotonic regression is the Pool-Adjacent-Violators (PAV) algorithm (Barlow et al., 1972, p.13; Hanson, Pledger, and Wright, 1973). Isotonic regression solution via the PAV algorithm is obtained as follows:

Step 1: Rank order the data ($\theta_j, P(\theta_j)$) for all examinees by $\theta$ into ($\theta_{(j)}, P(\theta_{(j)})$).

Step 2: Start with $P(\theta_{(1)})$, move to the right and stop if the pair ($P(\theta_{(j)}), P(\theta_{(j+1)})$) violates the monotonicity constraint. Pool $P(\theta_{(j)})$ and the adjacent $P(\theta_{(j+1)})$, and replace them both by their average:

$$P^*(\theta_{(j)}) = P^*(\theta_{(j+1)}) = \frac{P(\theta_{(j)}) + P(\theta_{(j+1)})}{2}.$$

Step 3: Check that $P(\theta_{(j-1)}) \leq P^*(\theta_{(j)})$. If not, pool $\{P(\theta_{(j-1)}), P(\theta_{(j)}), P(\theta_{(j+1)})\}$ into one average and set

$$P^*(\theta_{(j)}) = P^*(\theta_{(j-1)}) = P^*(\theta_{(j+1)}) = \frac{P(\theta_{(j-1)}) + P(\theta_{(j)}) + P(\theta_{(j+1)})}{3}.$$

Continue to the left, adding terms and averaging, until the monotonicity requirement is satisfied and then continue to the right. The final solution is the isotonic ICC estimates. However, the resulting isotonic function may produce a step function (i.e., a level set) over the corresponding interval of ability if a small group of examinees was analyzed or monotonicity in some $P(\theta)$ is violated because ICC probabilities were replaced by their averages. Under such situations, isotonic regression ICCs are usually not very smooth, but they are non-decreasing functions.

Smoothed isotonic regression ICC estimates are obtained by first isotonizing the data via the PAV algorithm and then smoothing the resulting isotonic regression function using a kernel function with an appropriate bandwidth (Lee, 2002). The resulting ICC is generally smoother than that obtained using the isotonic regression ICC estimation.

For this study, the bandwidth for kernel smoothing was selected as .4, .3 .2, and .1 for sample size 250, 500, 1000, and 2000, respectively. The smaller bandwidth (i.e., half of kernel smoothing bandwidth) was used for smoothed isotonic regression because it was pre-smoothed in some sense. Also, a Gaussian kernel function was used for both kernel smoothing and smoothed isotonic regression estimation.

## Method

### Data Generation

A simulation study was used to assess the accuracy of ICC recovery and model fit under various conditions of sample size and test length. The data were generated with three sample sizes ($N$ = 250, 500, 1000, and 2000 simulees), each under three test lengths ($n$ = 20, 40, and 80 items). Both the two-parameter logistic (2PL) and the three-parameter logistic (3PL) models were used to generate the data for either model fitting or non-model fitting items. For each test length, 20% of the items were simulated as non-model fitting items using 3PL. That is, the probability of a correct response for an examinee $j$ on item $i$ is given by

$$P_i(\theta_j) = c_i + (1 - c_i)\frac{\exp[1.702a_i(\theta_j - b_i)]}{1 + \exp[1.702a_i(\theta_j - b_i)]},$$

where $a_i$, $b_i$, and $c_i$ are the discrimination, difficulty, and pseudo-guessing parameters, respectively, for item *i*; $\theta_j$ is the ability level for examinee *j*; and 1.702 is the scaling factor used to transform the metric from logistic to normal. The remaining 80% of the items were simulated to fit the 2PL (i.e., $c_i$ parameters were set to zero).

Generating item parameters were obtained from the estimated item parameters on a college-level Mathematics Placement Test using *MULTILOG* (Thissen, 1991). Table 1 includes the item numbers, the generating item parameters for the 40-item test, and indication of the items that were used for the 20-item condition. In simulating misfitting (i.e., non-model fitting) items with the 3PL, $c_i$ parameters were at least .23 for all misfitting items.

-------------------------------

Insert Table 1 About Here

-------------------------------

For the 80-item condition, the 40-item test was duplicated twice as a whole. For the 20-item test, items 17-20 were the non-model fitting items and for the 40-item test, items 33-40 were the non-model fitting items. For the 80-item test, items 33-40 and 73-80 were non-model fitting items. The ability distribution was sampled from the standard normal distribution, $\theta_j \sim N(0,1)$. The program *GENIRV* (Baker, 1988) was used to simulate the item response vector for each condition of the study.

## Goodness-of-Fit Measures

The purpose of this study was to investigate the performance of three different nonparametric ICC estimation procedures -- isotonic regression, smoothed isotonic regression, and kernel smoothing -- under various simulation conditions and to assess the fit of parametric IRT models by comparing them to models fitted under nonparametric assumptions. This was done using a parametric bootstrap based on a selected parameter approximation to the nonparametric ICC to generate a reference distribution for testing the fit of the 2PL MMLE ICCs. The evaluation of the fit of the nonparametric ICC estimation procedures was broken down into three steps: inspecting ICCs, measuring the distance between ICCs, and testing for goodness-of-fit. These three procedures are discussed next.

### Inspecting ICCs

For a graphical representation of the results, we plotted each nonparametrically estimated ICC, $\hat{P}_{kernel,i}$, $\hat{P}_{isotonic,i}$, and $\hat{P}_{smo\text{-}isotonic,i}$ (kernel smoothing, isotonic regression, and smoothed isotonic regression, respectively), and the parametrically-estimated 2PL ICC, $\hat{P}_{2PL,i}(\theta)$, along with the generating $P_i(\theta)$. This provided a rough sense of whether the parametric and nonparametric ICCs were sufficiently similar to the true ICC for model fitting and non-model fitting items.

