Supplementary material for FDP control in multivariate linear models using the bootstrap

Samuel Davenport, Bertrand Thirion, Pierre Neuvial

September 22, 2023

Contents

S-1 Further theory for random fields

In this section will formalise some of the notation surrounding random fields. We will need this in order to establish our theoretical results.

S-1.1 Operations and convergence

Given two random fields g and $g' : \mathcal{V} \to \mathbb{R}^L$, operations of addition and subtraction can be performed pointwise and so $g + g'$ and $g - g'$ are well defined. Moreover if instead $g : \mathcal{V} \to \mathbb{R}$ then multiplication and division can also be performed pointwise and so, in that case, gg' and g'/g are well-defined.

Given $D \in \mathbb{N}$, suppose that $\mathcal{V} = \{u_1, \ldots, u_V\}$ for some $V \in \mathbb{N}$ and $u_1, \ldots, u_V \in \mathbb{R}^D$. For $L \in \mathbb{N}$, let $g : \mathcal{V} \to \mathbb{R}^L$ be a random field. Then we define $\text{vec}(g) \in \mathbb{R}^{LV}$ to be the vector whose $((i-1)L + j)$th element is $g_j(u_i)$ for $1 \leq i \leq V$ and $1 \leq j \leq L$. We refer this operation as **vectorization**. This allows us to easily define notions of convergence. Given a sequence: $((g_n)_{n\in\mathbb{N}}, g)$ of random fields from $\mathcal{V}$ to $\mathbb{R}^L$ we say that g_n converges to g in distribution (resp. probability/almost surely) if $\text{vec}(g_n)$ converges in distribution (resp. probability/almost surely) to $\text{vec}(g)$. We will write this as $g_n \xrightarrow{d} g$ (resp. $g_n \xrightarrow{\mathbb{P}} g / g_n \xrightarrow{a.s.} g$) - a notation that we will also use for random variables in what follows. Given such a sequence we will write $g_{n,j}$ $(1 \leq j \leq L)$ to denote its components.

Definition S-1.1. Given $L, L' \in \mathbb{N}$, a random field $g : \mathcal{V} \to \mathbb{R}^L$ and $M \in \mathbb{R}^{L' \times L}$ then we define the random field Mg which sends $v \in \mathcal{V}$ to $Mg(v) \in \mathbb{R}^{L'}$. Moreover, if $L = 1$ and $a \in \mathbb{R}^{L'}$ is a vector then we define the random field ag which sends $v \in \mathcal{V}$ to $ag(v) \in \mathbb{R}^{L'}$

Lemma S-1.2. *For $L, L' \in \mathbb{N}$ let $g : \mathcal{V} \to \mathbb{R}^L$ be a random field with covariance $\mathfrak{c}$ and let $M \in \mathbb{R}^{L' \times L}$, then Mg has covariance*

$$M\,\mathfrak{c}\,M^T.$$

Moreover if g is Gaussian then so is Mg.

S-1.2 Subsetting random fields

In the following sections we will want to restrict random fields to subsets.

Definition S-1.3. Suppose we have a random field $g : \mathcal{V} \to \mathbb{R}^L$, some $L \in \mathbb{N}$, and a set valued function: $\mathcal{N}$ on $\mathcal{V}$, such that for each $v \in \mathcal{V}$, $\mathcal{N}_v \subset \{1, \ldots, L\}$. Then we define the **restriction** of g to $\mathcal{N}$ to be the map $g|_{\mathcal{N}} : \Omega \to \left\{h : \mathcal{V} \to \bigcup_{1\leq j\leq L} \mathbb{R}^j\right\}$ such that $g|_{\mathcal{N}}(\omega)(v)$ is the vector $(g_k(\omega)(v) : k \in \mathcal{N}_v)^T \in \mathbb{R}^{|\mathcal{N}_v|}$.

Given a set function $\mathcal{N}$, defined as in Definition S-1.3, we can stack the entries of $g|_{\mathcal{N}}$ to create $\text{vec}(g|_{\mathcal{N}})$ and thus define $g_n|_{\mathcal{N}} \xrightarrow{d} g|_{\mathcal{N}}$,$g_n|_{\mathcal{N}} \xrightarrow{\mathbb{P}} g|_{\mathcal{N}}$ and $g_n|_{\mathcal{N}} \xrightarrow{a.s.} g|_{\mathcal{N}}$.

Definition S-1.4. Given an L-dimensional Gaussian field, $g \sim \mathcal{G}(\mu, \mathfrak{c})$ for some mean μ and covariance $\mathfrak{c}$ and a set function $\mathcal{N}$ as defined above, we shall write $\mathcal{G}(\mu, \mathfrak{c})|_{\mathcal{N}}$ to denote the distribution of the restricted random field. I.e. $g|_{\mathcal{N}} \sim \mathcal{G}(\mu, \mathfrak{c})|_{\mathcal{N}}$. Given

$$f : \left\{h : \mathcal{V} \to \mathbb{R}^L\right\} \to \mathbb{R}$$

we shall write $X \sim f(\mathcal{G}(\mu, \mathfrak{c}))$ to indicate that X is a real valued random variable which has the same distribution as $f(g)$. Given

$$f : \left\{ h : \mathcal{V} \to \bigcup_{1 \le j \le L} \mathbb{R}^j \right\} \to \mathbb{R} \tag{S-1}$$

we similarly define the notation $f(\mathcal{G}(\mu, \mathfrak{c})|_{\mathcal{N}})$.

S-2 Consistency of the bootstrap in the linear model

S-2.1 Lindeberg Central Limit Theorem

In order to prove our main results we require Proposition S-2.2 (stated below) which we prove using the Lindeberg CLT (see e.g. Van der Vaart (2000) Chapter 2.8). We will also require the following lemma.

Lemma S-2.1. *Let X and Y be random variables such that $\mathbb{E}\left[|X|^{2+\eta}\right] < \infty$ and $\mathbb{E}\left[|Y|^K\right] < \infty$ for some $K, \eta > 0$, then for all $a \in \mathbb{R}$,*

$$\mathbb{E}\left[X^2 1[a|Y| > \gamma]\right] \le \gamma^{-K/q} a^{K/q} \mathbb{E}\left[|X|^{2+\eta}\right]^{1/(1+\eta/2)} \mathbb{E}\left[|Y|^K\right]^{1/q}$$

where $q = 1 - (1 + \eta/2)^{-1}$.

Proof. By Holder's inequality for $p, q > 0$ such that $\frac{1}{p} + \frac{1}{q} = 1$,

$$\begin{aligned}\mathbb{E}\left[X^2 1[a|Y| > \gamma]\right] &\le \mathbb{E}\left[X^{2p}\right]^{1/p} \mathbb{E}[1[a|Y| > \gamma]]^{1/q} = \mathbb{E}\left[X^{2p}\right]^{1/p} \mathbb{P}(a|Y| > \gamma)^{1/q} \\ &\le \mathbb{E}\left[X^{2p}\right]^{1/p} \left(\frac{\mathbb{E}\left[a^K |Y|^K\right]}{\gamma^K} \right)^{1/q} = \gamma^{-K/q} a^{K/q} \mathbb{E}\left[X^{2p}\right]^{1/p} \mathbb{E}\left[|Y|^K\right]^{1/q}\end{aligned}$$

where the middle inequality holds by Markov's inequality. Taking $p = 1 + \eta/2$ and $q = 1 - \frac{1}{p}$, the result follows. □

