

Article

Weighted Sequential Construction of Goodness-of-Fit Statistics from Pairwise Concordance Marginal Information in Sparse Binary IRT Models with Symmetric Intercepts and Slopes

Jinhui Xu ^{1,*}, Runqi Li ^{2,†} and Mark Reiser ³

¹ School of Mathematics, Jinan University, 601 W. Huangpu Avenue, Guangzhou 510632, China

² Jinan University–University of Birmingham Joint Institute, Jinan University, 855 E. Xingye Avenue, Guangzhou 511443, China; rxl350@student.bham.ac.uk

³ School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ 85287, USA; mark.reiser@asu.edu

* Correspondence: xujinhui@jnu.edu.cn

† These authors contributed equally to this work.

Abstract

When many binary items are analyzed, most cells in the complete response table are empty. Full-table statistics may then be unreliable, and an omnibus result does not locate the misfit. This paper develops a local goodness-of-fit testing method for binary item response models by extending the weighted sequential sums-of-squares construction to pairwise concordance margins. The construction gives ordered one-degree-of-freedom components for individual item pairs. They are compared with Cholesky components and four local diagnostics under zero and wide symmetric intercepts. Omnibus statistics are examined separately under symmetric slopes. The sequential components have an empirical Type I error close to the nominal levels. With zero intercepts, all methods show substantial power for the target pairs, although pair order moves some shared discrepancy into non-target components. Cholesky gives higher target-pair power and broader moderate elevations across non-target pairs, whereas the sequential components show more concentrated, order-dependent peaks. With wide symmetric intercepts, power differs greatly among target pairs, and the Lagrange multiplier statistic is particularly sensitive to intercept distance. The main qualitative patterns persist across three nominal levels. Lower-order omnibus statistics generally maintain an empirical Type I error close to the nominal levels and detect the slope alternatives, whereas Pearson–Fisher is liberal in the sparse full table. A symmetric parameter vector therefore need not give uniform local diagnostic results.

Keywords: goodness-of-fit diagnostics; item response theory; symmetric intercept and slope patterns; sequential sums of squares; concordance margins; sparse binary data



Academic Editors: Yichuan Zhao and Chioneso Show Marange

Received: 11 August 2026

Revised: 4 September 2026

Accepted: 8 September 2026

Published: 11 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Goodness-of-fit tests are used to decide whether a statistical model gives an adequate description of observed data. Pearson’s chi-square statistic and the likelihood-ratio statistic are common choices for multinomial models [1–3]. Their use becomes difficult when many binary variables are cross-classified because the number of response patterns increases exponentially with the number of variables. The resulting table can contain many empty or nearly empty cells. The usual chi-square reference distribution may then be inaccurate, and the inclusion of many higher-order cells can dilute power against a lower-order model error [4–8].

Tests based on lower-order marginal frequencies were developed to address these problems. If a model fits the joint response distribution, it must also fit its univariate and bivariate margins. These margins combine many cells from the full table and have larger expected frequencies. A test based on them is therefore less affected by sparsity and can focus on the part of the distribution where lack of fit is expected [9–15]. Limited-information tests such as M_2 give a useful overall assessment, but a single statistic may still give little guidance about the item or item pair responsible for rejection [16–18].

The broader IRT model-fit literature includes item-fit statistics, residual diagnostics, and pairwise measures of local dependence [19–22]. These approaches address different aspects of model misfit, but they share a practical challenge: a large collection of correlated local diagnostics can be difficult to interpret. Pairwise limited-information procedures have also been extended to complex survey samples [23]. The concordance construction studied here belongs to this broader group of local methods and is intended to be used alongside an overall fit assessment.

Reiser [24] developed a more detailed approach based on orthogonal components of the Pearson–Fisher statistic. The method uses Pearson weights, adjusts for estimated model parameters, and applies sequential sums of squares to an ordered set of marginal directions. Each new component measures the additional lack of fit explained by one marginal direction after the preceding directions have been included. Under the null model, the components have asymptotic chi-square distributions with one degree of freedom and are mutually independent. Their sum gives the part of the adjusted Pearson–Fisher statistic represented by the selected margins. The method therefore links an overall test to local sources of misfit. Later work studied these components in sparse tables and used them to form orthogonal *GFfit* diagnostics [8,25–27].

Xu [28] introduced a concordance-based summary for binary pairs. For items i and j , concordance includes the two cells in which both responses are zero or both responses are one. The resulting statistic is denoted by X_{2c}^2 . It uses both types of pairwise agreement rather than the single cell in which both responses equal one. A later study extended concordance-based components to sparse longitudinal binary data [29]. The sequential sums-of-squares decomposition of X_{2c}^2 has not yet been studied in detail under symmetric item parameter structures or compared directly with a covariance-based Cholesky construction. This leaves a practical question about localization. When pairwise marginal directions are correlated, the component recorded for a pair may depend on the directions entered before it. Previous concordance studies have not examined how much the component labels change when the same discrepancy is decomposed in a different order.

The item parameters themselves may affect a directional test. Xu et al. [30] compared an LM statistic, an orthogonal component, $GFfit^{(ij)}$, and $\bar{X}_{ij}^2$ under zero, symmetric, and asymmetric intercept patterns. Type I error changed little across the intercept settings, but power decreased when the two intercepts in a pair were farther apart. Extreme thresholds have also affected the finite-sample behavior of goodness-of-fit statistics for categorical data [22,31]. These findings motivate the intercept comparison in the present study, where a wider symmetric range is used to examine whether the same loss of local information occurs for sequential concordance components. The slope comparison is treated separately because slopes control the strength and pattern of association, whereas intercepts determine where the item probabilities lie on the response scale.

This paper examines whether pairwise concordance information can identify local misfit in sparse binary IRT models. This study makes three contributions. First, we extend Reiser’s weighted sequential sums-of-squares construction to pairwise concordance margins, obtaining an ordered decomposition of X_{2c}^2 [24]. Second, we clarify that pair order does not change the total discrepancy but can change the individual components, and we

examine the consequences for localization. Third, we compare the resulting components with Cholesky components and four established local diagnostics under zero and wide symmetric intercept patterns. A separate study examines lower-order omnibus statistics under symmetric slope patterns. Together, the two studies evaluate componentwise Type I error, detection of model misspecification, and localization of the pairwise signal.

2. Materials and Methods

2.1. Binary Item Response Models

Let Y_j be the binary response to item j . Under a one-factor two-parameter logistic item response theory model, referred to as a 2PL model [32,33],

$$\Pr(Y_j = 1 \mid X_1 = x_1) = \frac{\exp(\beta_{0j} + \beta_{1j}x_1)}{1 + \exp(\beta_{0j} + \beta_{1j}x_1)}, \quad j = 1, \dots, q, \quad (1)$$

where $X_1 \sim N(0, 1)$. The parameter β_{0j} is the intercept of item j , and β_{1j} is its slope on the first factor. The two-factor 2PL model used to generate data under lack of fit is

$$\Pr(Y_j = 1 \mid X_1 = x_1, X_2 = x_2) = \frac{\exp(\beta_{0j} + \beta_{1j}x_1 + \beta_{2j}x_2)}{1 + \exp(\beta_{0j} + \beta_{1j}x_1 + \beta_{2j}x_2)}, \quad (2)$$

where X_1 and X_2 are independent standard normal variables. The parameter β_{2j} is the slope of item j on the second factor. The vectors β_0 , β_1 , and β_2 collect these item parameters.

The simulation uses symmetric parameter arrangements. A slope vector is called symmetric when items at equal distances from the two ends have equal slopes. For factor r , this condition is $\beta_{rj} = \beta_{r,q+1-j}$. The centered symmetric intercept pattern has equal magnitudes and opposite signs at matching positions, so $\beta_{0j} = -\beta_{0,q+1-j}$. These definitions describe the parameter settings used in the simulation.

2.2. Pairwise Concordance and Adjusted Pearson–Fisher Residuals

Let $T = 2^q$, and let $\mathbf{y}_s = (y_{s1}, \dots, y_{sq})^\top$ be response pattern s for $s = 1, \dots, T$. The vectors $\hat{\mathbf{p}}$ and $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}(\hat{\boldsymbol{\theta}})$ contain the observed and fitted response-pattern proportions. Write

$$\mathbf{r} = \hat{\mathbf{p}} - \hat{\boldsymbol{\pi}}, \quad \mathbf{D} = \text{diag}(\hat{\boldsymbol{\pi}}),$$

and define the Pearson residual vector

$$\mathbf{z} = \sqrt{N} \mathbf{D}^{-1/2} \mathbf{r}. \quad (3)$$

Let

$$\mathbf{J} = \left. \frac{\partial \boldsymbol{\pi}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} \right|_{\hat{\boldsymbol{\theta}}}, \quad \mathbf{A} = \mathbf{D}^{-1/2} \mathbf{J}.$$

The fitted-parameter and probability-sum directions are removed by

$$\hat{\mathbf{W}} = \mathbf{I} - \hat{\boldsymbol{\pi}}^{1/2} (\hat{\boldsymbol{\pi}}^{1/2})^\top - \mathbf{A} (\mathbf{A}^\top \mathbf{A})^{-} \mathbf{A}^\top, \quad (4)$$

where $(\cdot)^{-}$ is a generalized inverse [24,34].

For item pair (i, j) , define

$$P_{ij}^{11} = \Pr(Y_i = 1, Y_j = 1), \quad P_{ij}^{00} = \Pr(Y_i = 0, Y_j = 0),$$

and let $P_{ij}^{\text{PC}} = P_{ij}^{11} + P_{ij}^{00}$. The quantity P_{ij}^{PC} is the probability that the two responses agree. Following Xu [28], define

$$\mathbf{v}_j = (y_{1j}, \dots, y_{Tj})^\top, \quad \bar{\mathbf{v}}_j = \mathbf{1}_T - \mathbf{v}_j,$$

and let $\circ$ denote elementwise multiplication. The rows for pair (i, j) in the one-concordant and zero-concordant extraction matrices are

$$[\mathbf{H}_{[2]}]_{(ij),\cdot} = (\mathbf{v}_i \circ \mathbf{v}_j)^\top, \quad [\mathbf{H}^{[2]}]_{(ij),\cdot} = (\bar{\mathbf{v}}_i \circ \bar{\mathbf{v}}_j)^\top, \quad i < j. \quad (5)$$

The pairwise-concordance extraction matrix is

$$\mathbf{M}_{2c} = \mathbf{H}_{[2]} + \mathbf{H}^{[2]}. \quad (6)$$

Each row of $\mathbf{M}_{2c}$ collects the response patterns in which the corresponding pair agrees.

For a general extraction matrix $\mathbf{L}$, define

$$\mathbf{e}_L = \mathbf{L}\mathbf{r}, \quad \hat{\Sigma}_L = \mathbf{L}\hat{\Sigma}_r\mathbf{L}^\top, \quad (7)$$

where

$$\hat{\Sigma}_r = \mathbf{D} - \hat{\pi}\hat{\pi}^\top - \mathbf{J}(\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{J}^\top. \quad (8)$$

The corresponding marginal statistic is

$$X_L^2 = \mathbf{N}\mathbf{e}_L^\top \hat{\Sigma}_L^{-1} \mathbf{e}_L. \quad (9)$$

The choices $\mathbf{L} = \mathbf{M}_{2c}$ and $\mathbf{L} = \mathbf{H}_{[2]}$ give X_{2c}^2 and $X_{[2]}^2$, respectively.

2.3. Weighted Sequential Sums of Squares for Pairwise Concordance

We apply the weighted sequential sums-of-squares construction in Reiser [24] to the concordance margins. Let $\mathbf{L}$ be a $d_L \times T$ extraction matrix, let $\mathbf{b}_\ell^\top$ be row ℓ of $\mathbf{L}$, and let $[\mathbf{L}^\top]_\ell$ be column ℓ of $\mathbf{L}^\top$. Because $\mathbf{L}$ acts on raw probability margins in this paper, the adjusted direction matrix is

$$\hat{\mathbf{M}}_L = \hat{\mathbf{W}}\mathbf{D}^{1/2}\mathbf{L}^\top. \quad (10)$$

The columns of $\hat{\mathbf{M}}_L$ are the marginal directions after the probability-sum and fitted-parameter directions have been removed. This is the weighted-regression form of the sequential construction on the Pearson residual scale.

Orthogonal components arise from incremental sums of squares obtained by regressing $\mathbf{z}$ on the ordered columns of $\hat{\mathbf{M}}_L$ under the inner product induced by $\hat{\mathbf{W}}$. Under a fixed row ordering of $\mathbf{L}$, the first one-degree-of-freedom component is

$$\begin{aligned} SS([\mathbf{L}^\top]_1) &= \hat{\gamma}_1^2 = \mathbf{z}^\top \hat{\mathbf{W}}\mathbf{D}^{1/2}[\mathbf{L}^\top]_1 \\ &\quad \times \left(\mathbf{b}_1^\top \mathbf{D}^{1/2} \hat{\mathbf{W}}\mathbf{D}^{1/2} [\mathbf{L}^\top]_1 \right)^{-1} \mathbf{b}_1^\top \mathbf{D}^{1/2} \hat{\mathbf{W}}\mathbf{z}. \end{aligned} \quad (11)$$

To obtain the second component, the second column $[\mathbf{L}^\top]_2$ is first orthogonalized with respect to $[\mathbf{L}^\top]_1$ under the same weighted inner product. Its orthogonal complement is

$$[\mathbf{L}^\top]_2^\perp = [\mathbf{L}^\top]_2 - [\mathbf{L}^\top]_1 \left(\mathbf{b}_1^\top \mathbf{D}^{1/2} \hat{\mathbf{W}}\mathbf{D}^{1/2} [\mathbf{L}^\top]_1 \right)^{-1} \mathbf{b}_1^\top \mathbf{D}^{1/2} \hat{\mathbf{W}}\mathbf{D}^{1/2} [\mathbf{L}^\top]_2, \quad (12)$$

and the corresponding sequential sum of squares is defined in the same way as Equation (11). Further components repeat this orthogonalization and sum-of-squares step for $\ell = 3, \dots, d_L$. The resulting nondegenerate squared components are asymptotically independent χ_1^2 variables, and their sum equals X_L^2 [24].

A permutation of the pair order permutes the rows of $\mathbf{L}$ and, hence, the columns of $\hat{\mathbf{M}}_{\mathbf{L}}$. The selected column space is unchanged, so the sum of the nondegenerate components remains $X_{\mathbf{L}}^2$. Individual increments need not remain the same because the sequence of nested column spaces changes. Thus, order invariance applies to the total selected discrepancy, not to the component carrying a particular pair label.

For the proposed statistic, $\mathbf{L} = \mathbf{M}_{2c}$, and the pair order is fixed as $(1, 2), (1, 3), \dots, (14, 15)$. The component attached to pair k is denoted by $X_{2c, SS, k}^2$. Its label records the pair entering at that step, but the component is calculated after all earlier directions have been removed. A large value can therefore contain a signal shared with later pairs and should not be read as the unadjusted residual of the labeled pair.

In practice, an SS component is an ordered increment rather than the unadjusted residual of the labeled pair. The pair order should be stated explicitly. Evidence for a particular problematic pair is stronger when a direct diagnostic such as LM or $GFfit^{(ij)}$ gives a similar result, or when the pair remains important under another order chosen before examining the data.

2.4. Cholesky Components for Pairwise Concordance

The comparison construction works directly with the covariance matrix in Equation (7). If $\hat{\Sigma}_{\mathbf{L}} = \mathbf{C}_{\mathbf{L}}\mathbf{C}_{\mathbf{L}}^T$, where $\mathbf{C}_{\mathbf{L}}$ is lower triangular, define

$$\gamma_{\text{Chol}} = \mathbf{C}_{\mathbf{L}}^{-1}\sqrt{N}\mathbf{e}_{\mathbf{L}}, \quad X_{\mathbf{L}, \text{Chol}, k}^2 = \gamma_{\text{Chol}, k}^2. \quad (13)$$

Each component has one degree of freedom under the null model, and the sum of the component statistics equals $X_{\mathbf{L}}^2$. Sequential sums of squares compare nested sets of adjusted marginal directions. Cholesky instead transforms the extracted residuals through their estimated covariance. Both constructions use the same pair order in Study 1 [35–37].

The triangular Cholesky factor also changes when the residual vector and covariance matrix are permuted. Its individual components are therefore order-dependent, although the degree of sensitivity may differ from that of the sequential components.

2.5. Other Local Diagnostics

Study 1 uses five statistics as comparisons for the proposed $X_{2c, SS}^2$ component. The first is $X_{2c, \text{Chol}}^2$ from Equation (13). For item pair (i, j) , the Lagrange multiplier statistic tests the null hypothesis $(\beta_{i2}, \beta_{j2})^T = (0, 0)^T$ against a two-factor extension. Let $\mathbf{h}_2$ be the score for the two tested slopes under the null hypothesis. Partition the information matrix into a nuisance block $\mathbf{I}_{11}$ and blocks involving the tested slopes. The efficient information is

$$\mathbf{U} = \mathbf{I}_{22} - \mathbf{I}_{21}\mathbf{I}_{11}^{-1}\mathbf{I}_{12},$$

and the statistic is

$$LM(\beta_{i2}, \beta_{j2}) = \mathbf{h}_2^T \mathbf{U}^{-1} \mathbf{h}_2. \quad (14)$$

It has two degrees of freedom under the null model [30,38–40].