### Measuring the distance between ICCs

The distance between the estimated ICCs and the underlying true (i.e., generating) ICC was used to assess the quality of estimates of the different IRT estimation techniques. Although many choices are available to measure the distances between ICCs, we calculated two measures, the root integrated squared error (*RISE*) and the mean absolute deviation (*MAD*) as indices of estimation precision. *RISE*, employed in Douglas & Cohen (2001), is computed as follows:

$$RISE = d(P,\hat{P}) = \sqrt{\int [P(\theta) - \hat{P}(\theta)]^2 f(\theta)d\theta},$$

where $d(P,\hat{P})$ is a measure of the distance between the true and estimated ICCs (i.e., kernel smoothing, isotonic regression, smoothed isotonic regression, and 2PL MMLE ICC estimates), and $f(\theta)$ indicates the density function of an examinee's ability. For the purpose of this study, $f(\theta)$ was taken to be the standard normal distribution. *MAD* is computed as follows:

$$MAD = \int \left| [P(\theta) - \hat{P}(\theta) \right| f(\theta)d\theta.$$

## **Testing Goodness-of-Fit**

In practice, of course, one cannot know which items are fitting and which are non-fitting. However, if it can be demonstrated that the NIRT models provide more accurate approximations of the underlying ICCs, then the quality of fit of an estimated PIRT model can be assessed by comparing it with an estimated NIRT models, for each individual item. Using the bootstrapping procedures described in Azzalini, Bowman & Hardle (1989) and Douglas & Cohen (2001), an item-by-item test of goodness-of-fit for the parametric IRT model was performed as follows:

Step 1 – Fit nonparametric estimates and find 2PL MMLE estimates with $N \sim (0, 1)$ prior.

Step 2 – Compute three different measures of the differences between the curves,

$d_i(\hat{P}_{2PL}, \hat{P}_{nonpar})$, for each of the three nonparametric methods from the data.

Step 3 – Generate *K* datasets to obtain the reference distribution of test statistic, $d_i$, under the null hypothesis that the parametric model (i.e., 2PL) holds. In order to distinguish the "signal" from the "noise" in nonparametric regression estimates, 30 replications were done to resample the data set (i.e., $K = 30$).

Step 4 – For each of the $K$ data sets, refit estimates and obtain the distribution of departure measures $d_i^K(\hat{P}_{2PL}, \hat{P}_{nonpar})$ under the null hypothesis as in Step 1.

Step 5 – Construct approximate $t$ statistics to test the goodness of fit of each item. For this study, the critical value is set at $|t| \geq 2.045$ at a significance level $\alpha = .05$ with the degrees of freedom, $df = 29$, as a possible indicator of a poor fit.

## Results

### Visual inspection of the estimated ICCs

Some exemplary nonparametric ICC estimates and 2PL MMLE estimates along with true underlying ICC are presented graphically in Figure 1 and Figure 2. The shortest test length with the smallest sample size and the longest test length with the largest sample size in the study were chosen in Figure 1 and Figure 2, respectively, for diagnosing how nonparametric and parametric ICC estimates differ from the true underlying ICC from the impact of sample size and test length. Also, for these two conditions, two model fitting items and two non-model fitting items were shown. In each plot, the solid curve indicates the underlying true ICC and the dotted curve shows the kernel smoothing estimates. Isotonic regression estimates are plotted with the combination of dot and dash curve, smoothed isotonic regression estimates are presented with dashed curves. 2PL ICC estimates are presented with the combination of multiple dots and dashed curves. This item-by-item graphical inspection of ICC provides information as to how and where a selected parametric model (i.e., 2PL) does not fit an item. For example, by virtue of the way item misfit was simulated, items 18 and 20 in Figure 1 and items 39 and 40 in Figure 2 fail to asymptote properly to the correct value of $P_i(\theta)$. Thus, it is important to check how close the nonparametric ICCs and 2PL ICC are to the underlying ICC to determine the quality of estimates and the fit of an item.

The overall patterns in the performance of the ICC estimation procedures were similar across all conditions. It can be seen from these plots that, in general, for model fitting items, the fit of the nonparametric ICCs and 2PL ICC were fairly good. The two isotonic regression estimation procedures appeared to have performed better than the kernel smoothing estimation at the upper and lower ends of the ability scale. Kernel smoothing ICCs revealed somewhat larger discrepancies due to the failure of the method to accurately model the upper and lower asymptotes. As was anticipated, the 2PL ICC estimates were the best approximation to the underlying ICCs for model fitting items. For non-model fitting items in each simulation condition, however, nonparametric ICC estimates produced better approximations to the corresponding underlying ICCs than the 2PL ICCs especially when the sample size was large. The fit of the smoothed isotonic regression ICC estimates to 2PL was the best with large sample sizes. In general, fit for all methods was improved as sample size increased. Visual inspection did not reveal a clear effect of test length.

## **<u>Measures of distance between ICCs</u>**

1. *RISE*

The measure of *RISE* for three nonparametric estimation and 2PL MMLE procedures is presented in Table 2. Table 2 contains the marginal average, minimum, and maximum of *RISE* values for model fitting and non-model fitting items separately across each simulation condition.

For model fitting items, the average of *RISE* ranged from .017 to .073 for the small sample sizes (i.e., 250 and 500 examinees condition) and ranged from .015

to .046 for the large sample sizes (i.e., 1000 and 2000 examinees condition). The smallest average *RISE* was obtained from the 2PL MMLE while the largest average *RISE* was found from the isotonic regression estimation. Looking at the maximum values of *RISE*, all *RISE* values were less than .121 for the 250 examinees condition and less than .096 for the other sample size conditions. *RISE* values decreased as the sample size increased. Also, they decreased as the test length increased except in the isotonic regression with 500 examinees condition. 2PL MMLE procedure yielded smaller *RISE* values than the nonparametric ICC estimation procedures. Among the nonparametric ICC estimation techniques, kernel smoothing and smoothed isotonic regression methods had smaller *RISE* values than isotonic regression.

For non-model fitting items, the average value of *RISE* was anywhere from.031 to .072 for the small sample sizes and from .028 to .042 for the large sample sizes. Similar to model fitting items, the smallest and the largest *RISE* values were found from the 2PL MMLE and isotonic regression estimation, respectively. The largest *RISE* value (*RISE* = .103) was observed in the 250 examinee condition and all *RISE* values were less than .077 for the 500, 1000, and 2000 examinee condition. *RISE* values decreased as the sample size increased. However, *RISE* values increased as the test length increased. The smoothed isotonic regression estimation procedure produced consistently smaller values of the *RISE* among the nonparametric ICC estimation methods. Both kernel smoothing and smoothed isotonic regression performed similarly with the same sample size and test length, showing the differences were negligible. The differences between 2PL MMLE and the smoothed isotonic regression were all quite small, appearing primarily at the third decimal place (less than .01) for all conditions. Especially, for 2000 examinees with 40-item and

80-item, the smoothed isotonic regression produced smaller values of the average *RISE* than 2PL MMLE.

2. *MAD*

For each simulation condition, the *MAD* results for ICCs obtained from the nonparametric ICC estimation and 2PL MMLE procedures are summarized in Table 3. The marginal average, minimum, and maximum of *MAD* are presented for model fitting and non-model fitting items separately.