Proposition S-2.2. *Given a sequence $(k_n)_{n \in \mathbb{N}}$, let $\{\xi_{n,i} : n, i \in \mathbb{N}, 1 \le i \le k_n\}$ be a triangular array of mean-zero random fields on $\mathcal{V}$ which are i.i.d within rows and have finite covariance. Let $\{a_{ni} : n, i \in \mathbb{N}, 1 \le i \le n\}$ be a triangular array of $D-$dimensional vectors such that $\sum_{i=1}^n \|a_{ni}\|^{2+K/q} \to 0$ as $n \to \infty$ and $\sup_{i,n} \mathbb{E}\left[|\xi_{n,i}|^{\max(K, 2+\eta)}\right] < \infty$ for some $K > 0$, any $\eta > 0$ and $q = 1 - (1 + \eta/2)^{-1}$. Let $A_n = (a_{n1}, \dots, a_{nk_n}) \in \mathbb{R}^{D \times k_n}$ and suppose that $A_n^T A_n \to \Sigma \in \mathbb{R}^{D \times D}$. For $n \in \mathbb{N}$, let $\mathfrak{c}_n$ be the covariance function of $\xi_{n,1}$ and suppose that as $n \to \infty, \mathfrak{c}_n \to \mathfrak{c}$ (pointwise) for some covariance function $\mathfrak{c}$ on $\mathcal{V}$. Then as $n \to \infty$,*

$$\sum_{i=1}^{k_n} a_{ni} \xi_{n,i} \xrightarrow{d} \mathcal{G}(0, \mathfrak{c}\Sigma).$$

Proof. The proof is an application of the Lindeberg CLT (see e.g. van der Vaart (1998) Proposition 2.27) to the vectors $\mathrm{vec}(a_{ni}\xi_{n,i})$. There are two conditions to verify. The first is to show that the covariance converges. We can show this blockwise, i.e., for each $u, v \in \mathcal{V}$,

$$\begin{aligned}\sum_{i=1}^{k_n} \operatorname{cov}(a_{ni}\xi_{n,i}(u), a_{ni}\xi_{n,i}(v)) &= \sum_{i=1}^{k_n} \mathbb{E}\left[a_{ni}\xi_{n,i}(u)\xi_{n,i}(v)a_{ni}^T\right] \\ &= \mathfrak{c}_n(u,v)\sum_{i=1}^{k_n} a_{ni}a_{ni}^T = \mathfrak{c}_n(u,v)A_n^T A_n.\end{aligned}$$

which converges to $\mathfrak{c}(u,v)\Sigma$ as $n \to \infty$. For the second condition we need to show that for all $\gamma > 0$,

$$\sum_{i=1}^{k_n} \mathbb{E}\left[\|\operatorname{vec}(a_{ni}\xi_{n,i})\|^2 1[\|\operatorname{vec}(a_{ni}\xi_{n,i})\| > \gamma]\right] \underset{n\to\infty}{\longrightarrow} 0.$$

We can expand the left hand side as

$$\sum_{i=1}^{k_n}\mathbb{E}\left[\sum_{v\in\mathcal{V}}\|a_{ni}\xi_{n,i}(v)\|^2 1\left[\sum_{u\in\mathcal{V}}\|a_{ni}\xi_{n,i}(u)\|^2 > \gamma^2\right]\right] \tag{S-2}$$

$$\leq \sum_{i=1}^{k_n}\sum_{v\in\mathcal{V}}\|a_{ni}\|^2\mathbb{E}\left[\xi_{n,i}(v)^2\sum_{u\in\mathcal{V}}1\left[\|a_{ni}\||\xi_{n,i}(u)| > \gamma|\mathcal{V}|^{-1/2}\right]\right] \tag{S-3}$$

$$= \sum_{i=1}^{k_n}\|a_{ni}\|^2\sum_{u,v\in\mathcal{V}}\mathbb{E}\left[\xi_{n,i}(v)^2 1\left[\|a_{ni}\||\xi_{n,i}(u)| > \gamma|\mathcal{V}|^{-1/2}\right]\right] \tag{S-4}$$

$$\leq C\sum_{i=1}^{k_n}\|a_{ni}\|^{2+K/q} \tag{S-5}$$

for some fixed constant $C > 0$, chosen in accordance with Lemma S-2.1. This bound converges to zero as $n \to \infty$. □

S-2.2 Useful lemmas

In order to establish consistency of the bootstrap we first establish some preliminary lemmas. The first lemma we prove shows that the contribution of $X_n(X_n^T X_n)^{-1}X_n^T E_n$ is zero asymptotically.

Lemma S-2.3. *Suppose that Assumption 1a holds, that $(x_m)_{m\in\mathbb{N}}$ is independent of the i.i.d sequence $(\epsilon_m)_{m\in\mathbb{N}}$ and that $\sup_{v\in\mathcal{V}}\mathbb{E}(\epsilon_1(v)^2) < \infty$. Then, letting $P_n = X_n(X_n^T X_n)^{-1}X_n^T$, as $n \to \infty$,*

$$\|P_n E_n\| \xrightarrow{\mathbb{P}} 0.$$

Proof. Letting $\beta = (2+\delta/2)^{-1}$, we have

$$P_n E_n = X_n(X_n^T X_n)^{-1}X_n^T E_n = \frac{X_n}{n^\beta}\left(\frac{X_n^T X_n}{n}\right)^{-1}\left(\frac{X_n^T E_n}{n^{1-\beta}}\right).$$

Thus,

$$\|P_n E_n\| = \left\|\frac{X_n}{n^\beta}\left(\frac{X_n^T X_n}{n}\right)^{-1}\left(\frac{X_n^T E_n}{n^{1-\beta}}\right)\right\| \leq \left\|\frac{X_n}{n^\beta}\right\|\left\|\left(\frac{X_n^T X_n}{n}\right)^{-1}\right\|\left\|\left(\frac{X_n^T E_n}{n^{1-\beta}}\right)\right\|.$$

$\frac{X_n^T E_n}{\sqrt{n}}$ converges in distribution so $\left\|\left(\frac{X_n^T E_n}{n^{1-\beta}}\right)\right\| \xrightarrow{\mathbb{P}} 0$ since $1-\beta > \frac{1}{2}$ and $\left(\frac{X_n^T X_n}{n}\right)^{-1}$ converges almost surely to Σ_X^{-1} by Lemma S-3.2. Applying the Gershgorin circle theorem and the AM-RM inequality, we have

$$\|X_n\| \leq \max_{1\leq i\leq n} \sum_{j=1}^{p} |(X_n)_{ij}| = \max_{1\leq i\leq n} \sum_{j=1}^{p} |(x_i)_j| \leq \frac{p}{\sqrt{p}} \max_{1\leq i\leq n} \|x_i\|.$$

$n^{-\beta} \max_{1\leq i\leq n} \|x_i\| \xrightarrow{a.s.} 0$ since $\mathbb{E}(\|x_1\|^{2+\delta}) < \infty$, so in particular $\|n^{-\beta} X_n\| \xrightarrow{a.s.} 0$. Combining these results and using Slutsky, it follows that $\|P_n E_n\| \xrightarrow{\mathbb{P}} 0$. □

In order to apply the Lindeberg Central Limit theorem and subsequently the triangular law of large numbers we will need the following bound on the moments of the bootstrapped residuals.

