Web Appendices for "Tests of matrix structure for construct validation"

A. Rate of convergence

Existing results for the large sample behavior of permutation tests focus on the relationship between the conditional permutation distribution of a statistic and the unconditional limiting distribution as the number of observations increases (e.g. see Lehmann and Romano, 2005, Section 15.2.2). In particular, let $T(x_1, \ldots, x_n)$ be a test statistic of the n observations $x_1, \ldots, x_n$. Also, let $\hat{R}_n(t)$ be the permutation distribution of T, and let $R(t)$ be the unconditional asymptotic distribution of T. Then most existing results study the scenario in which $\hat{R}_n \to R(t)$ as $n \to \infty$, with the goal of understanding the large sample properties of the permutation test, such as power.

In this appendix, we address a related but different question. In our setup, we need to account for: 1) measurement error and 2) fixed number of inputs to the test statistic. Let $a_j^n = \rho_j + u_j^n$, where ρ_j is the true population quantity, a_j^n is our estimate of ρ_j from n observations, and u_j^n is measurement error, which is a function of the number of respondents n (throughout this appendix, we use superscript n to denote sample size). In our proposed method, we use a statistic of the form $T(\rho_1 + u_1^n, \ldots, \rho_N + u_N^n)$, where the number of correlations $N = p(p-1)/2$ is fixed by the questionnaire, which contains p items. In our setting, instead of letting $N \to \infty$, N is constant and we let $n \to \infty$. Assuming a_j^n are consistent estimators of ρ_j, $u_j^n \to 0$ as $n \to \infty$. Our goal is to understand the rate at which the p-value with the estimated quantities a_j^n converges to the p-value that would be obtained with the true quantities ρ_j.

Similar to before, we denote the $N \times 1$ vector of upper triangular elements of A as $\boldsymbol{a}^n = (a_1^n, a_2^n, \ldots, a_N^n)^T$. Let π be a permutation, or bijection, of the columns and rows of A, let Π be the set of all such permutations π, and let $|\Pi| = p!$ be the total number of permutations in Π. Let A_π be matrix A with the rows and columns permuted according to

π, and let $\boldsymbol{a}_\pi^n$ be the $N \times 1$ vector of upper triangular elements of A_π. Let $\Gamma_{\text{norm}}(\boldsymbol{a}_\pi^n)$ be Hubert's Γ computed with $\boldsymbol{a}_\pi^n$, and let $\boldsymbol{a}_0^n$ be the vector of correlation coefficients under the hypothesized ordering.

In data analyses, we use Monte Carlo methods to approximate the permutation p-value obtained with the estimated quantities $\boldsymbol{a}^n$. We denote the two-sided permutation p-value with the estimated quantities as $\hat{p}(\boldsymbol{a}^n) = |\Pi|^{-1} \sum_{\pi \in \Pi} \mathbb{1}\left[|\Gamma_{\text{norm}}(\boldsymbol{a}_\pi^n)| \geq |\Gamma_{\text{norm}}(\boldsymbol{a}_0^n)|\right]$. However, we would ideally approximate the permutation p-value obtained with the true population quantities, which we denote as $\hat{p}(\boldsymbol{\rho}) = |\Pi|^{-1} \sum_{\pi \in \Pi} \mathbb{1}\left[|\Gamma_{\text{norm}}(\boldsymbol{\rho}_\pi)| \geq |\Gamma_{\text{norm}}(\boldsymbol{\rho}_0)|\right]$. Fortunately, under general conditions specified in Theorem 1, if $|a_j^n - \rho_j| = O_p(g(n))$ for $j = 1, \ldots, N$, then we also have $|\hat{p}(\boldsymbol{a}^n) - \hat{p}(\boldsymbol{\rho})| = O_p(g(n))$. In other words, the rate of convergence for the permutation p-value is the same as the rate of convergence of the underlying elements of $\boldsymbol{a}^n$. As shown in Corollary 1, when $\boldsymbol{a}^n$ are Pearson's or Spearman's correlation coefficients, we have $g(n) = O(1/\sqrt{n})$. As shown in Corollary 2, the same rate of convergence holds when using the absolute values of Pearson's or Spearman's correlation coefficients.

As shown in Section 4.1, $\Gamma_{\text{norm}}(\boldsymbol{a}^n) = (\hat{\sigma}_\delta / \hat{\sigma}_{a^n})(\bar{a}_{\text{in}}^n - \bar{a}_{\text{out}}^n)$, where $\bar{a}_{\text{in}}^n$ is the mean of the within-block elements and $\bar{a}_{\text{out}}^n$ is the mean of the between-block elements. Because $\hat{\sigma}_\delta$ and $\hat{\sigma}_{a^n}$ are constant conditional on the data, this shows that $\Gamma_{\text{norm}}(\boldsymbol{a}^n)$ is permutationally equivalent to the difference in means, which we denote by $D(\boldsymbol{a}^n) = \bar{a}_{\text{in}}^n - \bar{a}_{\text{out}}^n$. Similarly, we denote the difference in means of the true population quantities as $D(\boldsymbol{\rho}) = \bar{\rho}_{\text{in}} - \bar{\rho}_{\text{out}}$.

In this appendix, we work with D instead of Γ_{norm} because the former simplifies the derivations. Since D and Γ_{norm} are permutationally equivalent, they produce identical permutation p-values. Consequently, the convergence rate of the permutation p-value must be the same for D as for Γ_{norm}.

Before focusing on our primary interest, $|\hat{p}(\boldsymbol{a}^n) - \hat{p}(\boldsymbol{\rho})|$, we state an inequality in Lemma 1 that we will use to prove our main result in Theorem 1.

Lemma 1. Let $\epsilon_j(n,\delta)$ be a decreasing, strictly positive function of n for all $\delta \in (0,1)$ such that: i) $\epsilon_j(n,\delta) = O(g(n))$, and ii) for all δ, there exist an $n_\delta \in \mathbb{N}$ such that $\Pr\{|a_j^n - \rho_j| \leq \epsilon_j(n,\delta)\} \geq 1-\delta$ for $n > n_\delta$, $j = 1,\ldots,N$. Then $\Pr\{|D(\boldsymbol{a}^n) - D(\boldsymbol{\rho})| \leq 2\epsilon_{\max}(n,\delta)\} \geq h(\delta)$ for $n > n_\delta$ where $\epsilon_{\max}(n,\delta) = \max_j \epsilon_j(n,\delta)$ and $h(\delta) = \Pr(\cap_j\{|a_j^n - \rho_j| \leq \epsilon_j(n,\delta)\})$. If we also have $(a_i^n - \rho_i) \perp (a_j^n - \rho_j)$ for $i \neq j$, then $h(\delta) = (1-\delta)^N$.

Proof of Lemma 1. Let $\mathcal{J}_{\text{in}}$ and $\mathcal{J}_{\text{out}}$ be the sets of indices of within-block and between-block elements, respectively, of $\boldsymbol{a}^n$. Also, let $N_{\text{in}} = |\mathcal{J}_{\text{in}}|$ and $N_{\text{out}} = |\mathcal{J}_{\text{out}}|$ be the number of within-block and between-block elements. Then for $n > n_\delta$ and using w.p. as shorthand for "with probability,"

$$
\begin{aligned}
\underbrace{|D(\boldsymbol{a}^n) - D(\boldsymbol{\rho})|}_{E_1} &= |\bar{a}_{\text{in}}^n - \bar{a}_{\text{out}}^n - (\bar{\rho}_{\text{in}} - \bar{\rho}_{\text{out}})| \\
&= |\bar{a}_{\text{in}}^n - \bar{\rho}_{\text{in}} + \bar{\rho}_{\text{out}} - \bar{a}_{\text{out}}^n| \\
&\leq |\bar{a}_{\text{in}}^n - \bar{\rho}_{\text{in}}| + |\bar{\rho}_{\text{out}} - \bar{a}_{\text{out}}^n| \\
&= \frac{1}{N_{\text{in}}}\left|\sum_{j\in\mathcal{J}_{\text{in}}}(a_j^n - \rho_j)\right| + \frac{1}{N_{\text{out}}}\left|\sum_{j\in\mathcal{J}_{\text{out}}}(a_j^n - \rho_j)\right| \\
&\leq \underbrace{\frac{1}{N_{\text{in}}}\sum_{j\in\mathcal{J}_{\text{in}}}\left|a_j^n - \rho_j\right| + \frac{1}{N_{\text{out}}}\sum_{j\in\mathcal{J}_{\text{out}}}\left|a_j^n - \rho_j\right|}_{E_2} \\
&\leq \underbrace{\frac{1}{N_{\text{in}}}\sum_{j\in\mathcal{J}_{\text{in}}}\epsilon_j(n,\delta) + \frac{1}{N_{\text{out}}}\sum_{j\in\mathcal{J}_{\text{out}}}\epsilon_j(n,\delta)}_{E_3} \qquad \text{(w.p. at least } h(\delta)) \qquad (1) \\
&\leq \underbrace{\frac{1}{N_{\text{in}}}N_{\text{in}}\epsilon_{\max}(n,\delta) + \frac{1}{N_{\text{out}}}N_{\text{out}}\epsilon_{\max}(n,\delta)}_{E_4} \\
&= 2\epsilon_{\max}(n,\delta).
\end{aligned}
$$

To see why the inequality in (1) holds with probability at least $h(\delta)$ (as opposed to an exact equality), note that $E_2 \leq E_3$ if $|a_j^n - \rho_j| \leq \epsilon_j(n)$ for all j. However, this is a subset of the conditions under which $E_2 \leq E_3$ (e.g. we could have $|a_j^n - \rho_j| > \epsilon_j(n)$ for some j, which are offset by $|a_j^n - \rho_j| < \epsilon_j(n)$ for other j). Consequently the inequality holds with probability at least $h(\delta)$.