In general, the kernel smoothing and smoothed isotonic regression procedures produced nearly the same pattern of measure of *MAD* for all conditions, although, *MAD* for the smoothed isotonic regression procedure was negligibly smaller. Between the two isotonic regression estimation procedures, the smoothed isotonic regression consistently provided smaller values of *MAD* than the isotonic regression estimation procedure regardless of the sample size, test length and types of items (either model fitting or non-model fitting). Also, the 2PL MMLE yielded smaller *MAD* values than the three nonparametric ICC estimation procedures. The values for the 2PL MMLE ranged from .014 to .30 compared to .023 to .059 for the nonparametric ICC estimation procedures. Increasing sample size was associated with a decrease in values of *MAD*. Increasing the number of items reduced the size of *MAD* for all three nonparametric ICC estimation procedures while the reverse was observed for the 2PL MMLE procedure.

For model fitting items, the three nonparametric estimation procedures yielded somewhat similar *MAD* results where the smoothed isotonic regression produced apparently smaller MAD than the two other nonparametric estimation procedures with large sample sizes. For non-model fitting items, all *MAD* values were less than .068

and decreased as the sample size increased. Kernel smoothing and smoothed isotonic regression estimation procedures yielded nearly the same size of *MAD* for the large sample sizes. Also, 2PL MMLE and the smoothed isotonic regression provided similar *MAD* values across the various sample size conditions. The 2PL MMLE procedure, however, did exhibit slightly smaller values of *MAD* in the 250, 500, and 1000 examinees condition compared to the smoothed isotonic regression estimation procedure.

## Test of Goodness-of-Fit

### 1. *RISE*

Table 4 shows the results of goodness-of-fit for each condition of the study. To see the behavior of *RISE* measure, Table 4 also contains the summary statistics (i.e., mean and standard deviation) of *RISE* values for model fitting items and non-model fitting items separately across each simulation condition. The items with large *RISE* values show poor fit resulting in *p*-values < .05.

For model fitting items, Type I probabilities were calculated separately for all sample size and test length conditions. Across all conditions, Type I probabilities ranged from 0 to .125; most were less than .05. For non-model fitting items, power of the NIRT procedures to detect misfit in the 2PL was 1.0 for all conditions, except the 250 examinee condition. Within this condition, as test length increased, the power of kernel smoothing increased. Power in both the isotonic and smoothed isotonic regression methods increased from 20 to 40 items, but decreased from 40 to 80 items. Power for the 250 examinee condition ranged from .5 to 1.0 across all three NIRT estimation methods, which indicates good detection of non-model fitting items, even with relatively small samples. Among the nonparametric ICC estimation procedures, isotonic regression-based estimates, and in particular, the smoothed isotonic regression procedure, detected more non-model fitting items than did the kernel smoothing estimates, in the 250 examinee condition. With larger sample sizes, all three methods were equally good at identifying misfitting items.

2. *MAD*

Table 5 shows the goodness-of-fit results for the *MAD* statistic for each condition in the study. The items with large *MAD* values might show the misfit item, where *p*-values are less than .05. In the same manner with *RISE*, Type I probabilities were calculated for model fitting items. Type I probabilities did not show a consistent pattern as sample size and test length changed. Type I probabilities were less than .0938 with the large sample sizes and were less .125 with the small sample sizes. For non-model fitting items, all three nonparametric ICC methods had a power of 1.0 to detect misfitting items, except in the 250 examinee condition. In that condition, the power of the three nonparametric ICC estimation methods ranged from .75 to 1.0, and generally increased as test length increased. Power to identify misfitting items was higher using the *MAD* criterion than when using the *RISE* criterion. Among the nonparametric ICC estimation procedures, in the 250-examinee condition, kernel smoothing and smoothed isotonic regression procedures detected more misfitting items than the isotonic regression method. For sample sizes of at least 500 examinees, each of the three nonparametric methods identified all of the misfitting items.

## Conclusion and Discussion

Results from this simulation study appear to have several implications for how practitioners use nonparametric ICC estimation methods to assess the fit of items when the underlying parametric model may not be appropriate for all items. First, an item-by-item visual inspection of parametric and nonparametric ICCs provides a graphical representation of misfitting items. Visual inspection suggests that nonparametric ICC estimation techniques are very good at reproducing underlying ICCs for all items, while the 2PL often may not. Isotonic regression-based estimates were all monotonic, thereby satisfying the popular monotonicity assumptions of ICCs, and showed the capability of asymptotic behavior. Second, both *RISE* and *MAD* results indicated that the overall patterns in the performance of the three nonparametric ICC estimation and the 2PL MMLE procedures were similar across all simulation conditions. In general, increasing the sample size decreased both *RISE* and *MAD* and increasing test length decreased both *RISE* and *MAD*. For small sample size conditions, the 2PL MMLE estimates yielded smaller *RISE* and *MAD* than estimates from the three nonparametric regression estimation procedures, regardless of model fitting and non-model fitting items. For large sample size conditions, all three nonparametric ICC estimation procedures yielded comparatively similar *RISE* and *MAD* results for non-model fitting items. Third, with respect to goodness-of-fit test in terms of *RISE* and *MAD*, Type I probability and power for the nonparametric estimation methods were very close to those expected for all sample sizes and test lengths in both model fitting items and non-model fitting items. Third, in terms of the factors influencing the fit of the items, increased sample size and test length should enhance the fit of ICC estimates for all methods. It has been shown that estimating examinee's ability, $\theta$, on a short test length based on the total test score is less reliable than estimates based on total scores from long tests (Douglas & Cohen, 2001). In addition, the result of this study showed that the smoothed isotonic regression estimation method provided a better fit than the kernel smoothing and isotonic regression estimation procedures at the two extremes of ability.

Parametric ICC estimation procedures are very useful when the model assumptions hold, but it is not clear how robust parametric models are to violations of these assumptions. Nonparametric ICC estimation procedures have been shown to be a nice alternative to the parametric approach in cases where monotonicity or model fit may not hold. Therefore, before routinely fitting PIRT models to analyze test data, researchers should check the fundamental modeling assumptions to make sure they are appropriate. In practice, monotonicity often does not strictly hold, which could cause serious estimation problems.

The results in this study are, to a certain extent, a function of the way in which misfit was simulated. Certainly there are other types of misfit, including non-monotone functions, or functions that do not have an easy-to-describe relationship between $\theta$ and $P(\theta)$. This becomes important because, in practice, it is hoped that the 2PL would not be used for items with substantial lower asymptotes. The advantages of NIRT procedures over sophisticated PIRT models such as the 3PL or nominal response model (Bock, 1972) for other, perhaps more realistic, types of item misfit must continue to be studied. Therefore, additional work related to other types of non-model fitting item can be extended to provide a general framework in the assessment of parametric item fit using nonparametric estimation procedures.