Lemma S-2.4. *Under Assumption 1, conditional on $(X_m, Y_m)_{m\in\mathbb{N}}$, for almost all sequences $(X_m, Y_m)_{m\in\mathbb{N}}$,*

$$\sup_{n\in\mathbb{N}, 1\leq i\leq n} \mathbb{E}(E_{n,i}^b)^4 < \infty.$$

Proof. Let $P_n = X_n(X_n^T X_n)^{-1} X_n^T$. For each $n \in \mathbb{N}$, conditional on $(X_m, Y_m)_{m\in\mathbb{N}}$, and $1 \leq i \leq n$ and $1 \leq b \leq B$,

$$\mathbb{E}(E_{n,i}^b)^4 = \sum_{j=1}^{n} \frac{1}{n}\left(\hat{E}_{n,j} - \frac{1}{n}\sum_{l=1}^{n} \hat{E}_{n,l}\right)^4 = \frac{1}{n}\sum_{j=1}^{n}\left(\epsilon_j - (P_n E_n)_j - \frac{1}{n}\sum_{l=1}^{n}(\epsilon_l - (P_n E_n)_l)\right)^4$$

Now $\|P_n E_n\|$ converges in probability to 0 unconditionally by Lemma S-2.3 and so

$$\max_{1\leq l\leq n} |(P_n E_n)_l| \xrightarrow{\mathbb{P}} 0,$$

since $\max_{1\leq l\leq n}(P_n E_n)_l^2 \leq \|P_n E_n\|^2$. In particular it follows that for $M > 0$,

$$\lim_{k\to\infty} \mathbb{P}\left(\max_{n\geq k} \max_{1\leq l\leq n} |(P_n E_n)_l| > M\right) \to 0 \tag{S-6}$$

as $k \to \infty$. For $k \in \mathbb{N}$, Let $A_k = \{\max_{n\geq k} \max_{1\leq l\leq n} |(P_n E_n)_l| \leq M\}$, then equation (S-6) implies that $\mathbb{P}(\cup_k A_k) = 1$ since the sets are nested. As such for $\omega \in \cup_k A_k$, ω is contained in A_K some $K = K(\omega) \in \mathbb{N}$. It follows that

$$\max_{n>K} \max_{1\leq l\leq n} |(P_n E_n)_l| \leq M$$

almost everywhere which implies that, almost surely,

$$\max_{n\in\mathbb{N}} \max_{1\leq l\leq n} |(P_n E_n)_l| \leq M' = M + \max_{1\leq n\leq K} \max_{1\leq l\leq n} |(P_n E_n)_l|.$$

We can thus almost surely bound $\mathbb{E}(E_{n,i}^b)^4$ by

$$\frac{1}{n}\sum_{j=1}^{n}\sum_{k=0}^{4}\left(\epsilon_j - \frac{1}{n}\sum_{l=1}^{n}\epsilon_l\right)^k (2M')^{4-k} \leq (2M')^4 \frac{1}{n}\sum_{j=1}^{n}\sum_{k=0}^{4}\left(\epsilon_j - \frac{1}{n}\sum_{l=1}^{n}\epsilon_l\right)^k.$$

The right hand side converges almost surely by the strong law of large numbers to a quantity that is the same for each i. It follows that the supremum over i, n of $\mathbb{E}(E_{n,i}^b)^4$ is bounded, a fact that is true almost everywhere since $\mathbb{P}(\cup_k A_k) = 1$. □

S-2.3 Proof of consistency of the bootstrap using the Lindeberg CLT

Theorem S-2.5. *Suppose $(X_n)_{n\in\mathbb{N}}$ and $(\epsilon_n)_{n\in\mathbb{N}}$ satisfy Assumption 1. Then consistency of the multivariate bootstrap implies that, conditional on $(X_m, Y_m)_{m\in\mathbb{N}}$ for almost all sequences $(X_m, Y_m)_{m\in\mathbb{N}}$, for each $1 \le b \le B$, as $n \to \infty$,*

$$\sqrt{n}(\hat{\beta}_n^b - \hat{\beta}_n) \xrightarrow{d} \mathcal{G}(0, \mathfrak{c}_\epsilon \Sigma_X^{-1})$$

$$\textit{and} \quad \hat{\sigma}_n^b \xrightarrow{\mathbb{P}} \sigma.$$

Proof. Expanding, we have that

$$\sqrt{n}(\hat{\beta}_n^b - \hat{\beta}_n) = \sqrt{n}(X_n^T X_n)^{-1} X_n^T E_n^b = \left(\frac{X_n^T X_n}{n}\right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i E_{n,i}^b.$$

Applying Lemma S-3.2, $\left(\frac{X_n^T X_n}{n}\right)^{-1}$ converges a.s. to Σ_X^{-1}. Moreover, $(E_{n,i}^b)_{n\in\mathbb{N},1\le i\le n}$ is a triangular array which is mean-zero and i.i.d within rows so result will follow by applying Proposition S-2.2 with $\eta = 2$ and $K < \frac{\delta}{2}$ and taking $a_{ni} = x_i/\sqrt{n}$ for $1 \le i \le k_n = n$ and $n \in \mathbb{N}$. Letting $A_n = (a_{n1}, \ldots, a_{nn})$ we have $A_n^T A_n = \frac{1}{n}\sum_{i=1}^n \|x_i\|^2 \xrightarrow{a.s.} \Sigma_X$ as $n \to \infty$. Moreover,

$$\sum_{i=1}^n \|a_{ni}\|^{2+K/q} = \frac{1}{n^{1+K/2q}} \sum_{i=1}^n \|x_i\|^{2+K/q} < \frac{1}{n^{1+K/2q}} \sum_{i=1}^n \|x_i\|^{2+\delta} \xrightarrow{a.s.} 0.$$

The requisite bounds on the moments are provided by Lemma S-2.4.

As such in order to apply Proposition S-2.2 it suffices to show that the covariance converges. In order to do so, for each $u, v \in \mathcal{V}$, conditional on $(X_m, Y_m)_{m\in\mathbb{N}}$,

$$\begin{aligned}\operatorname{cov}(E_{n,1}^b(u), E_{n,1}^b(v)) &= \sum_{j=1}^n \frac{1}{n}\left(\hat{E}_{n,j}(u) - \frac{1}{n}\sum_{l=1}^n \hat{E}_{n,l}(u)\right)\hat{E}_{n,j}(v) \\ &= \frac{1}{n}\hat{E}_n(u)^T \hat{E}_n(v) - \left(\frac{1}{n}\sum_{j=1}^n \hat{E}_{n,j}(u)\right)\left(\frac{1}{n}\sum_{j=1}^n \hat{E}_{n,j}(v)\right)\end{aligned}$$

Now, letting $P_n = X_n(X_n^T X_n)^{-1} X_n$ and letting I_n be the $n \times n$ identity matrix,

$$\frac{1}{n}\hat{E}_n(u)^T \hat{E}_n(v) = \frac{1}{n}E_n(u)^T(I_n - P_n)E_n(v) = \frac{1}{n}E_n(u)^T E_n(v) - \frac{1}{n}E_n(u)^T P_n E_n(v).$$

We can write $\frac{1}{n}E_n(u)^T E_n(v) = \frac{1}{n}\sum_{i=1}^n \epsilon_i(u)\epsilon_i(v)$, which converges almost surely to $\mathfrak{c}(u,v)$ by the strong law of large numbers. Moreover,

$$\begin{aligned}\frac{1}{n}E_n(u)^T P_n E_n(v) &= \frac{1}{n}E_n(u)^T X_n(X_n^T X_n)^{-1} X_n^T E_n(v) \\ &= \left(\frac{X_n^T E_n(u)}{n}\right)^T \left(\frac{X_n^T X_n}{n}\right)^{-1} \left(\frac{X_n^T E_n(u)}{n}\right)\end{aligned}$$

which converges almost surely to zero as $n \to \infty$. Finally,

$$\frac{1}{n}\sum_{j=1}^n \hat{E}_{n,j}(u) = \frac{1}{n}1_n^T(I_n - P_n)E_n(u) = \frac{1}{n}1_n^T E_n(u) - \frac{1}{n}1_n^T X_n(X_n^T X_n)^{-1} X_n^T E_n(u)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \epsilon_i(u) - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{X_n^T X_n}{n}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\epsilon_i(u)\right)$$

which converges almost surely to 0 as $n \to \infty$. This establishes the CLT.