Now, we can write $\Pr(E_1 \leq E_4) = \Pr(E_1 \leq E_2, E_2 \leq E_3, E_3 \leq E_4)$. Furthermore, the events $\{E_1 \leq E_2\}$ and $\{E_3 \leq E_4\}$ are deterministic. Consequently, $\Pr(E_1 \leq E_2) = 1$, $\Pr(E_3 \leq E_4) = 1$, and $\{E_1 \leq E_2\}$, $\{E_2 \leq E_3\}$, and $\{E_3 \leq E_4\}$ are mutually independent. It follows that

$$\begin{aligned}
\Pr(E_1 \leq E_4) &= \Pr(E_1 \leq E_2, E_2 \leq E_3, E_3 \leq E_4) \\
&= \Pr(E_1 \leq E_2) \Pr(E_2 \leq E_3) \Pr(E_3 \leq E_4) \\
&= \Pr(E_2 \leq E_3) \\
&\geq h(\delta).
\end{aligned}$$

This shows that

$$\Pr\{|D(\boldsymbol{a}^n) - D(\boldsymbol{\rho})| \leq 2\epsilon_{\max}(n, \delta)\} \geq h(\delta). \tag{2}$$

Furthermore, we have $\epsilon_{\max}(n, \delta) = O(g(n))$. Therefore, (2) implies that $|D(\boldsymbol{a}^n) - D(\boldsymbol{\rho})| = O_p(g(n))$. Finally, we note that if the errors are independent, then we have

$$\begin{aligned}
h(\delta) &= \Pr\left(\cap_{j=1}^N \{|a_j^n - \rho_j| \leq \epsilon_j(n)\}\right) \\
&= \prod_{j=1}^N \Pr\left(|a_j^n - \rho_j| \leq \epsilon_j(n)\right) \\
&= (1 - \delta)^N.
\end{aligned}$$

This proves the lemma. $\square$

We now turn to our primary interest, $|\hat{p}(\boldsymbol{a}^n) - \hat{p}(\boldsymbol{\rho})|$. To that end, for fixed $\epsilon > 0$, let

$B_\epsilon = (|D(\boldsymbol{\rho}_0)| - \epsilon, |D(\boldsymbol{\rho}_0)| + \epsilon)$ be the ϵ-ball centered around $|D(\boldsymbol{\rho}_0)|$. Also, let

$$\Pi_B(\epsilon) = \{\pi \in \Pi : |D(\boldsymbol{\rho}_\pi)| \in B_\epsilon\}$$
$$\Pi_{\bar{B}}(\epsilon) = \{\pi \in \Pi : |D(\boldsymbol{\rho}_\pi)| \notin B_\epsilon\}.$$

Note that for each ϵ, $\Pi_B(\epsilon)$ and $\Pi_{\bar{B}}(\epsilon)$ partition Π, i.e. $\Pi = \Pi_B(\epsilon) \cup \Pi_{\bar{B}}(\epsilon)$ and $\Pi_B(\epsilon) \cap \Pi_{\bar{B}}(\epsilon) = \emptyset$.

For fixed ϵ we have

$$\begin{aligned}
|\Pi|\,|\hat{p}(\boldsymbol{a}^n) - \hat{p}(\boldsymbol{\rho})| &= \left| \sum_{\pi\in\Pi} \mathbb{1}\left(|D(\boldsymbol{a}^n_\pi)| \geq |D(\boldsymbol{a}^n_0)|\right) - \sum_{\pi\in\Pi} \mathbb{1}\left(|D(\boldsymbol{\rho}_\pi)| \geq |D(\boldsymbol{\rho}_0)|\right) \right| \\
&= \left| \sum_{\pi\in\Pi} \left\{ \mathbb{1}\left(|D(\boldsymbol{a}^n_\pi)| \geq |D(\boldsymbol{a}^n_0)|\right) - \mathbb{1}\left(|D(\boldsymbol{\rho}_\pi)| \geq |D(\boldsymbol{\rho}_0)|\right) \right\} \right| \\
&\leq \underbrace{\left| \sum_{\pi\in\Pi_B(2\epsilon)} \left\{ \mathbb{1}\left(|D(\boldsymbol{a}^n_\pi)| \geq |D(\boldsymbol{a}^n_0)|\right) - \mathbb{1}\left(|D(\boldsymbol{\rho}_\pi)| \geq |D(\boldsymbol{\rho}_0)|\right) \right\} \right|}_{C_B} \qquad (3) \\
&\quad + \underbrace{\left| \sum_{\pi\in\Pi_{\bar{B}}(2\epsilon)} \left\{ \mathbb{1}\left(|D(\boldsymbol{a}^n_\pi)| \geq |D(\boldsymbol{a}^n_0)|\right) - \mathbb{1}\left(|D(\boldsymbol{\rho}_\pi)| \geq |D(\boldsymbol{\rho}_0)|\right) \right\} \right|}_{C_{\bar{B}}}. \qquad (4)
\end{aligned}$$

We proceed by bounding C_B (3) in Lemma 2 and $C_{\bar{B}}$ (4) in Lemma 3. We then combine these bounds with the ϵ given by Lemma 1 to prove our main result in Theorem 1. In the rest of this appendix, we use the notation $C_B = C_B(n)$ and $C_{\bar{B}} = C_{\bar{B}}(n)$ to explicitly write these quantities as functions of the sample size n.

Lemma 2. Let $\hat{R}_N(t)$ be the permutation distribution of $|D(\boldsymbol{\rho})|$. Suppose $\hat{R}_N(t) \approx R(t)$ for N sufficiently large, where $R(t)$ has density $f(t)$ such that $M = \sup_t f(t) < \infty$. Also, in (3) and (4) let $\epsilon = \epsilon(n)$ be a function of n and suppose $\epsilon(n) = O(g(n))$ for some strictly decreasing function g such that $g(n) \to 0$ as $n \to \infty$. Then for N sufficiently large, $C_B(n) = O_p(g(n))$.

Proof of Lemma 2. In the following, we use the convention that $f(t) = 0$ for $t \notin \text{supp}(f)$, where $\text{supp}(f)$ is the support of f. For fixed n, we have

$$\frac{C_B(n)}{|\Pi|} \leq \frac{|\Pi_B(2\epsilon(n))|}{|\Pi|} \tag{5}$$

$$= \hat{R}_N(|D(\boldsymbol{\rho}_0)| + 2\epsilon(n)) - \hat{R}_N(|D(\boldsymbol{\rho}_0)| - 2\epsilon(n)) \tag{6}$$

$$\approx R(|D(\boldsymbol{\rho}_0)| + 2\epsilon(n)) - R(|D(\boldsymbol{\rho}_0)| - 2\epsilon(n)) \qquad \text{(for large } N\text{)}$$

$$= \int_{|D(\boldsymbol{\rho}_0)|-2\epsilon(n)}^{|D(\boldsymbol{\rho}_0)|+2\epsilon(n)} f(s)ds$$

$$\leq 4M\epsilon(n). \tag{7}$$

Line (5) follows because each term in the sum of $C_B(n)$ inside the absolute values is equal to $-1, 0$, or 1, and there are $|\Pi_B(\epsilon(n))|$ terms in the sum; the inequality in (5) would be an equality if and only if all terms in the sum were equal to 1 or if all terms in the sum were equal to -1. Line (6) follows because line (5) is just the proportion of the permutation distribution between $|D(\boldsymbol{\rho}_0)| - 2\epsilon(n)$ and $|D(\boldsymbol{\rho}_0)| + 2\epsilon(n)$.

