# Detecting Exposed Test Items in Computer-Based Testing[1,2]

**Ning Han and Ronald Hambleton**
**University of Massachusetts at Amherst**

## Background and Purposes

Exposed test items are a major threat to the validity of computer-based testing. Historically, paper and pencil tests have maintained test security by (1) closely monitoring test forms (including their printing, distribution, administration, and collection), and (2) regularly introducing new test forms. However, because of the necessities of daily exposure of item pools to candidates in a computer-based testing environment such as the one that was initiated by the AICPA on April 5, 2004, standard methods for maintaining test security with paper-and-pencil administrations are no longer applicable. Failure to adequately solve the item security problem with computer-based testing will guarantee the demise of this approach to assessment.

Much of the research for limiting item exposure with computer-based tests has focused on finding ways to minimize item usage: expanding the number of test items in a bank (either by hiring extra item writers and/or using item generation forms and algorithms) (see Pitoniak, 2002), establishing conditional item exposure controls (see, for example, Revuelta & Ponsoda, 1998; Stocking & Lewis, 1998; Yi & Chang, 2003), rotating item banks, expanded initiatives to reduce sharing of test items on the internet (see, for example, the work of Caveon in spotting web-sites where test items are exposed to candidates might be found), shortening test administration windows (a strategy

[1] **Center for Educational Assessment Research Report No. 526**. Amherst, MA: University of Massachusetts.
[2] Paper presented at the meeting of the NCME, San Diego, April, 2004.

adopted by AICPA already), modifying the test design (with the intent of reducing the number of items that candidates are administered, without loss of precision—see for example the work of Luecht and Zenisky and others for the AICPA), better item bank utilization (see van der Linden and Veldkamp's work on item inventory control, and the work of Yi & Chang, 2003, on item bank usage), and so on.

A very different approach to addressing the problem is to focus attention on the generation and investigation of item statistics that can reveal whether test items have become known to candidates prior to seeing the items in the test they are administered (Lu & Hambleton, in press; Segall, 2001; Zhu & Liu, 2002). If these exposed items can be spotted statistically, they can be deleted from the item bank. Along these lines, several item statistics have been proposed (see, for example, Han, 2003; Lu & Hambleton, in press).

Han (2003) proposed the concept of "moving averages" for detecting exposed test items in an earlier study for the AICPA. The moving average is a form of average which has been adjusted to allow for periodic and random components of a time series data. A moving average is a smoothing technique used to make the long term trends of a time series clearer. Much like moving averages which are used on Wall Street to monitor stock price changes and in manufacturing industry to control product qualities, item performance can be monitored over time (e.g., after each item administration), and any changes can be noted and used to identify potentially exposed test items. Preliminary research has been encouraging. At the same time this research has been based upon the assumption that the examinees' ability distribution over time is stationary (Han, 2003) and a simple item exposure model was put in place. Several directions seemed worthy of

follow up research: investigating additional item exposure statistics, and evaluating these statistics under different conditions such as with shifting ability distributions over time and with various types of items (e.g., hard and easy, low and high discrimination), and for several exposure models.

More specifically then, the purposes of this research were (1) to evaluate several item exposure detection statistics in the presence of shifts in the ability distribution over time, (2) to address the suitability of the item exposure detection statistics under a number of item exposure models, and (3) to investigate item exposure detection for items with different statistical characteristics. The first purpose was essential because it simply is not reasonable to assume a fixed ability distribution at all times during a testing window. Some drift in the distribution might be expected—for example, the poorer candidates may come first, and higher ability candidates may follow later in the window. Several new item exposure statistics need to be investigated because the moving p-value statistic that Ning (2003) considered was sensitive to ability shifts and therefore, it is less suitable for use by the AICPA and other agencies doing computer-based testing: Shifts in ability distribution and detection of exposed items using moving p-value averages are confounded. While it may be true that the ability distribution of candidates will by-and-large be equivalent over time, item exposure detection statistics that are free of this questionable assumption should be studied.

Achieving the second purpose would provide data on competing item exposure detection statistics under various item exposure models. For example, in one simple model, after an item is exposed by a candidate one might conjecture that all candidates will have knowledge of the item and answer it correctly if it is selected for administration

again. Several other item exposure models need to be investigated too, several that are a bit more realistic.

The third purpose was added because we expected that the item exposure detection rate would depend not only on the choice of item exposure detection statistic, sample size, and nature of the exposure, but would also depend on the statistical characteristics of the exposed test items. For example, we expected it would be very difficult to detect exposed items when they were easy for candidates (after all, candidates are already expected to do well, and any improvements in item performance due to exposure then would be small); harder items should be considerably easier to spot because the shifts in item performance due to exposure are likely to be greater.

## Research Design

A great number of simulated data sets were considered in the study. Variables under study included (1) ability distribution (fixed or variable), (2) choice of item exposure detection statistic, (3) type of item exposure model, and (4) statistical characteristics of exposed test items.

In the present study, the level of item exposure was controlled by one parameter, $\rho$, and it was varied from no exposure ($\rho=0$) to full exposure ($\rho=1$) to either 10% or 100% of the candidates. An intermediate value of $\rho=.25$ applied to either 10% or 100% of the candidates was also considered in the simulations.

The study was implemented as follows:

(1) A linear test consisting of 75 items whose parameters were consistent with item statistics in a national credentialing exam were simulated. To roughly approximate

the actual testing condition we considered an item administration level of about 20% to candidates. Since the proposed item exposure detection statistics monitor examinee's response on an item over time, it is independent of the delivery mechanism of the test. Therefore, a simple linear test design was used without loss of generality of the findings.

(2) The number of candidates used in the study was 5000. We assumed 25,000 candidates in a testing window, with a 20% administration level, so up to 5000 examinees would see any set of 75 items. Three different ability distributions for the 5000 candidates were considered: Normal (0,1), drifting from a lesser ability group to a higher ability, $\theta \sim N(-1+i/2500, 1)$, and abrupt shift from $\theta \sim N(-1,1)$ for the first 2500 candidates and $\theta \sim N(1,1)$ for the next 2500 candidates. In simulating drift, we were assuming that the poorer candidates, generally, would take the test early (average ability = -1.0) and then gradually the ability distribution would shift from a mean of -1.0 to a mean of +1.0 by the end of the testing window. With the abrupt shift in ability distribution condition, after the first 2500 candidate abilities were sampled from a N(-1.0,1), for the last 2500 candidates, candidate abilities were sampled from a N(+1.0, 1) distribution.

(3) The probability that an examinee answers an item correctly is

$$P' = P + \rho(1 - P)$$

where: P: probability computed from the three-parameter logistic IRT model based on a candidate's ability level and item statistics.

$\rho$: a positive number $0 \leq \rho \leq 1$, was varied in the simulations, to reflect the item exposure model in place.

(4) Simulation variables:

1) Ability distributions:

a) normal;

b) drifting;

c) abrupt shift.

2) Extent to which an item is exposed:

$\rho = 0, 0.25, 1$

$\rho = 0$ is a base-line situation where the item is secure.

$\rho = 1$ is an extreme situation in which every candidate answers the item correctly.

$\rho=0.25$ is a situation where candidate performance, relative to ability and item statistics, is increased to reflect the fact that some general information is being disseminated about the item which gives candidates a boost in their likelihood of success, but not a guarantee they will answer the item correctly.

3) Statistics:

a) Moving P values;

b) Moving averages of item residuals (actual score –expected score based on the 3p model);

c) Moving averages of standardized item residuals (actual score – expected based on the 3p model/standard error)

(The idea with b and c here was to look at item performance compared to expected performance given an examinee's ability estimate. These ability estimates were calculated after test administration, and then used along

with the statistics for an item, and the candidate's item performance to calculate an item residual and the item standardized residual. It is only when these differences consistently exceeded what might be expected by chance for the item that the alarm would go off—that is, item exposure was suspected.)

4) The statistical characteristics of the items:

b = -1.0, 0.0, 1.0, 2.0

a = 0.40, 0.70, 1.20

These statistics were crossed to produce 12 item types to focus on in the research. These items were embedded into the 75 item test. Item exposure if was simulated always began with the 2501 candidate in the sequence.

5) Simulation times for each combination of the above situations:

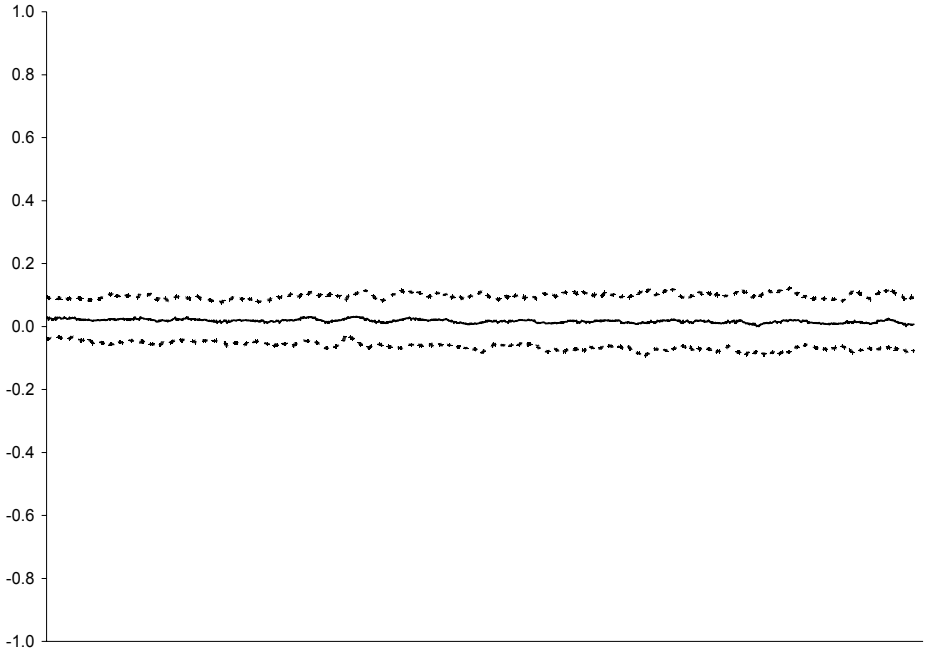
100

6) Detecting exposed test items:

Under the no exposure condition, it was possible for each of the 12 item types, to determine the empirical sampling distribution of each of the item statistics after each item administration (100 replications were carried out and the approximate .025, .975 percentiles were determined along with the mean of the 100 item statistics). What was used to approximate the percentiles was the mean + two standard deviations and the mean – two standard deviations. The graph below shows these values over many item administrations. These extremes were used in the flagging (i.e. detecting of exposed items). Whenever an item statistic

exceeded these boundaries, either a type I was made (if no exposure had been modeled) or exposure was detected (if exposure had been modeled).

**Example and explanation of the item exposure detection plot:**




The chart above is for the moving item residual and show that the situation when no exposure has been introduced.

A more formal explanation of what is happening follows. Given a sequence of examinees:

$$\{\theta_1, \theta_2, \ldots, \theta_t, \ldots, \theta_{5000}\}$$

where $\theta_t$ is the true ability of the examinee *t*.

For item *i*, the binary score for examinee *t* are obtained:

$$\{x_{i1}, x_{i2}, \ldots, x_{i5000}\}$$

Three item statistics are computed and plotted: moving *p* values, moving item residuals

and moving standardized item residual (called "K" here).

For example: when windows size $k$ equals to 100, the sequence of moving $p$ values is:

$$\{p_{100}, p_{101}, \ldots, p_{5000}\}$$

where

$$\begin{aligned} p_{100} &= \tfrac{1}{100}(x_{i1} + \cdots + x_{i100}) \\ p_{101} &= \tfrac{1}{100}(x_{i2} + \cdots + x_{i101}) \\ &\vdots \\ p_{n-k+1} &= \tfrac{1}{k}(x_{i,n-k+1} + x_{i,n-k+2} + \cdots + x_{i,n}) \end{aligned}$$

The sequence of moving item residuals is:

$$\{r_{100}, r_{101}, \ldots, r_{5000}\}$$

where

$$\begin{aligned} r_{100} &= \tfrac{1}{100}([x_{i1} - prob\ (a_i, b_i, c_i, \theta_1)] + \cdots + [x_{i100} - prob\ (a_i, b_i, c_i, \theta_{100})]) \\ r_{101} &= \tfrac{1}{100}([x_{i2} - prob\ (a_i, b_i, c_i, \theta_2)] + \cdots + [x_{i101} - prob\ (a_i, b_i, c_i, \theta_{101})]) \\ &\vdots \end{aligned}$$

The sequence of K indices is:

$$\{k_{100}, k_{101}, \ldots, k_{5000}\}$$

where

$$K_{100} = \frac{\sum_{j=1}^{100} (x_{ij} - prob\ (a_i, b_i, c_i, \theta_j))}{\sqrt{\sum_{j=1}^{100} prob\ (a_i, b_i, c_i, \theta_j)(1 - prob\ (a_i, b_i, c_i, \theta_j))}}$$

$$K_{101} = \frac{\sum_{j=2}^{101} \left( x_{ij} - prob\,(a_{i,} b_i, c_i, \theta_j) \right)}{\sqrt{\sum_{j=2}^{101} prob\,(a_i, b_i, c_i, \theta_j)(1 - prob\,(a_i, b_i, c_i, \theta_j))}}$$

For each simulation, we can obtain one sequence for each item statistic. The simulation process was replicated 100 times. Therefore, for each item statistic we can obtain 100 sequences. Three new sequences for each item statistic are obtained and plotted: Mean, Mean + 2*SD, Mean – 2*SD. For example, for moving p values, the means of the simulations are:

$$\left\{ \frac{\sum_{h=1}^{100} p_{h,100}}{100}, \frac{\sum_{h=1}^{100} p_{h,101}}{100}, \ldots, \frac{\sum_{h=1}^{100} p_{h,5000}}{100} \right\}$$

where *h* stands for the *h*th replication.

This sequence is plotted in the middle of the plot and the dotted lines are Mean + 2*SD and Mean – 2*SD. The vertical axis is the values of the sequence and the horizontal axis is the order of the sequence.

## Results

Our first task was to determine the window size, i.e., the amount of candidate data that would be used in calculating the rolling averages of item statistics. At one point this was going to be a variable in the study, but ultimately we determined from many practice simulations that a window size of 100 was large enough to provide stable statistical information, but not so large, that items might go for extended periods without being

spotted if they had been exposed. We will leave comprehensive study of the window size variable and its interactions with other variables in the study for another time and place. A review of Figures 1 to 3 shows the type of variability of these item exposure detection statistics associated with a window size of 100 for item 5 (b=0.0, a=.70). For the moving average p-values the standard deviation looks to be about .05. For the item residuals, the standard deviation looks to be about .05, and for the item standardized residuals, the standard deviation appears to be about 1.0 (recall that the upper and lower bands cover about four standard deviations).