It remains to show that the variance converges. To do so recall that

$$(\hat{\sigma}_n^b)^2 = \frac{1}{n}\sum_{i=1}^{n}(E_{n,i}^b)^2 - \left(\frac{1}{n}\sum_{i=1}^{n} E_{n,i}^b\right)^2.$$

The $E_{n,i}^b$ are i.i.d and mean-zero and the covariance of $E_{n,i}^b$ converges as shown above. As such by the Lindeberg CLT, conditional on $(X_m, Y_m)_{m\in\mathbb{N}}$, $\frac{1}{\sqrt{n}}\sum_{i=1}^n E_{n,i}^b$ converges in distribution. Thus dividing by $\sqrt{n}$, it follows that $\frac{1}{n}\sum_{i=1}^n E_{n,i}^b$ converges in probability to zero as $n \to \infty$, for almost all sequences $(X_m, Y_m)_{m\in\mathbb{N}}$. For the first term, note that

$$\mathbb{E}(E_{n,i}^b)^2 = \sum_{j=1}^{n}\frac{1}{n}\left(\hat{E}_{n,j} - \frac{1}{n}\sum_{l=1}^{n}\hat{E}_{n,l}\right)^2 \tag{S-7}$$

which converges to σ^2 almost surely as $n \to \infty$. As such the result follows by the triangular weak law of large numbers which we can apply because $\sup_{n\in\mathbb{N},1\le i\le n}\mathbb{E}(E_{n,i}^b)^4$ is bounded by Lemma S-2.4. □

Remark S-2.6. *Formally the randomness in the data can be separated from the randomness in the bootstrap resamples by considering a cross product of probability spaces (one for the bootstrap randomness and one for the randomness in the data). This allows the expectations conditional on $(X_m, Y_m)_{m\in\mathbb{N}}$ to be well defined. For further details of this decomposition, see e.g. Kosorok (2003) and Telschow et al. (2020).*

The above result can be used to show convergence of the test-statistic which, in our notation, we write formally as follows.

Theorem S-2.7. *(Bootstrap test-statistic convergence.) Suppose that $(X_n)_{n\in\mathbb{N}}$ and $(\epsilon_n)_{n\in\mathbb{N}}$ satisfy Assumption 1 and, for each $1 \le b \le B$, let $T_n^b : \mathcal{V} \to \mathbb{R}$ be the L-dimensional random field on $\mathcal{V}$ such that, for $1 \le l \le L$,*

$$T_{n,l}^b = \frac{c_l^T(\hat{\beta}_n^b - \hat{\beta}_n)}{\hat{\sigma}_n^b\sqrt{c_l^T(X_n^TX_n)^{-1}c_l}}. \tag{S-8}$$

Then conditional on $(X_m, Y_m)_{m\in\mathbb{N}}$, for almost every sequence $(X_m, Y_m)_{m\in\mathbb{N}}$, for each $1 \le b \le B$,

$$T_n^b \xrightarrow{d} \mathcal{G}(0, \mathfrak{c}')$$

as $n \to \infty$. Here $\mathfrak{c}' : \mathcal{V}\times\mathcal{V} \to \mathbb{R}$ takes $u, v \in \mathcal{V}$ to $\mathfrak{c}'(u,v) = \rho_\epsilon(u,v)AC\Sigma_X^{-1}C^TA^T$ where $A \in \mathbb{R}^{L\times L}$ is a diagonal matrix with $A_{ll} = (c_l^T\Sigma_X^{-1}c_l)^{-1/2}$ for $1 \le l \le L$.

Proof. We can write

$$T_n^b = \sqrt{n}A_nC(\hat{\beta}_n^b - \hat{\beta}_n)/\hat{\sigma}_n^b.$$

where A_n is defined as in the proof of S-3.4. Theorem 1 implies that as $n \to \infty$, $\hat{\sigma}_n^b \xrightarrow{\mathbb{P}} \sigma$. Moreover $A_n \xrightarrow{a.s.} A$ so applying Theorem S-2.5, Lemma S-1.2 and Slutsky, the first result holds. The second result immediately follows from the first. □

S-2.4 Proof of consistency of the bootstrap using the results of Eck (2018)

Eck (2018) recently proved consistency of the bootstrap by extending the work of Freedman (1981) to multiple dimensions. In what follows here we demonstrate that their Theorem 1 implies Theorem S-2.5 by translating their result into our notation.

Proof. Applying Eck (2018)'s Theorem 1a (conditioning on $(X_m, Y_m)_{m\in\mathbb{N}}$ and restricting to the probability 1 event that $(\frac{1}{n}X_n^T X_n)^{-1} \longrightarrow \Sigma_X^{-1}$ which exists by Lemma S-3.2), we see that

$$\sqrt{n}(\mathrm{vec}(\hat{\beta}_n^b) - \mathrm{vec}(\hat{\beta}_n)) \longrightarrow N(0, \Sigma \otimes \Sigma_X^{-1}),$$

where $\Sigma = \mathrm{cov}(\mathrm{vec}(\epsilon_1))$. It follows that $\sqrt{n}(\hat{\beta}_n^b - \hat{\beta}_n)$ converges in distribution to a Gaussian random field. The form of the covariance in the statement of the theorem follows as writing $\mathcal{V} = \{u_1, \dots, u_V\}$, for $1 \leq l, m \leq L$ and $1 \leq j, k \leq V$,

$$(\Sigma \otimes \Sigma_X^{-1}))_{L(j-1)+l, L(k-1)+m} = \mathfrak{c}_\epsilon(u_j, u_k)(\Sigma_X^{-1})_{lm}.$$

The convergence in probability of the bootstrapped variance follows from Eck (2018)'s Theorem 1b. □

Remark S-2.8. *Eck (2018)'s theorem needs to be applied with care as they write the model as $Y = \beta X + \epsilon$ rather than via the more standard formulation of $Y = X\beta + \epsilon$, i.e. they take β to be a row vector rather than a column vector. Their vec operation is thus the result of stacking a transposed matrix. As such in our notation the resulting distribution in the statement of their Theorem 1 is $N(0, \Sigma_X^{-1} \otimes \Sigma)$ rather than $N(0, \Sigma \otimes \Sigma_X^{-1})$ - we have transposed it in order to match our notation.*

Remark S-2.9. *Eck (2018)'s Theorem 1 is stated in terms of fixed design matrices which converge. Here we assume that the design is random but condition on it which allows us to apply their Theorem 1 because $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ are independent. Eck (2018) has an alternative result (their Theorem 2) which applies when $(x_n, y_n)_{n\in\mathbb{N}}$ has a joint distribution, however this requires an alternative form of the bootstrap first introduced in Freedman (1981).*

S-2.5 Consistency of the bootstrap quantile

When bootstrapping we use the bootstrap samples to estimate quantiles of the test-statistic under the null hypothesis. In what follows we will demonstrate that as the number of bootstraps and subjects tends to infinity, the derived quantiles converge to a limit. In order to do so given $G : \mathbb{R} \to [0,1]$, define $G^- : [0,1] \to \mathbb{R}$, which takes $y \in [0,1]$ to $G^-(y) = \inf\{x : G(x) \geq y\}$, to be the generalized inverse of G. Then we have the following result.