By assumption, $\epsilon(n) = O(g(n))$. Furthermore, $|\Pi|$ is constant so $4M|\Pi|\epsilon(n) = O(g(n))$. Now, $C_B(n)$ is a random variable and the preceding argument shows that $C_B(n) \leq 4M|\Pi|\epsilon(n)$ with probability one (as noted above, we must have $C_B(n) \leq |\Pi_B(2\epsilon(n))|$). It follows that for any $\lambda \in (0, 1)$, there exists an $n_\lambda \in \mathbb{N}$ and $\alpha_\lambda \in (0, 1)$ such that $\Pr\{C_B(n) \leq \alpha_\lambda 4M|\Pi|\epsilon(n)\} \geq 1 - \lambda$ for all $n > n_\lambda$. This shows that $C_B(n) = O_p(g(n))$, which proves the lemma. □

We note that the constraint on the limiting distribution R in Lemma 2 precludes distributions that concentrate on sets of measure zero, such as the dirac delta function. In other words, the limiting distribution cannot be degenerate. We also note that in Lemma 2, we could set $\epsilon(n) = 2\epsilon_{\max}(n, \delta)$ for fixed $\delta \in (0, 1)$, where $\epsilon_{\max}(n, \delta)$ is given in Lemma 1. In this case, (7) becomes $8M\epsilon_{\max}(n, \delta)$.

The proof of Lemma 2 assumes that $N = p(p - 1)/2$ is sufficiently large for the approximation $\hat{R}_N(t) \approx R(t)$ to hold, i.e. that the matrix A has many elements. In practice, N is determined by the number of items p on the questionnaire. Furthermore, since the total number of permutations $N!$ grows very quickly, we anticipate that $p > 10$ ($N > 45$) is sufficient in most applications for the permutation distribution to be approximated well by a limiting distribution for which the density exists and is bounded above. The bound on $C_B(n)$ is then a function of the number of subjects n who reply to the questionnaire.

We now turn to the $C_{\bar{B}}(n)$ term (4).

Lemma 3. Suppose that $|D(\boldsymbol{a}_\pi^n) - D(\boldsymbol{\rho}_\pi)| = O_p(g(n))$ for all permutations $\pi \in \Pi$ for some strictly decreasing, positive function $g(n)$. In particular, suppose that for all $\delta \in (0,1)$, there exists an $n_\delta \in \mathbb{N}$ and $\epsilon(n,\delta) > 0$ such that $\Pr\{|D(\boldsymbol{a}_\pi^n) - D(\boldsymbol{\rho}_\pi)| \leq \epsilon(n,\delta)\} \geq h(\delta)$ for all $n > n_\delta$ where $h(\delta) = \Pr(\cap_j \{|a_j^n - \rho_j| \leq \epsilon_j(n,\delta)\})$ and $\epsilon(n,\delta) = O(g(n))$. Then $C_{\bar{B}}(n) = O_p(1)$.

Proof of Lemma 3. We note that

$$\begin{aligned} C_{\bar{B}}(n) &= \left| \sum_{\pi \in \Pi_{\bar{B}}(2\epsilon(n,\delta))} \{\mathbb{1}\left(|D(\boldsymbol{a}_\pi^n)| \geq |D(\boldsymbol{a}_0^n)|\right) - \mathbb{1}\left(|D(\boldsymbol{\rho}_\pi)| \geq |D(\boldsymbol{\rho}_0)|\right)\} \right| \\ &\leq \sum_{\pi \in \Pi_{\bar{B}}(2\epsilon(n,\delta))} |\{\mathbb{1}\left(|D(\boldsymbol{a}_\pi^n)| \geq |D(\boldsymbol{a}_0^n)|\right) - \mathbb{1}\left(|D(\boldsymbol{\rho}_\pi)| \geq |D(\boldsymbol{\rho}_0)|\right)\}| \\ &= \sum_{\pi \in \Pi_{\bar{B}}(2\epsilon(n,\delta))} \underbrace{\mathbb{1}[\operatorname{sgn}(|D(\boldsymbol{a}_\pi^n)| - |D(\boldsymbol{a}_0^n)|) \neq \operatorname{sgn}(|D(\boldsymbol{\rho}_\pi)| - |D(\boldsymbol{\rho}_0)|)]}_{S(n,\pi)} \end{aligned} \tag{8}$$

where $\operatorname{sgn}(x) = 1$ if $x \geq 0$ and $\operatorname{sgn}(x) = -1$ otherwise. On a conceptual level, $C_{\bar{B}}(n)$ is bounded above by the sum of sign differences $S(n,\pi)$ in (8), where the sum is taken over all $\pi \in \Pi_{\bar{B}}(2\epsilon(n,\delta))$. Furthermore, $\Pi_{\bar{B}}(2\epsilon(n,\delta))$ is defined so that with high probability $S(n,\pi) = 0$ for each $n \in \Pi_{\bar{B}}(2\epsilon(n,\delta))$. This causes $C_{\bar{B}}(n)$ to be stochastically bounded with a constant rate of convergence, which is formalized below.

For fixed $\delta \in (0,1)$ and n, consider a permutation $\pi \in \Pi_{\bar{B}}(2\epsilon(n,\delta))$ and let $q_\pi = \big||D(\boldsymbol{\rho}_\pi)| - |D(\boldsymbol{\rho}_0)|\big| \geq 2\epsilon(n,\delta)$ be the distance between the observed and permuted test statistic computed with the true population values. Then for the term $S(n,\pi)$ in (8) we have

$$
\begin{aligned}
&\Pr\{S(n,\pi) = 0\} \\
&\quad = \Pr\{\operatorname{sgn}(|D(\boldsymbol{a}_\pi^n)| - |D(\boldsymbol{a}_0^n)|) = \operatorname{sgn}(|D(\boldsymbol{\rho}_\pi)| - |D(\boldsymbol{\rho}_0)|)\} \\
&\quad \geq \Pr\left\{\big||D(\boldsymbol{a}_0^n)| - |D(\boldsymbol{\rho}_0)|\big| \leq q_\pi/2, \big||D(\boldsymbol{a}_\pi^n)| - |D(\boldsymbol{\rho}_\pi)|\big| \leq q_\pi/2\right\} && (9) \\
&\quad \approx \Pr\left\{\big||D(\boldsymbol{a}_0^n)| - |D(\boldsymbol{\rho}_0)|\big| \leq q_\pi/2\right\} \Pr\left\{\big||D(\boldsymbol{a}_\pi^n)| - |D(\boldsymbol{\rho}_\pi)|\big| \leq q_\pi/2\right\} && (10) \\
&\quad \geq \Pr\{|D(\boldsymbol{a}_0^n) - D(\boldsymbol{\rho}_0)| \leq q_\pi/2\} \Pr\{|D(\boldsymbol{\rho}_\pi) - D(\boldsymbol{a}_\pi^n)| < q_\pi/2\} && (11) \\
&\quad \geq \Pr\{|D(\boldsymbol{a}_0^n) - D(\boldsymbol{\rho}_0)| \leq \epsilon(n,\delta)\} \Pr\{|D(\boldsymbol{\rho}_\pi) - D(\boldsymbol{a}_\pi^n)| < \epsilon(n,\delta)\} && (12) \\
&\quad \geq h(\delta)^2. && (13)
\end{aligned}
$$

Line (9) holds because the signs must be the same if $\big||D(\boldsymbol{a}_0^n)| - |D(\boldsymbol{\rho}_0)|\big| \leq q_\pi/2$ and $\big||D(\boldsymbol{a}_\pi^n)| - |D(\boldsymbol{\rho}_\pi)|\big| \leq q_\pi/2$, which can be verified by diagramming the quantities on a real number line. However, this is a subset of the conditions under which the signs must be the same, which gives the inequality. Line (10) follows from assuming the errors in the permuted test statistics are independent. Line (11) holds because: i) for $x, y \in \mathbb{R}$, $\big||x| - |y|\big| \leq |x - y|$, and ii) for random variable X, constant c, and function $f(\cdot)$, if $f(X) \geq X$ then $\Pr(X \leq c) \geq \Pr(f(X) \leq c)$. Line (12) holds because: i) $q_\pi \geq 2\epsilon(n,\delta)$, and ii) for random variable X and constants $c_1 \geq c_2$, $\Pr(X \leq c_1) \geq \Pr(X \leq c_2)$.