## References

Azzalini, A., Bowman, A. W., & Härdle, W. H. (1989). On the use of nonparametric regression for model checking. *Biometrika*, *76*, 1-11.

Baker, F. B. (1988). *GENIRV: Computer program for generating item responses* [Computer program]. Madison: University of Wisconsin, Department of Educational Psychology, Laboratory of Experimental Design.

Barlow, R. E., Bartholomew, D. J., Bremmer, J. M., & Brunk, H. D. (1972). *Statistical inference under order restrictions*. Wiley: New York.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, *37*, 29-51.

Douglas, J. (1997). Joint consistency of nonparametric item characteristic curves and ability estimation. *Psychometrika*, *62*, 7-28.

Douglas, J. (1999). *Asymptotic identifiability of nonparametric item response models* (Technical Report No. 142). University of Wisconsin, Department of Biostatistics and Medical Informatics.

Douglas, J., & Cohen, A. (2001). Nonparametric item response function estimation for assessing parametric model fit. *Applied Psychological Measurement*, *25*(3), 234-243.

Eubank, R. L. (1988). *Spline smoothing and nonparametric regression*. New York: Marcel Dekker.

Fischer, G. H., & Molenaar, I. W. (Eds.). (1995). *Rasch models: Foundations, recent developments, and applications*. New York: Springer-Verlag.

Hanson, D. L., Pledger, G., & Wright, F. T. (1973). On consistency in monotonic regression. *Ann. Statist., 1,* 401-421.

Härdle, W. (1990). *Applied nonparametric regression*. London: Chapman & Hall.

Kingston, N. M., & Dorans, N. J. (1985). The analysis of item-ability regressions: an exploratory IRT model fit tool. *Applied Psychological Measurement*, *9*, 281-288.

Lee, Y.-S. (2002). *Applications of isotonic regression in item response theory*. Ph.D. Dissertation. University of Wisconsin – Madison.

Mokken, R. J. (1971). *A Theory and Procedure of Scale Analysis, with Applications in Political Research.* New York/Berlin: Walter de Gruyter-Mouton.

Mokken, R. J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement*, *6,* 417-430.

Ramsay, J. O. (1988). Monotone regression splines in action (with discussion). *Statistical Science, 3,* 425-461.

Ramsay, J. O. (1991). Kernel smoothing approaches to nonparametric item characteristic curve estimation. *Psychometrika, 56,* 611-630.

Ramsay, J. O. (1995). A similarity-based smoothing approach to nondimensional item analysis. *Psychometrika*, *60,* 323-339.

Ramsay, J. O. (2000). *TESTGRAF: A computer program for nonparametric analysis of testing data*. Unpublished manuscript, McGill University.

Ramsay, J. O., & Abrahamowicz, M. (1989). Binomial regression with monotone splines: A psychometric application. *Journal of the American Statistical Association, 84,* 906-915.

Ramsay, J. O., & Winsberg, S. (1991). Maximum marginal likelihood estimation for semiparametric item analysis. *Psychometrika, 56,* 365-379.

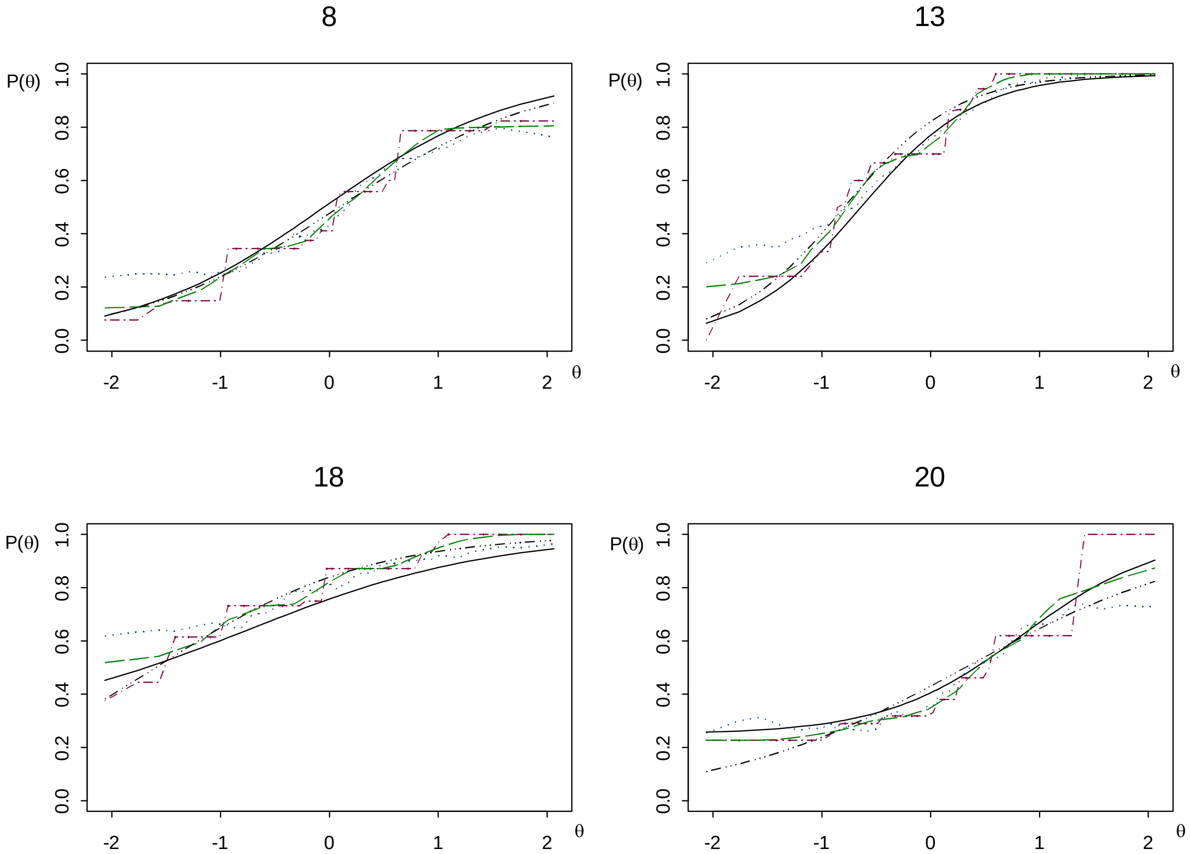
Robertson, T., Wright, F. T., & Dykstra, R. (1988). *Order restricted statistical inference*. New York: Wiley.