**Comparison of Item Exposure Detection Statistics in Presence of Ability Distribution Shifts**

Figures 1 to 3 highlight the functioning of the three item statistics for a medium difficult item (b=0.0, a=0.7) with three ability distributions—normal, shifting, and abrupt change, respectively. What is very clear is that with a fixed normal distribution, all three item exposure detection statistics are quite stable as they should be—both the item statistics and the 95% confidence bands. With a shift in the ability distribution—gradual or abrupt, the p-value statistic shifted too—substantially. Clearly, p-value shifts are confounded with shifts in ability distributions and not reflecting item exposure because there was no exposure. Obviously this finding is not surprising, but the figures do highlight this fact, as well as the stability of the two IRT-based item exposure statistics that take into account examinee ability.

**Speed of Detection, Type I Errors and Power of Detection for Items with Various Statistical Properties Under Four Exposure Models**

Tables 1 to 24 contain the relevant information. Tables 1 to 8 provide the data we obtained with a constant normal distribution of ability for the candidates. Here, all three item exposure detection statistics were expected to be potentially useful and they were. Table 1 shows that with $\rho$=1.0, with 100% of the examinees benefiting from the exposed information on the 12 items, that detection was very fast. Across 100 replications for example, Table 1 highlights that with b=-1.00 and a=0.40, the average number of examinees who saw the exposed item was 27.4 before the statistic exceeded the threshold. (Note that in the simulations, exposure always occurred with the 2501st student in the sequence of 5000 candidates who would see the item.) Detection was even faster with harder items. And, in general, more discriminating items were detected faster too, except when the items were on the easy side. There was very little, if any, differences among the item exposure detection statistics. They all functioned about the same and well.

Table 2 shows the type I and power statistics for the 12 items. Type I errors were based on data compiled from the 1500$^{th}$ administration of the item to the 2500$^{th}$ administration. In this portion of the window, there was no item exposure. It is seen in Table 1, that under the conditions simulated, the type I error rate varied from 1.5% to 2.7% with the low discriminating items and was somewhat closer to the 5% level with the more discriminating items (2.6% to 4.4% with a=.7, and 1.9 to 6.6% with a=1.2) which had been the goal. More important, was the level of power of detection. In the case with $\rho$=1.0 and 100% exposure, detection was very easy and the power of detection was 100% for all items. Figure 4 shows what was going on graphically with a normal distribution of ability. More interesting cases follow.

Table 3 presents the first set of interesting results for the case where only 10% of the candidates have exposure to the item. Again, the more difficult items are spotted after considerably less item administrations that than easier items. For example, with b= -1.0, a=0.40, 320.7 (on the average) candidates were administered the easy item prior to exposure being detected with the moving p value item exposure statistic. With the hardest item (b=+2.0), and with the same item exposure detection statistic, 98.5 (on the average) candidates were administered the item prior to exposure being detected. With the other item exposure statistics, exposure appeared to be a bit quicker. In general, more discriminating items were detected faster than less discriminating items if they were medium to high difficulty.

Table 4 shows, for example, that type I errors were in the 1.5% to 6.6% range across all of the combinations of runs. Choice of item exposure detection statistic was of no major significance in the findings. Perhaps the most noticeable result in Table 4 is the low power of detection of exposed easy items (b=-1.0 or b=0.0). 25.2% detection rate was the highest. Whereas for the more difficult items (b=1.0 and b=2.0), power of detecting exposure ran as high as 94.7%. Clearly too, for the more difficult items, detection rates were higher for the more discriminating items. For example, considering the most difficult item (b=2.0), with the standardized item residual statistic, the power rates for items with discrimination levels of .4, .7, and 1.2, were 49.4%, 74.9%, and 93.5%.

Table 5 presents the first set of results for the case where $\rho$=0.25 and 100% of the candidates had exposure to the 12 items. Detection of item exposure did not take very long. Here again, the more difficult items were spotted after considerably less

administrations that than easier items. For example, with a=0.40, 113.5 (on the average) candidates were administered the easy item (b=-1.0) prior to exposure being detected with the moving p value item exposure detection statistic. With the hardest item (b=+2.0), and with the same item exposure statistic, 39.5 (on the average) candidates were administered the item prior to exposure being detected. With the other item exposure detection statistics, detection of exposure appeared to be a bit quicker, but only marginally. In general, more discriminating items were detected faster than less discriminating items if they were medium to high difficulty.

Table 6 shows, for example, that type I errors were in the 1.5% to 6.6% range as noted before across all of the combinations of runs. Choice of item exposure detection statistic was of no major significance though the two IRT-based statistics appeared to function a bit better overall. This time, detection rates for exposed easy items ran about 35 to 40%, compared to a detection rate of 100% for the hardest items.

Table 7 presents the poorest detection rates of the four item exposure models ($\rho$=.25, 10% exposure). Even for the most difficult and discriminating items, nearly 200 administrations were needed. In the main though, trends were the same: More difficulty and more discriminating items took less time to detect than the easier items. In this condition, interestingly, the moving p value item exposure detection statistic actually functioned a bit better than the other two statistics. It was not clear why.

Table 8 shows that the likelihood of detecting exposure was very poor. Even for the most difficult and discriminating items, power of detection did not exceed 26%. Choice of item exposure detection statistic was of no major significance.

Figures 4 to 7 highlight the pattern of the item exposure detection statistics for item 5 (b=0.0, a=0.7) under the four item exposure models with a normal distribution of ability. What is seen is the following: For ρ=1, and 100% exposure, the item was very easy to detect (see Figure 4); for ρ= 0.25, 100% exposure, the item took somewhat longer to identify and the power was moderate (see Figure 6); for ρ=1.0, 10% exposure, the trend was clear but the item was not identified very often (Figure 5); and finally with ρ=.25, and 10% exposure, the exposure was barely detectable in the moving average lines. These figures were presented for illustrative purposes only, and for accurate information on power of detection associated with specific items, see Tables 1 to 8.

**Impact of Shifts in the Ability Distribution**

Tables 9 to 16 and Figures 8 to 11 contain the statistical results for the gradually shifting ability distribution; Tables 17 to 24 and Figures 12 to 15 contain the statistical results for the abrupt shift in ability distributions. All of the findings reported above for the normal distribution were observed again. The major problem is clear from the levels of power of detection with the moving average p-values. These are very high for easy and hard items and both low, moderate, and high discriminating power (and though not reported, but can be seen in Figure 2, type I error rates are very high too). Basically, the item p value is flagging "all" items regardless of exposure. This is because the statistics themselves were drifting higher because of the increase in ability. Notice, for example, that the number of administrations needed for detection were substantially lower for the moving average p-value statistic compared to the other two exposure detection statistics. This is because the p-values were already drifting off to one because of the shift in distribution and well before the exposure had even been introduced into the simulation.

As the cutscores were set under the ρ=0 case, everything looked fine for type I error. But had they been set under this particular set of simulations they would have been unstable and inaccurate. As can easily be seen in Figures 8 through 11, the item p-values were already drifting off to 1.0 before any exposure was introduced. The problem was not seen with this statistic in Figures 12 through 15 because the shift in ability did not take place until after the cutoff scores had been set.

Looking at the big picture, and by-passing some of the irregularities and minor trends in the findings, we were struck by the similarity of results for the two IRT-based exposure detection statistics across the three ability distributions compared to the very different results observed with the moving average p-value statistic.

## Conclusions

The results from the study were revealing for all of the variables studied: (1) ability shifts, (2) item exposure models, (3) item exposure detection statistics, and (4) item statistics. First, the ability shifts were consequential. As a starter, it was easy to see that the moving p values produced unacceptable results when shifts in the ability distribution took place over the testing window—basically **all** items would be flagged with shifts in the ability distribution, regardless of whether or not they were exposed. In those situations, clearly, the other two statistics would be preferred. With a normal distribution of ability over the testing window all three statistics produced comparable results.

With respect to the item exposure models, putting aside the somewhat unrealistic first case (ρ=1, 100%) where detection was easy, one finding was that the ρ=.25, 10%

case produced quite unacceptable levels of exposed item detection. This is the case where 10% of the candidates have a small boost in the their performance level because of prior knowledge. For an examinee with a 50% probability of success on an item, that success was upped to 62.5% under the item exposure model. For a better candidate with a probability of success of 75%, that success would be upped to 81.2%. For an examinee operating at chance level based on their ability (25%) that probability would be increased to 43.75%, far from any assurance of a correct response to the item. And in this condition, these increased probabilities would be applied to the item level performance of only 10% of the candidates. Clearly, this level of exposure would be very difficult to spot in practice. The levels of detection of exposure were substantially higher in the other two cases, but especially so for the case $\rho$=.25 and 100% exposure. How realistic this case might be in practice is not certain, but the detection rates were quite good, and certainly preferable to not taking any action at all.

As for the item exposure detection statistics, our research showed a strong advantage to the two IRT-based statistics. They were applicable across all conditions simulated whereas the item p-value was not. And, they typically identified exposed items except in the cases where a small amount of exposure was simulated. We noticed too, that what ever the detection rates, it was always easiest to detect the more difficult items, and generally the more discriminating items. Some reversals were seen in the data however.

Interestingly and importantly, the findings about the item exposure detection statistics and how they functioned are applicable to all forms of computer-based testing from linear or linear-on-to-fly to multi-stage, to fully adaptive tests. Once an item is

administered in whatever design is operative in the testing program, the candidate performance data can be added to the string of data being collected on each item, and the item detection statistics can be updated, and tested for significance. An item remains in the bank until it is retired or identified as being exposed. The likelihood of detection of exposed items obviously depends on the confidence bands that have been established (which depend on the window size, in this study the number of candidates used in the statistics was 100), the statistical characteristics of the test items, and the type of exposure taking place. For the two IRT-based statistics, that considered ability in the calculation of statistics, the nature of the ability distribution was irrelevant. We were pleased too to discover that the harder more discriminating items are the ones that can be detected fastest. These are the same items that influence the ability estimates the most, and therefore they raise the most questions about the validity of candidate scores.

We were pleased with the results from the study and expect to continue on with the work. Obviously, we are looking forward to seeing the statistics actually used in practice which we expect to happen soon. Also, next steps in this research probably will focus on just one of the item detection statistics—item residuals, and investigate additional item exposure models. Other detection flags are also possible too. For example, candidate time information on items is being compiled. Were candidates to answer an item correctly using substantially less time than other candidates, a question would be raised about the validity of the candidate's response. Possibly, this information can be combined with the item detection statistic to more rapidly identify exposed items. Clearly there is lots of work to be done.

## References


Han, N. (2003). Using moving averages to assess test and item security in computer-based testing (Center for Educational Assessment Research Report No. 468). Amherst, MA: University of Massachusetts, School of Education.

Lu, Y., & Hambleton, R. K. (in press). Statistics for detecting disclosed items in a CAT environment. Metodologia de Las Ciencias del Comportamiento.

Pitoniak, M. (2002). Automatic item generation methodology in theory and practice (Center for Educational Assessment Research Report No. 444). Amherst, MA: University of Massachusetts, School of Education.

Revuelta, J., & Ponsoda, V. (1998). A comparison of item exposure control methods in computerized adaptive testing. Journal of Educational Measurement, 35, 311-327.

Segall, D. O. (2001, April). Measuring test compromise in high-stakes computerized adaptive testing: a Bayesian strategy for surrogate test-taker detection. Paper presented at the meeting of the National Council on Measurement in Education, Seattle, WA.

Stocking, M. L., & Lewis, C. (1998). Controlling item exposure conditional on ability in computerized adaptive testing. Journal of Educational and Behavioral Statistics, 23, 57-75.

Yi, Q., & Chang, H. H. (2003). A-stratified CAT design with content blocking. British Journal of Mathematical and Statistical Psychology, 56, 359-378.

Zhu, R., Yu, F., & Liu, S. (2002, April). Statistical indexes for monitoring item behavior under computer adaptive testing environment. Paper presented at the meeting of the American Educational Research Association, New Orleans.

Version: April 8, 2004

Table 1. Number of times of item administration after exposure. (ρ = 1.0, for 100%, normal distribution of ability)

| | | a=0.40 | a=0.70 | a=1.20 |
|---|---|---|---|---|
| Moving P values | b=-1.00 | 27.4 | 22.0 | 28.6 |
| | b= 0.00 | 15.5 | 10.4 | 9.0 |
| | b= 1.00 | 11.9 | 7.3 | 4.5 |
| | b= 2.00 | 9.2 | 4.7 | 2.6 |
| Moving Item Residuals | b=-1.00 | 25.3 | 22.9 | 24 |
| | b= 0.00 | 16.3 | 12.4 | 11.2 |
| | b= 1.00 | 12.5 | 8.7 | 7.5 |
| | b= 2.00 | 10.4 | 6.4 | 3.6 |
| Standardized Item Residuals | b=-1.00 | 25.2 | 22.6 | 23.5 |
| | b= 0.00 | 16.3 | 12.4 | 10.9 |
| | b= 1.00 | 12.4 | 8.6 | 7.5 |
| | b= 2.00 | 10.4 | 6.7 | 4.6 |

Table 2. Type I errors and power. (ρ = 1.0, for 100%, normal distribution of ability)

| | | a=0.40 | | a=0.70 | | a=1.20 | |
|---|---|---|---|---|---|---|---|
| | | I | II | I | II | I | II |
| Moving P Values | b=-1.00 | 1.50 | 100.0 | 3.36 | 100.0 | 2.33 | 100.0 |
| | b= 0.00 | 2.68 | 100.0 | 4.42 | 100.0 | 3.60 | 100.0 |
| | b= 1.00 | 2.16 | 100.0 | 2.86 | 100.0 | 5.55 | 100.0 |
| | b= 2.00 | 1.99 | 100.0 | 4.08 | 100.0 | 6.61 | 100.0 |
| Moving Item Residuals | b=-1.00 | 2.14 | 100.0 | 2.78 | 100.0 | 1.97 | 100.0 |
| | b= 0.00 | 2.55 | 100.0 | 3.27 | 100.0 | 1.94 | 100.0 |
| | b= 1.00 | 2.36 | 100.0 | 2.56 | 100.0 | 2.85 | 100.0 |
| | b= 2.00 | 2.02 | 100.0 | 2.63 | 100.0 | 3.11 | 100.0 |
| Standardized Item Residuals | b=-1.00 | 2.15 | 100.0 | 2.78 | 100.0 | 2.16 | 100.0 |
| | b= 0.00 | 2.55 | 100.0 | 3.10 | 100.0 | 1.94 | 100.0 |
| | b= 1.00 | 2.45 | 100.0 | 2.74 | 100.0 | 2.88 | 100.0 |
| | b= 2.00 | 2.09 | 100.0 | 2.59 | 100.0 | 2.99 | 100.0 |