This shows that for all $\pi \in \Pi_{\bar{B}}(2\epsilon(n,\delta))$,

$$\Pr\{S(n,\pi) < 1\} \geq h(\delta)^2 \ \forall n. \tag{14}$$

Here, δ does not appear on the left-hand side of (14), though it does affect the rate at which the set $\Pi_{\bar{B}}(2\epsilon(n,\delta))$ grows. It follows that for all $\pi \in \Pi_{\bar{B}}(2\epsilon(n,\delta))$, $S(n,\pi) = O_p(1)$. Because $C_{\bar{B}}(n) \leq \sum_{\pi \in \Pi_{\bar{B}}(2\epsilon(n,\delta))} S(n,\pi)$ and the sum of $O_p(1)$ terms is also $O_p(1)$, it follows that $C_{\bar{B}}(n)$ is bounded above by an $O_p(1)$ term. Consequently, we must also have that $C_{\bar{B}}(n) = O_p(1)$, which proves the lemma. $\square$

We note that in Lemma 3, the rate of convergence is constant, i.e. $C_{\bar{B}}(n) = O_p(1)$. However, the number of permuted test statistics included in the sum of $C_{\bar{B}}(n)$ in (4) is controlled by the size of the $\Pi_{\bar{B}}(\epsilon(n,\delta))$, which we can grow at rate $\epsilon(n,\delta) = O(g(n))$. In particular, for each $n \in \mathbb{N}$, we can set $\epsilon(n,\delta) = 2\epsilon_{\max}(n,\delta)$, where $\epsilon_{\max}(n,\delta)$ is given in Lemma 1.

We now state our main result in Theorem 1 followed by Corollaries 1 and 2, which focus on the special case of Pearson's and Spearman's correlations.

Theorem 1. Let a_j^n be the sample estimates of ρ_j, $j = 1, \ldots, N$, and suppose that for all j, $|a_j^n - \rho_j| = O_p(g(n))$ for some strictly decreasing function g such that $g(n) \to 0$ as $n \to \infty$. Also suppose that the permutation distribution $\hat{R}_N(t)$ has limiting distribution $R(t)$ such that the density of $R(t)$, denoted as $f(t)$, exists and $\sup_t f(t) < \infty$. Then for N sufficiently large, $|\hat{p}(\boldsymbol{a}^n) - \hat{p}(\boldsymbol{\rho})| = O_p(g(n))$.

Proof of Theorem 1. From (3) and (4), we have $|\hat{p}(\boldsymbol{a}^n) - \hat{p}(\boldsymbol{\rho})| \leq |\Pi|^{-1}\left(C_B(n) + C_{\bar{B}}(n)\right)$. By assumption, for all $\delta \in (0,1)$ there exists an $n_\delta \in \mathbb{N}$ such that $\Pr\{|a_j^n - \rho_j| \leq \epsilon_j(n,\delta)\} \geq 1-\delta$ for all $n > n_\delta$, where $\epsilon_j(n,\delta) = O(g(n))$, $j = 1, \ldots, N$. Then by Lemma 1, $|D(\boldsymbol{a}^n) - D(\boldsymbol{\rho})| = O_p(g(n))$. In particular, Lemma 1 gives that $\Pr\{|D(\boldsymbol{a}^n) - D(\boldsymbol{\rho})| \leq 2\epsilon_{\max}(n,\delta)\} \geq h(\delta)$ where $h(\delta) = \Pr(\cap_j\{|a_j^n - \rho_j| \leq \epsilon_j(n)\})$ and $\epsilon_{\max}(n,\delta) = \max_j \epsilon_j(n,\delta)$. By setting $\epsilon = 2\epsilon_{\max}(n,\delta)$ in (3) and (4), we have $C_B(n) = O_p(g(n))$ by Lemma 2 and $C_{\bar{B}}(n) = O_p(1)$ by Lemma 3. It follows that $|\hat{p}(\boldsymbol{a}^n) - \hat{p}(\boldsymbol{\rho})| = O_p(g(n)) + O_p(1) = O_p(g(n))$, which proves the theorem. $\square$

Corollary 1. Let $\boldsymbol{a}^n$ be Pearson's or Spearman's correlation coefficients estimated from n independent and identically distributed (i.i.d.) observations. Let $\tau_j^2 = \mathrm{Var}(a_j^n)$ and assume $\tau_j^2 < \infty$ for $j = 1, \ldots, N$. Also suppose that the permutation distribution $\hat{R}_N(t)$ has limiting distribution $R(t)$ such that the density of $R(t)$, denoted as $f(t)$, exists and $\sup_t f(t) < \infty$. Then for N sufficiently large, $|\hat{p}(\boldsymbol{a}^n) - \hat{p}(\boldsymbol{\rho})| = O_p(1/\sqrt{n})$.

Proof of Corollary 1. Suppose that $\boldsymbol{a}^n$ are Pearson's correlation coefficients. Then under these assumptions and by the central limit theorem and delta method, $\sqrt{n}(a_j^n - \rho_j)$ is asymptotically normal for $j = 1, \ldots, N$ (Lehmann and Romano, 2005, p. 438). Then for n sufficiently large and finite $\epsilon > 0$,

$$\begin{aligned}
\Pr\left(|a_j^n - \rho_j| > \epsilon\right) &= \Pr\left(\frac{\sqrt{n}|a_j^n - \rho_j|}{\tau_j} > \frac{\sqrt{n}\epsilon}{\tau_j}\right) \\
&\approx \Pr\left(|Z| > \sqrt{n}\epsilon/\tau_j\right) \qquad\qquad (Z \sim N(0,1)) \\
&= 2\left[1 - \Phi\left(\frac{\sqrt{n}\epsilon}{\tau_j}\right)\right],
\end{aligned}$$

where Φ is the standard normal CDF. Setting $\delta = 2\left(1 - \Phi(\sqrt{n}\epsilon/\tau_j)\right)$ and solving for $\delta \in (0,1)$, we get that with probability $1 - \delta$,

$$|a_j^n - \rho_j| \le \tau_j \Phi^{-1}\left(1 - \delta/2\right)/\sqrt{n}. \tag{15}$$

This shows that $a_j^n = O_p(1/\sqrt{(}n))$, $j = 1, \ldots, N$. Then by Theorem 1, we have $|\hat{p}(\boldsymbol{a}^n) - \hat{p}(\boldsymbol{\rho})| = O_p(1/\sqrt{n})$. Because Spearman's correlation is Pearson's correlation of the ranks, the above argument carries over to Spearman's correlation. □

Corollary 2. Under the same conditions as Corollary 1, but with $\boldsymbol{a}^n$ and $\boldsymbol{\rho}$ replaced with absolute values of Pearson's or Spearman's correlations, we also have that $|\hat{p}(\boldsymbol{a}^n) - \hat{p}(\boldsymbol{\rho})| = O_p(1/\sqrt{n})$.

Proof of Corollary 2. Let $\boldsymbol{a}_{\text{abs}}^n$ and $\boldsymbol{\rho}_{\text{abs}}$ be $N \times 1$ vectors of the absolute values of the estimated correlation coefficients and the true correlations, respectively. We have that with probability at least $1 - \delta$, $|a_{\text{abs},j}^n - \rho_{\text{abs},j}| = \left||a_j^n| - |\rho_j|\right| \le |a_j^n - \rho_j| \le \tau_j \Phi^{-1}(1 - \delta/2)/\sqrt{n}$, where the last inequality follows from (15) in the proof of Corollary 1. Hence, $|a^n_{\text{abs},j} - \rho_{\text{abs},j}| = O_p(1/\sqrt{n})$, $j = 1, \dots, N$. Then by Theorem 1, $|\hat{p}(\boldsymbol{a}^n_{\text{abs}}) - \hat{p}(\boldsymbol{\rho}_{\text{abs}})| = O_p(1/\sqrt{n})$. □

We believe that the regularity conditions in these proofs are sufficiently general to be applicable to most data encountered in practice. However, in future work, we plan to investigate alternative proofs that relax the constraint that $\hat{R}_N(t)$ has a limiting distribution $R(t)$. We also plan to extend these results to the block-specific tests, and provide corollaries for other common correlations.