Lemma S-2.10. *Let $(f_n)_{n\in\mathbb{N}}, f$ be functions from $\{h : \mathcal{V} \to \mathbb{R}^L\}$ to $\mathbb{R}$ such that conditional on $(X_m, Y_m)_{m\in\mathbb{N}}$ for almost all sequences $(X_m, Y_m)_{m\in\mathbb{N}}$, for each $b \in \mathbb{N}$,*

$$f_n(T_n^b) \xrightarrow{d} f(\mathcal{G}(0, \mathfrak{c}')).$$

For each $n, B \in \mathbb{N}$ and $0 < \alpha < 1$, let

$$\lambda^*_{\alpha,n,B} = \inf\left\{\lambda : \frac{1}{B}\sum_{b=1}^{B} 1\left[f_n(T_n^b) \leq \lambda\right] \geq \alpha\right\}.$$

Take F to be the CDF of $f(\mathcal{G}(0,\mathfrak{c}'))$ conditional on $(X_m, Y_m)_{m\in\mathbb{N}}$, i.e. for $\lambda \in [0,1]$, $F(\lambda) = \mathbb{P}(f(\mathcal{G}(0,\mathfrak{c}')) \le \lambda | (X_m, Y_m)_{m\in\mathbb{N}})$ and assume that F is strictly increasing and continuous. Then, letting $\lambda_\alpha = F^{-1}(\alpha)$, conditional on $(X_m, Y_m)_{m\in\mathbb{N}}$ for almost all sequences $(X_m, Y_m)_{m\in\mathbb{N}}$,

$$\lim_{n\to\infty}\lim_{B\to\infty} \lambda^*_{\alpha,n,B} = \lambda_\alpha$$

almost surely.

Proof. Let $F_n : \mathbb{R} \to [0,1]$ send $\lambda \in \mathbb{R}$ to $\mathbb{P}(f(T_n^1) \le \lambda | (X_m, Y_m)_{m\in\mathbb{N}})$. Define a sequence $(\eta_n)_{n\in\mathbb{N}} \ge 0$ such that $\alpha \pm \eta_n$ are continuity points of F_n^- and $\eta_n \to 0$ as $n \to \infty$. To do so let $\eta_n = 0$ if α is a continuity point of F_n^- and take $\eta_n = \frac{1}{2n^n}$ otherwise. Note that there are at most n^n distinct values that $f(T_n^1)$ can take and each value of T_n^1 is equally likely by construction. As such F_n is a step function with at most n^n steps, meaning that the height difference between steps is at least $\frac{1}{n^n}$. The points of discontinuity of F_n^- are the values in the range of F_n and so if α is not a point of continuity of F_n^- then $\alpha \pm \frac{1}{2n^n}$ must be. Now,

$$\lambda^*_{\alpha-\eta_n,n,B} \le \lambda^*_{\alpha,n,B} \le \lambda^*_{\alpha+\eta_n,n,B}. \tag{S-9}$$

The values $\alpha \pm \eta_n$ are continuity points of F_n^- and for $\lambda \in \mathbb{R}$, conditional on $(X_m, Y_m)_{m\in\mathbb{N}}$, by the SLLN, $\frac{1}{B}\sum_{b=1}^{B} 1\left[f(T_n^b) \le \lambda\right]$ converges almost surely to $F_n(\lambda)$ as $B \to \infty$. As such, applying Lemma 1.1.1 from De Haan and Ferreira (2006), $\lambda^*_{\alpha\pm\eta_n,n,B} \to F_n^-(\alpha \pm \eta_n)$ almost surely as $B \to \infty$. Moreover, as $n \to \infty$, F_n converges pointwise to F (as $f(T_n^1)|(X_m, Y_m)_{m\in\mathbb{N}} \xrightarrow{d} f(\mathcal{G}(0,\mathfrak{c}'))$) which is an increasing invertible function with continuous inverse. As such $F_n^-(\alpha \pm \eta_n) \to \lambda_\alpha$ as $n \to \infty$. To see this note that for all $\delta > 0$, there exists $N \in \mathbb{N}$ such that for all $n \ge N, \eta_n < \delta$ and

$$F_n^-(\alpha) \le F_n^-(\alpha + \eta_n) \le F_n^-(\alpha + \delta)$$

and $F_n^-(\alpha + \delta)$ converges to $F^{-1}(\alpha + \delta)$ as $n \to \infty$ by applying Lemma 1.1.1 from De Haan and Ferreira (2006) once again (since F^{-1} is continuous and so $\alpha + \delta$ is a point of continuity of F^{-1}). Continuity of F^{-1} implies that $F^{-1}(\alpha + \delta) \to F^{-1}(\alpha)$ as $\delta \to 0$. Arguing similarly for the sequence $\alpha - \eta_n$ the result follows. Taking limits and using the bound in equation (S-9), it almost surely follows that

$$\lim_{n\to\infty}\lim_{B\to\infty} \lambda^*_{\alpha,n,B} = \lambda_\alpha.$$

□

We refer to $\lambda^*_{\alpha,n,B}$ as the α-quantile of the bootstrap distribution of $f_n(T_n)$ based on B bootstraps. By Theorem S-2.5, for suitable choices of f_n, f, $f_n(T_n^b) \xrightarrow{d} f(\mathcal{G}(0,\mathfrak{c}'))$ as $n \to \infty$ and so Lemma S-2.10 shows that the bootstrapped t-statistics can be used to provide consistent estimates of the quantiles of the limiting distribution of $f(\mathcal{G}(0,\mathfrak{c}'))$. The easiest example of such a suitable sequence of functions is to take f to be continuous and let $f_n = f$ for all $n \in \mathbb{N}$, because of the Continuous Mapping Theorem. A more general result that provides a sufficient condition, based on uniform convergence of f_n to f, is given in Lemma S-4.3.

S-3 Proofs for the main text

S-3.1 Proofs for Section 2

S-3.1.1 Proof of Lemma 2.4

Proof. For $1 \leq k \leq K$, we have

$$\begin{aligned}\{|R_k(\lambda) \cap \mathcal{N}| > k-1\} &= \{|\{(l,v) \in \mathcal{N} : p_{n,l}(v) \leq t_k(\lambda)\}| > k-1\} \\ &= \left\{p^n_{(k:\mathcal{N})} \leq t_k(\lambda)\right\} = \left\{t_k^{-1}(p^n_{(k:\mathcal{N})}) \leq \lambda\right\}\end{aligned}$$

As such,

$$\bigcup_{1\leq k\leq K} \{|R_k(\lambda) \cap \mathcal{N}| > k-1\} = \left\{\min_{1\leq k\leq K} t_k^{-1}(p^n_{(k:\mathcal{N})}) \leq \lambda\right\} = \left\{\min_{1\leq k\leq K\wedge|\mathcal{H}|} t_k^{-1}(p^n_{(k:\mathcal{N})}) \leq \lambda\right\}.$$

□

Remark S-3.1. *Lemma 2.4 can be generalized to arbitrary ζ_k. The statement of the result in that case is that*

$$JER((R_k(\lambda), \zeta_k)_{1\leq k\leq K}) = \mathbb{P}\left(\min_{1\leq k\leq K\wedge|\mathcal{H}|} t_k^{-1}(p^n_{(\zeta_k+1:\mathcal{N})}) \leq \lambda\right).$$

Throughout the main text we take $\zeta_k = k-1$, this can be motivated by the fact that it implies that each individual rejection region $R_k(\lambda)$ controls the k-familywise error rate. However other choices provide valid inference, see Blanchard et al. (2020) for a discussion of the different choices of ζ_k. In particular this implies that the results in Section 3 can trivially be generalized to arbitrary ζ_k.