Table 3. Number of times of item administration after exposure. ($\rho = 1.0$, for 10%, normal distribution of ability)

| | | a=0.40 | a=0.70 | A=1.20 |
|---|---|---|---|---|
| Moving P values | b=-1.00 | 320.7 | 301.2 | 292.3 |
| | b= 0.00 | 173.8 | 160.2 | 140.3 |
| | b= 1.00 | 169.0 | 115.9 | 61.5 |
| | b= 2.00 | 98.5 | 57.2 | 44.8 |
| Moving Item Residuals | b=-1.00 | 283.7 | 313.9 | 329.8 |
| | b= 0.00 | 191.7 | 143.5 | 188.5 |
| | b= 1.00 | 140.8 | 113.8 | 66.6 |
| | b= 2.00 | 98.1 | 61.2 | 48.8 |
| Standardized Item Residuals | b=-1.00 | 283.8 | 315.4 | 307.2 |
| | b= 0.00 | 192.9 | 149.1 | 189.5 |
| | b= 1.00 | 135.0 | 112.4 | 67.8 |
| | b= 2.00 | 96.9 | 60.8 | 48.8 |

Table 4. Type I errors and power. ($\rho = 1.0$, for 10%, normal distribution of ability)

| | | a=0.40 | | a=0.70 | | a=1.20 | |
|---|---|---|---|---|---|---|---|
| | | I | II | I | II | I | II |
| Moving P Values | b=-1.00 | 1.50 | 8.3 | 3.36 | 10.5 | 2.33 | 9.80 |
| | b= 0.00 | 2.68 | 16.8 | 4.41 | 23.7 | 3.63 | 25.2 |
| | b= 1.00 | 2.16 | 26.6 | 2.86 | 38.4 | 5.55 | 64.7 |
| | b= 2.00 | 1.99 | 47.5. | 4.08 | 77.9 | 6.60 | 94.7 |
| Moving Item Residuals | b=-1.00 | 2.14 | 9.8 | 2.78 | 10.0 | 1.97 | 8.7 |
| | b= 0.00 | 2.55 | 16.7 | 3.27 | 23.8 | 1.94 | 24.1 |
| | b= 1.00 | 2.36 | 29.5 | 2.56 | 41.1 | 2.84 | 63.6 |
| | b= 2.00 | 2.02 | 49.0 | 2.62 | 75.5 | 3.10 | 94.0 |
| Standardized Item Residuals | b=-1.00 | 2.15 | 9.9 | 2.78 | 10.0 | 2.16 | 9.1 |
| | b= 0.00 | 2.54 | 16.7 | 3.10 | 23.4 | 1.94 | 24.3 |
| | b= 1.00 | 2.45 | 29.9 | 2.74 | 42.2 | 2.88 | 63.5 |
| | b= 2.00 | 2.08 | 49.4 | 2.59 | 74.9 | 2.99 | 93.5 |

Table 5. Number of times of item administration after exposure. ($\rho = 0.25$, for 100%, normal distribution of ability)

| | | a=0.40 | a=0.70 | A=1.20 |
|---|---|---|---|---|
| Moving P values | b=-1.00 | 113.5 | 123.5 | 118.8 |
| | b= 0.00 | 67.9 | 55.3 | 55.4 |
| | b= 1.00 | 53.8 | 49.7 | 24.5 |
| | b= 2.00 | 39.5 | 21.0 | 16.1 |
| Moving Item Residuals | b=-1.00 | 99.4 | 119.5 | 109.9 |
| | b= 0.00 | 64.9 | 52.6 | 56.0 |
| | b= 1.00 | 47.1 | 46.2 | 29.6 |
| | b= 2.00 | 38.4 | 23.3 | 18.7 |
| Standardized Item Residuals | b=-1.00 | 99.3 | 119.1 | 109.3 |
| | b= 0.00 | 64.9 | 52.9 | 56.0 |
| | b= 1.00 | 46.3 | 45.6 | 29.7 |
| | b= 2.00 | 38.3 | 25.1 | 20.8 |

Table 6. Type I errors and power. ($\rho = 0.25$, for 100%, normal distribution of ability)

| | | a=0.40 | | a=0.70 | | a=1.20 | |
|---|---|---|---|---|---|---|---|
| | | I | II | I | II | I | II |
| Moving P Values | b=-1.00 | 1.50 | 40.9 | 3.36 | 39.0 | 2.33 | 33.9 |
| | b= 0.00 | 2.68 | 71.6 | 4.41 | 78.0 | 3.63 | 85.5 |
| | b= 1.00 | 2.16 | 88.8 | 2.86 | 97.3 | 5.55 | 99.8 |
| | b= 2.00 | 1.99 | 99.3 | 4.08 | 100.0 | 6.60 | 100.0 |
| Moving Item Residuals | b=-1.00 | 2.14 | 46.7 | 2.78 | 39.5 | 1.97 | 41.0 |
| | b= 0.00 | 2.55 | 74.0 | 3.27 | 80.8 | 1.94 | 89.1 |
| | b= 1.00 | 2.36 | 91.2 | 2.56 | 98.2 | 2.84 | 99.9 |
| | b= 2.00 | 2.02 | 99.4 | 2.62 | 100.0 | 3.10 | 100.0 |
| Standardized Item Residuals | b=-1.00 | 2.15 | 47.2 | 2.78 | 39.8 | 2.16 | 42.0 |
| | b= 0.00 | 2.54 | 74.0 | 3.10 | 80.5 | 1.94 | 89.2 |
| | b= 1.00 | 2.45 | 91.4 | 2.74 | 98.3 | 2.88 | 99.8 |
| | b= 2.00 | 2.08 | 99.4 | 2.59 | 100.0 | 2.99 | 100.0 |

Table 7. Number of times of item administration after exposure. ($\rho = 0.25$, for 10%, normal distribution of ability)

| | | a=0.40 | a=0.70 | A=1.20 |
|---|---|---|---|---|
| Moving P values | b=-1.00 | 517.6 | 473.3 | 393.2 |
| | b= 0.00 | 530.9 | 420.6 | 310.1 |
| | b= 1.00 | 539.2 | 340.4 | 186.2 |
| | b= 2.00 | 424.2 | 173.1 | 136.8 |
| Moving Item Residuals | b=-1.00 | 666.5 | 622.5 | 721.6 |
| | b= 0.00 | 482.3 | 538.2 | 478.2 |
| | b= 1.00 | 558.9 | 415.1 | 270.0 |
| | b= 2.00 | 480.9 | 271.6 | 179.3 |
| Standardized Item Residuals | b=-1.00 | 650.5 | 671.9 | 674.7 |
| | b= 0.00 | 482.9 | 591.6 | 479 |
| | b= 1.00 | 573.2 | 388.0 | 282.3 |
| | b= 2.00 | 474.4 | 255.3 | 180.5 |

Table 8. Type I errors and power. ($\rho = 0.25$, for 10%, normal distribution of ability)

| | | a=0.40 | | a=0.70 | | a=1.20 | |
|---|---|---|---|---|---|---|---|
| | | I | II | I | II | I | II |
| Moving P Values | b=-1.00 | 1.50 | 3.34 | 3.36 | 4.2 | 2.33 | 4.5 |
| | b= 0.00 | 2.68 | 4.53 | 4.41 | 7.6 | 3.63 | 8.0 |
| | b= 1.00 | 2.16 | 5.34 | 2.86 | 7.7 | 5.55 | 16.0 |
| | b= 2.00 | 1.99 | 6.81 | 4.08 | 15.5 | 6.60 | 26.1 |
| Moving Item Residuals | b=-1.00 | 2.14 | 3.72 | 2.78 | 3.4 | 1.97 | 3.5 |
| | b= 0.00 | 2.55 | 4.68 | 3.27 | 6.0 | 1.94 | 4.8 |
| | b= 1.00 | 2.36 | 6.12 | 2.56 | 7.4 | 2.84 | 10.8 |
| | b= 2.00 | 2.02 | 7.03 | 2.62 | 11.8 | 3.10 | 21.1 |
| Standardized Item Residuals | b=-1.00 | 2.15 | 3.78 | 2.78 | 3.4 | 2.16 | 3.6 |
| | b= 0.00 | 2.54 | 4.67 | 3.10 | 5.8 | 1.94 | 4.9 |
| | b= 1.00 | 2.45 | 6.23 | 2.74 | 7.8 | 2.88 | 10.6 |
| | b= 2.00 | 2.08 | 7.24 | 2.59 | 11.5 | 2.99 | 20.6 |

Table 9. Number of times of item administration after exposure. ($\rho$ = 1.0, for 100%, gradual change in ability from a mean of -1.0 to a mean of +1.0)

| | | a=0.40 | a=0.70 | A=1.20 |
|---|---|---|---|---|
| Moving P values | b=-1.00 | 21.7 | 18.5 | 18.8 |
| | b= 0.00 | 15.6 | 14.4 | 7.9 |
| | b= 1.00 | 12.0 | 9.6 | 7.0 |
| | b= 2.00 | 9.5 | 6.8 | 3.0 |
| Moving Item Residuals | b=-1.00 | 23.2 | 22.7 | 26.3 |
| | b= 0.00 | 16.4 | 13.5 | 11.9 |
| | b= 1.00 | 12.8 | 10.4 | 9.0 |
| | b= 2.00 | 10.7 | 6.8 | 4.1 |
| Standardized Item Residuals | b=-1.00 | 23.2 | 22.8 | 26.7 |
| | b= 0.00 | 16.8 | 13.4 | 12.1 |
| | b= 1.00 | 13.1 | 10.4 | 9.0 |
| | b= 2.00 | 10.9 | 7.8 | 4.3 |

Table 10. Type I errors and power. ($\rho$ = 1.0, for 100%, gradual change in ability from a mean of -1.0 to a mean of +1.0)

| | | a=0.40 | | a=0.70 | | A=1.20 | |
|---|---|---|---|---|---|---|---|
| | | I | II | I | II | I | II |
| Moving P Values | b=-1.00 | 1.9 | 100.0 | 1.7 | 100.0 | 4.5 | 100.0 |
| | b= 0.00 | 1.3 | 100.0 | 3.4 | 100.0 | 3.2 | 100.0 |
| | b= 1.00 | 1.2 | 100.0 | 6.1 | 100.0 | 1.0 | 100.0 |
| | b= 2.00 | 1.1 | 100.0 | 2.8 | 100.0 | 1.2 | 100.0 |
| Moving Item Residuals | b=-1.00 | 3.4 | 100.0 | 1.9 | 100.0 | 3.4 | 100.0 |
| | b= 0.00 | 3.5 | 100.0 | 2.9 | 100.0 | 3.4 | 100.0 |
| | b= 1.00 | 2.8 | 100.0 | 1.9 | 100.0 | 2.4 | 100.0 |
| | b= 2.00 | 2.2 | 100.0 | 1.4 | 100.0 | 2.2 | 100.0 |
| Standardized Item Residuals | b=-1.00 | 3.1 | 100.0 | 1.5 | 100.0 | 2.4 | 100.0 |
| | b= 0.00 | 3.2 | 100.0 | 2.8 | 100.0 | 2.7 | 100.0 |
| | b= 1.00 | 2.5 | 100.0 | 2.2 | 100.0 | 2.5 | 100.0 |
| | b= 2.00 | 2.3 | 100.0 | 2.0 | 100.0 | 3.6 | 100.0 |

Table 11. Number of times of item administration after exposure. ($\rho$ = 1.0, for 10%, gradual change in ability from a mean of -1.0 to a mean of +1.0)

| | | a=0.40 | a=0.70 | A=1.20 |
|---|---|---|---|---|
| Moving P values | b=-1.00 | 212.8 | 178.6 | 124.5 |
| | b= 0.00 | 180.2 | 150.3 | 91.6 |
| | b= 1.00 | 140.5 | 115.0 | 102.2 |
| | b= 2.00 | 110.9 | 107.9 | 57.8 |
| Moving Item Residuals | b=-1.00 | 285.3 | 312.3 | 278.2 |
| | b= 0.00 | 182.7 | 163.6 | 133.1 |
| | b= 1.00 | 130.1 | 88.5 | 73.5 |
| | b= 2.00 | 106.2 | 74.9 | 47.3 |
| Standardized Item Residuals | b=-1.00 | 280.0 | 304.4 | 320.9 |
| | b= 0.00 | 190.0 | 170.5 | 149.1 |
| | b= 1.00 | 134.8 | 89.4 | 74.1 |
| | b= 2.00 | 114.2 | 73.7 | 47.9 |

Table 12. Type I errors and power. ($\rho$ = 1.0, for 10%, gradual change in ability from a mean of -1.0 to a mean of +1.0)

| | | a=0.40 | | a=0.70 | | a=1.20 | |
|---|---|---|---|---|---|---|---|
| | | I | II | I | II | I | II |
| Moving P Values | b=-1.00 | 1.9 | 68.4 | 1.7 | 90.7 | 4.5 | 99.3 |
| | b= 0.00 | 1.3 | 79.9 | 3.4 | 100.1 | 3.2 | 100.0 |
| | b= 1.00 | 1.2 | 82.7 | 6.1 | 97.1 | 1.0 | 99.8 |
| | b= 2.00 | 1.1 | 92.5 | 2.8 | 97.9 | 1.2 | 99.6 |
| Moving Item Residuals | b=-1.00 | 3.4 | 5.1 | 1.9 | 3.8 | 3.4 | 2.4 |
| | b= 0.00 | 3.5 | 12.7 | 2.9 | 14.2 | 3.4 | 12.8 |
| | b= 1.00 | 2.8 | 24.2 | 1.9 | 30.6 | 2.4 | 48.7 |
| | b= 2.00 | 2.2 | 45.9 | 1.4 | 66.0 | 2.2 | 86.9 |
| Standardized Item Residuals | b=-1.00 | 3.1 | 7.1 | 1.5 | 6.9 | 2.4 | 6.1 |
| | b= 0.00 | 3.2 | 12.1 | 2.8 | 15.2 | 2.7 | 12.9 |
| | b= 1.00 | 2.5 | 19.9 | 2.2 | 25.3 | 2.5 | 37.2 |
| | b= 2.00 | 2.3 | 38.0 | 2.0 | 54.4 | 3.6 | 74.7 |