The orthogonal statistic $GFfit_{\perp}^{(ij)}$ is the one-degree-of-freedom SS component obtained by setting $\mathbf{L} = \mathbf{H}_{[2]}$. It is the procedure called “Orth. Comp.” by Xu et al. [30]. The perpendicular symbol distinguishes it from the direct pairwise statistic. For the latter, let $\hat{p}_{ij}^{ab}$ and $\hat{\pi}_{ij}^{ab}$ be the observed and fitted probabilities for cell (a, b) of the binary pair. Then,

$$GFfit^{(ij)} = N \sum_{a=0}^1 \sum_{b=0}^1 \frac{(\hat{p}_{ij}^{ab} - \hat{\pi}_{ij}^{ab})^2}{\hat{\pi}_{ij}^{ab}}. \quad (15)$$

The statistic is compared with χ^2_1 for binary pairs and is not orthogonal across pairs [27,30,41]. The statistic $\bar{X}_{ij}^2$ applies a mean and variance correction to Equation (15). If $\hat{\Sigma}_{ij}$ is the adjusted four-cell covariance and $\hat{\mathbf{D}}_{ij} = \text{diag}(\hat{\pi}_{ij})$, set

$$\mu_1 = \text{tr}(\hat{\mathbf{D}}_{ij}^{-1} \hat{\Sigma}_{ij}), \quad \mu_2 = 2 \text{tr}\{(\hat{\mathbf{D}}_{ij}^{-1} \hat{\Sigma}_{ij})^2\}, \quad \nu = \frac{2\mu_1^2}{\mu_2},$$

and $\bar{X}_{ij}^2 = (\nu/\mu_1) \text{GFfit}^{(ij)}$, compared with χ^2_ν .

2.6. Omnibus Statistics

Study 2 compares five statistics based on lower-order margins. They are X_{2c}^2 , $X_{[2]}^2$, M_2 , Y_2 , and X_{red}^2 . The full-table Pearson–Fisher statistic X_{PF}^2 is included only in the Type I error study. Let $\hat{\mathbf{u}}$ contain the 15 item proportions followed by the 105 pair proportions for two responses equal to one. Let $\boldsymbol{\mu}$ be the fitted value and $\mathbf{e} = \hat{\mathbf{u}} - \boldsymbol{\mu}$. Write $\hat{\mathbf{T}}$ for the unadjusted multinomial covariance of these margins, $\mathbf{B}_u = \partial \boldsymbol{\mu} / \partial \boldsymbol{\theta}^T$, and $\mathbf{I}_\theta$ for the fitted-model information. The covariance adjusted for parameter estimation is

$$\hat{\Sigma}_u = \hat{\mathbf{T}} - \mathbf{B}_u \mathbf{I}_\theta^{-1} \mathbf{B}_u^T. \quad (16)$$

The pair block of $\hat{\Sigma}_u$ gives $X_{[2]}^2$. Replacing the (1, 1) margins by the concordance margins in Equation (6) gives X_{2c}^2 . Both statistics have 105 degrees of freedom.

The remaining statistics are

$$M_2 = N \mathbf{e}^T \left[\hat{\mathbf{T}}^- - \hat{\mathbf{T}}^- \mathbf{B}_u (\mathbf{B}_u^T \hat{\mathbf{T}}^- \mathbf{B}_u)^- \mathbf{B}_u^T \hat{\mathbf{T}}^- \right] \mathbf{e}, \quad (17)$$

$$X_{\text{red}}^2 = N \mathbf{e}^T \hat{\mathbf{T}}^- \mathbf{e}, \quad (18)$$

$$Y_{2,\text{raw}} = N \sum_k \frac{e_k^2}{\mu_k(1 - \mu_k)}, \quad (19)$$

$$X_{\text{PF}}^2 = \sum_s \frac{(n_s - N \hat{\pi}_s)^2}{N \hat{\pi}_s}. \quad (20)$$

Both M_2 and X_{red}^2 use 104 degrees of freedom. For Y_2 , let $\mathbf{V} = \text{diag}\{\mu_k(1 - \mu_k)\}$. The positive eigenvalues of $\mathbf{V}^{-1/2} \hat{\Sigma}_u \mathbf{V}^{-1/2}$ give the scale $c = \sum_k \lambda_k^2 / \sum_k \lambda_k$ and degrees of freedom $\nu_Y = (\sum_k \lambda_k)^2 / \sum_k \lambda_k^2$. The adjusted value $Y_2 = Y_{2,\text{raw}} / c$ is compared with $\chi^2_{\nu_Y}$. These definitions follow the lower-order marginal formulations in Maydeu-Olivares and Joe [12,13], Cai et al. [14], and Tollenaar and Mooijaart [15]. The full-table statistic has 32,751 degrees of freedom in the present model.

3. Simulation Design

Two simulation studies were conducted with $q = 15$ binary items, sample sizes of $N = 300$ and $N = 1000$, and 1000 replications per condition. The full table contains $2^{15} = 32,768$ response patterns, giving only 0.009 and 0.031 observations per cell on average. Lower-order margins are much less sparse than this full table.

The 15-item design and the two sample sizes follow earlier sparse-table and directional studies that used 10 or 15 items and samples between 300 and 1000 [30,42]. This choice allows direct comparison with those results. For larger q , the full response table becomes still sparser, whereas the number of bivariate margins grows as $q(q - 1)/2$. Lower-order diagnostics remain feasible, although pair ordering and multiplicity become more important as the number of local components increases.

Study 1 fixes the slope pattern and compares zero and wide symmetric intercepts. Its purpose is to examine Type I error, power, and pairwise localization for six local diagnostics.

Study 2 fixes all intercepts at zero and changes the symmetric slope pattern. It compares five lower-order omnibus statistics and uses the full-table Pearson–Fisher statistic only to assess Type I error under sparsity.

3.1. Common Settings and Model Fitting

Normal integrals were evaluated by 21-point Gauss–Hermite quadrature. The detailed pairwise tables use the 5% nominal level. The additional analysis applies the 1%, 5%, and 10% rejection levels to the same fitted replications. The generated samples, fitted models, and test statistics are identical across the three levels. Only the critical values change.

For a rejection estimate $\hat{p}$ based on $R = 1000$ replications, the Monte Carlo standard error is $\{\hat{p}(1 - \hat{p})/R\}^{1/2}$. It is 0.0069 when $\hat{p} = 0.05$ and cannot exceed 0.0158. Differences of about one percentage point near 0.05 are therefore interpreted cautiously. Selected pairwise estimates are accompanied by 95% confidence intervals for binomial proportions. Monte Carlo standard errors and intervals for the added analyses are reported in the Supplementary File.

In Study 1, item parameters were estimated by separate conditional logistic regressions using the generated first-factor values. Under the power model, the second factor was omitted from these regressions. Equation (4) was used for the SS projection, while the conditional information supplied the parameter-covariance adjustment for Cholesky and the other pairwise diagnostics. Treating the first-factor values as known makes Study 1 an idealized, oracle-like comparison of the local statistics. It isolates differences among the diagnostics but does not include the additional uncertainty that arises when the latent trait is estimated in an applied IRT analysis. In Study 2, a one-factor 1PL model was fitted by marginal maximum likelihood. The fitted model contains 15 item intercepts and one common slope [43].

3.2. Study 1: Fixed Symmetric Slopes and Varying Intercepts

The first-factor slope vector is fixed at

$$\beta_1 = (0.1, 0.1, 0.1, 1, 1, 1, 1, 1, 1, 1, 1, 0.1, 0.1, 0.1)^T. \quad (21)$$

Here, β_{1j} is the slope of item j on the first factor. The vector is symmetric because items in matching positions have the same value. Its mean is 0.64. Two intercept settings are considered:

$$\beta_{0,Z} = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^T, \quad (22)$$

$$\beta_{0,S} = (-3.5, -3.0, -2.5, -2.0, -1.5, -1.0, -0.5, 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5)^T. \quad (23)$$

The first setting places every item at the center of the response scale. The second spreads the intercepts evenly from -3.5 to 3.5 . It is centered at zero and has equal values with opposite signs in matching positions. Comparing these two settings shows whether a symmetric arrangement with a wide range behaves like the all-zero arrangement.

For the Type I error part, data were generated from Equation (1) with Equation (21) and fitted by the same one-factor model. For the power part, data were generated from Equation (2). The second slope vector was

$$\beta_2 = (1, 1, 1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 1, 1, 1)^T. \quad (24)$$

Here, β_{2j} is the slope of item j on the second factor, and its mean is 0.46. Items 1, 2, 3, 13, 14, and 15 have a slope of 0.1 on the first factor and a slope of 1 on the second factor. The other nine items have the reverse assignment. For every item, the two slopes add to 1.1, and their

squared sum is 1.01. All items have the same overall loading size, but the larger loading is assigned to a different factor. The six selected items therefore depend mainly on the factor that is left out of the fitted model. The target pairs are (1, 2), (1, 3), (1, 13), (1, 14), (1, 15), (2, 3), (2, 13), (2, 14), (2, 15), (3, 13), (3, 14), (3, 15), (13, 14), (13, 15), and (14, 15). They are expected to contain the clearest local signal. The remaining pairs are called non-target pairs. Their rejection rates show how closely each statistic locates the designed signal. A rate near 0.05 is used only as a practical localization benchmark for these pairs, rather than as a formal Type I error rate, because the fitted model is misspecified in the power conditions.

The conditions were the two intercept vectors crossed with $N = 300$ and $N = 1000$. The same six local statistics were calculated for all 105 pairs.

3.3. Study 1 Evaluation: Multiple Testing and Localization

Each local procedure in Study 1 produces one p -value for each of the $\binom{15}{2} = 105$ item pairs. The main pairwise tables report componentwise rejection rates at the stated nominal level. To examine the selection of several pairs at once, we also applied Holm's procedure for controlling the familywise error rate and the Benjamini–Hochberg procedure for controlling the false discovery rate [44,45]. The adjusted results were compared with the unadjusted results.

The power design has 15 target pairs and 90 non-target pairs. For each replication, target sensitivity is the proportion of target pairs selected, and the non-target selection rate is the proportion of non-target pairs selected. Precision is the proportion of selected pairs that belong to the target set and is averaged over replications with at least one selection. The false discovery proportion is the proportion of selected pairs outside the target set and is set to zero when no pair is selected. The remaining measures are averaged over all 1000 replications.

Two additional analyses were carried out for Study 1. Pair-order sensitivity was examined at $N = 1000$ under both intercept settings. The original lexicographic order was compared with its reverse and with one random permutation generated before the simulation and then kept unchanged for every replication and condition. The same generated sample and fitted model were used for all three orders within a replication. SS, Cholesky, and $GFit_{\perp}^{(ij)}$ were then recomputed after permuting the rows of the relevant extraction matrix. Each method was compared only with itself across orders.

The second analysis also used $N = 1000$, with the original pair order and all six local procedures at the 5% level. The 105 p -values in each replication were considered without adjustment and after Holm and Benjamini–Hochberg adjustment. Under the null model, we recorded whether at least one pair was selected. Under misspecification, the measures defined above were used to compare localization.

3.4. Study 2: Fixed Zero Intercepts and Symmetric Slope Patterns

Study 2 fixes all intercepts at zero. It uses the following weak and strong symmetric slope vectors.

$$\mathbf{W} = (0.35, 0.25, 0.80, 0.50, 0.60, 0.45, 0.55, 0.50, 0.55, 0.45, 0.60, 0.50, 0.80, 0.25, 0.35)^T, \quad (25)$$

$$\mathbf{S} = (2.00, 1.35, 1.65, 1.00, 1.50, 1.40, 1.60, 1.50, 1.60, 1.40, 1.50, 1.00, 1.65, 1.35, 2.00)^T. \quad (26)$$

The mean slope is 0.5 for $\mathbf{W}$ and 1.5 for $\mathbf{S}$. Both vectors have equal values in matching positions. The three two-factor combinations are (W, W) , (W, S) , and (S, S) . Because the two factors have the same standard normal distribution, exchanging their labels does not create a new condition. The combination (S, W) is therefore not repeated.

The Type I error design needs a correctly specified null model. For that part, every item was assigned the common slopes (0.5, 0.5), (0.5, 1.5), or (1.5, 1.5) according to the condition label. Since

$$a_1X_1 + a_2X_2 \stackrel{d}{=} \sqrt{a_1^2 + a_2^2}X, \quad X \sim N(0, 1),$$

each data-generating model is exactly a one-factor 1PL model in its observable response distribution. The fitted common slope is therefore correctly specified.

For power, the full vectors in Equations (25) and (26) are used as the two columns of item slopes. Data are generated from a two-factor 2PL model and fitted by a one-factor 1PL model with one common slope. In (W, W) and (S, S), both factors use the same item pattern. These conditions mainly test whether the omnibus statistics detect the false common-slope restriction. In (W, S), the two factors use different patterns. This condition includes both unequal item slopes and a two-dimensional structure that cannot be reduced to one common loading pattern. The three conditions separate slope size from the form of model misspecification.

4. Results

4.1. Study 1: Type I Error

Under the correctly specified one-factor model, the mean rejection rates of $X_{2c,SS}^2$, $X_{2c,Chol}^2$, LM, $GFfit_{\perp}^{(ij)}$, and $\bar{X}_{ij}^2$ remain close to the nominal level across both intercept settings and sample sizes. Tables 1–4 show no systematic Type I error inflation under the wide symmetric intercept pattern. The direct non-orthogonal $GFfit^{(ij)}$ statistic is mildly conservative, with mean rejection rates from 0.038 to 0.039. Given a Monte Carlo standard error of about 0.0069 near a rejection probability of 0.05, isolated differences of about one percentage point are not emphasized.

Figure 1 summarizes the pair-level values in Table 4. The first five statistics fluctuate around 0.05, while the direct $GFfit^{(ij)}$ values tend to lie slightly below the reference line.

Table 1. Empirical Type I error rates of the six local statistics for the zero-intercept model with $q = 15$ and $N = 300$ at the 5% nominal level.

Pair	$X_{2c,SS}^2$	$X_{2c,Chol}^2$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}_{ij}^2$	$GFfit^{(ij)}$	Pair	$X_{2c,SS}^2$	$X_{2c,Chol}^2$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}_{ij}^2$	$GFfit^{(ij)}$
(1, 2)	0.042	0.041	0.038	0.042	0.041	0.040	(5, 9)	0.043	0.043	0.050	0.044	0.049	0.037
(1, 3)	0.047	0.047	0.041	0.047	0.047	0.047	(5, 10)	0.041	0.044	0.048	0.041	0.038	0.026
(1, 4)	0.054	0.062	0.054	0.055	0.059	0.042	(5, 11)	0.059	0.051	0.050	0.061	0.058	0.039
(1, 5)	0.050	0.051	0.049	0.050	0.051	0.039	(5, 12)	0.049	0.047	0.058	0.048	0.058	0.046
(1, 6)	0.045	0.049	0.049	0.045	0.047	0.036	(5, 13)	0.053	0.047	0.049	0.051	0.048	0.032
(1, 7)	0.056	0.046	0.053	0.056	0.045	0.034	(5, 14)	0.059	0.042	0.048	0.051	0.045	0.031
(1, 8)	0.040	0.064	0.063	0.043	0.062	0.042	(5, 15)	0.041	0.036	0.045	0.046	0.037	0.028
(1, 9)	0.045	0.050	0.051	0.046	0.049	0.035	(6, 7)	0.046	0.041	0.040	0.046	0.045	0.031
(1, 10)	0.051	0.048	0.056	0.052	0.048	0.038	(6, 8)	0.055	0.055	0.060	0.055	0.056	0.036
(1, 11)	0.044	0.038	0.048	0.045	0.039	0.029	(6, 9)	0.046	0.037	0.049	0.044	0.034	0.022
(1, 12)	0.057	0.050	0.053	0.059	0.050	0.043	(6, 10)	0.048	0.056	0.057	0.046	0.063	0.041
(1, 13)	0.036	0.045	0.044	0.031	0.045	0.044	(6, 11)	0.060	0.057	0.049	0.061	0.056	0.042
(1, 14)	0.062	0.054	0.051	0.051	0.054	0.054	(6, 12)	0.046	0.060	0.059	0.045	0.057	0.039
(1, 15)	0.049	0.049	0.049	0.042	0.049	0.049	(6, 13)	0.045	0.043	0.038	0.050	0.039	0.025
(2, 3)	0.048	0.055	0.046	0.048	0.055	0.052	(6, 14)	0.043	0.048	0.040	0.042	0.045	0.031
(2, 4)	0.052	0.055	0.050	0.051	0.059	0.042	(6, 15)	0.049	0.050	0.046	0.053	0.052	0.034
(2, 5)	0.054	0.055	0.051	0.052	0.054	0.038	(7, 8)	0.040	0.048	0.049	0.041	0.052	0.042
(2, 6)	0.059	0.063	0.050	0.056	0.063	0.048	(7, 9)	0.042	0.044	0.050	0.042	0.042	0.034
(2, 7)	0.054	0.056	0.048	0.057	0.054	0.043	(7, 10)	0.052	0.062	0.049	0.052	0.061	0.046
(2, 8)	0.050	0.048	0.047	0.048	0.048	0.034	(7, 11)	0.046	0.051	0.038	0.048	0.053	0.038
(2, 9)	0.052	0.039	0.045	0.051	0.040	0.028	(7, 12)	0.046	0.047	0.050	0.047	0.047	0.034
(2, 10)	0.049	0.058	0.047	0.048	0.056	0.034	(7, 13)	0.056	0.044	0.043	0.056	0.045	0.035
(2, 11)	0.046	0.034	0.049	0.048	0.037	0.026	(7, 14)	0.050	0.054	0.045	0.046	0.057	0.041
(2, 12)	0.047	0.057	0.056	0.045	0.056	0.039	(7, 15)	0.038	0.045	0.044	0.041	0.047	0.032

Table 1. Cont.

Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}^2_{ij}$	$GFfit^{(ij)}$	Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}^2_{ij}$	$GFfit^{(ij)}$
(2, 13)	0.056	0.059	0.055	0.061	0.060	0.059	(8, 9)	0.063	0.073	0.062	0.063	0.071	0.051
(2, 14)	0.058	0.046	0.048	0.055	0.046	0.043	(8, 10)	0.046	0.053	0.058	0.047	0.050	0.035
(2, 15)	0.053	0.057	0.045	0.051	0.057	0.057	(8, 11)	0.044	0.055	0.059	0.043	0.056	0.040
(3, 4)	0.066	0.063	0.056	0.066	0.063	0.048	(8, 12)	0.054	0.047	0.055	0.051	0.052	0.034
(3, 5)	0.051	0.049	0.047	0.050	0.052	0.040	(8, 13)	0.056	0.056	0.053	0.053	0.051	0.035
(3, 6)	0.049	0.051	0.052	0.048	0.052	0.036	(8, 14)	0.048	0.052	0.048	0.045	0.050	0.030
(3, 7)	0.041	0.045	0.052	0.042	0.047	0.036	(8, 15)	0.037	0.049	0.056	0.040	0.052	0.036
(3, 8)	0.044	0.049	0.040	0.044	0.050	0.038	(9, 10)	0.045	0.053	0.048	0.043	0.055	0.038
(3, 9)	0.061	0.050	0.051	0.061	0.050	0.041	(9, 11)	0.053	0.047	0.052	0.054	0.063	0.041
(3, 10)	0.052	0.066	0.060	0.052	0.062	0.051	(9, 12)	0.046	0.043	0.047	0.048	0.044	0.023
(3, 11)	0.057	0.046	0.043	0.061	0.044	0.033	(9, 13)	0.046	0.054	0.062	0.045	0.052	0.034
(3, 12)	0.041	0.044	0.048	0.045	0.041	0.030	(9, 14)	0.045	0.046	0.041	0.052	0.046	0.034
(3, 13)	0.051	0.055	0.055	0.053	0.055	0.054	(9, 15)	0.045	0.060	0.049	0.041	0.065	0.048
(3, 14)	0.052	0.050	0.053	0.047	0.050	0.048	(10, 11)	0.051	0.057	0.058	0.048	0.062	0.048
(3, 15)	0.048	0.053	0.058	0.049	0.052	0.049	(10, 12)	0.053	0.048	0.045	0.053	0.051	0.036
(4, 5)	0.061	0.049	0.052	0.058	0.061	0.040	(10, 13)	0.045	0.040	0.052	0.046	0.039	0.033
(4, 6)	0.060	0.061	0.051	0.058	0.063	0.045	(10, 14)	0.054	0.045	0.040	0.058	0.051	0.035
(4, 7)	0.048	0.050	0.051	0.048	0.056	0.038	(10, 15)	0.055	0.052	0.051	0.059	0.054	0.038
(4, 8)	0.048	0.054	0.056	0.049	0.047	0.039	(11, 12)	0.054	0.053	0.057	0.052	0.063	0.042
(4, 9)	0.051	0.056	0.049	0.052	0.054	0.040	(11, 13)	0.044	0.048	0.050	0.043	0.051	0.033
(4, 10)	0.053	0.051	0.047	0.050	0.054	0.040	(11, 14)	0.046	0.059	0.049	0.048	0.061	0.043
(4, 11)	0.050	0.052	0.041	0.053	0.055	0.046	(11, 15)	0.064	0.048	0.048	0.061	0.049	0.037
(4, 12)	0.045	0.049	0.060	0.042	0.056	0.043	(12, 13)	0.058	0.052	0.049	0.060	0.052	0.038
(4, 13)	0.058	0.051	0.051	0.056	0.054	0.040	(12, 14)	0.049	0.045	0.047	0.043	0.046	0.032
(4, 14)	0.050	0.046	0.052	0.056	0.049	0.034	(12, 15)	0.048	0.049	0.036	0.045	0.044	0.035
(4, 15)	0.052	0.054	0.040	0.054	0.055	0.040	(13, 14)	0.059	0.053	0.046	0.056	0.052	0.051
(5, 6)	0.058	0.049	0.049	0.057	0.056	0.038	(13, 15)	0.053	0.046	0.049	0.051	0.046	0.046
(5, 7)	0.038	0.053	0.056	0.038	0.053	0.042	(14, 15)	0.035	0.053	0.054	0.041	0.052	0.052
(5, 8)	0.052	0.040	0.039	0.052	0.046	0.033	Mean	0.050	0.050	0.050	0.050	0.052	0.039

Note: In Tables 1–4, bold values mark rejection rates significantly different from 0.05 based on a two-sided exact binomial test (5% level).

Table 2. Empirical Type I error rates of the six local statistics for the zero-intercept model with $q = 15$ and $N = 1000$ at the 5% nominal level.

Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}^2_{ij}$	$GFfit^{(ij)}$	Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}^2_{ij}$	$GFfit^{(ij)}$
(1, 2)	0.042	0.043	0.045	0.042	0.043	0.043	(5, 9)	0.051	0.051	0.050	0.051	0.051	0.037
(1, 3)	0.041	0.041	0.037	0.041	0.041	0.041	(5, 10)	0.048	0.051	0.052	0.047	0.048	0.032
(1, 4)	0.046	0.046	0.059	0.043	0.046	0.036	(5, 11)	0.050	0.056	0.047	0.052	0.062	0.037
(1, 5)	0.053	0.068	0.057	0.052	0.066	0.051	(5, 12)	0.040	0.056	0.060	0.040	0.054	0.044
(1, 6)	0.054	0.059	0.057	0.054	0.057	0.046	(5, 13)	0.042	0.033	0.048	0.046	0.031	0.024
(1, 7)	0.054	0.051	0.040	0.053	0.048	0.038	(5, 14)	0.066	0.056	0.059	0.071	0.055	0.045
(1, 8)	0.056	0.054	0.059	0.060	0.051	0.035	(5, 15)	0.045	0.056	0.047	0.052	0.056	0.042
(1, 9)	0.045	0.051	0.046	0.049	0.049	0.034	(6, 7)	0.052	0.057	0.050	0.052	0.049	0.039
(1, 10)	0.052	0.046	0.049	0.051	0.047	0.037	(6, 8)	0.043	0.052	0.051	0.042	0.049	0.035
(1, 11)	0.047	0.054	0.044	0.046	0.052	0.039	(6, 9)	0.047	0.045	0.050	0.046	0.044	0.030
(1, 12)	0.049	0.050	0.047	0.056	0.053	0.037	(6, 10)	0.054	0.052	0.040	0.054	0.047	0.029
(1, 13)	0.053	0.044	0.057	0.057	0.044	0.044	(6, 11)	0.037	0.045	0.043	0.038	0.045	0.025
(1, 14)	0.053	0.050	0.050	0.049	0.050	0.050	(6, 12)	0.049	0.045	0.049	0.049	0.047	0.034
(1, 15)	0.047	0.048	0.039	0.041	0.050	0.047	(6, 13)	0.052	0.062	0.050	0.052	0.059	0.039
(2, 3)	0.038	0.038	0.045	0.038	0.037	0.036	(6, 14)	0.047	0.055	0.059	0.041	0.056	0.038
(2, 4)	0.052	0.052	0.053	0.052	0.055	0.042	(6, 15)	0.044	0.049	0.040	0.045	0.044	0.034
(2, 5)	0.048	0.038	0.036	0.049	0.039	0.034	(7, 8)	0.051	0.058	0.054	0.051	0.062	0.045
(2, 6)	0.052	0.059	0.047	0.053	0.059	0.043	(7, 9)	0.064	0.053	0.055	0.063	0.053	0.038
(2, 7)	0.053	0.056	0.052	0.053	0.055	0.044	(7, 10)	0.038	0.050	0.056	0.038	0.046	0.030
(2, 8)	0.059	0.060	0.048	0.061	0.056	0.045	(7, 11)	0.047	0.048	0.045	0.048	0.041	0.025
(2, 9)	0.063	0.064	0.058	0.061	0.057	0.043	(7, 12)	0.050	0.045	0.047	0.055	0.039	0.025
(2, 10)	0.054	0.052	0.050	0.055	0.058	0.036	(7, 13)	0.056	0.056	0.053	0.057	0.056	0.045
(2, 11)	0.051	0.047	0.044	0.055	0.048	0.036	(7, 14)	0.050	0.052	0.048	0.048	0.055	0.039
(2, 12)	0.053	0.052	0.053	0.052	0.059	0.047	(7, 15)	0.042	0.049	0.040	0.042	0.046	0.032
(2, 13)	0.042	0.051	0.055	0.045	0.051	0.051	(8, 9)	0.037	0.045	0.046	0.037	0.042	0.031
(2, 14)	0.053	0.054	0.051	0.053	0.054	0.054	(8, 10)	0.050	0.043	0.047	0.050	0.045	0.031
(2, 15)	0.048	0.062	0.053	0.048	0.062	0.062	(8, 11)	0.036	0.049	0.057	0.037	0.048	0.038
(3, 4)	0.050	0.041	0.049	0.051	0.043	0.031	(8, 12)	0.043	0.046	0.042	0.043	0.039	0.028
(3, 5)	0.058	0.058	0.057	0.058	0.054	0.041	(8, 13)	0.049	0.054	0.051	0.048	0.054	0.040

Table 2. Cont.

Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}_{ij}^2$	$GFfit^{(ij)}$	Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}_{ij}^2$	$GFfit^{(ij)}$
(3, 6)	0.038	0.049	0.051	0.039	0.045	0.029	(8, 14)	0.044	0.056	0.060	0.043	0.058	0.044
(3, 7)	0.046	0.042	0.050	0.042	0.043	0.036	(8, 15)	0.042	0.041	0.042	0.043	0.041	0.030
(3, 8)	0.054	0.056	0.058	0.054	0.055	0.046	(9, 10)	0.060	0.051	0.046	0.060	0.041	0.036
(3, 9)	0.048	0.042	0.043	0.048	0.044	0.031	(9, 11)	0.051	0.041	0.041	0.048	0.035	0.028
(3, 10)	0.048	0.053	0.051	0.051	0.050	0.041	(9, 12)	0.059	0.054	0.068	0.064	0.055	0.047
(3, 11)	0.045	0.051	0.048	0.043	0.048	0.035	(9, 13)	0.048	0.056	0.054	0.047	0.058	0.039
(3, 12)	0.042	0.035	0.050	0.046	0.038	0.022	(9, 14)	0.050	0.062	0.050	0.046	0.065	0.046
(3, 13)	0.051	0.054	0.048	0.046	0.054	0.054	(9, 15)	0.044	0.046	0.051	0.041	0.047	0.038
(3, 14)	0.051	0.052	0.057	0.051	0.052	0.051	(10, 11)	0.038	0.049	0.058	0.039	0.050	0.033
(3, 15)	0.054	0.042	0.044	0.056	0.042	0.042	(10, 12)	0.053	0.046	0.047	0.052	0.040	0.030
(4, 5)	0.059	0.048	0.047	0.058	0.057	0.035	(10, 13)	0.042	0.036	0.040	0.043	0.035	0.027
(4, 6)	0.058	0.056	0.055	0.055	0.056	0.040	(10, 14)	0.049	0.043	0.045	0.052	0.044	0.030
(4, 7)	0.055	0.060	0.047	0.054	0.059	0.049	(10, 15)	0.062	0.052	0.051	0.065	0.049	0.040
(4, 8)	0.042	0.063	0.056	0.039	0.059	0.041	(11, 12)	0.054	0.042	0.037	0.058	0.048	0.033
(4, 9)	0.060	0.046	0.052	0.059	0.052	0.036	(11, 13)	0.048	0.037	0.038	0.052	0.036	0.023
(4, 10)	0.045	0.057	0.058	0.044	0.053	0.043	(11, 14)	0.041	0.048	0.059	0.044	0.053	0.038
(4, 11)	0.056	0.047	0.066	0.054	0.047	0.032	(11, 15)	0.061	0.055	0.049	0.056	0.051	0.036
(4, 12)	0.062	0.055	0.047	0.068	0.052	0.035	(12, 13)	0.050	0.056	0.052	0.045	0.052	0.035
(4, 13)	0.049	0.041	0.041	0.057	0.047	0.030	(12, 14)	0.046	0.047	0.057	0.047	0.041	0.033
(4, 14)	0.043	0.063	0.050	0.052	0.060	0.051	(12, 15)	0.056	0.049	0.045	0.062	0.046	0.034
(4, 15)	0.048	0.049	0.046	0.058	0.049	0.035	(13, 14)	0.050	0.054	0.064	0.052	0.054	0.054
(5, 6)	0.050	0.063	0.047	0.048	0.055	0.046	(13, 15)	0.050	0.058	0.047	0.052	0.058	0.058
(5, 7)	0.048	0.051	0.055	0.048	0.048	0.042	(14, 15)	0.052	0.052	0.057	0.061	0.053	0.052
(5, 8)	0.061	0.044	0.047	0.062	0.046	0.034	Mean	0.050	0.051	0.050	0.050	0.050	0.038

Table 3. Empirical Type I error rates of the six local statistics for the wide symmetric-intercept model with $q = 15$ and $N = 300$ at the 5% nominal level.

Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}_{ij}^2$	$GFfit^{(ij)}$	Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}_{ij}^2$	$GFfit^{(ij)}$
(1, 2)	0.056	0.051	0.083	0.055	0.055	0.055	(5, 9)	0.046	0.048	0.046	0.047	0.046	0.036
(1, 3)	0.038	0.037	0.060	0.038	0.040	0.039	(5, 10)	0.044	0.041	0.044	0.044	0.045	0.034
(1, 4)	0.050	0.054	0.059	0.048	0.047	0.039	(5, 11)	0.057	0.055	0.053	0.058	0.048	0.042
(1, 5)	0.048	0.065	0.044	0.049	0.049	0.035	(5, 12)	0.051	0.044	0.043	0.055	0.051	0.046
(1, 6)	0.052	0.054	0.045	0.052	0.050	0.033	(5, 13)	0.055	0.043	0.054	0.055	0.049	0.041
(1, 7)	0.047	0.046	0.041	0.045	0.047	0.031	(5, 14)	0.043	0.047	0.036	0.037	0.041	0.035
(1, 8)	0.048	0.051	0.056	0.047	0.047	0.035	(5, 15)	0.056	0.048	0.055	0.054	0.044	0.037
(1, 9)	0.062	0.048	0.045	0.068	0.053	0.038	(6, 7)	0.041	0.044	0.044	0.040	0.042	0.029
(1, 10)	0.043	0.058	0.053	0.047	0.050	0.039	(6, 8)	0.056	0.049	0.039	0.057	0.044	0.029
(1, 11)	0.060	0.061	0.051	0.060	0.045	0.038	(6, 9)	0.045	0.047	0.044	0.045	0.041	0.035
(1, 12)	0.042	0.041	0.068	0.043	0.045	0.039	(6, 10)	0.043	0.059	0.059	0.044	0.060	0.047
(1, 13)	0.051	0.051	0.073	0.045	0.050	0.049	(6, 11)	0.045	0.037	0.039	0.041	0.041	0.029
(1, 14)	0.040	0.039	0.073	0.043	0.040	0.040	(6, 12)	0.054	0.057	0.053	0.061	0.061	0.051
(1, 15)	0.068	0.071	0.093	0.064	0.073	0.072	(6, 13)	0.044	0.047	0.049	0.045	0.045	0.036
(2, 3)	0.038	0.037	0.066	0.038	0.036	0.036	(6, 14)	0.053	0.054	0.047	0.046	0.051	0.045
(2, 4)	0.047	0.040	0.054	0.048	0.035	0.032	(6, 15)	0.045	0.051	0.050	0.047	0.050	0.035
(2, 5)	0.042	0.045	0.056	0.042	0.044	0.034	(7, 8)	0.057	0.050	0.038	0.057	0.043	0.029
(2, 6)	0.048	0.052	0.055	0.047	0.055	0.047	(7, 9)	0.046	0.049	0.050	0.045	0.053	0.040
(2, 7)	0.058	0.049	0.057	0.059	0.055	0.042	(7, 10)	0.061	0.054	0.046	0.063	0.057	0.042
(2, 8)	0.055	0.052	0.059	0.053	0.046	0.041	(7, 11)	0.057	0.051	0.048	0.056	0.052	0.041
(2, 9)	0.043	0.053	0.049	0.044	0.043	0.031	(7, 12)	0.053	0.062	0.046	0.053	0.061	0.050
(2, 10)	0.046	0.057	0.049	0.044	0.057	0.046	(7, 13)	0.055	0.051	0.056	0.055	0.052	0.038
(2, 11)	0.042	0.054	0.057	0.040	0.054	0.042	(7, 14)	0.056	0.052	0.044	0.049	0.052	0.042
(2, 12)	0.052	0.038	0.052	0.047	0.032	0.027	(7, 15)	0.046	0.050	0.042	0.053	0.045	0.034
(2, 13)	0.040	0.034	0.050	0.042	0.037	0.037	(8, 9)	0.063	0.047	0.052	0.062	0.056	0.042
(2, 14)	0.059	0.063	0.089	0.058	0.065	0.064	(8, 10)	0.063	0.055	0.065	0.064	0.058	0.040
(2, 15)	0.053	0.056	0.091	0.052	0.060	0.060	(8, 11)	0.049	0.058	0.053	0.047	0.063	0.045
(3, 4)	0.053	0.053	0.053	0.053	0.052	0.043	(8, 12)	0.066	0.056	0.054	0.071	0.055	0.043
(3, 5)	0.056	0.051	0.059	0.056	0.055	0.040	(8, 13)	0.055	0.062	0.055	0.052	0.060	0.048
(3, 6)	0.040	0.039	0.040	0.040	0.035	0.023	(8, 14)	0.059	0.053	0.039	0.059	0.043	0.027
(3, 7)	0.041	0.044	0.036	0.040	0.045	0.034	(8, 15)	0.055	0.056	0.046	0.054	0.051	0.035
(3, 8)	0.039	0.046	0.052	0.038	0.048	0.038	(9, 10)	0.054	0.047	0.048	0.056	0.054	0.036
(3, 9)	0.053	0.053	0.060	0.052	0.053	0.042	(9, 11)	0.051	0.049	0.050	0.055	0.051	0.033
(3, 10)	0.064	0.057	0.053	0.059	0.055	0.040	(9, 12)	0.047	0.058	0.061	0.052	0.055	0.040
(3, 11)	0.046	0.046	0.046	0.046	0.045	0.038	(9, 13)	0.046	0.054	0.052	0.044	0.051	0.035
(3, 12)	0.054	0.049	0.054	0.048	0.049	0.041	(9, 14)	0.046	0.053	0.048	0.046	0.048	0.039

Table 3. Cont.

Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}_{ij}^2$	$GFfit^{(ij)}$	Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}_{ij}^2$	$GFfit^{(ij)}$
(3, 13)	0.040	0.038	0.045	0.038	0.039	0.039	(9, 15)	0.043	0.045	0.042	0.050	0.047	0.034
(3, 14)	0.048	0.040	0.054	0.042	0.040	0.040	(10, 11)	0.048	0.042	0.056	0.049	0.051	0.038
(3, 15)	0.039	0.038	0.062	0.040	0.037	0.037	(10, 12)	0.054	0.058	0.065	0.050	0.050	0.035
(4, 5)	0.054	0.052	0.045	0.054	0.058	0.044	(10, 13)	0.049	0.055	0.042	0.043	0.050	0.039
(4, 6)	0.046	0.049	0.040	0.045	0.046	0.032	(10, 14)	0.039	0.051	0.048	0.041	0.048	0.034
(4, 7)	0.059	0.046	0.040	0.059	0.041	0.029	(10, 15)	0.047	0.037	0.037	0.053	0.041	0.029
(4, 8)	0.052	0.053	0.046	0.055	0.057	0.044	(11, 12)	0.047	0.044	0.041	0.053	0.045	0.036
(4, 9)	0.053	0.057	0.057	0.055	0.057	0.048	(11, 13)	0.047	0.053	0.037	0.049	0.052	0.040
(4, 10)	0.055	0.040	0.045	0.060	0.046	0.037	(11, 14)	0.040	0.043	0.048	0.040	0.045	0.034
(4, 11)	0.068	0.060	0.052	0.065	0.054	0.049	(11, 15)	0.039	0.033	0.051	0.044	0.033	0.023
(4, 12)	0.047	0.053	0.045	0.040	0.050	0.046	(12, 13)	0.040	0.041	0.029	0.037	0.041	0.035
(4, 13)	0.051	0.043	0.053	0.054	0.044	0.039	(12, 14)	0.052	0.041	0.042	0.055	0.039	0.035
(4, 14)	0.052	0.048	0.056	0.048	0.041	0.034	(12, 15)	0.054	0.039	0.049	0.054	0.037	0.032
(4, 15)	0.056	0.042	0.056	0.040	0.030	0.026	(13, 14)	0.040	0.044	0.061	0.041	0.045	0.043
(5, 6)	0.045	0.042	0.054	0.046	0.049	0.030	(13, 15)	0.048	0.046	0.077	0.054	0.045	0.045
(5, 7)	0.048	0.050	0.049	0.047	0.053	0.039	(14, 15)	0.039	0.052	0.077	0.040	0.050	0.049
(5, 8)	0.065	0.060	0.055	0.064	0.061	0.054	Mean	0.050	0.049	0.052	0.050	0.048	0.039

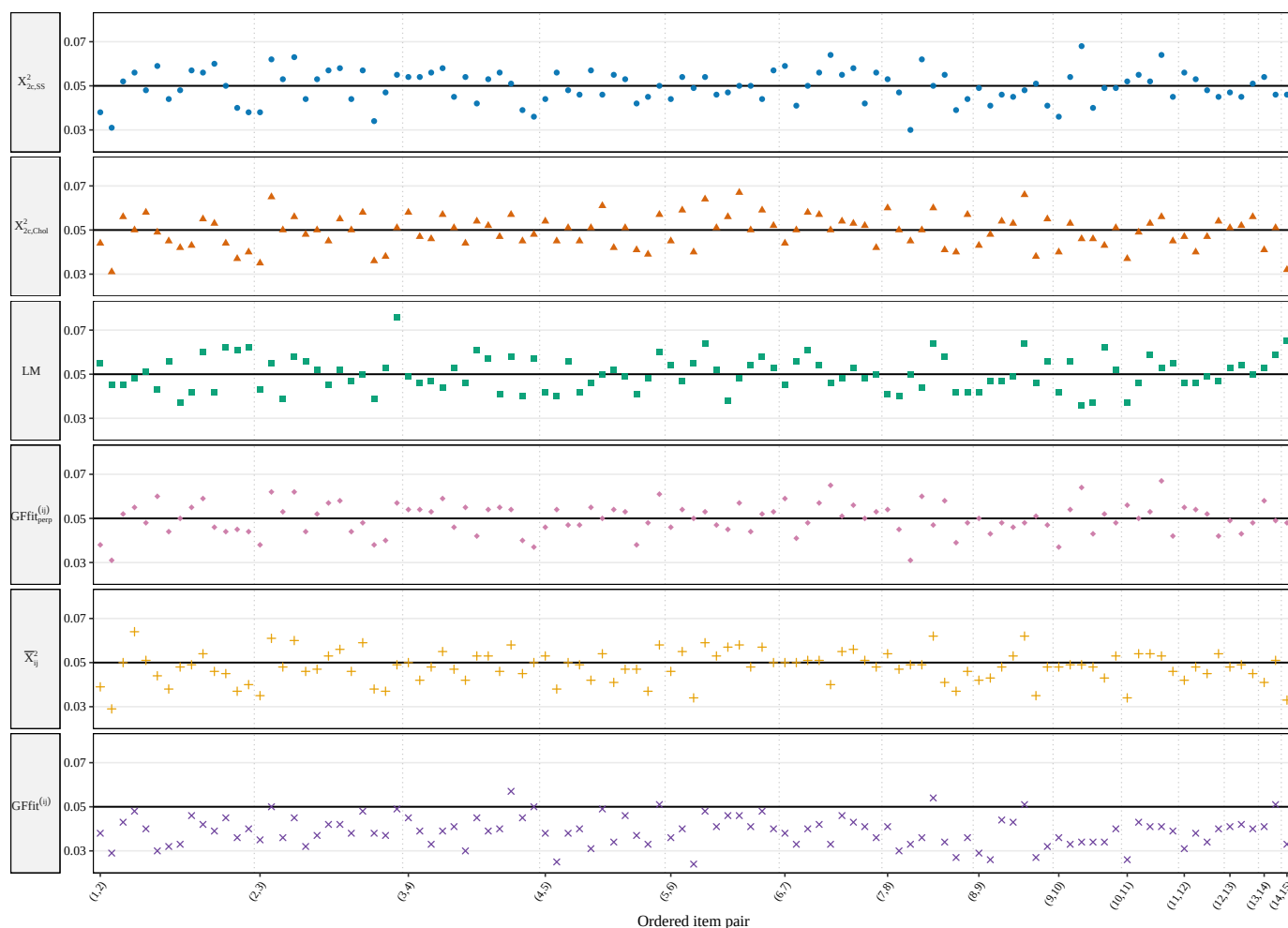


Figure 1. Empirical Type I error rates of the six local statistics under the wide symmetric-intercept model with $q = 15$ and $N = 1000$ at the 5% nominal level. Each panel shows one statistic. Pairs follow the table order, the horizontal line marks 0.05, and the vertical dotted lines separate groups with the same first item.

Table 4. Empirical Type I error rates of the six local statistics for the wide symmetric-intercept model with $q = 15$ and $N = 1000$ at the 5% nominal level.

Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}^2_{ij}$	$GFfit^{(ij)}$	Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}^2_{ij}$	$GFfit^{(ij)}$
(1, 2)	0.038	0.044	0.055	0.038	0.039	0.038	(5, 9)	0.054	0.064	0.064	0.053	0.059	0.048
(1, 3)	0.031	0.031	0.045	0.031	0.029	0.029	(5, 10)	0.046	0.051	0.052	0.047	0.053	0.041
(1, 4)	0.052	0.056	0.045	0.052	0.050	0.043	(5, 11)	0.047	0.056	0.038	0.045	0.057	0.046
(1, 5)	0.056	0.050	0.048	0.055	0.064	0.048	(5, 12)	0.050	0.067	0.048	0.057	0.058	0.046
(1, 6)	0.048	0.058	0.051	0.048	0.051	0.040	(5, 13)	0.050	0.050	0.054	0.044	0.048	0.041
(1, 7)	0.059	0.049	0.043	0.060	0.044	0.030	(5, 14)	0.044	0.059	0.058	0.052	0.057	0.048
(1, 8)	0.044	0.045	0.056	0.044	0.038	0.032	(5, 15)	0.057	0.052	0.053	0.053	0.050	0.040
(1, 9)	0.048	0.042	0.037	0.050	0.048	0.033	(6, 7)	0.059	0.044	0.045	0.059	0.050	0.038
(1, 10)	0.057	0.043	0.042	0.055	0.049	0.046	(6, 8)	0.041	0.050	0.056	0.041	0.050	0.033
(1, 11)	0.056	0.055	0.060	0.059	0.054	0.042	(6, 9)	0.050	0.058	0.061	0.048	0.051	0.040
(1, 12)	0.060	0.053	0.042	0.046	0.046	0.039	(6, 10)	0.056	0.057	0.054	0.057	0.051	0.042
(1, 13)	0.050	0.044	0.062	0.044	0.045	0.045	(6, 11)	0.064	0.050	0.046	0.065	0.040	0.033
(1, 14)	0.040	0.037	0.061	0.045	0.037	0.036	(6, 12)	0.055	0.054	0.048	0.051	0.055	0.046
(1, 15)	0.038	0.040	0.062	0.044	0.040	0.040	(6, 13)	0.058	0.053	0.053	0.056	0.056	0.043
(2, 3)	0.038	0.035	0.043	0.038	0.035	0.035	(6, 14)	0.042	0.052	0.048	0.050	0.051	0.041
(2, 4)	0.062	0.065	0.055	0.062	0.061	0.050	(6, 15)	0.056	0.042	0.050	0.053	0.048	0.036
(2, 5)	0.053	0.050	0.039	0.053	0.048	0.036	(7, 8)	0.053	0.060	0.041	0.054	0.054	0.041
(2, 6)	0.063	0.056	0.058	0.062	0.060	0.045	(7, 9)	0.047	0.050	0.040	0.045	0.047	0.030
(2, 7)	0.044	0.048	0.056	0.044	0.046	0.032	(7, 10)	0.030	0.045	0.050	0.031	0.049	0.033
(2, 8)	0.053	0.050	0.052	0.052	0.047	0.037	(7, 11)	0.062	0.050	0.044	0.060	0.049	0.036
(2, 9)	0.057	0.045	0.045	0.057	0.053	0.042	(7, 12)	0.050	0.060	0.064	0.047	0.062	0.054
(2, 10)	0.058	0.055	0.052	0.058	0.056	0.042	(7, 13)	0.055	0.041	0.058	0.058	0.041	0.034
(2, 11)	0.044	0.050	0.047	0.044	0.046	0.038	(7, 14)	0.039	0.040	0.042	0.039	0.037	0.027
(2, 12)	0.057	0.058	0.050	0.048	0.059	0.048	(7, 15)	0.044	0.057	0.042	0.048	0.046	0.036
(2, 13)	0.034	0.036	0.039	0.038	0.038	0.038	(8, 9)	0.049	0.043	0.042	0.050	0.042	0.029
(2, 14)	0.047	0.038	0.053	0.040	0.037	0.037	(8, 10)	0.041	0.048	0.047	0.043	0.043	0.026
(2, 15)	0.055	0.051	0.076	0.057	0.049	0.049	(8, 11)	0.046	0.054	0.047	0.048	0.048	0.044
(3, 4)	0.054	0.058	0.049	0.054	0.050	0.045	(8, 12)	0.045	0.053	0.049	0.046	0.053	0.043
(3, 5)	0.054	0.047	0.046	0.054	0.042	0.039	(8, 13)	0.048	0.066	0.064	0.048	0.062	0.051
(3, 6)	0.056	0.046	0.047	0.053	0.048	0.033	(8, 14)	0.051	0.038	0.046	0.051	0.035	0.027
(3, 7)	0.058	0.057	0.044	0.059	0.055	0.039	(8, 15)	0.041	0.055	0.056	0.047	0.048	0.032
(3, 8)	0.045	0.051	0.053	0.046	0.047	0.041	(9, 10)	0.036	0.040	0.042	0.037	0.048	0.036
(3, 9)	0.054	0.044	0.046	0.055	0.042	0.030	(9, 11)	0.054	0.053	0.056	0.054	0.049	0.033
(3, 10)	0.042	0.054	0.061	0.042	0.053	0.045	(9, 12)	0.068	0.046	0.036	0.064	0.049	0.034
(3, 11)	0.053	0.052	0.057	0.054	0.053	0.039	(9, 13)	0.040	0.046	0.037	0.043	0.048	0.034
(3, 12)	0.056	0.047	0.041	0.055	0.046	0.040	(9, 14)	0.049	0.043	0.062	0.052	0.043	0.034
(3, 13)	0.051	0.057	0.058	0.054	0.058	0.057	(9, 15)	0.049	0.051	0.052	0.048	0.053	0.040
(3, 14)	0.039	0.045	0.040	0.040	0.045	0.045	(10, 11)	0.052	0.037	0.037	0.056	0.034	0.026
(3, 15)	0.036	0.048	0.057	0.037	0.050	0.050	(10, 12)	0.055	0.049	0.046	0.050	0.054	0.043
(4, 5)	0.044	0.054	0.042	0.046	0.053	0.038	(10, 13)	0.052	0.053	0.059	0.053	0.054	0.041
(4, 6)	0.056	0.045	0.040	0.054	0.038	0.025	(10, 14)	0.064	0.056	0.053	0.067	0.053	0.041
(4, 7)	0.048	0.051	0.056	0.047	0.050	0.038	(10, 15)	0.045	0.045	0.055	0.042	0.046	0.039
(4, 8)	0.046	0.045	0.042	0.047	0.049	0.040	(11, 12)	0.056	0.047	0.046	0.055	0.042	0.031
(4, 9)	0.057	0.051	0.046	0.055	0.042	0.031	(11, 13)	0.053	0.040	0.046	0.054	0.048	0.038
(4, 10)	0.046	0.061	0.050	0.050	0.054	0.049	(11, 14)	0.048	0.047	0.049	0.052	0.045	0.034
(4, 11)	0.055	0.042	0.052	0.054	0.041	0.034	(11, 15)	0.045	0.054	0.047	0.042	0.054	0.040
(4, 12)	0.053	0.051	0.049	0.053	0.047	0.046	(12, 13)	0.047	0.051	0.053	0.049	0.048	0.041
(4, 13)	0.042	0.041	0.041	0.038	0.047	0.037	(12, 14)	0.045	0.052	0.054	0.043	0.049	0.042
(4, 14)	0.045	0.039	0.048	0.048	0.037	0.033	(12, 15)	0.051	0.056	0.050	0.048	0.045	0.040
(4, 15)	0.050	0.057	0.060	0.061	0.058	0.051	(13, 14)	0.054	0.041	0.053	0.058	0.041	0.041
(5, 6)	0.044	0.045	0.054	0.046	0.046	0.036	(13, 15)	0.046	0.051	0.059	0.049	0.051	0.051
(5, 7)	0.054	0.059	0.047	0.054	0.055	0.040	(14, 15)	0.046	0.032	0.065	0.048	0.033	0.033
(5, 8)	0.049	0.040	0.055	0.050	0.034	0.024	Mean	0.050	0.049	0.050	0.050	0.048	0.039

4.2. Study 1: Power and Localization

This section reports rejection rates when the fitted model does not include the second factor. Items 1, 2, 3, 13, 14, and 15 have the largest slopes on that factor. All pairs formed among these six items make up the target set. The intended pattern is a high rejection rate for target pairs and a rate near 0.05 for most non-target pairs. Departures from this pattern show how the method distributes the lack-of-fit signal.

Under zero intercepts, rejection rates are high for most target pairs. They increase further when the sample size rises from 300 to 1000, and most are then close to 1. Rejection rates are much lower for most non-target pairs, although their patterns differ across methods.

The SS results contain several concentrated non-target peaks. At $N = 300$, rejection rates for (1, 12), (2, 12), and (3, 12) are 0.614, 0.618, and 0.636. They rise to 0.907, 0.911, and 0.902 at $N = 1000$. Other high values occur at (11, 12) and at pairs joining item 12 to items 13, 14, and 15. Under the correctly specified model, the Type I error rates for the same pairs are close to 0.05. The high rejection rates therefore appear only when the fitted model is false.

This pattern follows from the sequential calculation. Pairs enter one at a time, and each component is based on what remains after the earlier pairs have been taken into account. Pairwise marginal directions are correlated. After the earlier directions are removed, the direction left for the current pair may overlap with directions assigned to later target pairs. A component can therefore be large even when its label is outside the target set. The peaks for $GFit_{\perp}^{(ij)}$ occur at similar positions because it uses the same order and the same sequential calculation.

SS and Cholesky show a clear trade-off under zero intercepts. Cholesky gives higher rejection rates for the target pairs. At $N = 1000$, it also gives moderate increases across a wider group of non-target pairs, with many rates between 0.07 and 0.10. Most SS rates for non-target pairs remain closer to 0.05, apart from the concentrated peaks created by the fixed order, but its target-pair power is lower. Some moderate non-target increases also appear for LM and $\bar{X}_{ij}^2$. The difference between SS and Cholesky is less marked at $N = 300$. The stronger contrast at $N = 1000$ is an empirical finding that warrants further study. This comparison is conditional on the stated parameter settings and the lexicographic order.