S-3.1.2 Convergence in the Linear Model

We will need the following useful Lemma which is Davenport et al. (2021)'s Lemma 8.2.

Lemma S-3.2. *Suppose that $(X_n)_{n\in\mathbb{N}}$ satisfies Assumption 1a and let $\Sigma_X = \mathbb{E}\left[x_1x_1^T\right]$, then Σ_X is invertible and*

$$\left(\frac{X_n^TX_n}{n}\right)^{-1} \xrightarrow{a.s.} \Sigma_X^{-1}.$$

In this section we establish results for asymptotics of coefficients and test-statistics in the linear model, written in terms of the framework of random fields.

Lemma S-3.3. *Suppose that $(X_n)_{n\in\mathbb{N}}$ and $(\epsilon_n)_{n\in\mathbb{N}}$ satisfy Assumption 1. Then*

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathcal{G}(0, \mathfrak{c}_\epsilon\Sigma_X^{-1}).$$

Proof. For each $n \in \mathbb{N}$,

$$\sqrt{n}(\hat{\beta}_n - \beta) = \sqrt{n}(X_n^TX_n)^{-1}X_n^T\epsilon_n = \left(\frac{X_n^TX_n}{n}\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i\epsilon_i.$$

By the Central Limit Theorem, $\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i\epsilon_i$ converges to a p-dimensional Gaussian random field with covariance

$$\mathrm{cov}(x_1\epsilon_1(u), x_1\epsilon_1(v)) = \mathbb{E}\left[x_1\epsilon_1(u)\epsilon_1(v)x_1^T\right] = \mathbb{E}[\epsilon_1(u)\epsilon_1(v)]\mathbb{E}\left[x_1x_1^T\right] = \mathfrak{c}_\epsilon(u,v)\Sigma_X$$

for $u, v \in \mathcal{V}$. $\left(\frac{X_n^T X_n}{n}\right)^{-1}$ converges almost surely to Σ_X^{-1} by Lemma S-3.2 and so the result follows by applying Lemma S-1.2 and Slutsky as the limiting distribution has covariance (for each $u, v, \in \mathcal{V}$)

$$\Sigma_X^{-1}(\mathfrak{c}_\epsilon(u,v)\Sigma_X)\Sigma_X^{-1} = \mathfrak{c}_\epsilon(u,v)\Sigma_X^{-1}.$$

□

Let $\mathfrak{c}' : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$ be the covariance function such that for all $u, v \in \mathcal{V}$

$$\mathfrak{c}'(u,v) = \rho_\epsilon(u,v) A C \Sigma_X^{-1} C^T A^T \tag{S-10}$$

where $A \in \mathbb{R}^{L\times L}$ is a diagonal matrix with $A_{ll} = (c_l^T \Sigma_X^{-1} c_l)^{-1/2}$ for $1 \le l \le L$. Then we having the following results (using the subsetting notation defined in Section S-1.2).

Theorem S-3.4. *For $n \in \mathbb{N}$, let S_n be the L-dimensional random field on $\mathcal{V}$ defined by*

$$S_{n,l} = \frac{c_l^T(\hat{\beta}_n - \beta)}{\hat{\sigma}_n \sqrt{c_l^T (X_n^T X_n)^{-1} c_l}}.$$

for $1 \le l \le L$. Then, under the conditions of Lemma S-3.3, as $n \to \infty$,

$$S_n \xrightarrow{d} \mathcal{G}(0, \mathfrak{c}')$$

and it follows that

$$T_n|_\mathcal{N} \xrightarrow{d} \mathcal{G}(0, \mathfrak{c}')|_\mathcal{N}.$$

Proof. We can write

$$S_n = \sqrt{n} A_n C(\hat{\beta}_n - \beta_n)/\hat{\sigma}_n.$$

where A_n is a diagonal matrix with $(A_n)_{ll} = \left(c_l^T \left(\frac{X_n^T X_n}{n}\right)^{-1} c_l\right)^{-1/2}$. $A_n \xrightarrow{a.s.} A$ by Lemma S-3.2 and $\hat{\sigma}_n \xrightarrow{a.s.} \sigma$ as $n \to \infty$. So applying Lemmas S-3.3 and S-1.2 and Slutsky, the first result follows. For $(v, l) \in \mathcal{N}$, $c_l^T \beta(v) = 0$. As such $S_n|_\mathcal{N} = T_n|_\mathcal{N}$ and it follows that

$$T_n|_\mathcal{N} \xrightarrow{d} \mathcal{G}|_\mathcal{N}.$$

□

S-3.2 Proofs for Section 3

S-3.2.1 Setup

In what follows we will require the following lemma.

Lemma S-3.5. *Let $(F_n)_{n\in\mathbb{N}}, F$ be CDFs such that F_n converges to F pointwise and F is continuous. Let $(\lambda_n)_{n\in\mathbb{N}} \in \mathbb{R}$ be a sequence such that $\lambda_n \to \lambda \in \mathbb{R}$ as $n \to \infty$, then*

$$F_n(\lambda_n) \to F(\lambda).$$

Proof. We can write

$$F_n(\lambda_n) - F(\lambda) = F_n(\lambda_n) - F(\lambda_n) + F(\lambda_n) - F(\lambda).$$

F_n converges uniformly to F (as CDFs which converge pointwise to a continuous limit do so uniformly) so $F_n(\lambda_n) - F(\lambda_n) \to 0$ as $n \to \infty$ and $F(\lambda_n) - F(\lambda) \to 0$ because F is continuous. □

S-3.2.2 Proof of Theorem 3.1

In order to facilitate the proof we will first make some further definitions. Firstly, given $H \subseteq \mathcal{H}$ and $T : \mathcal{V} \to \mathbb{R}^L$ define $p_{(k:H)}(T)$ to be the minimum value in the set

$$\{2 - 2\Phi(|T_l(v)|) : (l, v) \in H\} \tag{S-11}$$

where Φ is the CDF of a standard normal distribution. Secondly, given $H \subseteq \mathcal{H}$, let $f_H : \left\{h : \mathcal{V} \to \mathbb{R}^L\right\} \to \mathbb{R}$ send $T \in \left\{h : \mathcal{V} \to \mathbb{R}^L\right\}$ to $\min_{1 \le k \le K \wedge |\mathcal{H}|} t_k^{-1}(p_{(k:H)}(T))$. Thirdly given a function S such that

$$S : \mathcal{V} \to \bigcup_{0 \le j \le L} \mathbb{R}^j,$$

$n \in \mathbb{N}$ and $1 \le k \le |\mathcal{H}|$ we shall define $q_k^n(S)$ to be the kth minimum value in the set

$$\{2 - 2\Phi_{n-r_n}(|S_l(v)|) : v \in \mathcal{V}, l \le \dim(S(v))\}$$

when this is well defined and take $q_k^n(S)$ to be 1 when it is not (i.e. when k is larger than the size of the set). Here for $z \in \bigcup_{1 \le j \le L} \mathbb{R}^j$, $\dim(z)$ denotes the dimension of z. Similarly define $q_k(S)$ to be the kth minimum value in the set

$$\{2 - 2\Phi(|S_l(v)|) : v \in \mathcal{V}, l \le \dim(S(v))\}.$$

Finally we define functions $\phi_n : \left\{h : \mathcal{V} \to \bigcup_{0 \le j \le L} \mathbb{R}^j\right\} \to \mathbb{R}$ which send

$$S \mapsto \min_{1 \le k \le K \wedge |\mathcal{H}|} t_k^{-1}(q_k^n(S))$$

and $\phi : \left\{h : \mathcal{V} \to \bigcup_{0 \le j \le L} \mathbb{R}^j\right\} \to \mathbb{R}$ which sends $S \mapsto \min_{1 \le k \le K \wedge |\mathcal{H}|} t_k^{-1}(q_k(S))$.