The derivations in this appendix apply to the true permutation p-value $\hat{p}(\boldsymbol{a}^n)$. However, in practice $\hat{p}(\boldsymbol{a}^n)$ is typically approximated with Monte Carlo methods, denoted as $\tilde{p}(\boldsymbol{a}^n)$. With simple Monte Carlo, $\tilde{p}(\boldsymbol{a}^n)$ converges to $\hat{p}(\boldsymbol{a}^n)$ at the rate $O(1/\sqrt{B})$ where B is the number of Monte Carlo resamples. Therefore, by selecting B sufficiently large, the error between $\tilde{p}(\boldsymbol{a}^n)$ and $\hat{p}(\boldsymbol{a}^n)$ is small relative to the error between $\hat{p}(\boldsymbol{a}^n)$ and $\hat{p}(\boldsymbol{\rho})$, and so the results in this appendix would also apply to the Monte Carlo approximation. For example, for Pearson's or Spearman's correlation, by setting $B \geq n$, the Monte Carlo approximation $\tilde{p}(\boldsymbol{a}^n)$ also converges to $\hat{p}(\boldsymbol{\rho})$ at rate $O_p(1/\sqrt{n})$.

B. Additional simulations

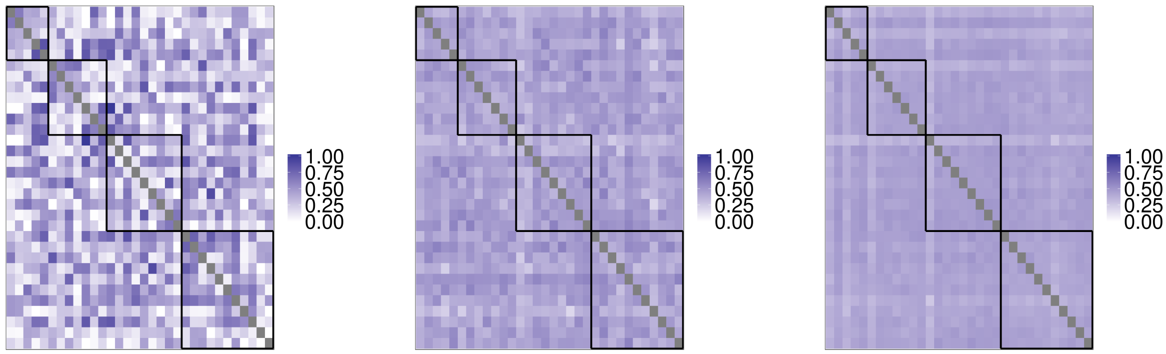
In this appendix, we simulated data under four additional scenarios: 1) constant off-diagonal values, 2) block diagonal structure on a subset of the matrix and white noise on the rest of the matrix (partial block diagonal structure), 3) a true CFA data generating process, and 4) a true CFA generating process with subsequently discretized outcomes. As before, for each scenario we generated 1,000 datasets for each sample size. For simulations under the permutation null hypothesis, we used sample sizes of n = 10, 100, and 1,000 with $B = 1,000$ resamples. For simulations under the permutation alternative hypothesis, we used samples sizes of n = 10, 50, 100, and 1,000 with $B = 10,000$ resamples to better approximate small p-values and statistical power. For all simulations, we used $K = 4$ blocks of sizes $p_1 = 5$, $p_2 = 7$, $p_3 = 9$, $p_4 = 11$, so that the total number of variables was $p = \sum_k p_k = 32$. In all figures, the block numbers begin in the upper left and end in the lower right, i.e., block $k = 1$ is in the top left corner, and block $k = 4$ is in the bottom right corner.

In the matrix structure testing framework, Appendix B.1 is under the null hypothesis (H_0 is true), and Appendices B.2, B.3, ad B.4 are under the alternative hypothesis (H_1 is true). In the GOF framework, the model is misspecified in Appendices B.1 and B.2 (H_1 is true), correctly specified in Appendix B.3, and correctly specified in Appendix B.4 apart from the discretized outcomes. This allows Appendix B.4 to serve as a check on the robustness of CFI, TLI, and RMSEA to continuous versus discrete outcomes.

B.1. Constant off-diagonal correlation

For the scenario of constant off-diagonal correlation, we set $\Sigma_{t,ij} = 0.5$ if $i \neq j$ and 1 if $i = j$. We used $B = 1,000$ MC resamples for each test. The rest of the simulation is as described in Section 5.1.

Figure S1 shows the estimated Spearman's absolute correlation matrices A from a single simulation at sample sizes of $n = 10, 100$, and 1,000.



(a) $n = 10$ (b) $n = 100$ (c) $n = 1,000$

Figure S1: Constant off-diagonal: Estimated Spearman's correlation coefficient (absolute values) from a single simulation at sample sizes of $n = 10, 100$, and 1,000.

Figure S2 shows the distribution of p-values from a permutation test with Γ_{norm} and $B = 1,000$ MC resamples, p-values from the X_2 pattern hypothesis test, and CFI values from a CFA model. Figure S3 shows the distribution of RMSEA values. As seen in Figures S2 and S3, the distribution of p-values from Γ_{norm} is uniform, which is as expected under the null hypothesis. The p-values from the X_2 statistic move from close to one to close to zero as the sample size increases, though not as quickly as in the block diagonal scenario, and the CFI values cluster close to 1 for all sample sizes. The RMSEA values tend to be near zero for $n = 10$ and are exactly zero for $n = 100$ and $n = 1,000$. In this scenario, the CFA model is misspecified, so large CFI values and small RMSEA values, indicating good model fit, represent a GOF false alarm.

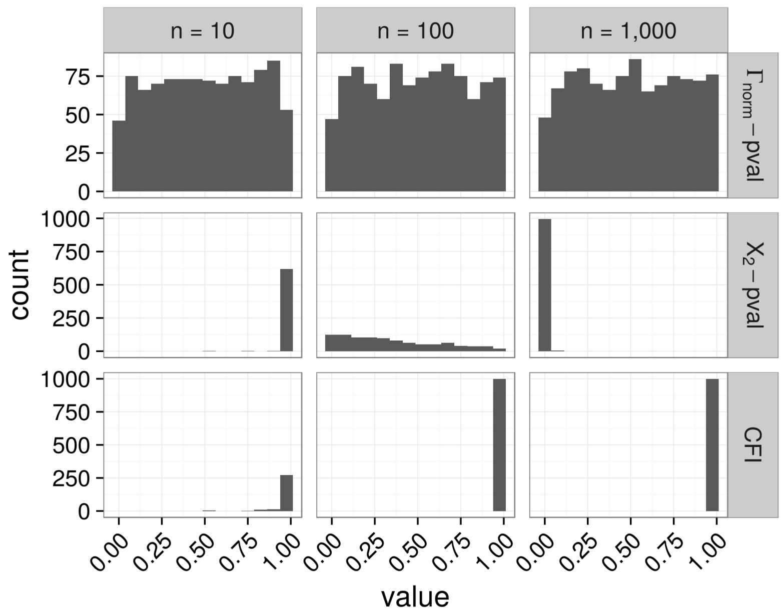


Figure S2: Overall test for constant off-diagonal scenario: permutation p-values with Γ_{norm} and $B = 1,000$ MC resamples, p-values from the X_2 pattern hypothesis test, and CFI values from a CFA model. For each sample size we did 1,000 simulations. Results with TLI are similar to those for CFI and are not shown.

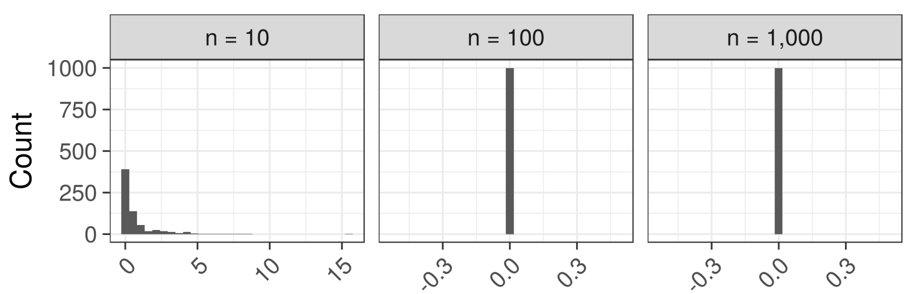


Figure S3: Overall test for constant off-diagonal scenario: RMSEA. For each sample size we did 1,000 simulations.

Table S1 shows the type I error rates for Γ_{norm} and the permutation test for statistical significance levels of $\alpha = 0.01$ and 0.05. As seen in Table S1, the error rates are near their nominal level for all sample sizes.