Sijtsma, K., & Molenaar, I. W. (1987). Reliability of test scores in nonparametric item response theory. *Psychometrika, 52,* 79-97.

Stone, C. A. (2000). Monte Carlo based null distribution for an alternative goodness-of-fit test statistic in IRT models. *Journal of Educational Measurement*, *37*, 58-75.

Thissen, D. (1991). *MULTILOG user's guide* [Computer program]. Chicago: Scientific Software.

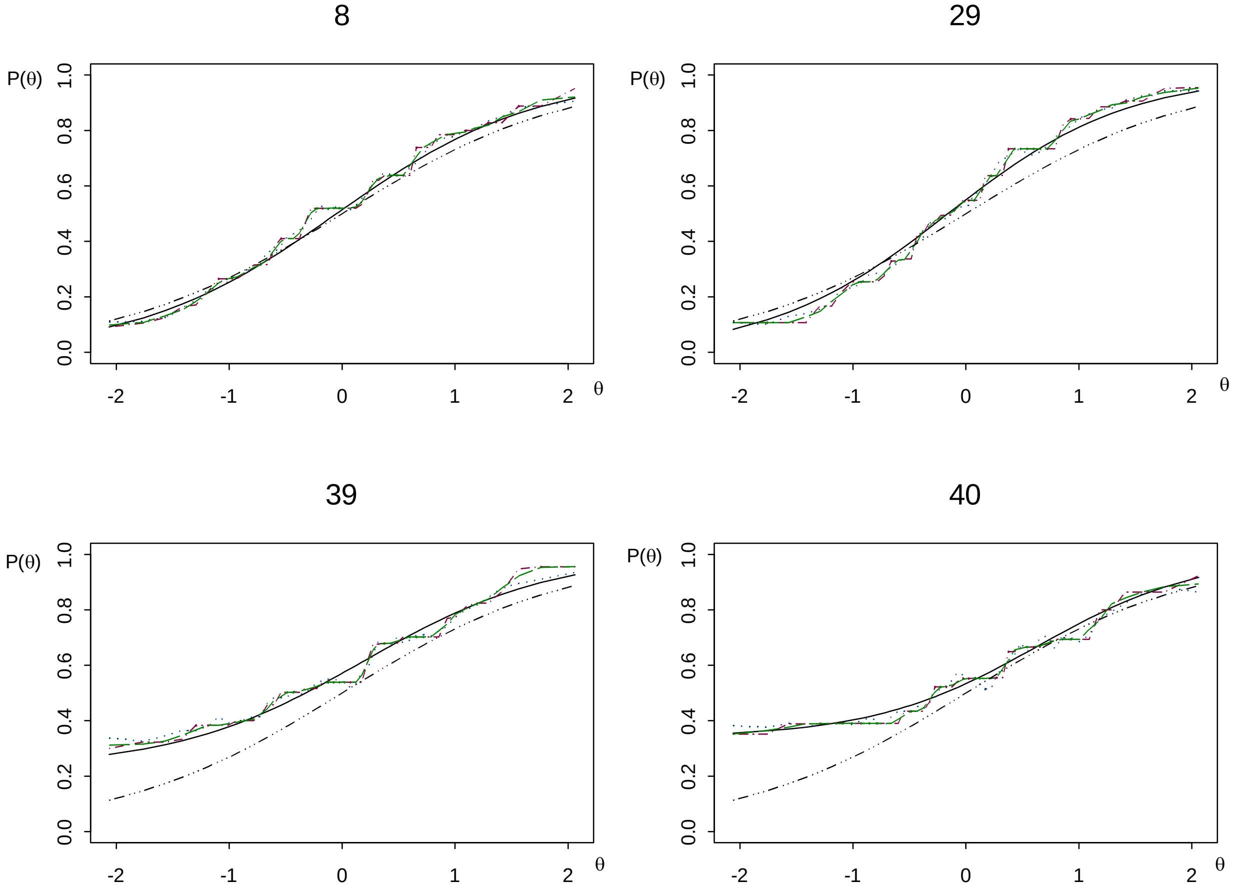
Figure 1. $P(\theta), \hat{P}_{nonpar}(\theta)$, and $\hat{P}_{2PL}(\theta)$ for 4 items in 250 examinees and 20-item condition



Note: Items 8 and 13 are model fitting items and items 18 and 20 are non-model fitting items.

| | |
|---|---|
| ________________ | True underlying ICC |
| …………………... | Kernel smoothing ICC |
| _ ·_· _ ·_· _ · _ · | Isotonic Regression ICC |
| _ _ _ _ _ _ _ _ _ _ | Smooth Isotonic Regression ICC |
| _ … _ … _ … _ … | 2PL ICC |

Figure 2. $P(\theta), \hat{P}_{nonpar}(\theta)$, and $\hat{P}_{2PL}(\theta)$ for 4 items in 2000 examinees and 80-item condition



Note: Items 8 and 13 are model fitting items and items 18 and 20 are non-model fitting items.

| | |
|---|---|
| ________________ | True underlying ICC |
| ………………….. | Kernel smoothing ICC |
| _ ·_· _ ·_· _ ·_ · | Isotonic Regression ICC |
| _ _ _ _ _ _ _ _ _ _ | Smooth Isotonic Regression ICC |
| _ … _ … _ … _ … | 2PL ICC |

Table 1. Generating item parameters and an indication of items used in 20-item test