With these definitions in mind we are ready to prove Theorem 3.1.

Proof. Defining $\mathfrak{c}'$ as in Section 2.4, $T_n|_{\mathcal{N}}$ converges to $\mathcal{G}(0, \mathfrak{c}')|_{\mathcal{N}}$ in distribution by Theorem S-3.4. As such, using the fact that ϕ_n is the composition of functions which are either continuous or converge uniformly with range $[0, 1]$, by Lemma S-4.3 and the Continuous Mapping Theorem,

$$f_{n,\mathcal{N}}(T_n) = \phi_n(T_n|_{\mathcal{N}}) \xrightarrow{d} \phi(\mathcal{G}(0, \mathfrak{c}')|_{\mathcal{N}}) = f_{\mathcal{N}}(\mathcal{G}(0, \mathfrak{c}')). \tag{S-12}$$

By the same logic, and applying Theorem S-2.7, for sets H such that $\mathcal{N} \subseteq H \subseteq \mathcal{H}$,

$$f_{n,H}(T_n^b) \xrightarrow{d} f_H(\mathcal{G}(0, \mathfrak{c}')). \tag{S-13}$$

This convergence occurs conditional on the data, a fact that we take as implicit in (S-13) and in the rest of the proof. As such, applying Lemma S-2.10, it follows almost surely that $\lambda^*_{\alpha,n,B}(H) \to \lambda_\alpha = F^{-1}(\alpha)$, where F is the CDF of $f_H(\mathcal{G}(0, \mathfrak{c}'))$ using the fact that F is strictly increasing (which follows from the form of f_H and the fact that the density of the multivariate normal distribution is positive everywhere) and continuous, by Lemma S-4.4. Letting F_n be the CDF of $f_{n,\mathcal{N}}(T_n)$ and F_0 be the CDF of $f_{\mathcal{N}}(\mathcal{G}(0, \mathfrak{c}'))$, we have $F_n \longrightarrow F_0$ pointwise using (S-12) and the fact that F_0 is continuous (which follows from Lemma S-4.4). As such, applying Lemma S-3.5 (since F_n and F_0 are CDFs and F_0 is continuous), it follows that for all $\epsilon > 0$,

$$\lim_{n \to \infty} \lim_{B \to \infty} \mathbb{P}\big(f_{n,\mathcal{N}}(T_n) \le \lambda^*_{\alpha,n,B}(H)\big)$$

$$
\begin{aligned}
&\le \lim_{n\to\infty}\lim_{B\to\infty} \mathbb{P}(f_{n,\mathcal{N}}(T_n) \le \lambda_\alpha + \epsilon) + \mathbb{P}(|\lambda^*_{\alpha,n,B}(H) - \lambda_\alpha| > \epsilon)\\
&= \lim_{n\to\infty}\lim_{B\to\infty} F_n(\lambda_\alpha + \epsilon) + \lim_{n\to\infty}\lim_{B\to\infty} \mathbb{P}(|\lambda^*_{\alpha,n,B}(H) - \lambda_\alpha| > \epsilon)\\
&= F_0(\lambda_\alpha + \epsilon) \le F(\lambda_\alpha + \epsilon).
\end{aligned}
$$

Taking ϵ to zero proves the upper bound since F is continuous and so $F(\lambda_\alpha + \epsilon) \to F(\lambda_\alpha) = \alpha$. When $H = \mathcal{N}$ arguing similarly but with $\lambda_\alpha - \epsilon$ yields a lower bound and the desired equality. Note that the final inequality holds because

$$
f_H(\mathcal{G}(0,\mathfrak{c}')) = \min_{1\le k\le K\wedge|\mathcal{H}|} t_k^{-1}(p_{(k:H)}(\mathcal{G}(0,\mathfrak{c}'))) \le \min_{1\le k\le K\wedge|\mathcal{H}|} t_k^{-1}(p_{(k:\mathcal{N})}(\mathcal{G}(0,\mathfrak{c}'))) = f_{\mathcal{N}}(\mathcal{G}(0,\mathfrak{c}'))
$$

and so

$$
F_0(\lambda_\alpha + \epsilon) = \mathbb{P}(f_{\mathcal{N}}(\mathcal{G}(0,\mathfrak{c}')) \le \lambda_\alpha + \epsilon) \le \mathbb{P}(f_H(\mathcal{G}(0,\mathfrak{c}')) \le \lambda_\alpha + \epsilon) = F(\lambda_\alpha + \epsilon).
$$

□

S-3.2.3 Proof of Corollary 3.2

Proof. For any $\epsilon > 0$, and all large enough n and B, we have

$$
\mathbb{P}\left(\min_{1\le k\le K\wedge|\mathcal{H}|} t_k^{-1}(p_{(k:\mathcal{N})}(T_n)) \le \lambda^*_{\alpha,n,B}(\mathcal{H})\right) \le \alpha + \epsilon
$$

and so, arguing as in Blanchard et al. (2020),

$$
\mathbb{P}\big(|H \cap \mathcal{N}| \le \overline{V}_{\alpha,n,B}(H),\ \forall H \subseteq \mathcal{H}\big) \le 1 - \alpha - \epsilon.
$$

The result follows by sending ϵ to zero. □

S-3.2.4 Proof of Theorem 3.3

The proof is similar to that of Proposition 4.5 of Blanchard et al. (2020).

Proof. Let

$$
\Omega_n = \big\{p^n_{(k:\mathcal{N})}(T_n) \ge t_k(\lambda^*_{\alpha,n,B}(\mathcal{N})) \text{ for all } 1 \le k \le K\big\}.
$$

Then by Theorem 3.1,

$$
\lim_{n\to\infty}\lim_{B\to\infty} \mathbb{P}(\Omega_n) = 1 - \alpha.
$$

We claim that on the event Ω_n, $\mathcal{N} \subseteq \hat{H}_n$. We prove this inductively, using the notation from Algorithm 1. $\mathcal{N} \subseteq H^{(0)}$ trivially. Assuming that $\mathcal{N} \subseteq H^{(j-1)}$ for some $j \in \mathbb{N}$, it follows that $p^n_{(k:H^{(j-1)})} \le p^n_{(k:\mathcal{N})}$ i.e. that $f_{n,H^{(j-1)}} \le f_{n,\mathcal{N}}$. In particular,

$$
\lambda^*_{\alpha,n,B}(H^{(j-1)}) \le \lambda^*_{\alpha,n,B}(\mathcal{N})
$$

and thus (since we are on Ω_n and t_1 is increasing),

$$
p^n_{(1:\mathcal{N})}(T_n) \ge t_1(\lambda^*_{\alpha,n,B}(H^{(j-1)}))
$$

which implies that $\mathcal{N} \subseteq H^{(j)}$. Thus $\mathcal{N} \subseteq \hat{H}_n$ and so for all $1 \le k \le K$,