Under the wide symmetric intercept pattern, rejection rates differ sharply within the target set. Power is generally low at $N = 300$ and rises at $N = 1000$, but the increase is not the same for every pair. At $N = 1000$, pair (2, 3) has similar intercepts and rejection rates of 0.675 for SS, 0.682 for Cholesky, and 0.643 for LM. Pair (1, 15) has the largest intercept difference. Its corresponding rates are only 0.021, 0.001, and 0.005. Rejection rates also decline from (1, 13) to (1, 15) as the intercept difference increases. A large slope on the second factor marks a structural source of misfit, but the pair can reveal that misfit only when its responses contain enough local information. LM shows the clearest sensitivity to this loss. Its Type I error remains close to the selected nominal level, yet its power falls from 0.643 for (2, 3) to 0.005 for (1, 15). The result indicates a loss of detectable pairwise information rather than poor null calibration. Xu et al. [30] found the same general relation between intercept distance and local power.

The intercept setting also changes the non-target pattern. At $N = 1000$, the mean non-target Cholesky rate falls from 0.073 under zero intercepts to 0.057 under wide symmetric intercepts. The largest SS values remain concentrated around items 11 and 12, although their location and size change. Wide symmetric intercepts reduce the broad non-target increase, but they also remove much of the target-pair power. The apparently cleaner separation is therefore not necessarily better localization because less local signal is available. A larger sample makes both the target signal and any stable signal assigned outside the target set easier to detect.

Table 5 summarizes target sensitivity and mean non-target rejection. Table 6 reports 95% confidence intervals for the three pairs highlighted above, showing the Monte Carlo uncertainty in their estimated rejection rates. The corresponding complete pair-specific results are reported in Tables 7–10.

Table 5 shows the overall localization pattern in a compact form. Under zero intercepts, target sensitivity ranges from 0.646 to 0.858 at $N = 300$ and from 0.928 to 1.000 at $N = 1000$. It is substantially lower under the wide symmetric intercept pattern, while most mean

non-target rates remain much closer to the nominal level. This contrast shows that the main loss under the wide pattern occurs within the target set.

Table 5. Localization summaries for Study 1 at the 5% nominal level. Each entry is target sensitivity/mean non-target rejection rate. Target sensitivity is the mean rejection rate over 15 target pairs, and the second value is the mean over the remaining 90 pairs.

Statistic	Zero Intercept		Wide Symmetric Intercept	
	<i>N</i> = 300	<i>N</i> = 1000	<i>N</i> = 300	<i>N</i> = 1000
$X^2_{2c,SS}$	0.646/0.088	0.928/0.121	0.102/0.055	0.256/0.066
$X^2_{2c,Chol}$	0.858/0.058	1.000/0.073	0.118/0.051	0.351/0.057
<i>LM</i>	0.777/0.055	0.999/0.067	0.117/0.050	0.264/0.053
$GFfit^{(ij)}_{\perp}$	0.677/0.085	0.962/0.118	0.106/0.055	0.261/0.066
$\bar{X}^2_{ij}$	0.858/0.059	1.000/0.073	0.119/0.050	0.353/0.055
$GFfit^{(ij)}$	0.856/0.044	1.000/0.056	0.119/0.040	0.352/0.043

Table 6. Empirical rejection rates and 95% confidence intervals for selected Study 1 pairs with *N* = 1000. Each value before the brackets is the rejection proportion over 1000 replications, and the values in brackets give the corresponding 95% binomial confidence interval.

Intercept Setting	Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	<i>LM</i>
Zero	(1, 12), non-target	0.907 [0.887, 0.923]	0.086 [0.070, 0.105]	0.070 [0.056, 0.088]
Wide symmetric	(2, 3), target	0.675 [0.645, 0.703]	0.682 [0.652, 0.710]	0.643 [0.613, 0.672]
Wide symmetric	(1, 15), target	0.021 [0.014, 0.032]	0.001 [0.000, 0.006]	0.005 [0.002, 0.012]

The interval for SS at the non-target pair (1, 12) lies well above those for Cholesky and LM at the same pair in Table 6. The difference between the target pairs (2, 3) and (1, 15) is also much larger than the Monte Carlo uncertainty from 1000 replications and is consistent with the intercept-distance pattern described above.

Figure 2 summarizes the *N* = 1000 wide symmetric-intercept results in Table 10. It shows the large differences among target pairs and the smaller number of non-target peaks produced by the sequential methods.

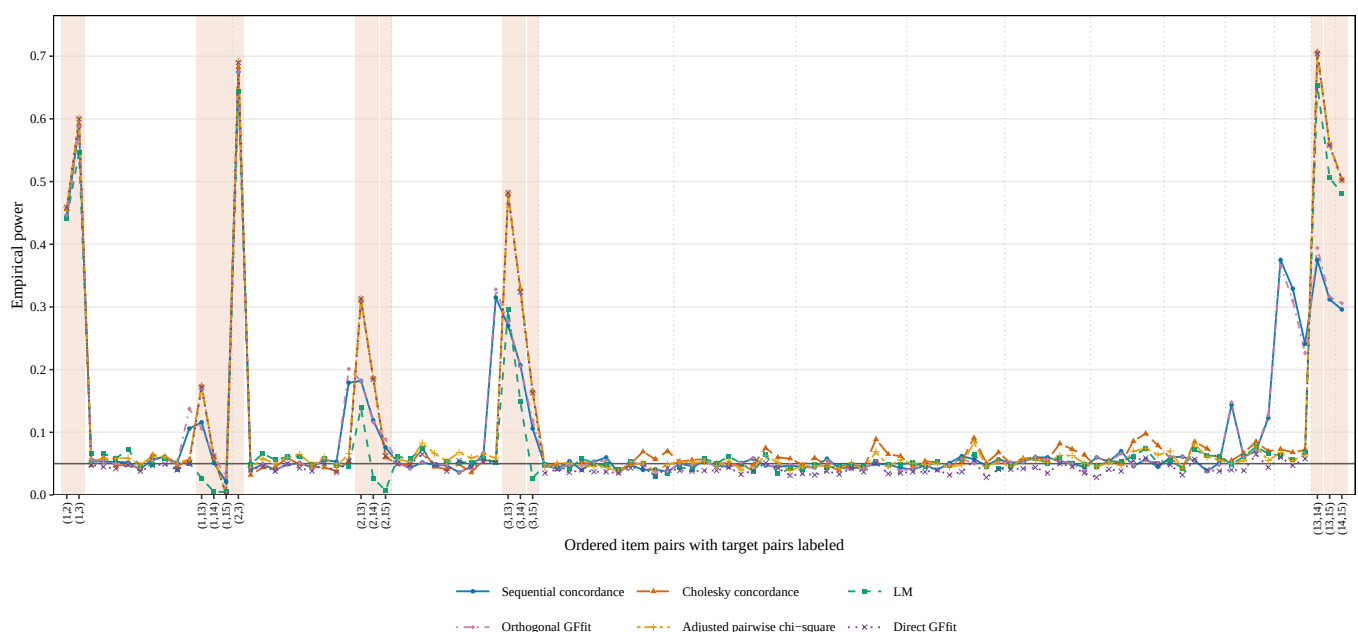


Figure 2. Empirical power of the six local statistics under the wide symmetric-intercept model with *q* = 15 and *N* = 1000 at the 5% nominal level. Shaded positions mark target pairs, and only target-pair labels are shown on the horizontal axis.

Table 7. Empirical power of the six local statistics for the zero-intercept model with $q = 15$ and $N = 300$ at the 5% nominal level.

Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit^{(ij)}_{\perp}$	$\bar{X}^2_{ij}$	$GFfit^{(ij)}$	Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit^{(ij)}_{\perp}$	$\bar{X}^2_{ij}$	$GFfit^{(ij)}$
(1, 2)	0.835	0.857	0.791	0.835	0.857	0.854	(5, 9)	0.047	0.063	0.058	0.046	0.059	0.041
(1, 3)	0.821	0.843	0.760	0.821	0.843	0.841	(5, 10)	0.060	0.061	0.053	0.064	0.065	0.048
(1, 4)	0.052	0.066	0.060	0.054	0.062	0.048	(5, 11)	0.051	0.057	0.052	0.050	0.061	0.045
(1, 5)	0.049	0.050	0.064	0.050	0.045	0.031	(5, 12)	0.054	0.052	0.045	0.059	0.054	0.039
(1, 6)	0.053	0.059	0.057	0.054	0.054	0.043	(5, 13)	0.054	0.057	0.058	0.058	0.059	0.038
(1, 7)	0.055	0.052	0.056	0.054	0.051	0.038	(5, 14)	0.035	0.061	0.058	0.041	0.062	0.049
(1, 8)	0.056	0.053	0.061	0.055	0.056	0.039	(5, 15)	0.046	0.059	0.050	0.047	0.062	0.048
(1, 9)	0.074	0.065	0.068	0.074	0.066	0.053	(6, 7)	0.056	0.051	0.055	0.057	0.052	0.042
(1, 10)	0.073	0.057	0.050	0.069	0.059	0.047	(6, 8)	0.052	0.050	0.046	0.052	0.052	0.039
(1, 11)	0.111	0.062	0.066	0.106	0.060	0.044	(6, 9)	0.048	0.047	0.042	0.047	0.044	0.033
(1, 12)	0.614	0.070	0.056	0.521	0.063	0.052	(6, 10)	0.038	0.042	0.048	0.038	0.041	0.026
(1, 13)	0.561	0.865	0.771	0.628	0.865	0.865	(6, 11)	0.053	0.044	0.045	0.053	0.044	0.030
(1, 14)	0.613	0.843	0.770	0.680	0.843	0.842	(6, 12)	0.062	0.054	0.052	0.061	0.052	0.038
(1, 15)	0.684	0.847	0.763	0.736	0.847	0.845	(6, 13)	0.047	0.080	0.068	0.050	0.077	0.056
(2, 3)	0.844	0.862	0.797	0.844	0.860	0.858	(6, 14)	0.056	0.061	0.049	0.057	0.060	0.048
(2, 4)	0.045	0.053	0.048	0.046	0.055	0.041	(6, 15)	0.050	0.061	0.061	0.053	0.062	0.050
(2, 5)	0.063	0.067	0.060	0.063	0.067	0.048	(7, 8)	0.043	0.055	0.041	0.044	0.056	0.043
(2, 6)	0.054	0.055	0.054	0.054	0.053	0.040	(7, 9)	0.051	0.059	0.063	0.051	0.052	0.031
(2, 7)	0.058	0.073	0.064	0.058	0.068	0.057	(7, 10)	0.035	0.046	0.047	0.034	0.047	0.033
(2, 8)	0.065	0.069	0.064	0.066	0.071	0.053	(7, 11)	0.052	0.055	0.057	0.048	0.056	0.041
(2, 9)	0.058	0.059	0.059	0.059	0.057	0.043	(7, 12)	0.050	0.049	0.059	0.049	0.063	0.043
(2, 10)	0.072	0.069	0.068	0.069	0.068	0.056	(7, 13)	0.054	0.051	0.047	0.054	0.055	0.040
(2, 11)	0.106	0.061	0.066	0.102	0.068	0.046	(7, 14)	0.046	0.064	0.061	0.043	0.062	0.045
(2, 12)	0.618	0.054	0.039	0.572	0.058	0.040	(7, 15)	0.046	0.061	0.064	0.044	0.061	0.050
(2, 13)	0.585	0.874	0.779	0.623	0.874	0.872	(8, 9)	0.046	0.048	0.054	0.048	0.048	0.036
(2, 14)	0.634	0.861	0.795	0.671	0.863	0.863	(8, 10)	0.045	0.058	0.051	0.046	0.054	0.044
(2, 15)	0.675	0.847	0.754	0.700	0.847	0.843	(8, 11)	0.048	0.052	0.047	0.048	0.054	0.042
(3, 4)	0.051	0.063	0.045	0.052	0.066	0.049	(8, 12)	0.052	0.045	0.053	0.052	0.055	0.037
(3, 5)	0.055	0.069	0.057	0.053	0.066	0.054	(8, 13)	0.057	0.070	0.068	0.053	0.072	0.053
(3, 6)	0.053	0.055	0.058	0.051	0.055	0.049	(8, 14)	0.046	0.062	0.056	0.046	0.062	0.049
(3, 7)	0.060	0.076	0.063	0.060	0.073	0.057	(8, 15)	0.054	0.054	0.059	0.058	0.059	0.043
(3, 8)	0.060	0.058	0.058	0.059	0.062	0.048	(9, 10)	0.065	0.067	0.046	0.067	0.055	0.039
(3, 9)	0.065	0.074	0.065	0.063	0.076	0.057	(9, 11)	0.037	0.043	0.051	0.039	0.044	0.031
(3, 10)	0.074	0.059	0.063	0.073	0.063	0.047	(9, 12)	0.057	0.050	0.046	0.058	0.044	0.037
(3, 11)	0.109	0.050	0.049	0.096	0.058	0.040	(9, 13)	0.049	0.071	0.060	0.045	0.072	0.055
(3, 12)	0.636	0.067	0.060	0.593	0.064	0.049	(9, 14)	0.059	0.060	0.062	0.056	0.061	0.045
(3, 13)	0.561	0.859	0.779	0.592	0.860	0.857	(9, 15)	0.064	0.078	0.079	0.066	0.079	0.064
(3, 14)	0.624	0.873	0.789	0.655	0.873	0.872	(10, 11)	0.068	0.048	0.049	0.067	0.061	0.045
(3, 15)	0.681	0.861	0.772	0.708	0.862	0.862	(10, 12)	0.075	0.050	0.056	0.070	0.051	0.038
(4, 5)	0.037	0.047	0.050	0.036	0.053	0.033	(10, 13)	0.045	0.052	0.044	0.042	0.052	0.037
(4, 6)	0.052	0.044	0.043	0.054	0.054	0.040	(10, 14)	0.061	0.075	0.071	0.063	0.071	0.058
(4, 7)	0.043	0.055	0.052	0.043	0.053	0.036	(10, 15)	0.058	0.062	0.065	0.057	0.058	0.049
(4, 8)	0.051	0.038	0.044	0.052	0.043	0.032	(11, 12)	0.232	0.056	0.055	0.245	0.065	0.047
(4, 9)	0.049	0.045	0.047	0.049	0.050	0.032	(11, 13)	0.071	0.058	0.060	0.070	0.057	0.045
(4, 10)	0.056	0.047	0.046	0.050	0.053	0.038	(11, 14)	0.102	0.063	0.051	0.101	0.065	0.048
(4, 11)	0.036	0.046	0.043	0.039	0.045	0.034	(11, 15)	0.144	0.058	0.052	0.144	0.060	0.045
(4, 12)	0.047	0.043	0.045	0.057	0.047	0.033	(12, 13)	0.352	0.059	0.069	0.324	0.061	0.045
(4, 13)	0.052	0.054	0.057	0.047	0.053	0.040	(12, 14)	0.386	0.074	0.064	0.361	0.075	0.057
(4, 14)	0.060	0.071	0.063	0.062	0.070	0.055	(12, 15)	0.327	0.061	0.052	0.286	0.060	0.044
(4, 15)	0.063	0.071	0.065	0.060	0.074	0.059	(13, 14)	0.513	0.873	0.788	0.557	0.875	0.872
(5, 6)	0.047	0.050	0.044	0.047	0.052	0.044	(13, 15)	0.502	0.857	0.768	0.534	0.859	0.855
(5, 7)	0.056	0.062	0.055	0.056	0.056	0.041	(14, 15)	0.562	0.845	0.778	0.573	0.846	0.844
(5, 8)	0.052	0.052	0.052	0.052	0.057	0.042							

Table 8. Empirical power of the six local statistics for the zero-intercept model with $q = 15$ and $N = 1000$ at the 5% nominal level.

Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}^2_{ij}$	$GFfit^{(ij)}$	Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}^2_{ij}$	$GFfit^{(ij)}$
(1, 2)	1.000	1.000	1.000	1.000	1.000	1.000	(5, 9)	0.046	0.065	0.054	0.047	0.061	0.044
(1, 3)	1.000	1.000	0.999	1.000	1.000	1.000	(5, 10)	0.049	0.050	0.044	0.051	0.045	0.028
(1, 4)	0.054	0.079	0.070	0.055	0.077	0.061	(5, 11)	0.042	0.060	0.059	0.042	0.060	0.046
(1, 5)	0.054	0.092	0.072	0.057	0.094	0.073	(5, 12)	0.084	0.050	0.057	0.082	0.048	0.039
(1, 6)	0.052	0.084	0.071	0.051	0.087	0.064	(5, 13)	0.047	0.103	0.086	0.046	0.104	0.081
(1, 7)	0.070	0.094	0.089	0.069	0.096	0.073	(5, 14)	0.045	0.080	0.079	0.047	0.082	0.058
(1, 8)	0.068	0.082	0.079	0.067	0.078	0.064	(5, 15)	0.055	0.090	0.075	0.057	0.096	0.071
(1, 9)	0.076	0.093	0.088	0.078	0.099	0.072	(6, 7)	0.054	0.048	0.047	0.055	0.048	0.040
(1, 10)	0.099	0.090	0.071	0.095	0.088	0.071	(6, 8)	0.054	0.045	0.046	0.053	0.046	0.034
(1, 11)	0.178	0.094	0.059	0.173	0.097	0.075	(6, 9)	0.055	0.043	0.061	0.054	0.045	0.034
(1, 12)	0.907	0.086	0.070	0.867	0.088	0.069	(6, 10)	0.049	0.062	0.052	0.050	0.051	0.041
(1, 13)	0.838	1.000	0.999	0.937	1.000	1.000	(6, 11)	0.052	0.048	0.053	0.052	0.043	0.033
(1, 14)	0.917	1.000	1.000	0.980	1.000	1.000	(6, 12)	0.069	0.055	0.043	0.069	0.058	0.042
(1, 15)	0.986	1.000	0.999	0.998	1.000	1.000	(6, 13)	0.035	0.086	0.082	0.035	0.087	0.076
(2, 3)	1.000	1.000	0.999	1.000	1.000	1.000	(6, 14)	0.042	0.098	0.083	0.043	0.096	0.079
(2, 4)	0.057	0.095	0.069	0.056	0.089	0.064	(6, 15)	0.054	0.089	0.074	0.055	0.093	0.066
(2, 5)	0.042	0.103	0.078	0.043	0.095	0.078	(7, 8)	0.058	0.060	0.057	0.058	0.055	0.044
(2, 6)	0.046	0.085	0.069	0.045	0.086	0.068	(7, 9)	0.040	0.049	0.058	0.040	0.045	0.039
(2, 7)	0.055	0.098	0.078	0.055	0.098	0.079	(7, 10)	0.051	0.059	0.067	0.050	0.061	0.044
(2, 8)	0.072	0.083	0.070	0.073	0.087	0.066	(7, 11)	0.050	0.057	0.056	0.050	0.050	0.039
(2, 9)	0.061	0.076	0.067	0.060	0.078	0.060	(7, 12)	0.074	0.045	0.044	0.075	0.043	0.028
(2, 10)	0.084	0.074	0.071	0.082	0.075	0.055	(7, 13)	0.034	0.090	0.084	0.040	0.096	0.071
(2, 11)	0.196	0.092	0.089	0.188	0.096	0.076	(7, 14)	0.063	0.104	0.081	0.061	0.103	0.077
(2, 12)	0.911	0.084	0.080	0.890	0.087	0.066	(7, 15)	0.053	0.080	0.084	0.055	0.076	0.061
(2, 13)	0.841	1.000	0.999	0.910	1.000	1.000	(8, 9)	0.045	0.059	0.049	0.045	0.055	0.037
(2, 14)	0.919	1.000	0.999	0.962	1.000	1.000	(8, 10)	0.048	0.046	0.040	0.047	0.047	0.041
(2, 15)	0.988	1.000	0.999	0.995	1.000	1.000	(8, 11)	0.044	0.047	0.052	0.044	0.041	0.028
(3, 4)	0.055	0.073	0.066	0.054	0.073	0.064	(8, 12)	0.072	0.050	0.049	0.073	0.053	0.032
(3, 5)	0.055	0.100	0.090	0.054	0.100	0.077	(8, 13)	0.047	0.078	0.063	0.046	0.079	0.058
(3, 6)	0.067	0.100	0.090	0.068	0.100	0.073	(8, 14)	0.050	0.070	0.072	0.050	0.074	0.058
(3, 7)	0.061	0.099	0.082	0.061	0.105	0.079	(8, 15)	0.058	0.091	0.082	0.058	0.092	0.072
(3, 8)	0.074	0.084	0.081	0.074	0.089	0.061	(9, 10)	0.041	0.047	0.042	0.040	0.043	0.029
(3, 9)	0.076	0.095	0.081	0.079	0.100	0.079	(9, 11)	0.047	0.043	0.047	0.047	0.038	0.027
(3, 10)	0.094	0.077	0.082	0.097	0.081	0.063	(9, 12)	0.079	0.047	0.057	0.083	0.046	0.034
(3, 11)	0.209	0.098	0.080	0.209	0.099	0.082	(9, 13)	0.042	0.079	0.067	0.044	0.076	0.062
(3, 12)	0.902	0.092	0.079	0.885	0.092	0.072	(9, 14)	0.054	0.090	0.086	0.053	0.086	0.073
(3, 13)	0.849	1.000	0.999	0.904	1.000	1.000	(9, 15)	0.059	0.091	0.073	0.058	0.090	0.071
(3, 14)	0.922	1.000	1.000	0.950	1.000	1.000	(10, 11)	0.082	0.056	0.042	0.077	0.046	0.035
(3, 15)	0.984	1.000	1.000	0.995	1.000	1.000	(10, 12)	0.108	0.050	0.047	0.108	0.040	0.029
(4, 5)	0.058	0.043	0.037	0.058	0.038	0.032	(10, 13)	0.044	0.095	0.071	0.049	0.094	0.079
(4, 6)	0.054	0.059	0.056	0.051	0.057	0.047	(10, 14)	0.060	0.083	0.080	0.057	0.085	0.065
(4, 7)	0.051	0.046	0.041	0.049	0.047	0.036	(10, 15)	0.060	0.079	0.072	0.060	0.082	0.068
(4, 8)	0.051	0.044	0.048	0.051	0.041	0.020	(11, 12)	0.419	0.053	0.062	0.419	0.044	0.027
(4, 9)	0.051	0.058	0.057	0.050	0.057	0.041	(11, 13)	0.086	0.101	0.089	0.081	0.107	0.083
(4, 10)	0.058	0.050	0.049	0.058	0.038	0.029	(11, 14)	0.159	0.099	0.091	0.161	0.103	0.079
(4, 11)	0.045	0.060	0.053	0.047	0.063	0.036	(11, 15)	0.343	0.096	0.091	0.337	0.096	0.079
(4, 12)	0.045	0.034	0.048	0.052	0.042	0.028	(12, 13)	0.728	0.077	0.067	0.699	0.077	0.056
(4, 13)	0.048	0.077	0.072	0.053	0.075	0.065	(12, 14)	0.719	0.086	0.082	0.668	0.085	0.066
(4, 14)	0.055	0.080	0.084	0.049	0.084	0.066	(12, 15)	0.695	0.082	0.075	0.633	0.082	0.063
(4, 15)	0.053	0.087	0.080	0.058	0.088	0.069	(13, 14)	0.831	0.999	0.997	0.896	0.999	0.999
(5, 6)	0.053	0.046	0.062	0.053	0.044	0.033	(13, 15)	0.888	1.000	1.000	0.930	1.000	1.000
(5, 7)	0.054	0.048	0.058	0.054	0.039	0.031	(14, 15)	0.950	1.000	1.000	0.972	1.000	1.000
(5, 8)	0.070	0.062	0.053	0.069	0.064	0.051							

Table 9. Empirical power of the six local statistics for the wide symmetric-intercept model with $q = 15$ and $N = 300$ at the 5% nominal level.

Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}^2_{ij}$	$GFfit^{(ij)}$	Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}^2_{ij}$	$GFfit^{(ij)}$
(1, 2)	0.213	0.221	0.238	0.214	0.222	0.222	(5, 9)	0.059	0.056	0.047	0.058	0.055	0.043
(1, 3)	0.265	0.268	0.271	0.266	0.266	0.265	(5, 10)	0.042	0.046	0.038	0.041	0.045	0.033
(1, 4)	0.059	0.061	0.057	0.058	0.054	0.046	(5, 11)	0.056	0.060	0.042	0.057	0.061	0.050
(1, 5)	0.060	0.061	0.052	0.061	0.062	0.050	(5, 12)	0.057	0.041	0.054	0.058	0.046	0.043
(1, 6)	0.058	0.052	0.054	0.059	0.052	0.041	(5, 13)	0.036	0.044	0.048	0.041	0.045	0.041
(1, 7)	0.060	0.054	0.062	0.057	0.069	0.058	(5, 14)	0.049	0.052	0.047	0.052	0.043	0.033
(1, 8)	0.042	0.059	0.056	0.046	0.052	0.041	(5, 15)	0.047	0.046	0.036	0.052	0.034	0.028
(1, 9)	0.044	0.049	0.048	0.046	0.038	0.024	(6, 7)	0.047	0.056	0.060	0.048	0.056	0.046
(1, 10)	0.046	0.049	0.039	0.049	0.054	0.041	(6, 8)	0.063	0.057	0.052	0.063	0.053	0.039
(1, 11)	0.049	0.039	0.039	0.052	0.034	0.027	(6, 9)	0.049	0.043	0.052	0.049	0.044	0.036
(1, 12)	0.045	0.045	0.055	0.033	0.037	0.031	(6, 10)	0.056	0.047	0.050	0.057	0.045	0.040
(1, 13)	0.023	0.005	0.004	0.023	0.005	0.005	(6, 11)	0.047	0.051	0.040	0.047	0.053	0.040
(1, 14)	0.012	0.004	0.010	0.012	0.004	0.004	(6, 12)	0.053	0.052	0.047	0.057	0.052	0.044
(1, 15)	0.014	0.006	0.024	0.029	0.006	0.006	(6, 13)	0.051	0.047	0.047	0.050	0.048	0.042
(2, 3)	0.297	0.296	0.291	0.296	0.304	0.301	(6, 14)	0.046	0.047	0.038	0.046	0.044	0.036
(2, 4)	0.050	0.053	0.048	0.048	0.057	0.048	(6, 15)	0.055	0.042	0.046	0.048	0.039	0.034
(2, 5)	0.053	0.062	0.065	0.053	0.057	0.047	(7, 8)	0.038	0.049	0.040	0.038	0.050	0.043
(2, 6)	0.048	0.048	0.056	0.048	0.052	0.042	(7, 9)	0.040	0.057	0.056	0.042	0.061	0.047
(2, 7)	0.069	0.050	0.063	0.068	0.063	0.048	(7, 10)	0.054	0.041	0.049	0.055	0.044	0.031
(2, 8)	0.052	0.052	0.052	0.054	0.049	0.043	(7, 11)	0.043	0.036	0.041	0.044	0.047	0.034
(2, 9)	0.039	0.051	0.051	0.041	0.050	0.039	(7, 12)	0.050	0.036	0.038	0.049	0.036	0.027
(2, 10)	0.037	0.047	0.041	0.037	0.053	0.034	(7, 13)	0.039	0.047	0.039	0.042	0.044	0.033
(2, 11)	0.046	0.038	0.037	0.047	0.041	0.030	(7, 14)	0.046	0.072	0.049	0.051	0.065	0.046
(2, 12)	0.051	0.042	0.039	0.050	0.041	0.032	(7, 15)	0.062	0.058	0.050	0.059	0.056	0.041
(2, 13)	0.037	0.018	0.007	0.046	0.019	0.019	(8, 9)	0.038	0.046	0.060	0.039	0.044	0.029
(2, 14)	0.020	0.002	0.012	0.027	0.003	0.003	(8, 10)	0.050	0.046	0.041	0.051	0.057	0.042
(2, 15)	0.017	0.003	0.015	0.026	0.004	0.004	(8, 11)	0.055	0.058	0.060	0.055	0.057	0.049
(3, 4)	0.057	0.055	0.056	0.057	0.061	0.052	(8, 12)	0.051	0.052	0.047	0.047	0.049	0.047
(3, 5)	0.045	0.056	0.065	0.045	0.053	0.045	(8, 13)	0.060	0.073	0.062	0.061	0.066	0.053
(3, 6)	0.064	0.055	0.061	0.064	0.060	0.050	(8, 14)	0.037	0.046	0.041	0.042	0.042	0.028
(3, 7)	0.048	0.057	0.070	0.050	0.061	0.047	(8, 15)	0.052	0.050	0.050	0.051	0.044	0.035
(3, 8)	0.048	0.046	0.053	0.049	0.051	0.036	(9, 10)	0.045	0.041	0.042	0.043	0.038	0.027
(3, 9)	0.037	0.042	0.052	0.038	0.047	0.034	(9, 11)	0.046	0.056	0.046	0.047	0.052	0.036
(3, 10)	0.037	0.045	0.050	0.038	0.055	0.045	(9, 12)	0.058	0.046	0.053	0.056	0.047	0.036
(3, 11)	0.049	0.043	0.040	0.050	0.048	0.039	(9, 13)	0.062	0.065	0.054	0.062	0.058	0.043
(3, 12)	0.066	0.043	0.033	0.069	0.042	0.036	(9, 14)	0.052	0.048	0.042	0.048	0.039	0.026
(3, 13)	0.082	0.089	0.010	0.077	0.092	0.091	(9, 15)	0.048	0.048	0.042	0.050	0.044	0.035
(3, 14)	0.039	0.012	0.007	0.037	0.012	0.012	(10, 11)	0.068	0.049	0.054	0.067	0.050	0.036
(3, 15)	0.028	0.002	0.014	0.034	0.003	0.003	(10, 12)	0.059	0.050	0.057	0.059	0.045	0.034
(4, 5)	0.037	0.050	0.043	0.038	0.044	0.036	(10, 13)	0.058	0.057	0.060	0.057	0.051	0.038
(4, 6)	0.058	0.059	0.052	0.054	0.060	0.050	(10, 14)	0.051	0.064	0.054	0.052	0.061	0.052
(4, 7)	0.059	0.056	0.048	0.060	0.049	0.041	(10, 15)	0.055	0.063	0.047	0.054	0.054	0.037
(4, 8)	0.046	0.047	0.056	0.051	0.051	0.040	(11, 12)	0.115	0.049	0.045	0.104	0.044	0.037
(4, 9)	0.053	0.051	0.052	0.054	0.053	0.043	(11, 13)	0.066	0.075	0.061	0.066	0.069	0.052
(4, 10)	0.054	0.059	0.049	0.052	0.068	0.060	(11, 14)	0.069	0.053	0.052	0.070	0.049	0.034
(4, 11)	0.033	0.044	0.055	0.035	0.044	0.037	(11, 15)	0.086	0.059	0.056	0.085	0.055	0.043
(4, 12)	0.058	0.049	0.047	0.051	0.048	0.041	(12, 13)	0.153	0.049	0.061	0.151	0.051	0.041
(4, 13)	0.051	0.047	0.047	0.059	0.045	0.039	(12, 14)	0.139	0.055	0.054	0.133	0.052	0.045
(4, 14)	0.050	0.047	0.035	0.047	0.041	0.028	(12, 15)	0.137	0.053	0.068	0.137	0.060	0.048
(4, 15)	0.040	0.046	0.037	0.047	0.037	0.029	(13, 14)	0.203	0.317	0.297	0.216	0.318	0.317
(5, 6)	0.052	0.046	0.064	0.051	0.045	0.033	(13, 15)	0.160	0.284	0.296	0.158	0.281	0.280
(5, 7)	0.061	0.069	0.057	0.063	0.064	0.051	(14, 15)	0.120	0.249	0.259	0.130	0.249	0.248
(5, 8)	0.058	0.053	0.048	0.058	0.045	0.033							

Table 10. Empirical power of the six local statistics for the wide symmetric-intercept model with $q = 15$ and $N = 1000$ at the 5% nominal level.

Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}^2_{ij}$	$GFfit^{(ij)}$	Pair	$X^2_{2c,SS}$	$X^2_{2c,Chol}$	LM	$GFfit_{\perp}^{(ij)}$	$\bar{X}^2_{ij}$	$GFfit^{(ij)}$
(1, 2)	0.446	0.456	0.441	0.446	0.459	0.459	(5, 9)	0.047	0.050	0.051	0.044	0.047	0.039
(1, 3)	0.588	0.592	0.546	0.588	0.601	0.600	(5, 10)	0.045	0.051	0.061	0.045	0.054	0.044
(1, 4)	0.054	0.047	0.066	0.058	0.054	0.048	(5, 11)	0.051	0.047	0.051	0.052	0.040	0.033
(1, 5)	0.053	0.060	0.066	0.052	0.058	0.045	(5, 12)	0.058	0.048	0.037	0.059	0.044	0.039
(1, 6)	0.053	0.047	0.058	0.050	0.058	0.042	(5, 13)	0.047	0.075	0.064	0.048	0.063	0.051
(1, 7)	0.052	0.048	0.073	0.049	0.059	0.048	(5, 14)	0.045	0.060	0.034	0.044	0.052	0.043
(1, 8)	0.041	0.042	0.045	0.042	0.048	0.038	(5, 15)	0.047	0.058	0.042	0.049	0.039	0.031
(1, 9)	0.057	0.064	0.048	0.055	0.062	0.047	(6, 7)	0.045	0.048	0.040	0.044	0.046	0.034
(1, 10)	0.061	0.061	0.058	0.060	0.061	0.049	(6, 8)	0.044	0.059	0.049	0.044	0.047	0.032
(1, 11)	0.051	0.046	0.041	0.049	0.053	0.040	(6, 9)	0.058	0.048	0.044	0.055	0.046	0.038
(1, 12)	0.106	0.056	0.051	0.138	0.056	0.051	(6, 10)	0.043	0.039	0.046	0.043	0.042	0.033
(1, 13)	0.116	0.174	0.027	0.106	0.174	0.172	(6, 11)	0.043	0.049	0.049	0.041	0.051	0.043
(1, 14)	0.051	0.060	0.005	0.062	0.062	0.062	(6, 12)	0.044	0.043	0.045	0.043	0.041	0.037
(1, 15)	0.021	0.001	0.005	0.029	0.001	0.001	(6, 13)	0.050	0.089	0.054	0.053	0.069	0.051
(2, 3)	0.675	0.682	0.643	0.675	0.691	0.690	(6, 14)	0.049	0.065	0.048	0.046	0.046	0.034
(2, 4)	0.040	0.032	0.049	0.040	0.044	0.040	(6, 15)	0.043	0.062	0.049	0.039	0.056	0.035
(2, 5)	0.049	0.044	0.066	0.049	0.058	0.046	(7, 8)	0.041	0.045	0.052	0.042	0.050	0.037
(2, 6)	0.040	0.045	0.056	0.039	0.050	0.038	(7, 9)	0.047	0.054	0.045	0.048	0.049	0.037
(2, 7)	0.049	0.059	0.061	0.049	0.060	0.050	(7, 10)	0.040	0.052	0.053	0.040	0.051	0.040
(2, 8)	0.051	0.049	0.061	0.051	0.065	0.043	(7, 11)	0.051	0.047	0.047	0.048	0.045	0.032
(2, 9)	0.044	0.045	0.048	0.043	0.048	0.038	(7, 12)	0.062	0.054	0.059	0.058	0.048	0.037
(2, 10)	0.056	0.044	0.058	0.058	0.060	0.050	(7, 13)	0.057	0.091	0.064	0.050	0.084	0.055
(2, 11)	0.053	0.038	0.047	0.050	0.045	0.036	(7, 14)	0.047	0.051	0.046	0.046	0.045	0.028
(2, 12)	0.179	0.054	0.045	0.201	0.066	0.054	(7, 15)	0.057	0.068	0.042	0.052	0.054	0.041
(2, 13)	0.182	0.310	0.139	0.182	0.313	0.313	(8, 9)	0.051	0.050	0.046	0.052	0.045	0.041
(2, 14)	0.119	0.187	0.027	0.116	0.186	0.185	(8, 10)	0.054	0.058	0.054	0.053	0.058	0.042
(2, 15)	0.076	0.060	0.007	0.089	0.066	0.062	(8, 11)	0.060	0.060	0.059	0.060	0.056	0.044
(3, 4)	0.052	0.050	0.061	0.051	0.057	0.049	(8, 12)	0.060	0.056	0.051	0.057	0.052	0.035
(3, 5)	0.043	0.049	0.058	0.042	0.053	0.049	(8, 13)	0.054	0.082	0.060	0.054	0.063	0.050
(3, 6)	0.052	0.073	0.074	0.052	0.083	0.065	(8, 14)	0.051	0.073	0.051	0.050	0.063	0.045
(3, 7)	0.048	0.045	0.048	0.049	0.066	0.048	(8, 15)	0.044	0.064	0.048	0.037	0.050	0.035
(3, 8)	0.047	0.041	0.054	0.048	0.054	0.038	(9, 10)	0.060	0.044	0.045	0.060	0.046	0.028
(3, 9)	0.036	0.049	0.052	0.036	0.068	0.055	(9, 11)	0.053	0.055	0.054	0.053	0.051	0.041
(3, 10)	0.045	0.036	0.052	0.044	0.059	0.044	(9, 12)	0.070	0.051	0.053	0.066	0.047	0.038
(3, 11)	0.066	0.056	0.058	0.067	0.065	0.053	(9, 13)	0.046	0.086	0.062	0.047	0.070	0.055
(3, 12)	0.315	0.053	0.052	0.328	0.059	0.052	(9, 14)	0.057	0.098	0.074	0.055	0.074	0.059
(3, 13)	0.270	0.479	0.296	0.276	0.482	0.482	(9, 15)	0.045	0.079	0.051	0.051	0.064	0.048
(3, 14)	0.207	0.329	0.149	0.205	0.326	0.324	(10, 11)	0.060	0.059	0.055	0.061	0.070	0.048
(3, 15)	0.106	0.166	0.027	0.119	0.164	0.163	(10, 12)	0.061	0.043	0.043	0.061	0.043	0.032
(4, 5)	0.047	0.047	0.048	0.048	0.044	0.035	(10, 13)	0.055	0.085	0.072	0.055	0.081	0.057
(4, 6)	0.042	0.048	0.043	0.042	0.051	0.041	(10, 14)	0.038	0.074	0.063	0.038	0.061	0.040
(4, 7)	0.054	0.046	0.042	0.048	0.051	0.036	(10, 15)	0.051	0.058	0.062	0.048	0.056	0.038
(4, 8)	0.040	0.053	0.058	0.047	0.047	0.040	(11, 12)	0.144	0.055	0.049	0.148	0.050	0.040
(4, 9)	0.053	0.048	0.051	0.053	0.046	0.037	(11, 13)	0.061	0.067	0.061	0.061	0.055	0.039
(4, 10)	0.060	0.048	0.048	0.055	0.044	0.037	(11, 14)	0.070	0.085	0.077	0.071	0.077	0.065
(4, 11)	0.036	0.039	0.041	0.036	0.038	0.035	(11, 15)	0.123	0.070	0.066	0.127	0.055	0.044
(4, 12)	0.048	0.050	0.054	0.053	0.047	0.043	(12, 13)	0.375	0.073	0.063	0.366	0.068	0.060
(4, 13)	0.040	0.069	0.050	0.048	0.051	0.044	(12, 14)	0.329	0.068	0.056	0.310	0.055	0.047
(4, 14)	0.041	0.057	0.029	0.037	0.038	0.031	(12, 15)	0.241	0.072	0.068	0.226	0.064	0.058
(4, 15)	0.037	0.070	0.034	0.038	0.049	0.043	(13, 14)	0.375	0.707	0.654	0.394	0.705	0.704
(5, 6)	0.050	0.053	0.044	0.050	0.053	0.039	(13, 15)	0.312	0.559	0.506	0.316	0.558	0.558
(5, 7)	0.046	0.056	0.041	0.047	0.051	0.038	(14, 15)	0.296	0.504	0.481	0.306	0.502	0.502
(5, 8)	0.054	0.057	0.058	0.056	0.053	0.039							

4.3. Study 1: Order Sensitivity and Multiple Testing

The order analysis compared each of the three procedures across the same three pair sequences at $N = 1000$. The original lexicographic order reproduced the main results. Changing the order left the sum of the SS components unchanged to numerical precision, with a maximum difference of 4.05×10^{-12} across all replications. This agrees with the invariance of the total statistic described in Section 2. The mean componentwise null rejection rates of SS, Cholesky, and $GFfit_{\perp}^{(ij)}$ remained close to 0.05 under all three orders.

The rejection rates for individual pairs changed much more. Pair (2, 12) was one of the largest non-target SS peaks under zero intercepts in the original lexicographic analysis, with a rejection rate of 0.911. Its rate fell to 0.044 under the reverse order and 0.042 under the fixed random order. Comparable peaks appeared at (4, 13) under the reverse order and at (8, 15) under the fixed random order. The concentration of large values near item 12 in the lexicographic analysis is therefore order-dependent and reflects the allocation of shared discrepancy in that sequence. A large SS component alone does not establish that the labeled raw pair is the source of the discrepancy.

For each pair, the range of rejection rates across the three orders measures how much its result changes with the sequence. Under zero intercepts, the largest such range was 0.869 for SS and 0.847 for $GFfit_{\perp}^{(ij)}$, compared with 0.009 for Cholesky. Under wide symmetric intercepts, the corresponding ranges were 0.418, 0.400, and 0.039. Thus, the Cholesky rejection rates were less sensitive to these three orders, although the components themselves also depend on order.

Table 11 shows how these changes affect the target and non-target groups as a whole. For SS under zero intercepts, target sensitivity is 0.928 in the lexicographic order and 0.913 in the reverse order. The corresponding mean non-target rates are 0.121 and 0.118. These summaries are similar even though the rate for pair (2, 12) changes sharply. Similar averages can therefore conceal substantial changes in which pairs are selected. The fixed random order gives higher target sensitivity and a lower mean non-target rate for SS and $GFfit_{\perp}^{(ij)}$ in this analysis. Because only one random permutation was examined, this result does not establish that random ordering is generally preferable. Pair-specific results and ranges are reported in Supplementary Tables S26 and S27.

Table 11. Order sensitivity of the three order-dependent local procedures in Study 1 with $N = 1000$. Each entry gives the target sensitivity followed by the mean rejection rate among non-target pairs at the 5% level. Monte Carlo standard errors and 95% intervals are reported in Supplementary Tables S24 and S25.

Intercepts	Statistic	Lexicographic	Reverse	Fixed Random
Zero	$X_{2c,SS}^2$	0.928/0.121	0.913/0.118	0.979/0.074
Zero	$X_{2c,Chol}^2$	1.000/0.073	1.000/0.073	1.000/0.073
Zero	$GFfit_{\perp}^{(ij)}$	0.962/0.118	0.950/0.114	0.983/0.080
Wide symmetric	$X_{2c,SS}^2$	0.256/0.066	0.259/0.065	0.326/0.052
Wide symmetric	$X_{2c,Chol}^2$	0.351/0.057	0.349/0.057	0.352/0.056
Wide symmetric	$GFfit_{\perp}^{(ij)}$	0.261/0.066	0.263/0.064	0.327/0.052

The multiple-testing analysis examines what happens when all 105 pairs are inspected together in the original order at $N = 1000$. At the 5% level, the probability of selecting at least one pair under the correctly specified model ranged from 0.975 to 0.996 across the six procedures with zero intercepts and from 0.979 to 0.998 with wide symmetric intercepts. Thus, almost every null replication contained at least one rejection when no adjustment was used. Rates near 0.05 for individual pairs do not give a 5% error rate for the full list of selected pairs.

Holm adjustment substantially reduced this probability, but some empirical family-wise error rates still exceeded 0.05. Across the six procedures, the rates ranged from 0.019 to 0.074 under zero intercepts and from 0.045 to 0.118 under wide symmetric intercepts. The largest rate, 0.118, occurred for LM in the wide setting. Holm's error control requires valid individual p -values. The excess rejection may reflect limited accuracy of the asymptotic p -values at the small thresholds used for simultaneous testing, with nearly degenerate margins as one possible contributing factor. The Benjamini–Hochberg function was examined as an alternative targeting the false discovery rate. These results do not establish nominal

familywise error control for the Holm function or nominal false discovery rate control for the Benjamini–Hochberg function for all six local diagnostics in the settings studied.

The localization results show the cost of reducing selections outside the target set. For SS under zero intercepts, the unadjusted target sensitivity and mean non-target rate are 0.928 and 0.121. Holm reduces the non-target rate to 0.042, while target sensitivity falls to 0.725. Benjamini–Hochberg gives an intermediate result, with sensitivity 0.855 and a non-target rate of 0.066. Under wide symmetric intercepts, unadjusted SS sensitivity is already only 0.256 and falls to 0.049 with Holm and 0.062 with Benjamini–Hochberg. The weak pairwise signal in this condition is difficult to retain after simultaneous correction. Results for all six procedures, including precision and false discovery proportion, are reported in Supplementary Tables S28–S30.

4.4. Study 2: Type I Error

Table 12 reports the Type I error rates for the omnibus statistics. The five lower-order statistics, X_{2c}^2 , $X_{[2]}^2$, M_2 , Y_2 , and X_{red}^2 , remain close to the nominal level in all three symmetric slope conditions. Their rates range from 0.043 to 0.059 at $N = 300$ and from 0.040 to 0.068 at $N = 1000$. There is no systematic Type I error inflation for the lower-order tests as the symmetric slope strength changes [12,14,23].

The full-table Pearson–Fisher statistic behaves differently. Its rejection rates are far above the nominal level in every condition. With $2^{15} = 32,768$ response patterns and samples of 300 or 1000, most cells remain empty, and the full-table chi-square approximation is not reliable. We therefore do not report its power because a statistic with an inflated Type I error cannot provide a valid power comparison.

Table 12. Empirical Type I error rates of the six omnibus statistics under the symmetric slope conditions with $q = 15$ at the 5% nominal level.

N	Condition	X_{2c}^2	$X_{[2]}^2$	M_2	Y_2	X_{red}^2	X_{FF}^2
300	(W, W)	0.045	0.044	0.043	0.047	0.043	0.219
300	(W, S)	0.056	0.051	0.053	0.059	0.054	0.381
300	(S, S)	0.054	0.055	0.054	0.050	0.058	0.423
1000	(W, W)	0.063	0.062	0.062	0.068	0.062	0.119
1000	(W, S)	0.053	0.049	0.051	0.052	0.051	0.313
1000	(S, S)	0.044	0.042	0.040	0.065	0.044	0.374

Note: Bold values mark rejection rates significantly different from 0.05 based on a two-sided exact binomial test (5% level).

Figure 3 compares the $N = 1000$ results at all three nominal levels. The lower-order statistics follow the corresponding reference lines, whereas the Pearson–Fisher statistic remains much higher under the mixed and strong slope conditions.

4.5. Study 2: Power

Table 13 gives the omnibus power results under the symmetric slope alternatives. The four quadratic lower-order statistics, X_{2c}^2 , $X_{[2]}^2$, M_2 , and X_{red}^2 , behave similarly. At $N = 300$, their rejection rates are moderate and highest under (S, S). At $N = 1000$, all four statistics have power of at least 0.992 in every condition. Concordance margins and the usual (1, 1) pair margins retain almost the same overall information for these alternatives.

The ordering of the slope conditions shows that power is not determined only by the average size of the generating slopes. The mixed (W, S) condition does not have higher power than (W, W) at $N = 300$, even though one factor uses the stronger slope vector. The discrepancy left after fitting the common-slope model depends on both the magnitude and arrangement of the item slopes [42].

Y_2 follows a different pattern. At $N = 300$, its power is 0.748 for (W, W) and 0.783 for (W, S) , but only 0.227 for (S, S) . This behavior is associated with its moment-matched reference distribution. Under strong slopes, changes in the marginal variances and eigenvalue structure reduce the estimated reference degrees of freedom to about 6.4 in the (S, S) condition. The low Y_2 value should therefore be interpreted as a feature of the approximation rather than evidence that the alternative is weak. At $N = 1000$, its power rises to 1.000 for (W, W) and (W, S) and to 0.869 for (S, S) .

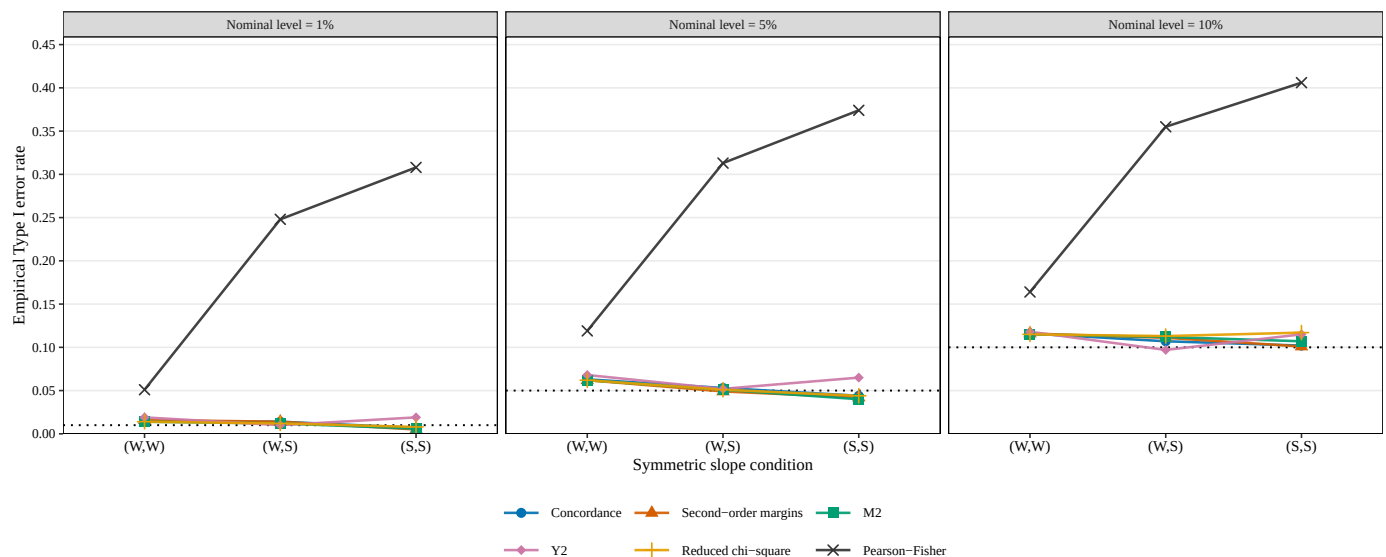


Figure 3. Empirical Type I error rates of the omnibus statistics under the three symmetric slope conditions with $q = 15$ and $N = 1000$ at the 1%, 5%, and 10% nominal levels. The horizontal dotted line in each panel marks the corresponding nominal level.

Table 13. Empirical power of the five lower-order omnibus statistics under the symmetric slope alternatives with $q = 15$ at the 5% nominal level.

N	Condition	X^2_{2c}	$X^2_{[2]}$	M_2	Y_2	X^2_{red}
300	(W, W)	0.447	0.446	0.455	0.748	0.455
300	(W, S)	0.426	0.421	0.423	0.783	0.424
300	(S, S)	0.623	0.620	0.615	0.227	0.625
1000	(W, W)	0.992	0.994	0.994	1.000	0.994
1000	(W, S)	0.992	0.993	0.993	1.000	0.993
1000	(S, S)	0.998	0.998	0.998	0.869	0.998

4.6. Sensitivity to the Nominal Level

Table 14 summarizes the additional results for $\alpha = 0.01, 0.05$, and 0.10 . Each entry gives the range across methods. The separate Supplementary File reports the method-specific values and the complete 105-pair Study 1 results at the additional 1% and 10% levels. In Study 1, the mean null rejection rates are close to the selected level under both intercept patterns. The direct $GFfit^{(ij)}$ remains somewhat conservative, while the other five local statistics are generally close to the stated level.

Power increases when the nominal level is raised. Under zero intercepts with $N = 300$, the average rejection rate over the 15 target pairs ranges across methods from 0.451 to 0.672 at $\alpha = 0.01$. At $\alpha = 0.10$, it ranges from 0.735 to 0.917. Under the wide symmetric intercept pattern, the corresponding ranges are only 0.045 to 0.062 and 0.157 to 0.176. Even with $N = 1000$, the range at $\alpha = 0.10$ is 0.351 to 0.481. A larger nominal level recovers some power but does not remove the loss produced by the wide intercept pattern.

Higher power is accompanied by more rejection outside the designed target set. For example, under zero intercepts with $N = 300$, the mean non-target rate for $X_{2c,SS}^2$ rises from 0.036 at $\alpha = 0.01$ to 0.144 at $\alpha = 0.10$. The corresponding Cholesky rates rise from 0.013 to 0.112. Changing α does not alter the main localization pattern, but it changes how often weaker or shared signals pass the rejection threshold.