| Item | *a* | *b* | *C* | Indication | Item | *a* | *b* | *c* | Indication |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.19 | -1.52 | | ✓ | 21 | 1.36 | -.35 | | |
| 2 | 1.81 | -.10 | | ✓ | 22 | 1.67 | -.23 | | |
| 3 | 1.13 | -.30 | | ✓ | 23 | 2.04 | -.13 | | |
| 4 | 1.41 | -.51 | | ✓ | 24 | .98 | 2.21 | | |
| 5 | 1.20 | .41 | | ✓ | 25 | 1.48 | .45 | | |
| 6 | 1.03 | .31 | | ✓ | 26 | 1.38 | .32 | | |
| 7 | 2.36 | -.17 | | ✓ | 27 | .85 | .30 | | |
| 8 | 1.14 | -.05 | | ✓ | 28 | 1.29 | .51 | | |
| 9 | .92 | -.04 | | ✓ | 29 | 1.26 | -.16 | | |
| 10 | 1.54 | -.25 | | ✓ | 30 | 1.49 | .77 | | |
| 11 | 2.01 | -.57 | | ✓ | 31 | 1.35 | .01 | | |
| 12 | 1.51 | .18 | | ✓ | 32 | .69 | .96 | | |
| 13 | 1.89 | -.64 | | ✓ | 33 | 1.35 | .06 | .27 | ✓ |
| 14 | 1.37 | -.01 | | ✓ | 34 | .89 | -.76 | .28 | ✓ |
| 15 | 1.86 | .07 | | ✓ | 35 | 1.05 | .31 | .25 | ✓ |
| 16 | 1.61 | -.22 | | ✓ | 36 | 1.58 | .85 | .25 | ✓ |
| 17 | 1.58 | -.29 | | | 37 | 1.47 | -.16 | .25 | |
| 18 | 1.92 | -.78 | | | 38 | 1.42 | -.35 | .24 | |
| 19 | 1.30 | .54 | | | 39 | 1.20 | .19 | .23 | |
| 20 | 1.16 | .77 | | | 40 | 1.37 | .64 | .34 | |

Table 2. *RISE*

| | | 2PL MMLE | | | | Kernel Smoothing | | | | Isotonic Regression | | | | Smooth Isotonic Regression | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *N* | *n* | Model Fitting | Min. Max. | Nonmodel Fitting | Min. Max. | Model Fitting | Min. Max. | Nonmodel Fitting | Min. Max. | Model Fitting | Min. Max. | Nonmodel Fitting | Min. Max. | Model Fitting | Min. Max. | Nonmodel Fitting | Min. Max. |
| 250 | 20 | .035 | .006<br>.071 | .050 | .033<br>.066 | .061 | .033<br>.092 | .060 | .043<br>.083 | .073 | .043<br>.105 | .064 | .043<br>.083 | .053 | .033<br>.080 | .051 | .040<br>.063 |
| | 40 | .033 | .008<br>.081 | .051 | .031<br>.069 | .063 | .034<br>.107 | .057 | .029<br>.077 | .069 | .041<br>.099 | .072 | .050<br>.103 | .050 | .026<br>.087 | .053 | .029<br>.078 |
| | 80 | .033 | .001<br>.089 | .045 | .028<br>.068 | .057 | .021<br>.121 | .053 | .024<br>.076 | .070 | .042<br>.107 | .063 | .050<br>.093 | .048 | .017<br>.097 | .050 | .027<br>.068 |
| 500 | 20 | .028 | .007<br>.070 | .031 | .015<br>.046 | .056 | .029<br>.088 | .043 | .021<br>.057 | .058 | .039<br>.096 | .047 | .032<br>.061 | .043 | .017<br>.085 | .034 | .018<br>.052 |
| | 40 | .019 | .005<br>.052 | .042 | .026<br>.063 | .044 | .024<br>.086 | .047 | .025<br>.063 | .053 | .038<br>.070 | .056 | .043<br>.068 | .039 | .022<br>.060 | .044 | .032<br>.057 |
| | 80 | .017 | .005<br>.065 | .035 | .013<br>.064 | .045 | .019<br>.081 | .044 | .018<br>.056 | .059 | .035<br>.089 | .057 | .036<br>.077 | .045 | .025<br>.081 | .044 | .020<br>.060 |
| 1000 | 20 | .019 | .004<br>.052 | .029 | .011<br>.051 | .041 | .029<br>.052 | .040 | .030<br>.051 | .045 | .032<br>.057 | .040 | .030<br>.049 | .038 | .024<br>.052 | .037 | .027<br>.045 |
| | 40 | .017 | .006<br>.037 | .033 | .019<br>.055 | .037 | .026<br>.049 | .036 | .027<br>.048 | .046 | .033<br>.060 | .042 | .033<br>.050 | .036 | .024<br>.049 | .034 | .024<br>.043 |
| | 80 | .017 | .001<br>.038 | .035 | .022<br>.052 | .031 | .019<br>.047 | .036 | .018<br>.052 | .041 | .020<br>.054 | .042 | .029<br>.049 | .031 | .017<br>.045 | .035 | .019<br>.043 |
| 2000 | 20 | .015 | .004<br>.030 | .028 | .011<br>.048 | .034 | .023<br>.044 | .036 | .034<br>.040 | .035 | .028<br>.048 | .035 | .025<br>.034 | .031 | .021<br>.044 | .031 | .023<br>.031 |
| | 40 | .017 | .002<br>.038 | .035 | .025<br>.057 | .033 | .023<br>.050 | .033 | .025<br>.039 | .036 | .027<br>.053 | .033 | .029<br>.040 | .031 | .022<br>.050 | .029 | .024<br>.035 |
| | 80 | .018 | .004<br>.044 | .031 | .009<br>.056 | .030 | .020<br>.050 | .029 | .025<br>.037 | .035 | .025<br>.054 | .032 | .025<br>.046 | .029 | .020<br>.050 | .027 | .021<br>.040 |