Table 13. Number of times of item administration after exposure. ($\rho = 0.25$, for 100%, gradual change in ability from a mean of -1.0 to a mean of +1.0)

| | | a=0.40 | a=0.70 | A=1.20 |
|---|---|---|---|---|
| Moving P values | b=-1.00 | 93.0 | 96.3 | 67.7 |
| | b= 0.00 | 65.7 | 77.8 | 38.1 |
| | b= 1.00 | 49.3 | 50.1 | 33.8 |
| | b= 2.00 | 45.7 | 30.9 | 15.6 |
| Moving Item Residuals | b=-1.00 | 89.6 | 122.6 | 128.9 |
| | b= 0.00 | 60.1 | 55.6 | 50 |
| | b= 1.00 | 45.0 | 42.2 | 29.6 |
| | b= 2.00 | 45.0 | 27.9 | 17.3 |
| Standardized Item Residuals | b=-1.00 | 90.1 | 124.3 | 128.8 |
| | b= 0.00 | 62.2 | 56.0 | 54.2 |
| | b= 1.00 | 46.8 | 42.3 | 30.2 |
| | b= 2.00 | 45.6 | 28.4 | 17.7 |

Table 14. Type I errors and power. ($\rho = 0.25$, for 100%, gradual change in ability from a mean of -1.0 to a mean of +1.0)

| | | a=0.40 | | a=0.70 | | a=1.20 | |
|---|---|---|---|---|---|---|---|
| | | I | II | I | II | I | II |
| Moving P Values | b=-1.00 | 1.9 | 92.9 | 1.7 | 98.5 | 4.5 | 100.0 |
| | b= 0.00 | 1.3 | 98.4 | 3.4 | 100.0 | 3.2 | 100.0 |
| | b= 1.00 | 1.2 | 99.9 | 6.1 | 100.0 | 1.0 | 100.0 |
| | b= 2.00 | 1.1 | 100.0 | 2.8 | 100.0 | 1.2 | 100.0 |
| Moving Item Residuals | b=-1.00 | 3.4 | 28.7 | 1.9 | 12.2 | 3.4 | 7.30 |
| | b= 0.00 | 3.5 | 58.8 | 2.9 | 55.7 | 3.4 | 56.94 |
| | b= 1.00 | 2.8 | 85.0 | 1.9 | 92.0 | 2.4 | 98.03 |
| | b= 2.00 | 2.2 | 98.0 | 1.4 | 99.9 | 2.2 | 100.0 |
| Standardized Item Residuals | b=-1.00 | 3.1 | 33.7 | 1.5 | 19.5 | 2.4 | 16.3 |
| | b= 0.00 | 3.2 | 58.0 | 2.8 | 57.7 | 2.7 | 57.4 |
| | b= 1.00 | 2.5 | 81.3 | 2.2 | 89.4 | 2.5 | 95.7 |
| | b= 2.00 | 2.3 | 96.8 | 2.0 | 99.5 | 3.6 | 100.0 |

Table 15. Number of times of item administration after exposure. ($\rho = 0.25$, for 10%, gradual change in ability from a mean of -1.0 to a mean of +1.0)

| | | a=0.40 | a=0.70 | A=1.20 |
|---|---|---|---|---|
| Moving P values | b=-1.00 | 353.8 | 264.0 | 160.3 |
| | b= 0.00 | 328.3 | 291.0 | 151.5 |
| | b= 1.00 | 322.9 | 344.3 | 240.1 |
| | b= 2.00 | 293.6 | 295.4 | 201.1 |
| Moving Item Residuals | b=-1.00 | 586.4 | 499.0 | 397.9 |
| | b= 0.00 | 470.7 | 496.0 | 348.8 |
| | b= 1.00 | 449.8 | 462.3 | 282.3 |
| | b= 2.00 | 364.8 | 282.9 | 169.4 |
| Standardized Item Residuals | b=-1.00 | 609.8 | 528.9 | 518.7 |
| | b= 0.00 | 548.5 | 497.7 | 375.9 |
| | b= 1.00 | 477.2 | 493.0 | 296.5 |
| | b= 2.00 | 409.6 | 306.9 | 177.6 |

Table16. Type I errors and power. ($\rho = 0.25$, for 10%, gradual change in ability from a mean of -1.0 to a mean of +1.0)

| | | a=0.40 | | a=0.70 | | a=1.20 | |
|---|---|---|---|---|---|---|---|
| | | I | II | I | II | I | II |
| Moving P Values | b=-1.00 | 1.9 | 53.2 | 1.7 | 84.7 | 4.5 | 98.5 |
| | b= 0.00 | 1.3 | 57.0 | 3.4 | 100.0 | 3.2 | 99.5 |
| | b= 1.00 | 1.2 | 54.5 | 6.1 | 83.6 | 1.0 | 96.3 |
| | b= 2.00 | 1.1 | 62.0 | 2.8 | 74.2 | 1.2 | 85.8 |
| Moving Item Residuals | b=-1.00 | 3.4 | 1.9 | 1.9 | 1.3 | 3.4 | 1.1 |
| | b= 0.00 | 3.5 | 3.0 | 2.9 | 4.5 | 3.4 | 4.3 |
| | b= 1.00 | 2.8 | 5.3 | 1.9 | 7.9 | 2.4 | 11.3 |
| | b= 2.00 | 2.2 | 9.7 | 1.4 | 12.1 | 2.2 | 23.6 |
| Standardized Item Residuals | b=-1.00 | 3.1 | 2.7 | 1.5 | 2.6 | 2.4 | 3.1 |
| | b= 0.00 | 3.2 | 2.8 | 2.8 | 4.9 | 2.7 | 4.4 |
| | b= 1.00 | 2.5 | 4.1 | 2.2 | 5.8 | 2.5 | 6.9 |
| | b= 2.00 | 2.3 | 6.4 | 2.0 | 6.5 | 3.6 | 12.3 |

Table 17. Number of times of item administration after exposure. ($\rho$ = 1.0, for 100%, abrupt change in the mean of the ability distribution)

| | | a=0.40 | a=0.70 | A=1.20 |
|---|---|---|---|---|
| Moving P values | b=-1.00 | 47.7 | 52.4 | 56.5 |
| | b= 0.00 | 38.2 | 40.6 | 45.3 |
| | b= 1.00 | 28.8 | 30.4 | 29.3 |
| | b= 2.00 | 21.8 | 18.4 | 14.2 |
| Moving Item Residuals | b=-1.00 | 38.9 | 48.9 | 63.9 |
| | b= 0.00 | 23.6 | 22.3 | 24.9 |
| | b= 1.00 | 15.4 | 14.3 | 10.1 |
| | b= 2.00 | 12.7 | 7.7 | 4.9 |
| Standardized Item Residuals | b=-1.00 | 40.6 | 50.5 | 70.0 |
| | b= 0.00 | 23.5 | 22.3 | 25.3 |
| | b= 1.00 | 15.0 | 12.5 | 8.0 |
| | b= 2.00 | 11.2 | 5.3 | 2.2 |

Table 18. Type I errors and power. ($\rho$ = 1.0, for 100%, abrupt change in the mean of the ability distribution)

| | | a=0.40 | | a=0.70 | | A=1.20 | |
|---|---|---|---|---|---|---|---|
| | | I | II | I | II | I | II |
| Moving P Values | b=-1.00 | 0 | 100.0 | 0 | 100.0 | 0 | 100.0 |
| | b= 0.00 | 0 | 100.0 | 0 | 100.0 | 0 | 100.0 |
| | b= 1.00 | 0 | 100.0 | 0 | 100.0 | 0 | 100.0 |
| | b= 2.00 | 0 | 100.0 | 0 | 100.0 | 0 | 100.0 |
| Moving Item Residuals | b=-1.00 | 4.1 | 100.0 | 4.0 | 100.0 | 6.3 | 77.8 |
| | b= 0.00 | 2.2 | 100.0 | 3.9 | 100.0 | 3.2 | 100.0 |
| | b= 1.00 | 1.9 | 100.0 | 1.3 | 100.0 | 0.5 | 100.0 |
| | b= 2.00 | 0.9 | 100.0 | 0.2 | 100.0 | 0.2 | 100.0 |
| Standardized Item Residuals | b=-1.00 | 2.7 | 100.0 | 1.7 | 100.0 | 2.2 | 100.0 |
| | b= 0.00 | 2.3 | 100.0 | 3.9 | 100.0 | 3.3 | 100.0 |
| | b= 1.00 | 2.6 | 100.0 | 3.3 | 100.0 | 2.9 | 100.0 |
| | b= 2.00 | 2.0 | 100.0 | 3.5 | 100.0 | 4.2 | 100.0 |

Table 19. Number of times of item administration after exposure. ($\rho$ = 1.0, for 10%, abrupt change in the mean of the ability distribution)

| | | a=0.40 | a=0.70 | A=1.20 |
|---|---|---|---|---|
| Moving P values | b=-1.00 | 81.1 | 71.9 | 65.4 |
| | b= 0.00 | 74.2 | 64.9 | 63.0 |
| | b= 1.00 | 68.9 | 65.8 | 57.6 |
| | b= 2.00 | 67.2 | 62.4 | 51.5 |
| Moving Item Residuals | b=-1.00 | 495.2 | 532.0 | 355 |
| | b= 0.00 | 249.3 | 206.4 | 271.9 |
| | b= 1.00 | 162.7 | 148.1 | 95.5 |
| | b= 2.00 | 131.4 | 80.1 | 54.5 |
| Standardized Item Residuals | b=-1.00 | 413.9 | 601.8 | 669.4 |
| | b= 0.00 | 248.9 | 236.2 | 274.2 |
| | b= 1.00 | 206.5 | 194.3 | 116.0 |
| | b= 2.00 | 149.9 | 95.0 | 57.4 |

Table 20. Type I errors and power. ($\rho$ = 1.0, for 10%, abrupt change in the mean of the ability distribution)

| | | a=0.40 | | a=0.70 | | a=1.20 | |
|---|---|---|---|---|---|---|---|
| | | I | II | I | II | I | II |
| Moving P Values | b=-1.00 | 0 | 946.5 | 0 | 100.0 | 0 | 100.0 |
| | b= 0.00 | 0 | 966.7 | 0 | 100.0 | 0 | 100.0 |
| | b= 1.00 | 0 | 978.5 | 0 | 99.8 | 0 | 100.0 |
| | b= 2.00 | 0 | 983.6 | 0 | 99.7 | 0 | 100.0 |
| Moving Item Residuals | b=-1.00 | 4.1 | 48.8 | 4.0 | 1.5 | 6.3 | 0.2 |
| | b= 0.00 | 2.2 | 103.8 | 3.9 | 9.9 | 3.2 | 7.1 |
| | b= 1.00 | 1.9 | 204.3 | 1.3 | 28.0 | 0.5 | 42.3 |
| | b= 2.00 | 0.9 | 419.3 | 0.2 | 61.1 | 0.2 | 82.2 |
| Standardized Item Residuals | b=-1.00 | 2.7 | 69.4 | 1.7 | 4.5 | 2.2 | 3.4 |
| | b= 0.00 | 2.3 | 109.4 | 3.9 | 9.9 | 3.3 | 7.6 |
| | b= 1.00 | 2.6 | 149.7 | 3.3 | 19.7 | 2.9 | 26.5 |
| | b= 2.00 | 2.0 | 312.5 | 3.5 | 43.7 | 4.2 | 60.6 |

Table 21. Number of times of item administration after exposure. ($\rho = 0.25$, for 100%, abrupt change in the mean of the ability distribution)

| | | a=0.40 | a=0.70 | a=1.20 |
|---|---|---|---|---|
| Moving P values | b=-1.00 | 72.1 | 67.0 | 62.3 |
| | b= 0.00 | 65.5 | 59.3 | 58.9 |
| | b= 1.00 | 55.9 | 56.1 | 50.1 |
| | b= 2.00 | 51.0 | 43.6 | 37.2 |
| Moving Item Residuals | b=-1.00 | 183.6 | 333.8 | 652.1 |
| | b= 0.00 | 105.8 | 88.7 | 102 |
| | b= 1.00 | 55.7 | 60.5 | 41.0 |
| | b= 2.00 | 50.4 | 32.8 | 24.0 |
| Standardized Item Residuals | b=-1.00 | 173.4 | 234.0 | 332.5 |
| | b= 0.00 | 104.7 | 89.3 | 103.3 |
| | b= 1.00 | 57.5 | 64.9 | 38.8 |
| | b= 2.00 | 50.9 | 26.6 | 17.7 |

Table 22. Type I errors and power. ($\rho = 0.25$, for 100%, abrupt change in the mean of the ability distribution)

| | | a=0.40 | | a=0.70 | | a=1.20 | |
|---|---|---|---|---|---|---|---|
| | | I | II | I | II | I | II |
| Moving P Values | b=-1.00 | 0 | 99.7 | 0 | 100.0 | 0 | 100.0 |
| | b= 0.00 | 0 | 100.0 | 0 | 100.0 | 0 | 100.0 |
| | b= 1.00 | 0 | 100.0 | 0 | 100.0 | 0 | 100.0 |
| | b= 2.00 | 0 | 100.0 | 0 | 100.0 | 0 | 100.0 |
| Moving Item Residuals | b=-1.00 | 4.1 | 19.9 | 4.0 | 6.7 | 6.3 | 1.4 |
| | b= 0.00 | 2.2 | 44.5 | 3.9 | 41.3 | 3.2 | 35.7 |
| | b= 1.00 | 1.9 | 76.2 | 1.3 | 86.0 | 0.5 | 94.3 |
| | b= 2.00 | 0.9 | 96.7 | 0.2 | 99.4 | 0.2 | 100.0 |
| Standardized Item Residuals | b=-1.00 | 2.7 | 25.6 | 1.7 | 15.6 | 2.2 | 8.7 |
| | b= 0.00 | 2.3 | 45.8 | 3.9 | 41.7 | 3.3 | 37.1 |
| | b= 1.00 | 2.6 | 69.5 | 3.3 | 79.7 | 2.9 | 86.8 |
| | b= 2.00 | 2.0 | 93.9 | 3.5 | 97.8 | 4.2 | 100.0 |

Table 23. Number of times of item administration after exposure. ($\rho = 0.25$, for 10%, abrupt change in the mean of the ability distribution)

| | | a=0.40 | a=0.70 | A=1.20 |
|---|---|---|---|---|
| Moving P values | b=-1.00 | 90.8 | 74.4 | 66.2 |
| | b= 0.00 | 85.0 | 68.6 | 64.8 |
| | b= 1.00 | 78.3 | 72.6 | 63.1 |
| | b= 2.00 | 96.0 | 76.4 | 64.8 |
| Moving Item Residuals | b=-1.00 | 870.9 | 538.6 | 222.6 |
| | b= 0.00 | 579.7 | 468.1 | 539.1 |
| | b= 1.00 | 438.0 | 357.0 | 288.4 |
| | b= 2.00 | 392.7 | 240.6 | 163.7 |
| Standardized Item Residuals | b=-1.00 | 768.7 | 719.0 | 659.4 |
| | b= 0.00 | 525.1 | 466.9 | 556.5 |
| | b= 1.00 | 505.8 | 521.8 | 338.9 |
| | b= 2.00 | 539.9 | 407.7 | 328.7 |