The lower-order omnibus statistics in Study 2 also stay reasonably close to the three nominal levels. Their empirical sizes range from 0.005 to 0.019 at $\alpha = 0.01$, from 0.040 to 0.068 at $\alpha = 0.05$, and from 0.087 to 0.118 at $\alpha = 0.10$. Their power increases with α in every condition. The Pearson–Fisher statistic remains too liberal at all three levels. Changing the nominal level does not correct its unreliable full-table reference distribution.

Table 14. Ranges across methods at the 1%, 5%, and 10% nominal levels. For Study 1, power is the mean over the 15 target pairs. For Study 2, the size range excludes the full-table Pearson–Fisher statistic. Exact method-specific results are given in the separate Supplementary File.

Study	Condition	N	Empirical Size			Empirical Power		
			$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
1	Zero intercept	300	0.005–0.011	0.039–0.052	0.089–0.102	0.451–0.672	0.646–0.858	0.735–0.917
1	Zero intercept	1000	0.006–0.011	0.038–0.051	0.086–0.101	0.860–0.998	0.928–1.000	0.951–1.000
1	Wide symmetric	300	0.007–0.014	0.039–0.052	0.086–0.099	0.045–0.062	0.102–0.119	0.157–0.176
1	Wide symmetric	1000	0.007–0.012	0.039–0.050	0.087–0.099	0.129–0.185	0.256–0.353	0.351–0.481
2	(W, W)	300	0.009–0.010	0.043–0.047	0.095–0.103	0.214–0.528	0.446–0.748	0.591–0.829
2	(W, W)	1000	0.014–0.019	0.062–0.068	0.115–0.118	0.958–1.000	0.992–1.000	0.997–1.000
2	(W, S)	300	0.009–0.018	0.051–0.059	0.101–0.106	0.218–0.601	0.421–0.783	0.562–0.864
2	(W, S)	1000	0.010–0.014	0.049–0.053	0.097–0.113	0.958–1.000	0.992–1.000	0.995–1.000
2	(S, S)	300	0.013–0.018	0.050–0.058	0.087–0.114	0.085–0.397	0.227–0.625	0.332–0.734
2	(S, S)	1000	0.005–0.019	0.040–0.065	0.101–0.117	0.537–0.997	0.869–0.998	0.963–0.999

5. Discussion

This study applies weighted sequential sums of squares to pairwise concordance margins. The resulting components give an ordered local decomposition of X_{2c}^2 and retain a direct link to the overall concordance discrepancy. Their Type I error rates remain close to the selected 1%, 5%, and 10% levels under both intercept settings.

An SS pair label is best interpreted in relation to the chosen order. Reordering leaves the total unchanged but can move a large incremental component to another label. A large SS value shows that the pair entering at that step adds discrepancy after the earlier directions have been removed. Evidence about a specific raw pair is stronger when it is also supported by a direct pairwise statistic or remains prominent under another pre-specified order.

The target-pair results show that structural misfit and observable local information are not the same. Large slopes on the second factor define the intended source of misfit, but a wide intercept range can make that source difficult to observe. Pairs with similar intercepts retain power, while pairs with a large intercept difference may have almost none. LM is particularly sensitive to this separation. It keeps the stated Type I error level, but its power can approach the nominal level when the two intercepts are far apart. This finding agrees with the relation reported by Xu et al. [30] and extends it to a wider symmetric pattern.

The non-target results show why the decomposition also matters. SS and $GFit_{\perp}^{(ij)}$ keep most non-target rates near 0.05, but the fixed pair order can place a shared signal in a few non-target components. Cholesky produces smaller increases across a broader set of non-target pairs under zero intercepts. Neither result gives a simple ranking of the two decompositions. SS gives a more concentrated pattern with somewhat lower target power, while Cholesky gives higher target power with a broader non-target increase. The contrast becomes clearer at $N = 1000$ and changes with the intercept pattern.

In the reordering analysis, the peaks near item 12 moved to other labels under the reverse and fixed random orders. This movement explains why a sequential label should be read together with the pair order. Cholesky rejection rates changed less across the three orders examined, although its components also depend on the order. Comparisons of localization in this study therefore refer to the stated orders and parameter settings.

The omnibus study supports lower-order assessment in the sparse full table. At the 5% level, the five marginal statistics keep Type I error close to 0.05. Four quadratic statistics, including X_{2c}^2 and $X_{[2]}^2$, have similar power against the symmetric slope alternatives. The moment adjustment makes Y_2 behave differently under the strongest slope setting. The full-table Pearson–Fisher statistic has severe Type I error inflation and was therefore excluded from the power comparison. A lower-order omnibus test can be used before the local components are examined [8,12,42].

The three nominal levels provide a separate check on these conclusions. The local and lower-order omnibus statistics generally follow the requested size as α changes. Power increases, but the main differences between the parameter settings remain. The wide symmetric intercept condition has weak target-pair power even at the 10% level. This result is not simply caused by the 5% cutoff. It reflects limited pairwise information when the intercepts are far apart. The nominal level changes the rejection threshold, but it cannot create a signal that is almost absent from the observed margin.

The multiple-testing analysis concerns the decision to report a list of pairs. Unadjusted results are useful for exploring the pattern of local discrepancy, but at least one rejection occurred in almost every null replication when 105 components were examined. Holm and Benjamini–Hochberg adjustments reduced selections outside the target set and also reduced target sensitivity. A confirmatory analysis should therefore report how the 105 p -values were handled.

This study has several limitations. Study 1 uses one item count, two sample sizes, two intercept patterns, and three nominal levels. It also uses only three data-independent pair orders in the added sensitivity analysis. The item parameters are estimated conditional on the generated first-factor values, so uncertainty from estimating the latent factor is not included. A full marginal maximum likelihood analysis is needed to examine the effect of latent-trait estimation. The choices $q = 15$, $N = 300$, and $N = 1000$ support comparison with previous studies but do not cover the full range of IRT applications. Study 2 uses only the selected weak and strong slope vectors. The results therefore describe the present simulation settings and should not be treated as general properties of every binary IRT model.

Extreme intercepts create an additional limitation. Fitted bivariate probabilities close to zero or one produce nearly degenerate 2×2 margins. They can reduce the information available for detecting misspecification and may also impair the accuracy of small p -values obtained from the asymptotic chi-square approximation. Componentwise rejection rates close to 0.05 do not guarantee that the much smaller tail probabilities needed for simultaneous testing are equally accurate.

Further work should repeat the local study with more sample sizes, including smaller samples, other intercept ranges, and a full latent-variable fit. A denser grid of nominal levels could be used to compare complete power curves rather than three cutoffs. The broad non-target increase found for Cholesky under zero intercepts at $N = 1000$ should be examined to determine when it occurs and whether it continues as the sample size grows. Many more pre-specified orders, including orders based on substantive item groupings, are needed before any ordering rule can be recommended. Calibrated resampling may also be useful when extreme margins make asymptotic tail probabilities unreliable. Other symmetric and asymmetric slope vectors can be added to the omnibus study to determine how far its results extend beyond the three conditions considered here.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/sym18091520/s1>, Supplementary File, Tables S1–S31: nominal-level, pair-order, multiplicity, and localization results; Data S1: complete pair-specific order-sensitivity results; Data S2: order and multiplicity summaries; Data S3: pair definitions and the three fixed orders.

Author Contributions: Conceptualization, J.X.; methodology, J.X. and M.R.; software, R.L.; validation, R.L. and J.X.; formal analysis, R.L.; investigation, R.L.; data curation, R.L.; writing—original draft preparation, R.L.; writing—review and editing, R.L., J.X. and M.R.; supervision, J.X.; funding acquisition, J.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Innovation and Entrepreneurship Training Program for Undergraduates of Guangdong Province, grant numbers S202610559078 and S202610559074, and the Fundamental Research Funds for the Central Universities, grant number 21624350.

Data Availability Statement: The simulation code and generated results are available from the corresponding author upon reasonable request. Complete machine-readable results for the added order and multiplicity analyses are supplied with the Supplementary File.

Acknowledgments: During the preparation of this manuscript, the authors used OpenAI Codex (GPT-5) to optimize English expression, correct grammatical errors, improve readability, and assist with LaTeX formatting. The authors reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Pearson, K. On the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling. *Philos. Mag.* **1900**, *50*, 157–175. [\[CrossRef\]](#)
2. Fisher, R.A. On a Property Connecting the χ^2 Measure of Discrepancy with the Method of Maximum Likelihood. In *Atti del Congresso Internazionale dei Matematici: Bologna 1928*; Nicola Zanichelli: Bologna, Italy, 1929; Volume 6, pp. 95–100.
3. Cochran, W.G. The χ^2 Test of Goodness of Fit. *Ann. Math. Stat.* **1952**, *23*, 315–345. [\[CrossRef\]](#)
4. Koehler, K.J.; Larntz, K. An Empirical Investigation of Goodness-of-Fit Statistics for Sparse Multinomials. *J. Am. Stat. Assoc.* **1980**, *75*, 336–344. [\[CrossRef\]](#)
5. Koehler, K.J. Goodness-of-Fit Tests for Log-Linear Models in Sparse Contingency Tables. *J. Am. Stat. Assoc.* **1986**, *81*, 483–493. [\[CrossRef\]](#)
6. Agresti, A.; Yang, M.C. An Empirical Investigation of Some Effects of Sparseness in Contingency Tables. *Comput. Stat. Data Anal.* **1987**, *5*, 9–21. [\[CrossRef\]](#)
7. Cressie, N.; Read, T.R.C. Multinomial Goodness-of-Fit Tests. *J. R. Stat. Soc. Ser. (Methodol.)* **1984**, *46*, 440–464. [\[CrossRef\]](#)
8. Reiser, M. Goodness-of-Fit Testing in Sparse Contingency Tables When the Number of Variables Is Large. *Wires Comput. Stat.* **2019**, *11*, e1470. [\[CrossRef\]](#)
9. Christofferson, A. Factor Analysis of Dichotomized Variables. *Psychometrika* **1975**, *40*, 5–32. [\[CrossRef\]](#)
10. Muthén, B. Contributions to Factor Analysis of Dichotomous Variables. *Psychometrika* **1978**, *43*, 551–560. [\[CrossRef\]](#)
11. Bartholomew, D.J.; Leung, S.O. A Goodness of Fit Test for Sparse 2^p Contingency Tables. *Br. J. Math. Stat. Psychol.* **2002**, *55*, 1–15. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Maydeu-Olivares, A.; Joe, H. Limited- and Full-Information Estimation and Goodness-of-Fit Testing in 2^n Contingency Tables: A Unified Framework. *J. Am. Stat. Assoc.* **2005**, *100*, 1009–1020. [\[CrossRef\]](#)
13. Maydeu-Olivares, A.; Joe, H. Limited Information Goodness-of-Fit Testing in Multidimensional Contingency Tables. *Psychometrika* **2006**, *71*, 713–732. [\[CrossRef\]](#)
14. Cai, L.; Maydeu-Olivares, A.; Coffman, D.L.; Thissen, D. Limited-Information Goodness-of-Fit Testing of Item Response Theory Models for Sparse 2^p Tables. *Br. J. Math. Stat. Psychol.* **2006**, *59*, 173–194. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Tollenaar, N.; Mooijaart, A. Type I Errors and Power of the Parametric Bootstrap Goodness-of-Fit Test: Full and Limited Information. *Br. J. Math. Stat. Psychol.* **2003**, *56*, 271–288. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Maydeu-Olivares, A. Goodness-of-Fit Assessment of Item Response Theory Models. *Meas. Interdiscip. Res. Perspect.* **2013**, *11*, 71–101. [\[CrossRef\]](#)
17. Maydeu-Olivares, A.; Liu, Y. Item Diagnostics in Multivariate Discrete Data. *Psychol. Methods* **2015**, *20*, 276–292. [\[CrossRef\]](#) [\[PubMed\]](#)

18. Sinharay, S.; Monroe, S. Assessment of Fit of Item Response Theory Models: A Critical Review of the Status Quo and Some Future Directions. *Br. J. Math. Stat. Psychol.* **2025**, *78*, e12378. [[CrossRef](#)] [[PubMed](#)]
19. Chen, W.H.; Thissen, D. Local Dependence Indexes for Item Pairs Using Item Response Theory. *J. Educ. Behav. Stat.* **1997**, *22*, 265–289. [[CrossRef](#)]
20. Orlando, M.; Thissen, D. Likelihood-Based Item-Fit Indices for Dichotomous Item Response Theory Models. *Appl. Psychol. Meas.* **2000**, *24*, 50–64. [[CrossRef](#)]
21. Haberman, S.J. Use of Generalized Residuals to Examine Goodness of Fit of Item Response Models. *Ets Res. Rep. Ser.* **2009**, 2009. [[CrossRef](#)]
22. Liu, Y.; Maydeu-Olivares, A. Identifying the Source of Misfit in Item Response Theory Models. *Multivar. Behav. Res.* **2014**, *49*, 354–371. [[CrossRef](#)] [[PubMed](#)]
23. Jamil, H.; Moustaki, I.; Skinner, C. Pairwise Likelihood Estimation and Limited-Information Goodness-of-Fit Test Statistics for Binary Factor Analysis Models under Complex Survey Sampling. *Br. J. Math. Stat. Psychol.* **2025**, *78*, 258–285. [[CrossRef](#)] [[PubMed](#)]
24. Reiser, M. Goodness-of-Fit Testing Using Components Based on Marginal Frequencies of Multinomial Data. *Br. J. Math. Stat. Psychol.* **2008**, *61*, 331–360. [[CrossRef](#)] [[PubMed](#)]
25. Milovanovic, J. Chi-Square Orthogonal Components for Assessing Goodness-of-Fit of Multidimensional Multinomial Data. Ph.D. Thesis, Arizona State University, Tempe, AZ, USA, 2011.
26. Dassanayake, M.M.K. A Study of Components of Pearson's Chi-Square Based on Marginal Distributions of Cross-Classified Tables for Binary Variables. Ph.D. Thesis, Arizona State University, Tempe, AZ, USA, 2018.
27. Reiser, M.; Cagnone, S.; Zhu, J. An Extended Gfit Statistic Defined on Orthogonal Components of Pearson's Chi-Square. *Psychometrika* **2023**, *88*, 208–240. [[CrossRef](#)] [[PubMed](#)]
28. Xu, J. Advances in Directional Goodness-of-Fit Testing of Binary Data under Model Misspecification in Case of Sparseness. Ph.D. Thesis, Arizona State University, Tempe, AZ, USA, 2022.
29. Xu, J.; Xu, X.; Reiser, M. Concordant Marginal Information-Based Goodness-of-Fit Testing for Binary Longitudinal Data in Generalized Linear Mixed Models. *Stat. Pap.* **2026**, *67*, 81. [[CrossRef](#)]
30. Xu, J.; Chen, J.; Reiser, M. An Empirical Power Comparison of Directional Goodness-of-Fit Tests for 2-PL IRT Model under Different Intercept Patterns. *J. Appl. Stat.* **2026**, *53*, 2203–2250. [[CrossRef](#)] [[PubMed](#)]
31. Savalei, V.; Rhemtulla, M. The Performance of Robust Test Statistics with Categorical Data. *Br. J. Math. Stat. Psychol.* **2013**, *66*, 201–223. [[CrossRef](#)] [[PubMed](#)]
32. Lord, F.M. *Applications of Item Response Theory to Practical Testing Problems*; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1980.
33. Reckase, M.D. *Multidimensional Item Response Theory*; Springer: New York, NY, USA, 2009. [[CrossRef](#)]
34. Moore, D.S. Generalized Inverses, Wald's Method, and the Construction of Chi-Squared Tests of Fit. *J. Am. Stat. Assoc.* **1977**, *72*, 131–137. [[CrossRef](#)]
35. Seber, G.A.F.; Lee, A.J. *Linear Regression Analysis*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2003.
36. Golub, G.H.; Van Loan, C.F. *Matrix Computations*, 4th ed.; Johns Hopkins University Press: Baltimore, MD, USA, 2013.
37. Kessy, A.; Lewin, A.; Strimmer, K. Optimal Whitening and Decorrelation. *Am. Stat.* **2018**, *72*, 309–314. [[CrossRef](#)]
38. Rao, C.R. Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications to Problems of Estimation. *Proc. Camb. Philos. Soc.* **1948**, *44*, 50–57. [[CrossRef](#)]
39. Silvey, S.D. The Lagrangian Multiplier Test. *Ann. Math. Stat.* **1959**, *30*, 389–407. [[CrossRef](#)]
40. Glas, C.A.W. Modification Indices for the 2-PL and the Nominal Response Model. *Psychometrika* **1999**, *64*, 273–294. [[CrossRef](#)]
41. Jöreskog, K.G.; Moustaki, I. Factor Analysis of Ordinal Variables: A Comparison of Three Approaches. *Multivar. Behav. Res.* **2001**, *36*, 347–387. [[CrossRef](#)] [[PubMed](#)]
42. Reiser, M.; Wang, L.; Khogeer, H.A. Power Comparisons for Testing Goodness of Fit for Constrained Models on High-Dimensional Cross-Classified Tables. *J. Appl. Stat.* **2025**, 1–18. [[CrossRef](#)]
43. Bock, R.D.; Aitkin, M. Marginal Maximum Likelihood Estimation of Item Parameters: Application of an EM Algorithm. *Psychometrika* **1981**, *46*, 443–459. [[CrossRef](#)]
44. Holm, S. A Simple Sequentially Rejective Multiple Test Procedure. *Scand. J. Stat.* **1979**, *6*, 65–70.
45. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. (Methodol.)* **1995**, *57*, 289–300. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.