Table S1: Type I error rates in constant off-diagonal scenario using Γ_{norm} in a permutation test for significance levels of $\alpha = 0.01$ and 0.05. 1,000 simulations were run for each sample size.

	n	Overall	Block-specific FWER
$\alpha = 0.01$			
	10	0.010	0.0010
	100	0.016	0.013
	1,000	0.008	0.010
$\alpha = 0.05$			
	10	0.055	0.052
	100	0.050	0.059
	1,000	0.041	0.053

Table S2 shows the percent of simulations with CFI and TLI above the cutoff value recommended by Hu and Bentler (1999) (0.95), as well as the more liberal cutoff values noted by Hooper et al. (2008) (0.9 and 0.8). As seen in Table S2, The GOF false alarm rates are high for CFI and TLI in this simulation, and increase with sample size.

Table S2: GOF false alarm rate for constant off-diagonal scenario: Percent of simulation results above the cutoff value (CFI and TLI above the cutoff indicate good model fit)

		Cutoff		
Fit index	n	0.95	0.9	0.8
CFI	10	0.89	0.92	0.98
	100	1.0	1.0	1.0
	1,000	1.0	1.0	1.0
TLI	10	0.88	0.92	0.97
	100	1.0	1.0	1.0
	1,000	1.0	1.0	1.0

Table S3 shows the percent of simulations with RMSEA below the cutoff values recommended by Steiger (2007) (0.07), as well as the alternative cutoff values recommended by Browne and Cudeck (1992) (0.05, 0.1). As can be seen in Table S3, the GOF false alarm rate is high for all cutoffs, and is equal to one for samples sizes of $n = 100$ and $n = 1,000$.

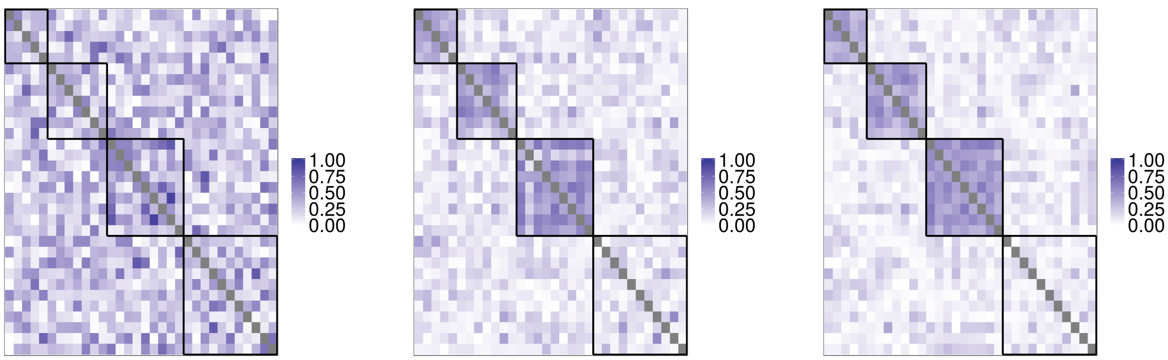
Table S3: GOF false alarm rate for constant off-diagonal scenario: Percent of simulation results below the cutoff value (RMSEA below the cutoff indicates good model fit)

		Cutoff		
Fit index	n	0.05	0.07	0.1
RMSEA	10	0.38	0.39	0.41
	100	1.0	1.0	1.0
	1,000	1.0	1.0	1.0

B.2. Partial block diagonal structure

For this scenario, we followed the simulation described in Section 5.1, but set $r_4 = 0$, i.e., the last hypothesized block is not a true block.

Figure S4 shows the estimated Spearman's absolute correlation matrices A from a single simulation at sample sizes of $n = 10, 100$, and 1,000.



(a) $n = 10$ (b) $n = 100$ (c) $n = 1,000$

Figure S4: Partial block diagonal: Estimated Spearman's correlation coefficient (absolute values) from a single simulation at sample sizes of $n = 10, 100$, and 1,000.

Figure S5 shows the distribution of p-values from a permutation test with Γ_{norm} and $B = 10,000$ MC resamples, p-values from the X_2 pattern hypothesis test, and CFI values from a CFA model. As seen in Figure S5, the distribution of p-values from Γ_{norm} is left-skewed, which is as expected under the alternative hypothesis. The p-values from the X_2 statistic move from close to one to close to zero as the sample size increases, and the CFI values cluster around 0.75 to 0.9 for all sample sizes.

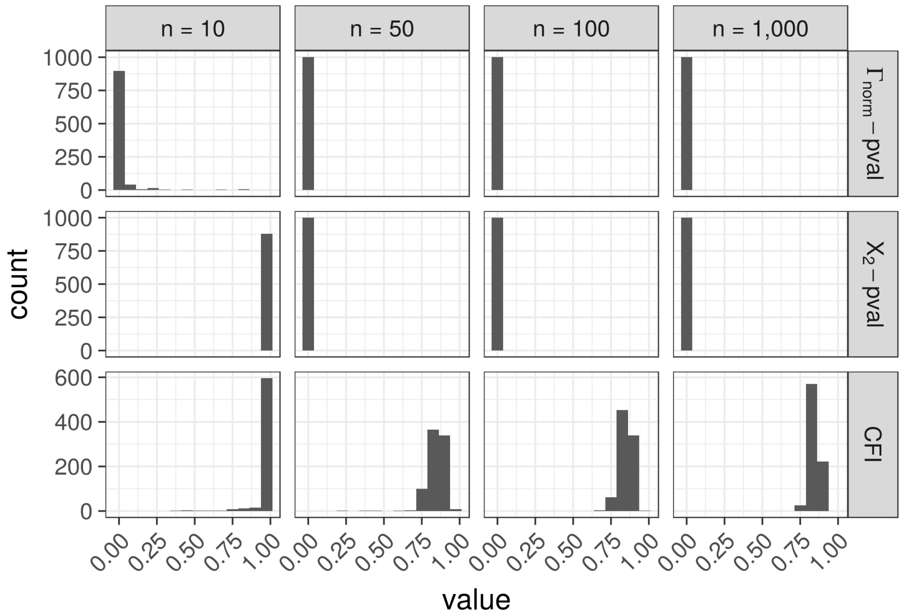


Figure S5: Overall test in partial block diagonal scenario: permutation p-values using Γ_{norm} and $B = 10,000$ MC resamples, p-values from the X_2 pattern hypothesis test, and CFI values from a CFA model. For each sample size we did 1,000 simulations. Results with TLI are similar to those for CFI and are not shown.

Table S4 shows the power (overall and blocks 1, 2, 3) and type I error rate (block 4) using Γ_{norm} in a permutation test for statistical significance levels of $\alpha = 0.01$ and 0.05. As seen in Table S4, the statistical power is high for blocks 1, 2, and 3, and the type I error rate is low for block $k = 4$.

Table S4: Partial block diagonal scenario: Power (overall and blocks 1, 2, 3) and type I error rate (block 4) using Γ_{norm} in a permutation test for significance levels of $\alpha = 0.01$ and 0.05. 1,000 simulations were run for each sample size.

$\alpha = 0.01$					
	Block				
n	Overall	$k = 1$	$k = 2$	$k = 3$	$k = 4$
$\alpha = 0.01$					
10	0.86	0.30	0.33	0.73	0.0
50	1.0	0.91	0.95	1.0	0.0
100	1.0	0.98	0.99	1.0	0.0
1,000	1.0	1.0	1.0	1.0	0.0
$\alpha = 0.05$					
10	0.93	0.48	0.49	0.83	0.0061
50	1.0	0.97	0.98	1.0	0.0010
100	1.0	1.0	1.0	1.0	0.0
1,000	1.0	1.0	1.0	1.0	0.0010

Table S5 shows the percent of simulations with CFI and TLI above the cutoff value recommended by Hu and Bentler (1999) (0.95), as well as the more liberal cutoff values noted by Hooper et al. (2008) (0.9 and 0.8). As seen in Table S5, The GOF false alarm rate decreases as sample size increases. However, these results do not by themselves show that three of the four block are correctly modeled, and only the fourth is incorrectly modeled.

Table S5: GOF false alarm rate for the partial block diagonal scenario: Percent of simulation results above the cutoff value (CFI and TLI above the cutoff indicate good model fit)

		Cutoff		
Fit index	n	0.95	0.9	0.8
CFI	10	0.92	0.94	0.97
	50	0.0024	0.14	0.82
	100	0.0	0.086	0.89
	1,000	0.0	0.020	0.93
TLI	10	0.92	0.93	0.96
	50	0.0024	0.095	0.75
	100	0.0	0.048	0.81
	1,000	0.0	0.0049	0.34

Table S6 shows the percent of simulations with RMSEA below the cutoff values recommended by Steiger (2007) (0.07), as well as the alternative cutoff values recommended by Browne and Cudeck (1992) (0.05, 0.1). As can be seen in Table S6, the GOF false alarm rate is zero for all cutoffs at $n = 50$, 100, and 1,000. However, as with CFI and TLI, these results do not by themselves show that three of the four block are correctly modeled, and only the fourth is incorrectly modeled.