Table 24. Type I errors and power. ($\rho = 0.25$, for 10%, abrupt change in the mean of the ability distribution)

| | | a=0.40 | | a=0.70 | | a=1.20 | |
|---|---|---|---|---|---|---|---|
| | | I | II | I | II | I | II |
| Moving P Values | b=-1.00 | 0 | 89.3 | 0 | 99.9 | 0 | 100.0 |
| | b= 0.00 | 0 | 89.8 | 0 | 99.9 | 0 | 100.0 |
| | b= 1.00 | 0 | 89.1 | 0 | 98.5 | 0 | 99.7 |
| | b= 2.00 | 0 | 86.9 | 0 | 94.3 | 0 | 96.6 |
| Moving Item Residuals | b=-1.00 | 4.1 | 2.1 | 4.0 | 0.5 | 6.3 | 0.12 |
| | b= 0.00 | 2.2 | 3.0 | 3.9 | 3.1 | 3.2 | 2.4 |
| | b= 1.00 | 1.9 | 5.2 | 1.3 | 8.2 | 0.5 | 1.2 |
| | b= 2.00 | 0.9 | 10.9 | 0.2 | 15.9 | 0.2 | 26.9 |
| Standardized Item Residuals | b=-1.00 | 2.7 | 3.2 | 1.7 | 2.2 | 2.2 | 2.0 |
| | b= 0.00 | 2.3 | 3.3 | 3.9 | 3.1 | 3.3 | 2.7 |
| | b= 1.00 | 2.6 | 3.5 | 3.3 | 5.2 | 2.9 | 5.4 |
| | b= 2.00 | 2.0 | 6.4 | 3.5 | 7.2 | 4.2 | 8.3 |

Figure 1. Plot of item exposure statistics for item 5. (normal ability distribution, ρ = 0.0)

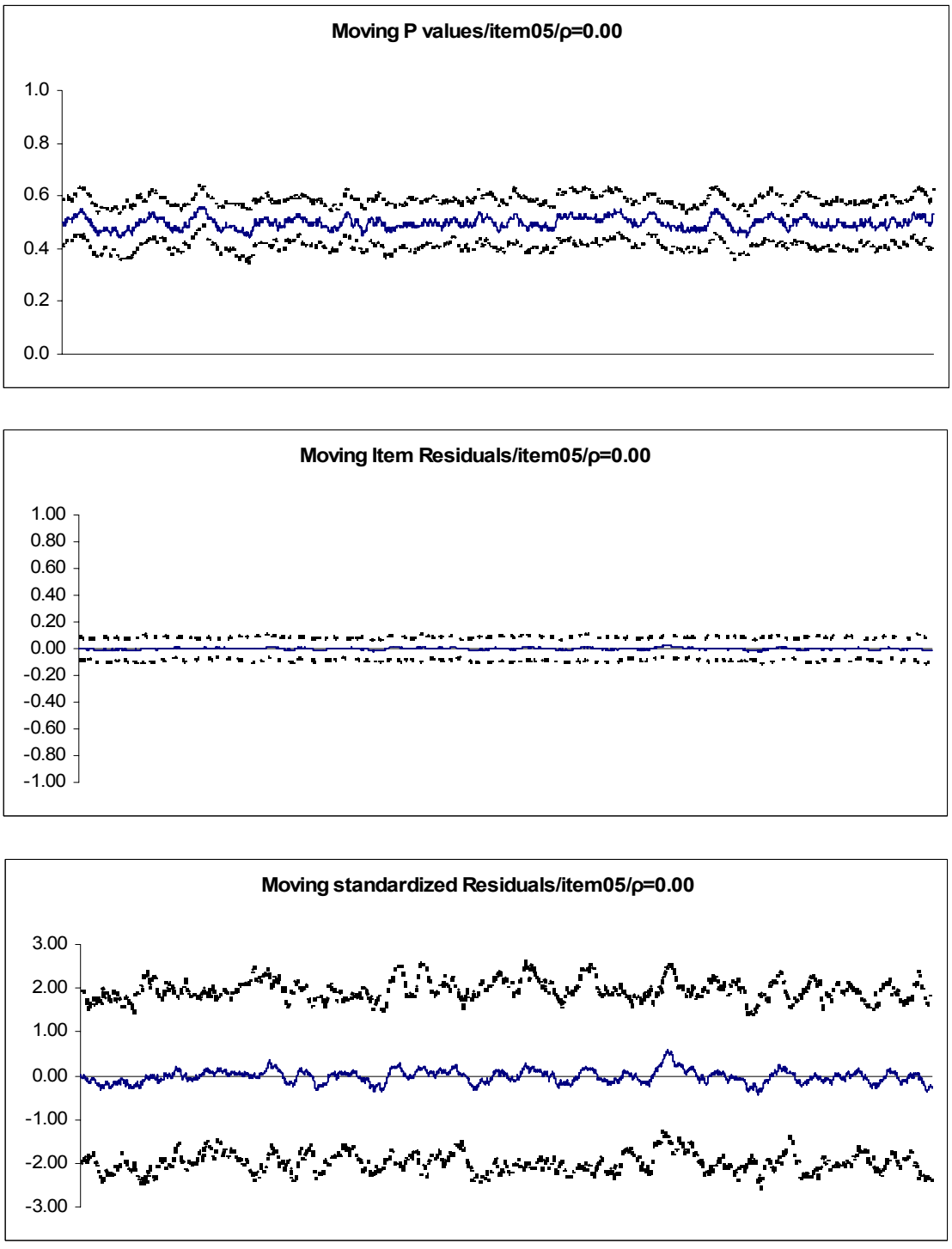


Figure 2. Plot of item exposure statistics for item 5. (gradually shifting ability distribution, ρ = 0.0)

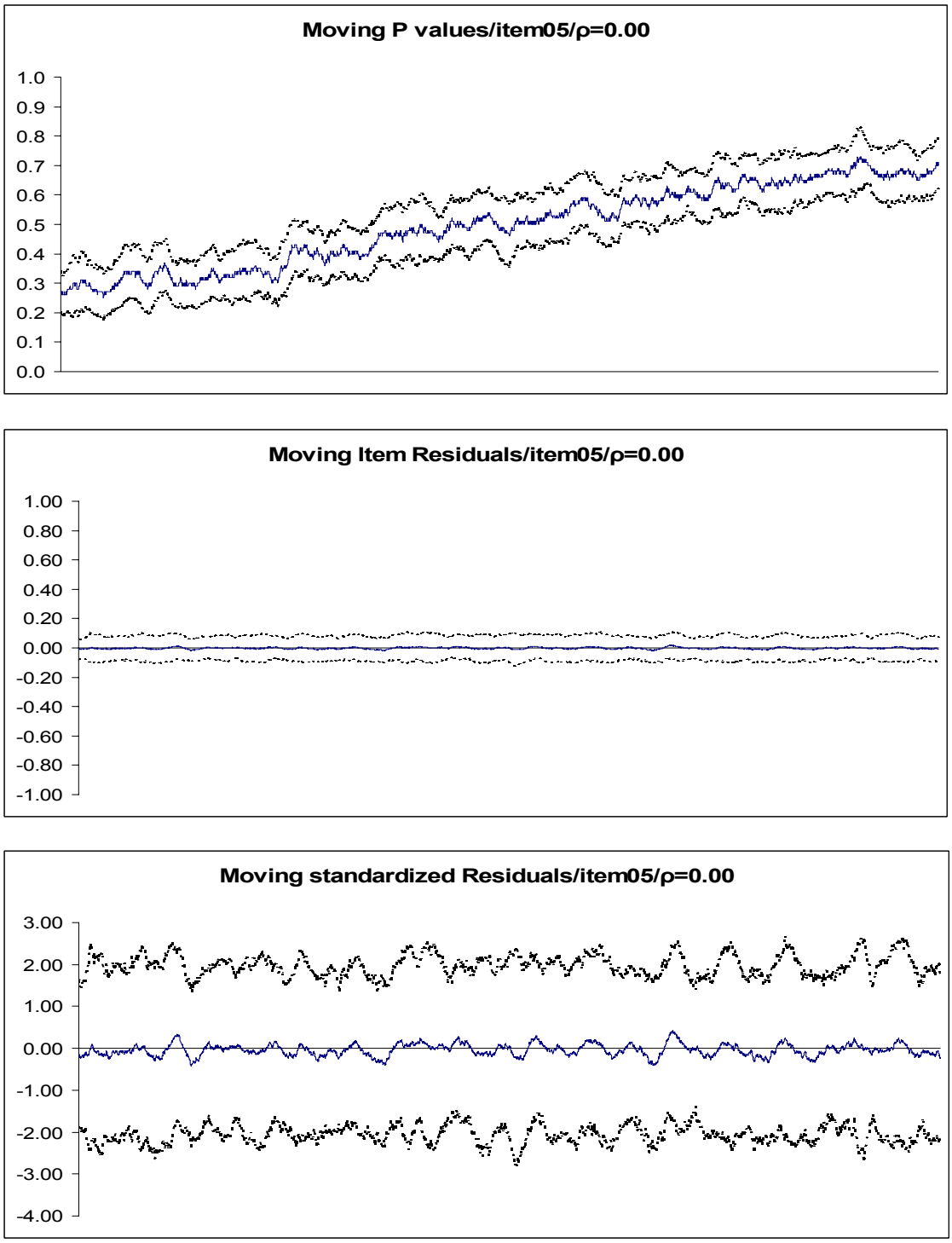


Figure 3. Plot of item exposure statistics for item 5. (abrupt shift in ability distribution, ρ = 0.0)

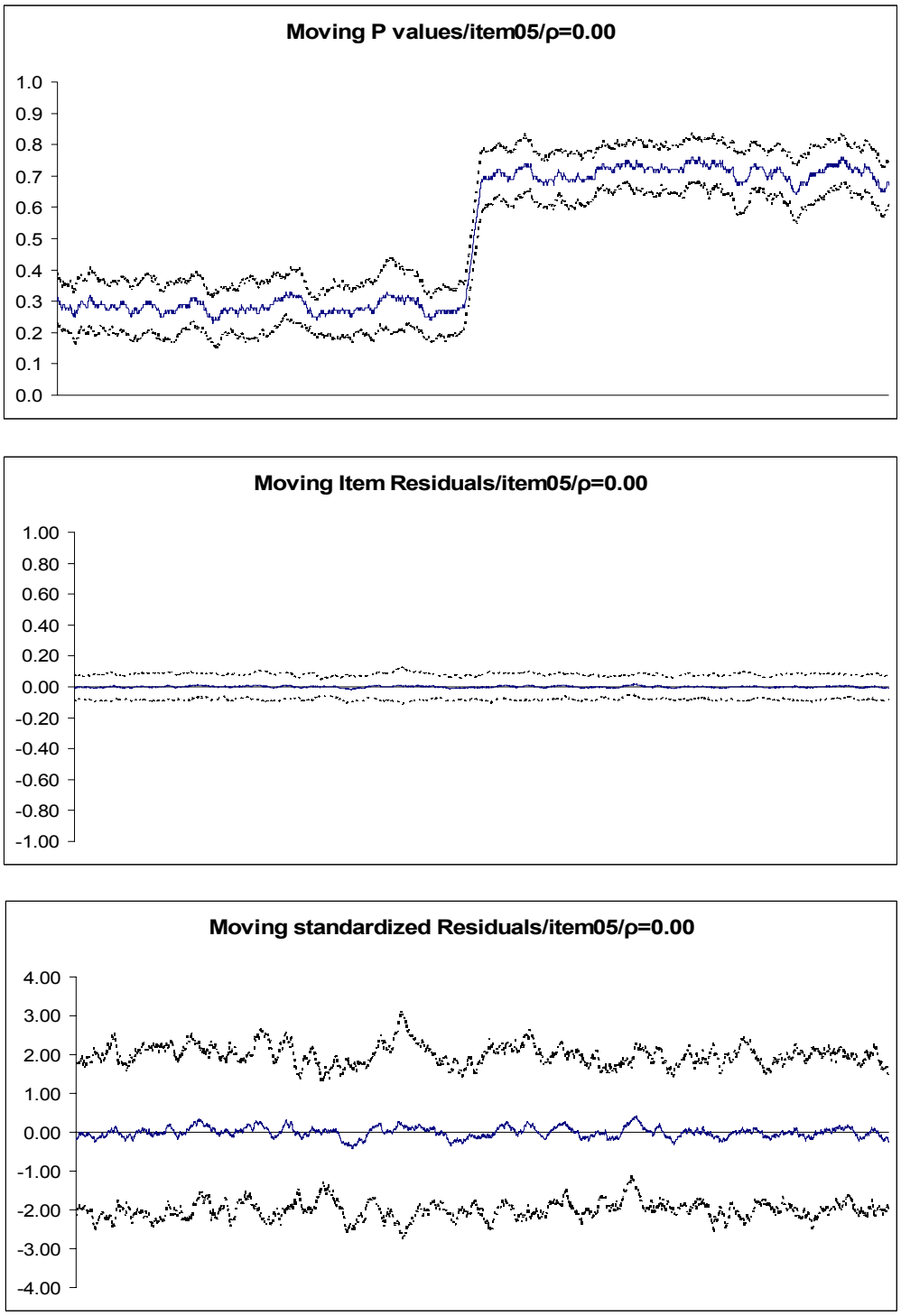


Figure 4. Plot of item exposure statistics for item 5. (normal ability distribution, ρ = 1.0, 100%)