*N* : Number of examinees
*n* : Number of items

Table 3. *MAD*

| | | 2PL MMLE | | | | Kernel Smoothing | | | | Isotonic Regression | | | | Smooth Isotonic Regression | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *N* | *n* | Model | Min. | Nonmodel | Min. | Model | Min. | Nonmodel | Min. | Model | Min. | Nonmodel | Min. | Model | Min. | Nonmodel | Min. |
| | | Fitting | Max. | Fitting | Max. | Fitting | Max. | Fitting | Max. | Fitting | Max. | Fitting | Max. | Fitting | Max. | Fitting | Max. |
| 250 | 20 | .030 | .006<br>.061 | .043 | .030<br>.063 | .048 | .029<br>.080 | .047 | .032<br>.064 | .059 | .033<br>.085 | .054 | .039<br>.066 | .043 | .024<br>.061 | .044 | .031<br>.061 |
| | 40 | .030 | .008<br>.080 | .041 | .016<br>.048 | .049 | .027<br>.10 | .047 | .024<br>.065 | .056 | .033<br>.085 | .058 | .038<br>.082 | .042 | .022<br>.082 | .045 | .018<br>.068 |
| | 80 | .029 | .001<br>.072 | .037 | .025<br>.057 | .044 | .018<br>.087 | .045 | .020<br>.058 | .056 | .032<br>.095 | .051 | .040<br>.068 | .039 | .014<br>.082 | .042 | .023<br>.059 |
| 500 | 20 | .024 | .006<br>.056 | .026 | .011<br>.039 | .043 | .018<br>.072 | .032 | .015<br>.047 | .046 | .031<br>.074 | .038 | .026<br>.051 | .034 | .014<br>.065 | .028 | .015<br>.043 |
| | 40 | .017 | .004<br>.051 | .033 | .019<br>.045 | .034 | .016<br>.068 | .038 | .021<br>.052 | .043 | .029<br>.058 | .045 | .036<br>.057 | .032 | .019<br>.054 | .036 | .025<br>.047 |
| | 80 | .015 | .004<br>.064 | .028 | .011<br>.058 | .036 | .014<br>.068 | .036 | .014<br>.050 | .048 | .028<br>.076 | .045 | .029<br>.061 | .037 | .020<br>.067 | .037 | .015<br>.055 |
| 1000 | 20 | .016 | .004<br>.044 | .024 | .010<br>.044 | .033 | .023<br>.042 | .032 | .025<br>.050 | .037 | .025<br>.049 | .031 | .023<br>.040 | .031 | .020<br>.045 | .028 | .021<br>.037 |
| | 40 | .015 | .005<br>.034 | .027 | .016<br>.049 | .029 | .020<br>.038 | .030 | .021<br>.042 | .037 | .026<br>.049 | .035 | .027<br>.044 | .029 | .019<br>.040 | .029 | .019<br>.038 |
| | 80 | .015 | .001<br>.033 | .028 | .012<br>.046 | .025 | .013<br>.043 | .029 | .018<br>.042 | .032 | .018<br>.046 | .034 | .021<br>.052 | .025 | .013<br>.038 | .028 | .016<br>.046 |
| 2000 | 20 | .014 | .004<br>.025 | .022 | .008<br>.037 | .026 | .019<br>.041 | .029 | .023<br>.033 | .028 | .023<br>.041 | .028 | .020<br>.029 | .024 | .019<br>.039 | .025 | .018<br>.027 |
| | 40 | .014 | .002<br>.032 | .027 | .019<br>.039 | .027 | .017<br>.041 | .026 | .019<br>.035 | .029 | .020<br>.043 | .026 | .021<br>.031 | .023 | .015<br>.041 | .023 | .016<br>.029 |
| | 80 | .016 | .003<br>.039 | .026 | .007<br>.049 | .024 | .015<br>.039 | .024 | .019<br>.031 | .028. | .016<br>.044 | 026 | .019<br>.038 | .023 | .013<br>.039 | .022 | .017<br>.033 |

*N* : Number of examinees
*n* : Number of items

Table 4. Type I Probability and Power of *RISE*

| | | Kernel Smoothing | | | | | | Isotonic Regression | | | | | | Smooth Isotonic Regression | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Model Fitting | | | Nonmodel Fitting | | | Model Fitting | | | Nonmodel Fitting | | | Model Fitting | | | Nonmodel Fitting | | |
| *N* | *N* | Type I | Min. | Mean | Power | Min. | Mean | Type I | Min. | Mean | Power | Min. | Mean | Type I | Min. | Mean | Power | Min. | Mean |
| | | Prob. | Max. | (Std.) | | Max. | (Std.) | Prob. | Max. | (Std.) | | Max. | (Std.) | Prob. | Max. | (Std.) | | Max. | (Std.) |
| 250 | 20 | .0000 | .033 | .067 | .5000 | .043 | .156 | .1250 | .043 | .075 | .7500 | .043 | .171 | .0000 | .033 | .057 | .7500 | .040 | .161 |
| | | | .092 | (.020) | | .083 | (.031) | | .105 | (.017) | | .083 | (.030) | | .080 | (.003) | | .063 | (.031) |
| | 40 | .0000 | .034 | .063 | .8750 | .029 | .157 | .0313 | .041 | .073 | 1.000 | .050 | .176 | .0313 | .026 | .053 | 1.000 | .029 | .165 |
| | | | .107 | (.019) | | .077 | (.031) | | .099 | (.016) | | .103 | (.029) | | .087 | (.019) | | .078 | (.030) |
| | 80 | .0469 | .021 | .059 | .9375 | .024 | .150 | .0469 | .042 | .072 | .8750 | .050 | .173 | .0781 | .017 | .051 | .9375 | .027 | .160 |
| | | | .121 | (.018) | | .076 | (.028) | | .107 | (.015) | | .093 | (.026) | | .097 | (.017) | | .068 | (.028) |
| 500 | 20 | .0625 | .029 | .052 | 1.000 | .021 | .149 | .0625 | .039 | .057 | 1.000 | .032 | .159 | .1250 | .017 | .046 | 1.000 | .018 | .154 |
| | | | .088 | (.013) | | .057 | (.020) | | .096 | (.011) | | .061 | (.02) | | .085 | (.012) | | .052 | (.021) |
| | 40 | .0313 | .024 | .047 | 1.000 | .025 | .152 | .0000 | .038 | .057 | 1.000 | .043 | .164 | .0000 | .022 | .044 | 1.000 | .032 | .158 |
| | | | .086 | (.013) | | .063 | (.022) | | .070 | (.011) | | .068 | (.021) | | .060 | (.013) | | .057 | (.022) |
| | 80 | .0625 | .019 | .046 | 1.000 | .018 | .152 | .0469 | .035 | .057 | 1.000 | .036 | .166 | .0313 | .025 | .043 | 1.000 | .020 | .159 |
| | | | .081 | (.012) | | .056 | .(.022) | | .089 | (.011) | | .077 | (.022) | | .081 | (.012) | | .060 | (.022) |
| 1000 | 20 | .0000 | .029 | .042 | 1.000 | .030 | .148 | .0000 | .032 | .047 | 1.000 | .030 | .154 | .0000 | .024 | .039 | 1.000 | .027 | .151 |
| | | | .052 | (.010) | | .051 | (.015) | | .057 | (.008) | | .049 | (.015) | | .052 | (.009) | | .045 | (.015) |
| | 40 | .0000 | .026 | .036 | 1.000 | .027 | .153 | .0938 | .033 | .045 | 1.000 | .033 | .160 | .0313 | .024 | .035 | 1.000 | .024 | .156 |
| | | | .049 | (.009) | | .048 | (.015) | | .060 | (.008) | | .050 | (.014) | | .049 | (.008) | | .043 | (.015) |
| | 80 | .0781 | .019 | .034 | 1.000 | .018 | .152 | .0313 | .020 | .044 | 1.000 | .029 | .160 | .0469 | .017 | .035 | 1.000 | .019 | .156 |
| | | | .047 | (.008) | | .052 | (.014) | | .054 | (.008) | | .049 | (.013) | | .045 | (.008) | | .043 | (.014) |
| 2000 | 20 | .0625 | .023 | .036 | 1.000 | .034 | .148 | .0000 | .028 | .038 | 1.000 | .025 | .151 | .0625 | .021 | .033 | 1.000 | .023 | .149 |
| | | | .044 | (.007) | | .040 | (.009) | | .048 | (.006) | | .034 | (.009) | | .044 | (.007) | | .031 | (.009) |
| | 40 | .0313 | .023 | .033 | 1.000 | .025 | .152 | .0313 | .027 | .036 | 1.000 | .029 | .155 | .0313 | .022 | .030 | 1.000 | .024 | .153 |
| | | | .050 | (.006) | | .039 | (.009) | | .053 | (.006) | | .040 | (.009) | | .050 | (.006) | | .035 | (.009) |
| | 80 | .0625 | .020 | .031 | 1.000 | .025 | .155 | .1094 | .025 | .035 | 1.000 | .025 | .157 | .0938 | .020 | .029 | 1.000 | .021 | .156 |
| | | | .050 | (.006) | | .037 | (.037) | | .054 | (.005) | | .046 | (.009) | | .050 | (.006) | | .040 | (.009) |