Table S6: GOF false alarm rate for partial block diagonal scenario: Percent of simulation results below the cutoff value (RMSEA below the cutoff indicates good model fit)

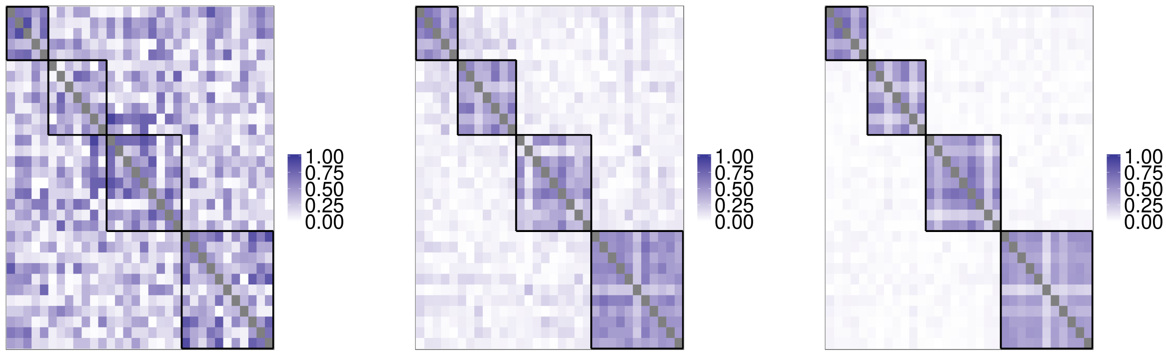
		Cutoff		
Fit index	n	0.05	0.07	0.1
	10	0.56	0.56	0.56
RMSEA	50	0.0	0.0	0.0
	100	0.0	0.0	0.0
	1,000	0.0	0.0	0.0

B.3. True CFA

For this scenario, we simulated data from a true CFA model using the `simulateData` function in the `lavaan` package (Rosseel, 2012) for R (R Core Team, 2017). In particular, we simulated data with four latent factors, $B_1, \ldots, B_4$, with loadings given by

```
B1 =~ a + 2*b + 1.5*c + 0.5*d + e
B2 =~ f + g + 0.4*h + 0.75*i + 2*j + 0.5*k + l
B3 =~ m + 0.5*n + o + 1.25*p + q + 3*r + s + 0.4*t + u
B4 =~ 1.25*v + w + 0.8*x + y + z + 0.4*z2 + z3 + 0.6*z4 + z5 + z6 + z7
```

Figure S6 shows the estimated Spearman's absolute correlation matrices A from a single simulation for sample sizes of $n = 10, 100,$ and $1{,}000$.



(a) $n = 10$ (b) $n = 100$ (c) $n = 1,000$

Figure S6: True CFA model: Estimated Spearman's correlation coefficient (absolute values) from a single simulation at sample sizes of $n = 10, 100$, and 1,000.

Figure S7 shows the distribution of p-values from a permutation test with Γ_{norm} and $B = 10,000$ MC resamples, p-values from the X_2 pattern hypothesis test, and CFI values from a CFA model. As seen in Figure S7, the distribution of p-values from Γ_{norm} is left-skewed, which is as expected under the alternative hypothesis. The p-values from the X_2 statistic move from close to one to close to zero as the sample size increases, and the CFI values cluster around 0.75 to 1 for all sample sizes. The CFA model did not converge for $n = 10$.

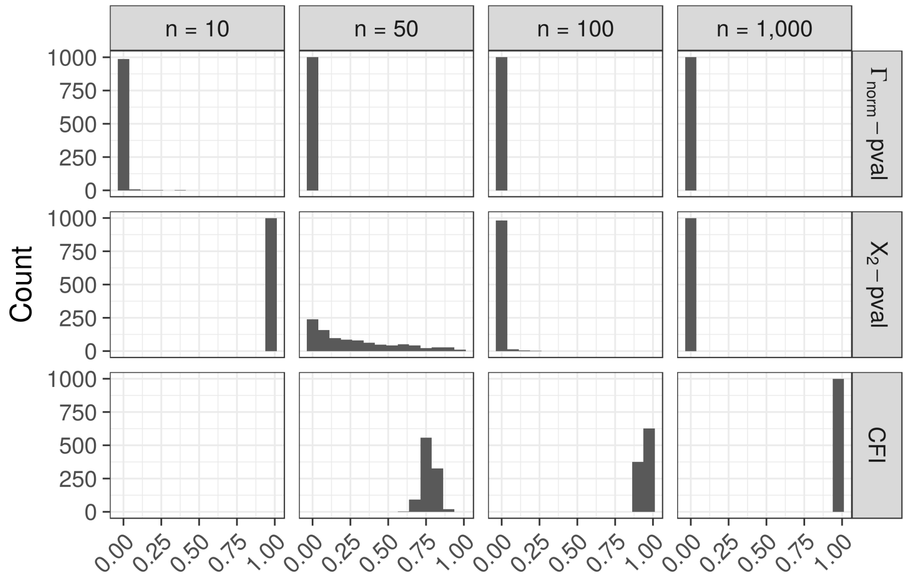


Figure S7: Overall test in true CFA scenario: permutation p-values using Γ_{norm} and $B = 10,000$ MC resamples, p-values from the X_2 pattern hypothesis test, and CFI values from a CFA model. For each sample size we did 1,000 simulations. Results with TLI are similar to those for CFI and are not shown. The CFA model did not converge for $n = 10$.

Table S7 shows the power using Γ_{norm} in a permutation test for statistical significance levels of $\alpha = 0.01$ and 0.05. As seen in Table S7, the statistical power is 1 for both the overall and block-specific tests for sample sizes of $n = 50$ and larger.

Table S7: True CFA scenario: Power using Γ_{norm} in a permutation test for significance levels of $\alpha = 0.01$ and 0.05. 1,000 simulations were run for each sample size.

$\alpha = 0.01$					
	Block				
n	Overall	$k = 1$	$k = 2$	$k = 3$	$k = 4$
$\alpha = 0.01$					
10	0.97	0.48	0.32	0.54	0.55
50	1.0	1.0	1.0	1.0	1.0
100	1.0	1.0	1.0	1.0	1.0
1,000	1.0	1.0	1.0	1.0	1.0
$\alpha = 0.05$					
10	0.99	0.66	0.49	0.69	0.69
50	1.0	1.0	1.0	1.0	1.0
100	1.0	1.0	1.0	1.0	1.0
1,000	1.0	1.0	1.0	1.0	1.0

Table S8 shows the percent of simulations with CFI and TLI above the cutoff value recommended by Hu and Bentler (1999) (0.95), as well as the more liberal cutoff values noted by Hooper et al. (2008) (0.9 and 0.8). As seen in Table S8, The type I power increases as sample size increases.

Table S8: Type I power for true CFA scenario with CFI and TLI: Percent of simulation results above the cutoff value (CFI and TLI above the cutoff indicate good model fit). CFA models did not converge for $n = 10$.

		Cutoff		
Fit index	n	0.95	0.9	0.8
CFI	10	–	–	–
	50	0.001	0.004	0.25
	100	0.44	0.97	1.0
	1,000	1.0	1.0	1.0
TLI	10	–	–	–
	50	0.001	0.003	0.15
	100	0.38	0.95	1.0
	1,000	1.0	1.0	1.0

Table S9 shows the percent of simulations with RMSEA below the cutoff values recommended by Steiger (2007) (0.07), as well as the alternative cutoff values recommended by Browne and Cudeck (1992) (0.05, 0.1). As can be seen in Table S9, the type I power increases with sample size, and is 1 for all cutoffs at $n = 1,000$.