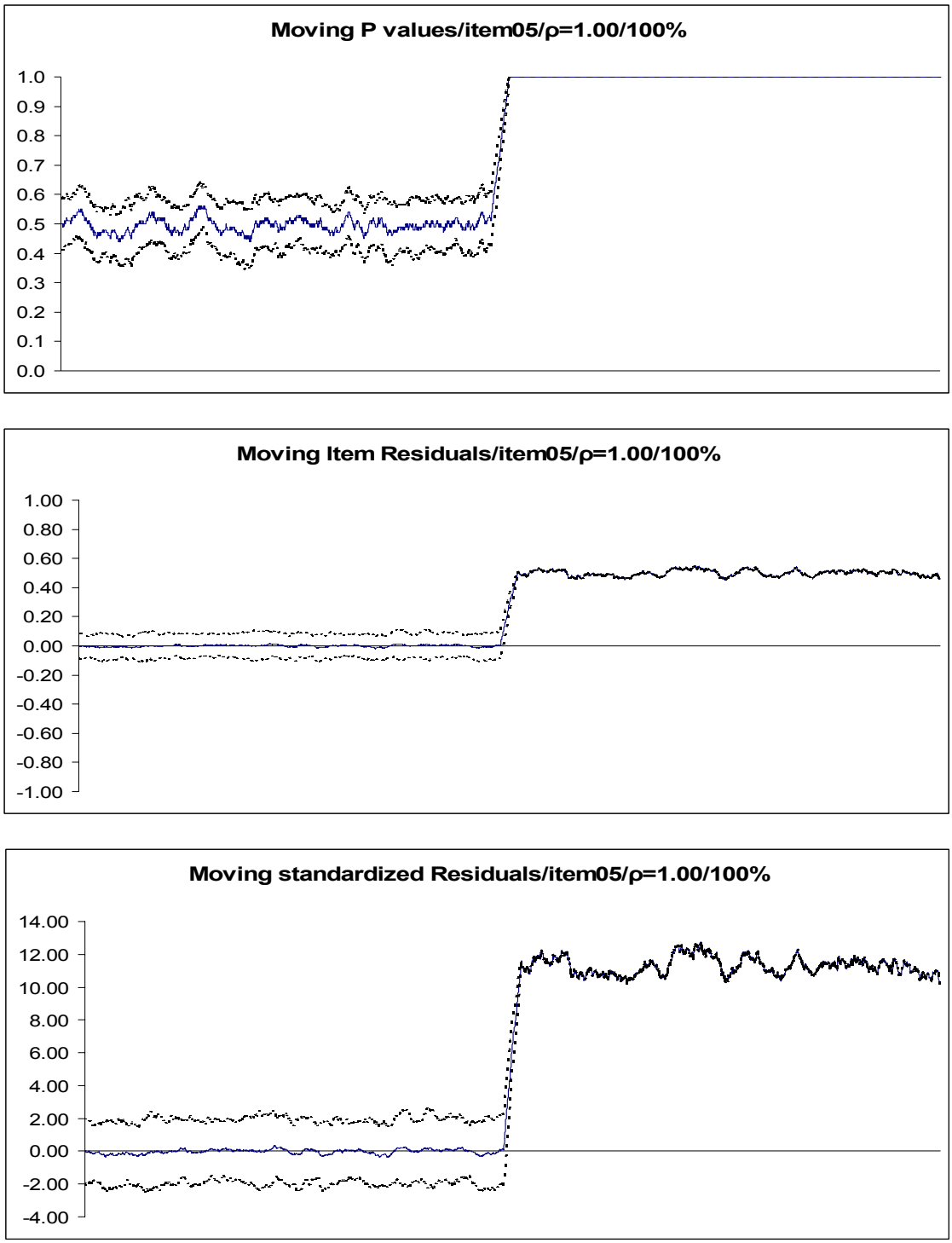


Figure 5. Plot of item exposure statistics for item 5. (normal ability distribution, ρ = 1.0, 10%)

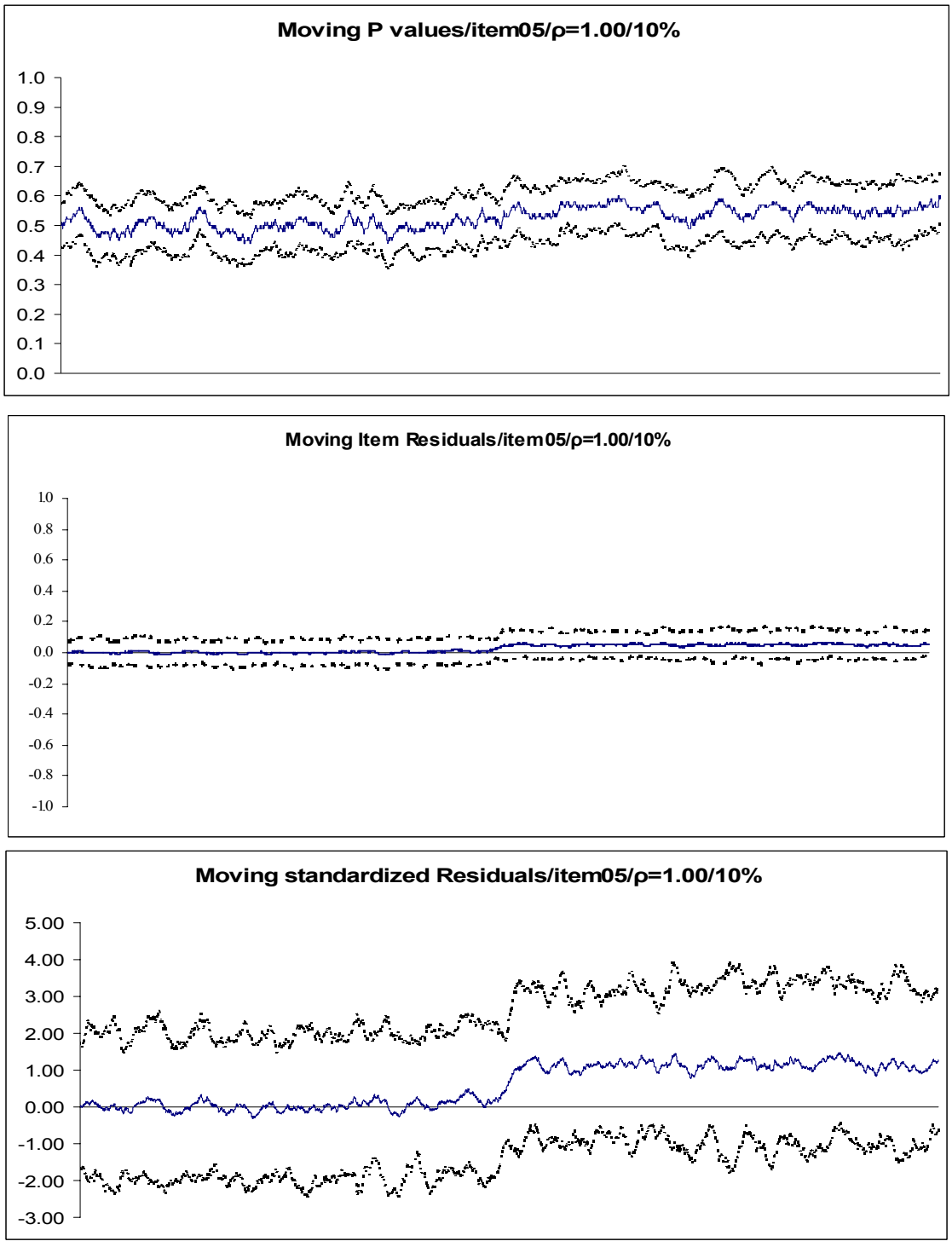


Figure 6. Plot of item exposure statistics for item 5. (normal ability distribution, ρ = 0.25, 100%)

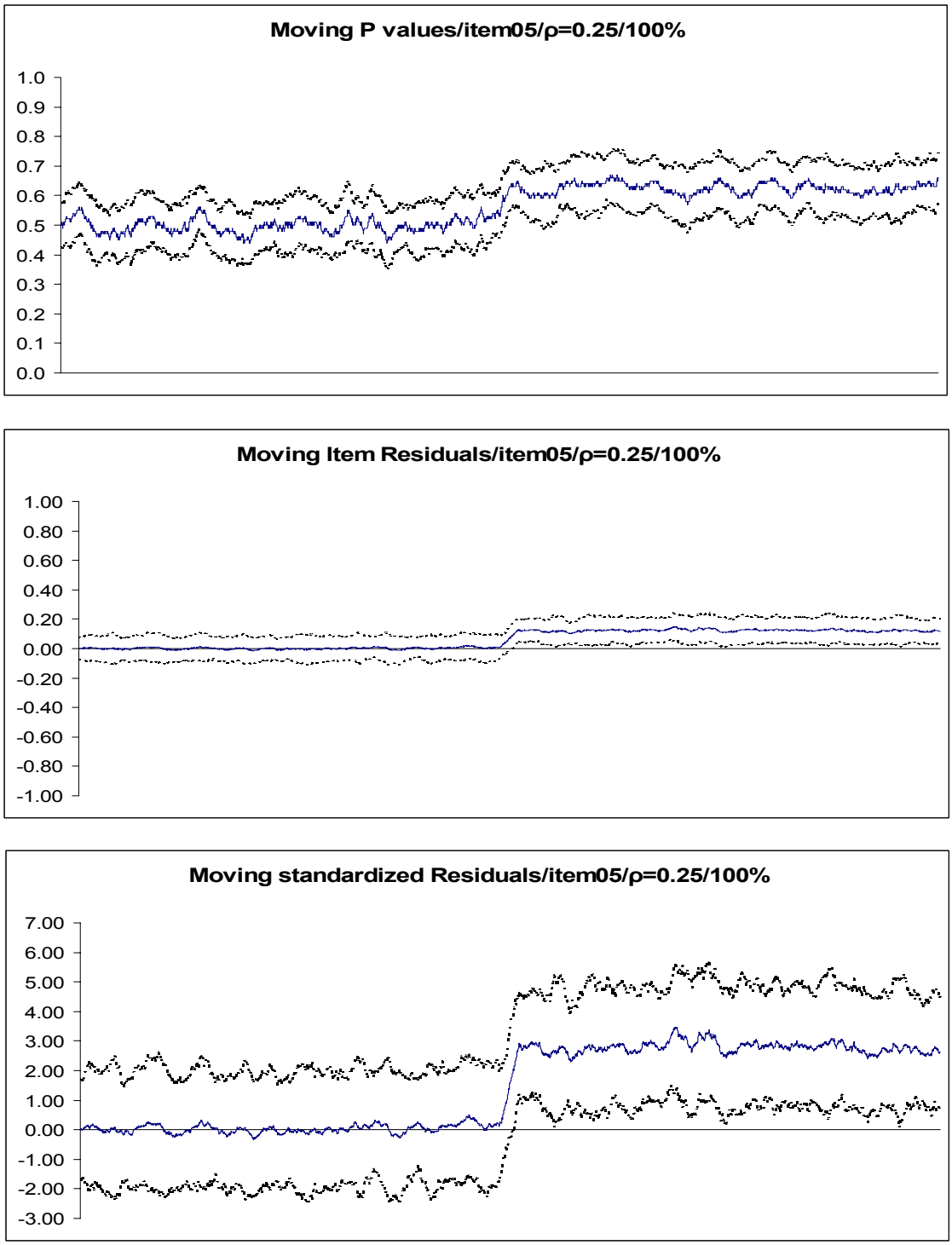
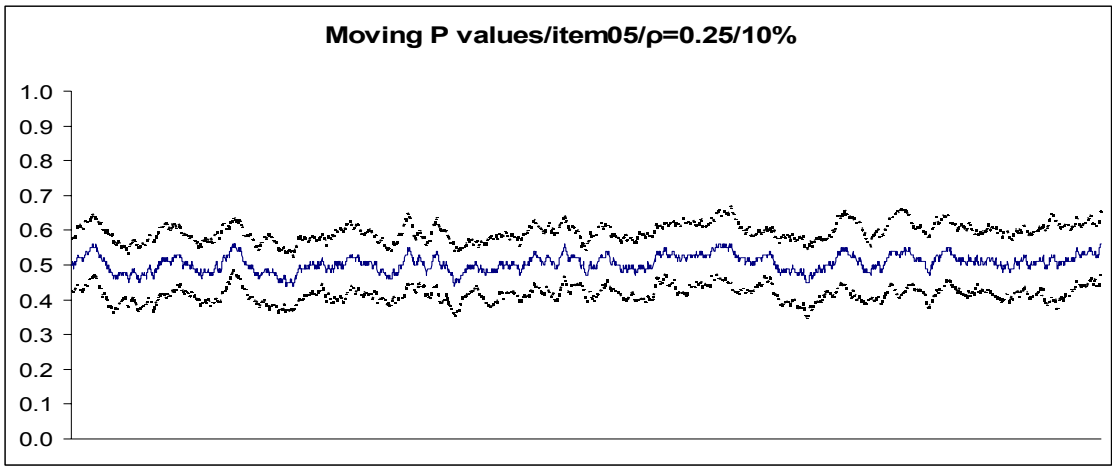
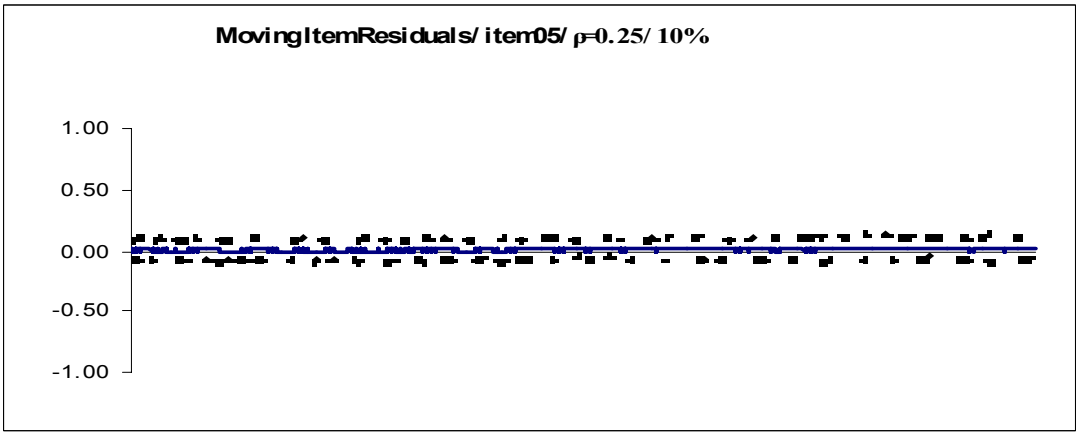


Figure 7. Plot of item exposure statistics for item 5. (normal ability distribution, ρ = 0.25, 10%)