*N* : Number of examinees
*n* : Number of items

Table 5. Type I Probability and Power of *MAD*

| | | Kernel Smoothing | | | | | | Isotonic Regression | | | | | | Smooth Isotonic Regression | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Model Fitting | | | Nonmodel Fitting | | | Model Fitting | | | Nonmodel Fitting | | | Model Fitting | | | Nonmodel Fitting | | |
| *N* | *N* | Type I | Min. | Mean | Power | Min. | Mean | Type I | Min. | Mean | Power | Min. | Mean | Type I | Min. | Mean | Power | Min. | Mean |
| | | Prob. | Max. | (Std.) | | Max. | (Std.) | Prob. | Max. | (Std.) | | Max. | (Std.) | Prob. | Max. | (Std.) | | Max. | (Std.) |
| 250 | 20 | .0000 | .029 | .054 | .7500 | .032 | .147 | .1250 | .033 | .060 | .7500 | .039 | .152 | .0000 | .024 | .047 | .7500 | .031 | .148 |
| | | | .080 | (.018) | | .064 | (.033) | | .085 | (.013) | | .066 | (.032) | | .061 | (.016) | | .061 | (.033) |
| | 40 | .0625 | .027 | .050 | .8750 | .024 | .148 | .0313 | .033 | .058 | .7500 | .038 | .156 | .0313 | .022 | .043 | .7500 | .018 | .150 |
| | | | .10 | (.017) | | .065 | (.032) | | .085 | (.014) | | .082 | (.029) | | .082 | (.016) | | .068 | (.031) |
| | 80 | .0313 | .018 | .046 | .8750 | .020 | .140 | .0469 | .032 | .058 | .8750 | .040 | .150 | .0469 | .014 | .042 | .9375 | .023 | .142 |
| | | | .087 | (.016) | | .058 | (.029) | | .095 | (.013) | | .068 | (.026) | | .082 | (.015) | | .059 | (.028) |
| 500 | 20 | .0625 | .018 | .041 | 1.000 | .015 | .141 | .0625 | .031 | .046 | 1.000 | .026 | .143 | .1250 | .014 | .037 | 1.000 | .015 | .141 |
| | | | .072 | (.011) | | .047 | (.021) | | .074 | (.009) | | .051 | (.020) | | .065 | (.010) | | .043 | (.021) |
| | 40 | .0313 | .016 | .037 | 1.000 | .021 | .142 | .0000 | .029 | .046 | 1.000 | .036 | .145 | .0000 | .019 | .035 | 1.000 | .025 | .143 |
| | | | .068 | (.012) | | .052 | (.023) | | .058 | (.009) | | .057 | (.022) | | .054 | (.011) | | .047 | (.022) |
| | 80 | .0469 | .014 | .036 | 1.000 | .014 | .142 | .0781 | .028 | .046 | 1.000 | .029 | .146 | .0938 | .020 | .035 | 1.000 | .015 | .143 |
| | | | .068 | (.011) | | .050 | (.023) | | .076 | (.009) | | .061 | (.022) | | .067 | (.010) | | .055 | (.023) |
| 1000 | 20 | .0000 | .023 | .033 | 1.000 | .025 | .141 | .0000 | .025 | .037 | 1.000 | .023 | .141 | .0000 | .020 | .031 | 1.000 | .021 | .141 |
| | | | .042 | (.008) | | .050 | (.016) | | .049 | (.007) | | .040 | (.016) | | .045 | (.008) | | .037 | (.016) |
| | 40 | .0313 | .020 | .029 | 1.000 | .021 | .142 | .0938 | .026 | .035 | 1.000 | .027 | .143 | .0625 | .019 | .028 | 1.000 | .019 | .142 |
| | | | .038 | (.008) | | .042 | (.015) | | .049 | (.006) | | .044 | (.015) | | .040 | (.007) | | .038 | (.016) |
| | 80 | .0313 | .013 | .027 | 1.000 | .018 | .141 | .0156 | .018 | .036 | 1.000 | .021 | .143 | .0156 | .013 | .028 | 1.000 | .016 | .141 |
| | | | .043 | (.007) | | .042 | (.015) | | .046 | (.006) | | .052 | (.015) | | .038 | (.007) | | .046 | (.015) |
| 2000 | 20 | .0625 | .019 | .030 | 1.000 | .023 | .139 | .0000 | .023 | .030 | 1.000 | .020 | .139 | .0625 | .019 | .027 | 1.000 | .018 | .139 |
| | | | .041 | (.006) | | .033 | (.010) | | .041 | (.005) | | .029 | (.010) | | .039 | (.006) | | .027 | (.010) |
| | 40 | .0313 | .017 | .026 | 1.000 | .019 | .140 | .0313 | .020 | .028 | 1.000 | .021 | .140 | .0625 | .015 | .024 | 1.000 | .016 | .140 |
| | | | .041 | (.005) | | .035 | (.010) | | .043 | (.005) | | .031 | (.010) | | .041 | (.005) | | .029 | (.010) |
| | 80 | .0625 | .015 | .024 | 1.000 | .019 | .141 | .0625 | .016 | .028 | 1.000 | .019 | .142 | .0781 | .013 | .023 | 1.000 | .017 | .141 |
| | | | .039 | (.005) | | .031 | (.010) | | .044 | (.004) | | .038 | (.010) | | .039 | (.005) | | .033 | (.010) |

*N* : Number of examinees
*n* : Number of items