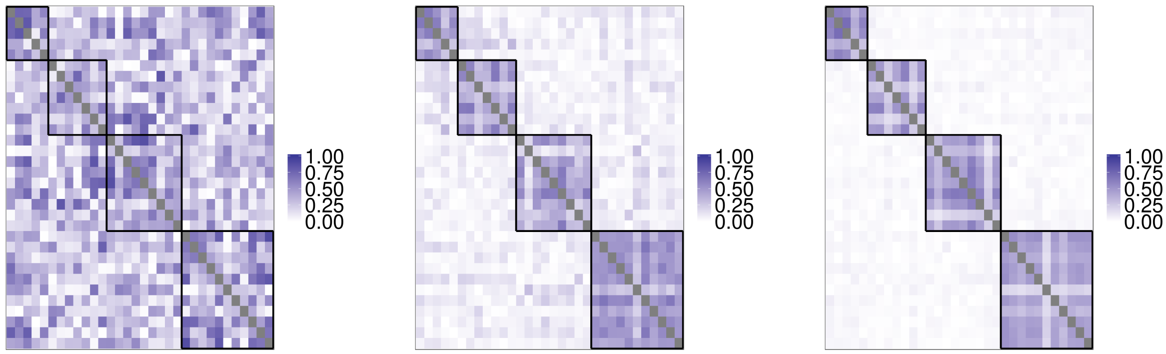
Table S9: Type I power for the true CFA scenario: Percent of simulation results below the cutoff value (RMSEA below the cutoff indicates good model fit)

		Cutoff		
Fit index	n	0.05	0.07	0.1
	10	–	–	–
RMSEA	50	0.003	0.023	0.78
	100	0.86	1.0	1.0
	1,000	1.0	1.0	1.0

B.4. True CFA with discretized outcome

For this scenario, we simulated data as in Appendix B.3 and then discretized the outcome as described in Section 5.1.

Figure S8 shows the estimated Spearman's absolute correlation matrices A from a single simulation for sample sizes of $n = 10, 100$, and 1,000.



(a) $n = 10$ (b) $n = 100$ (c) $n = 1,000$

Figure S8: True CFA model with discretized outcome: Estimated Spearman's correlation coefficient (absolute values) from a single simulation at sample sizes of $n = 10, 100$, and 1,000.

Figure S9 shows the distribution of p-values from a permutation test with Γ_{norm} and $B = 10,000$ MC resamples, p-values from the X_2 pattern hypothesis test, and CFI values from a CFA model. As seen in Figure S9, the distribution of p-values from Γ_{norm} is concentrated near 0, which is as expected under the alternative hypothesis. The p-values from the X_2 statistic move from being close to uniform to being close to zero as the sample size increases, and the CFI values cluster around 0.75 for $n = 50$ and near 1 for $n = 1,000$. The CFA model did not converge for $n = 10$.

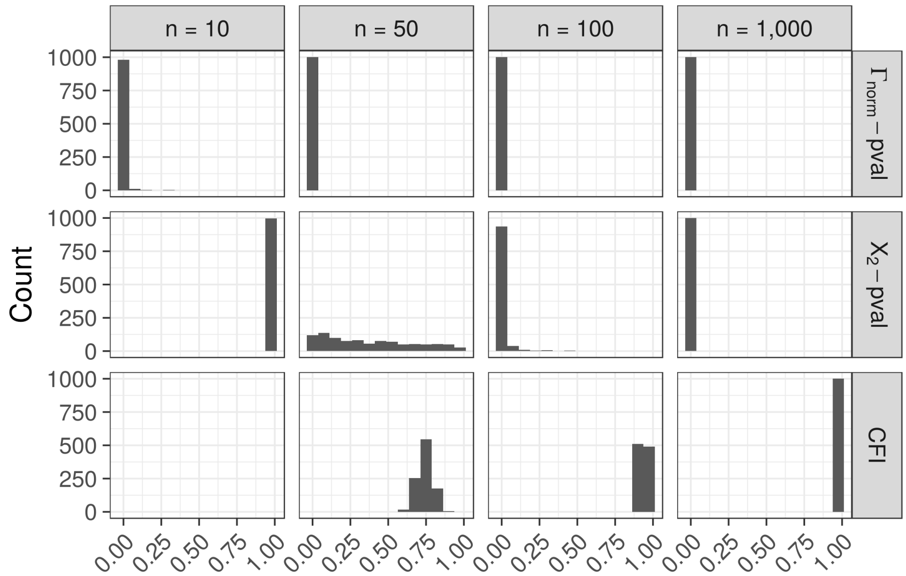


Figure S9: Overall test in true CFA scenario: permutation p-values using Γ_{norm} and $B = 10,000$ MC resamples, p-values from the X_2 pattern hypothesis test, and CFI values from a CFA model. For each sample size we did 1,000 simulations. Results with TLI are similar to those for CFI and are not shown. The CFA model did not converge for $n = 10$.

Table S10 shows the power using Γ_{norm} in a permutation test for statistical significance levels of $\alpha = 0.01$ and 0.05. As seen in Table S10, the statistical power is 1 for both the overall and block-specific tests for sample sizes of $n = 50$ and larger.

Table S10: True CFA scenario with discretized outcome: Power using Γ_{norm} in a permutation test for significance levels of $\alpha = 0.01$ and 0.05. 1,000 simulations were run for each sample size.

$\alpha = 0.01$					
	Block				
n	Overall	$k = 1$	$k = 2$	$k = 3$	$k = 4$
$\alpha = 0.01$					
10	0.97	0.48	0.34	0.56	0.52
50	1.0	1.0	1.0	1.0	1.0
100	1.0	1.0	1.0	1.0	1.0
1,000	1.0	1.0	1.0	1.0	1.0
$\alpha = 0.05$					
10	0.99	0.66	0.52	0.71	0.65
50	1.0	1.0	1.0	1.0	1.0
100	1.0	1.0	1.0	1.0	1.0
1,000	1.0	1.0	1.0	1.0	1.0

Table S11 shows the percent of simulations with CFI and TLI above the cutoff value recommended by Hu and Bentler (1999) (0.95), as well as the more liberal cutoff values noted by Hooper et al. (2008) (0.9 and 0.8). As seen in Table S11, The type I power increases as sample size increases.

Table S11: Type I power for the true CFA scenario with discretized outcomes (CFI and TLI): Percent of simulation results above the cutoff value (CFI and TLI above the cutoff indicate good model fit). CFA models did not converge for $n = 10$.

		Cutoff		
Fit index	n	0.95	0.9	0.8
CFI	10	–	–	–
	50	0.001	0.001	0.119
	100	0.32	0.92	1.0
	1,000	1.0	1.0	1.0
TLI	10	–	–	–
	50	0.0	0.001	0.06
	100	0.28	0.87	1.0
	1,000	1.0	1.0	1.0

Table S12 shows the percent of simulations with RMSEA below the cutoff values recommended by Steiger (2007) (0.07), as well as the alternative cutoff values recommended by Browne and Cudeck (1992) (0.05, 0.1). As can be seen in Table S12, the type I power increases with sample size, and is 1 for all cutoffs at $n = 1,000$.

Table S12: Type I power for true CFA scenario with discretized outcomes: Percent of simulation results below the cutoff value (RMSEA below the cutoff indicates good model fit)

		Cutoff		
Fit index	n	0.05	0.07	0.1
	10	–	–	–
RMSEA	50	0.002	0.023	0.77
	100	0.87	1.0	1.0
	1,000	1.0	1.0	1.0

C. Big five questionnaire items

As described by Smith et al. (2013) selected respondents to the 2010 Health and Retirement Survey were asked to rate how well 31 items described them on the following four point scale: 1) A lot, 2) Some, 3) A little, 4) Not at all.

The items were as follows (letters match those shown in Figure 1): a) Outgoing, b) Helpful, c) Reckless, d) Moody, e) Organized, f) Friendly, g) Warm, h) Worrying, i) Responsible, j) Lively, k) Caring, l) Nervous, m) Creative, n) Hardworking, o) Imaginative, p) Softhearted, q) Calm, r) Self-disciplined, s) Intelligent, t) Curious, u) Active, v) Careless, w) Broad-minded, x) Impulsive, y) Sympathetic, z) Cautious, z2) Talkative, z3) Sophisticated, z4) Adventurous, z5) Thorough, and z6) Thrifty.

The items were grouped into five sub-dimensions:

1. Neuroticism: d, h, l, q
2. Extroversion: a, f, j, u, z2
3. Agreeableness: b, g, k, p, y
4. Openness to experience: m, o, s, t, w, z3, z4
5. Conscientiousness: c, e, i, n, r, v, x, z, z5, z6

and all but c, q, v, and x were reverse coded.

References

Browne, M. W. and Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21(2):230–258.

Hooper, D., Coughlan, J., and Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods*, 6(1):53–60.

Hu, L.-t. and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1):1–55.

Lehmann, E. L. and Romano, J. P. (2005). *Testing statistical hypotheses*. Springer Science & Business Media, New York, NY, 3rd edition.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2):1–36.

Smith, J., Fisher, G., Ryan, L., Clarke, P., House, J., and Weir, D. (2013). Psychosocial and lifestyle questionnaire 2006–2010: Documentation report core section LB. Technical report, University of Michigan.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, 42(5):893–898.