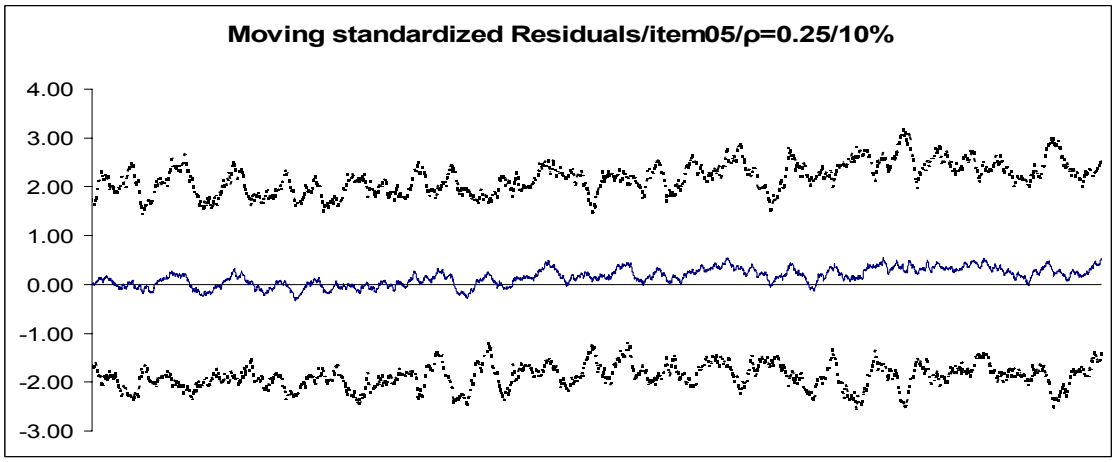



Figure 8. Plot of item exposure statistics for item 5. (gradually shifting ability distribution, ρ = 1.0, 100%)

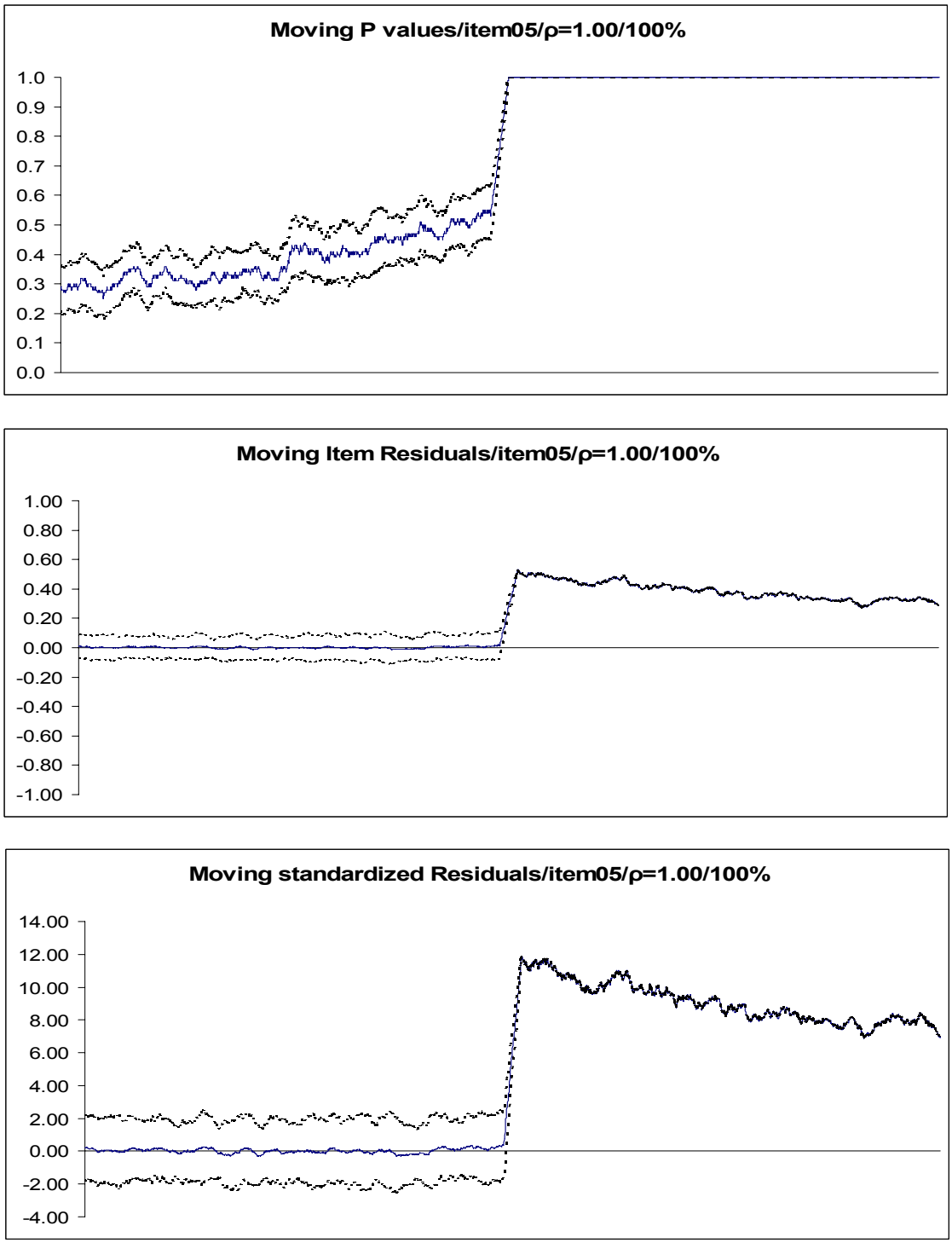


Figure 9. Plot of item exposure statistics for item 5. (gradually shifting ability distribution, ρ = 1.0, 10%)

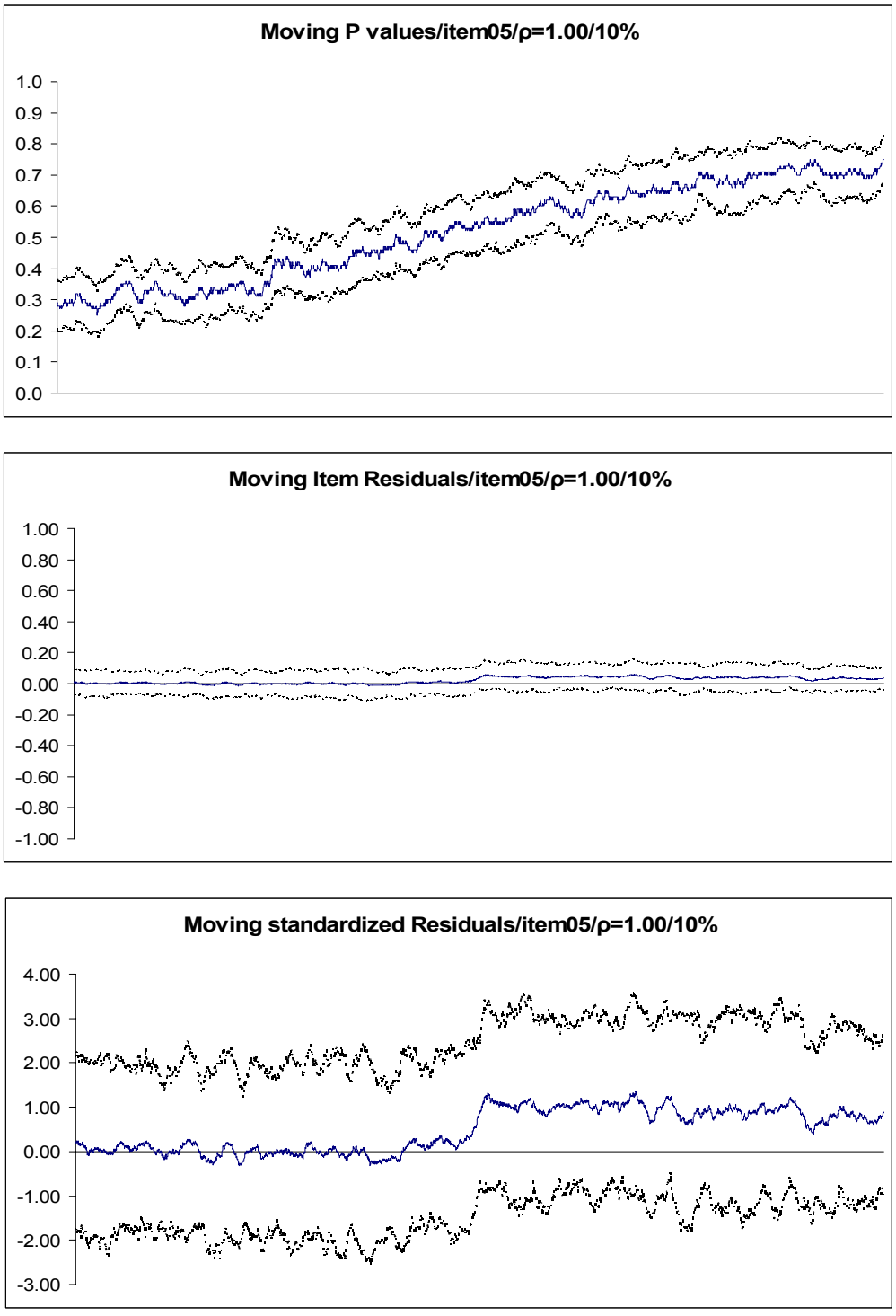


Figure 10. Plot of item exposure statistics for item 5. (gradually shifting ability distribution, ρ = 0.25, 100%)

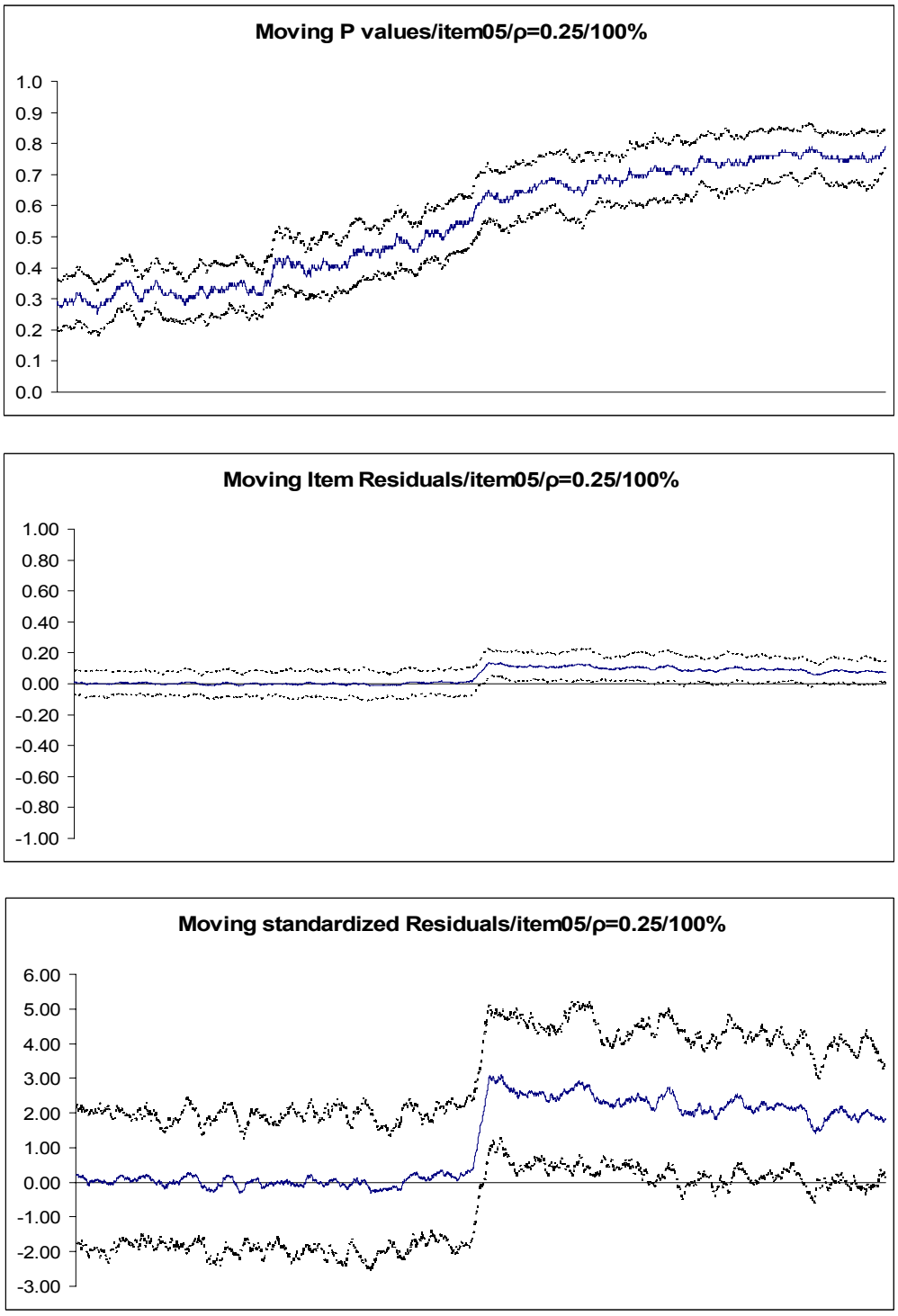


Figure 11. Plot of item exposure statistics for item 5. (gradually shifting ability distribution, ρ = 0.25, 10%)

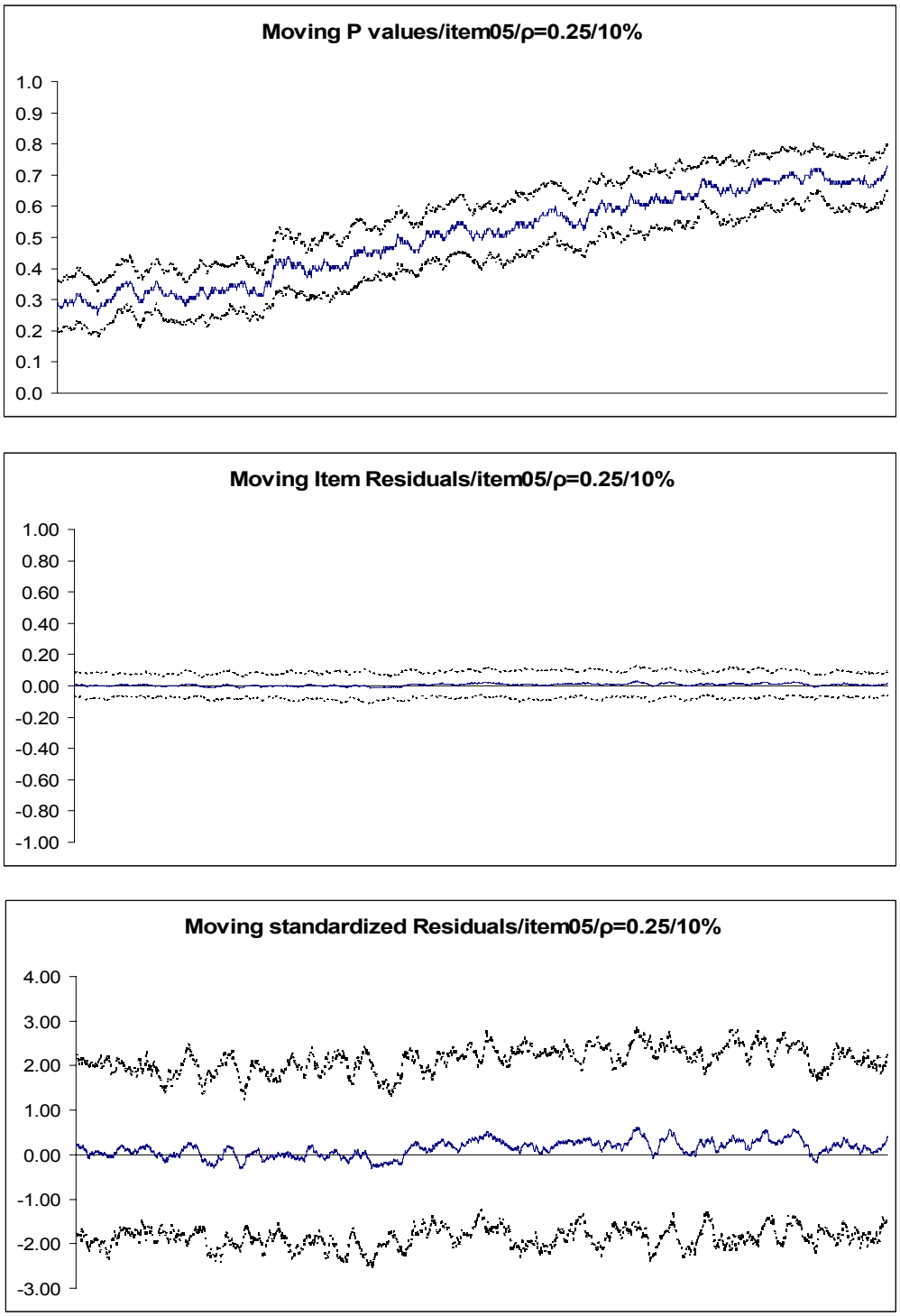
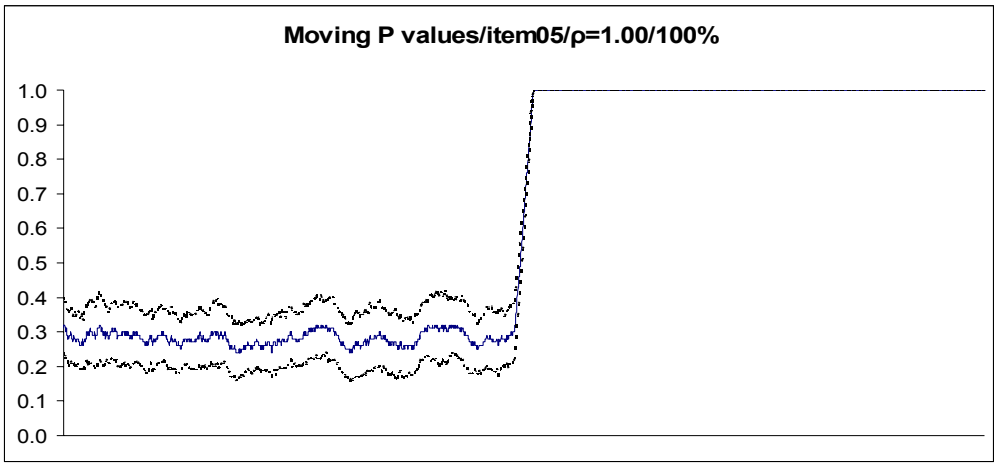
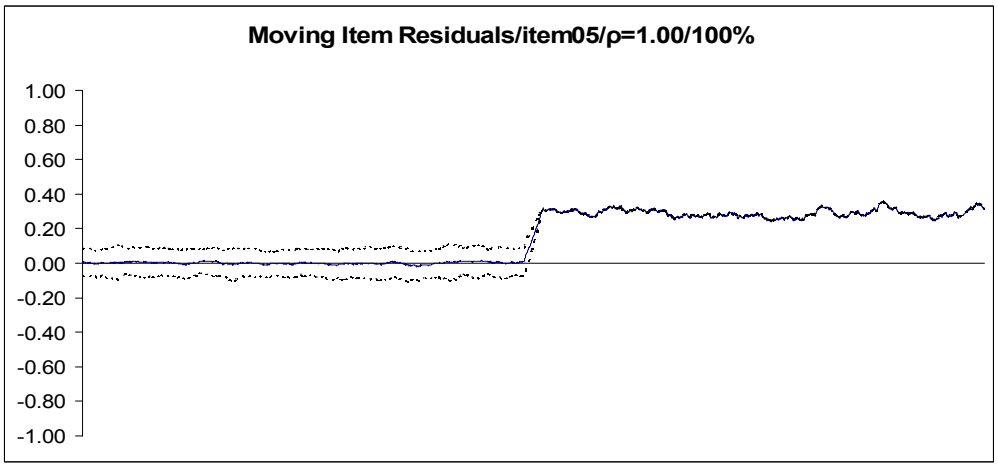


Figure 12. Plot of item exposure statistics for item 5. (abrupt shifting ability distribution, ρ = 1.0, 100%)



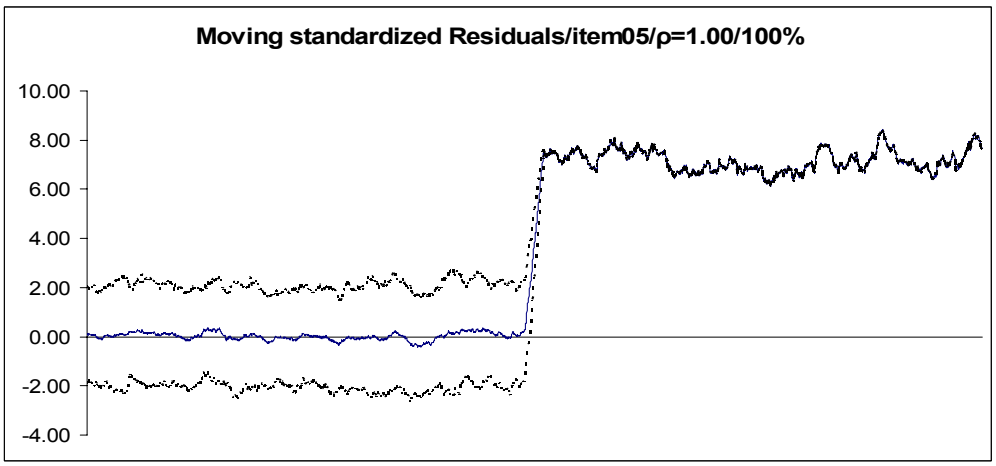



Figure 13. Plot of item exposure statistics for item 5. (abrupt shifting ability distribution, ρ = 1.0, 10%)

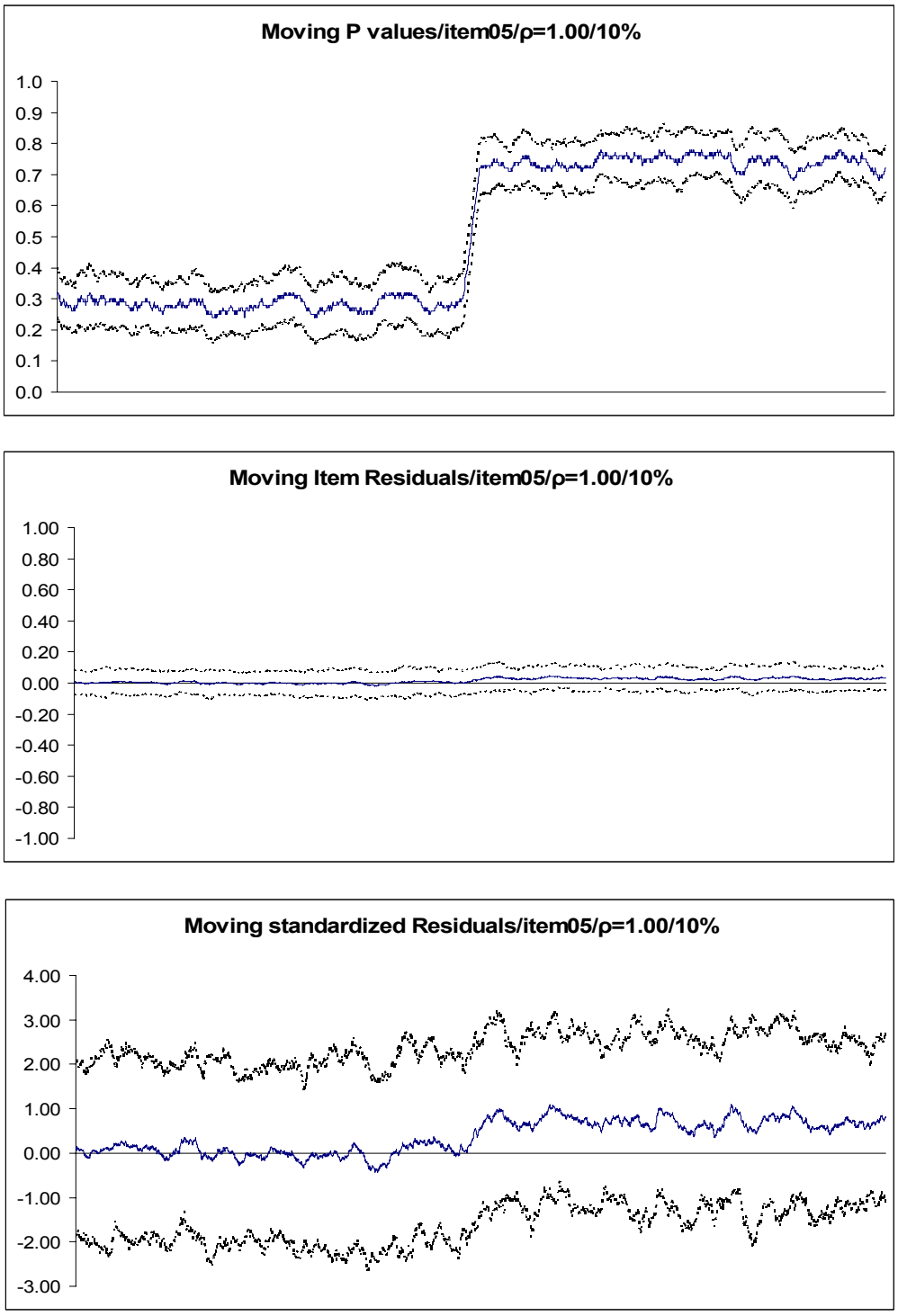


Figure 14. Plot of item exposure statistics for item 5. (abrupt shifting ability distribution, ρ = 0.25, 100%)

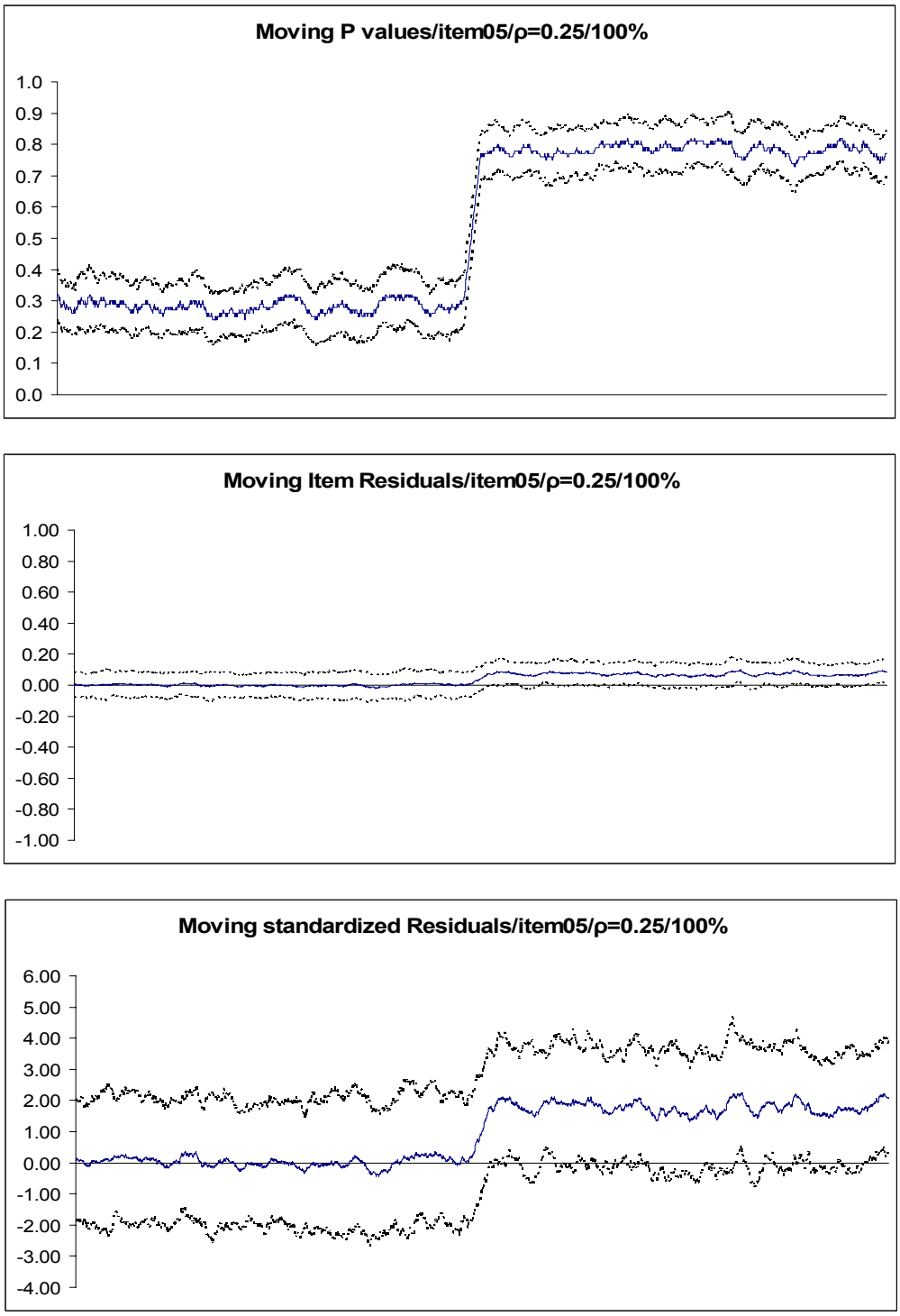


Figure 15. Plot of item exposure statistics for item 5. (abrupt shifting ability distribution, ρ = 0.25, 10%)