$$
p^n_{(k:\mathcal{N})}(T_n) \ge t_k(\lambda^*_{\alpha,n,B}(\mathcal{N})) \ge t_k(\lambda^*_{\alpha,n,B}(\hat{H}_n))
$$

and so

$$
f_{n,\mathcal{N}}(T_n) \ge \lambda^*_{\alpha,n,B}(\hat{H}_n).
$$

The post hoc bound follows as in the proof of Corollary 3.2. □

S-4 Further Theory

S-4.1 Parametric Approaches

In this section we will discuss two existing parametric[1] approaches to simultaneous FDP inference which are based on the Simes inequality (S-15). The first one is the original Simes post hoc bound introduced in Goeman and Solari (2011). The second one is the method of Rosenblatt et al. (2018) and Goeman et al. (2019). It corresponds to an improvement on the basic Simes bound that is adaptive to the proportion of true null hypotheses - i.e. it is a step-down version of the Simes bound. This method has been applied to brain imaging data in Rosenblatt et al. (2018), and is called **ARI** which stands for "**All Resolutions Inference**". Both methods can be conveniently formulated in terms of the bound $\overline{V}$ defined in (6), associated to the linear template family $(t_k)_{1\le k\le m}$, where $t_k(\lambda) = \lambda k/m$, i.e.

$$\overline{V}_\lambda(S) = \min_{1\le k\le m}\left\{\sum_{i\in S} 1\left[p_i \ge \frac{\lambda k}{m}\right] + k - 1\right\}. \tag{S-14}$$

As noted by Blanchard et al. (2020), the Simes post hoc bound of Goeman and Solari (2011) is simply $\overline{V}_\alpha$. Moreover, letting $\overline{\alpha} = \alpha m/h(\alpha)$, where

$$h(\alpha) = \max\left\{i \in \{1,\dots,m\}, \forall j \in \{1,\dots,i\}, p_{(m-i+j)} > \frac{\alpha j}{i}\right\},$$

the ARI bound of Goeman et al. (2019) is $\overline{V}_{\overline{\alpha}}$. The quantity $h(\alpha)$ is called the Hommel factor (Hommel, 1988) and can be interpreted as a $(1-\alpha)$-level upper confidence bound on $|\mathcal{N}|$, the number of true null hypotheses.

If the null p-values satisfy positive regression dependence then both of these methods result in simultaneous $(1-\alpha)$-level FDP control. This is shown formally in Goeman and Solari (2011) and Goeman et al. (2019) via closed testing and can also be shown to hold by combining the Simes inequality with the joint error rate framework of Section 2.3. To see this note that if the null p-values are positive regression dependent (Sarkar et al., 2008), then the Simes inequality is satisfied, that is:

$$\mathbb{P}\left(\exists 1 \le k \le |\mathcal{N}| : p^n_{(k:\mathcal{N})} \le \frac{\alpha k}{|\mathcal{N}|}\right) \le \alpha, \tag{S-15}$$

with equality if the null p-values are independent.

In particular taking $\lambda = \alpha$, and noting that $|\mathcal{N}| \le m$, the Simes inequality implies that (7) holds (taking $K = m$ and $(t_k)_{1\le k\le m}$ to be the linear reference family). Moreover, Goeman et al. (2019)'s Lemma 2 implies that if the null p-values satisfy positive regression dependence, then

$$\mathbb{P}\left(\exists 1 \le k \le |\mathcal{N}| : p^n_{(k:\mathcal{N})} \le \frac{\alpha k}{h(\alpha)}\right) \le \alpha. \tag{S-16}$$

Thus taking $\lambda = \overline{\alpha} = \alpha m/h(\alpha)$, it follows that (7) holds with respect to the linear reference family. In particular the Simes procedure, which uses $\overline{V}_\alpha$ as a bound, and ARI, which uses $\overline{V}_{\overline{\alpha}}$, provide simultaneous $(1-\alpha)$-level control of the FDP.

In our results, presented in the following sections, we compare the performance of the non-parametric bootstrap approach to these parametric alternatives.

[1] Here we use the term parametric to indicate that dependency assumptions on the data are required in order for the methods to be valid.

S-4.2 FWER inference

In brain imaging it is also desirable to control the familywise error rate (FWER) over the hypotheses. This corresponds to performing multiple testing inference on the data and returning a set of active hypotheses $R \subseteq \mathcal{H}$ such that

$$\text{FWER} = \mathbb{P}(R \cap \mathcal{N}) \leq \alpha.$$

When a single test is being used (for a single contrast or an F-test at each voxel), brain imaging studies typically use a permutation based procedure (Winkler et al., 2014) in order to control these error rates. In the case of multiple contrasts this approach is not always applicable because exchangeability can break down - see Section S-4.3. However the residual bootstrap can be used instead. In particular we have the following theorem which follows as a corollary of Theorem 3.1 by taking $K = 1$ and taking $t_1(x) = x$ for $x \in [0, 1]$.

Theorem S-4.1. *For $\alpha \in (0, 1)$ and $n, B \in \mathbb{N}$, let $\lambda'_{\alpha,n,B}$ be the α-quantile of the bootstrap distribution (based on B bootstraps) of*

$$p_{1:\mathcal{H}}(T_n) = \min_{(l,v)\in\mathcal{H}} p_{n,l}(v).$$

Let $R_{n,B} = \left\{(l, v) \in \mathcal{H} : p_{n,l}(v) \leq \lambda'_{\alpha,n,B}\right\}$. Then

$$\lim_{n\to\infty} \lim_{B\to\infty} \mathbb{P}(R_{n,B} \cap \mathcal{N}) = \lim_{n\to\infty} \lim_{B\to\infty} \mathbb{P}\left(\min_{(l,v)\in\mathcal{N}} p_{n,l}(v) \leq \lambda'_{\alpha,n,B} \right) \leq \alpha.$$

So choosing $R_{n,B}$ as the rejection set provides asymptotic strong control of the FWER.

Moreover, a step-down version of this result follows as a corollary to Theorem 3.3. For these results the control of the FWER does not occur simultaneously with the control of the joint error rate. However let us revert to the general setting of the paper and take $\lambda^*_{\alpha,n,B}$ to be the $\alpha-$quantile of the bootstrap distribution of $f_{n,\mathcal{N}}$ (as defined in the statement of Theorem 3.1). Then FWER control is automatically entailed with control of the joint error rate by using the rejection set $R = \left\{(l, v) \in \mathcal{H} : p_{n,l} \leq t_1^{-1}(\lambda^*_{\alpha,n,B})\right\}$. When $K > 1$, typically $t_1^{-1}(\lambda^*_{\alpha,n,B})$ will be less than the value of $\lambda'_{\alpha,n,B}$ from Theorem S-4.1 so this will result in less power but comes with the advantage of holding jointly with control of the joint error rate. Which version is to be preferred depends on which error rate one desires to control.

S-4.3 Permutation in the Linear Model

Here we show that under the alternative that $\beta \neq 0$ at a given point (e.g. voxel or gene), naively permuting the data (Manly (1986)) does not generate data under the global null even when the noise is exchangeable.

Claim S-4.2. *Suppose that the global null is not true, i.e. $\beta \neq 0$, then permuting Y is not equivalent to generating data under the global null (and so cannot be used to generate under the null and provide strong control over contrasts).*

Proof. Let P be a permutation matrix, then

$$PY = P(X\beta + \epsilon) = PX\beta + P\epsilon.$$

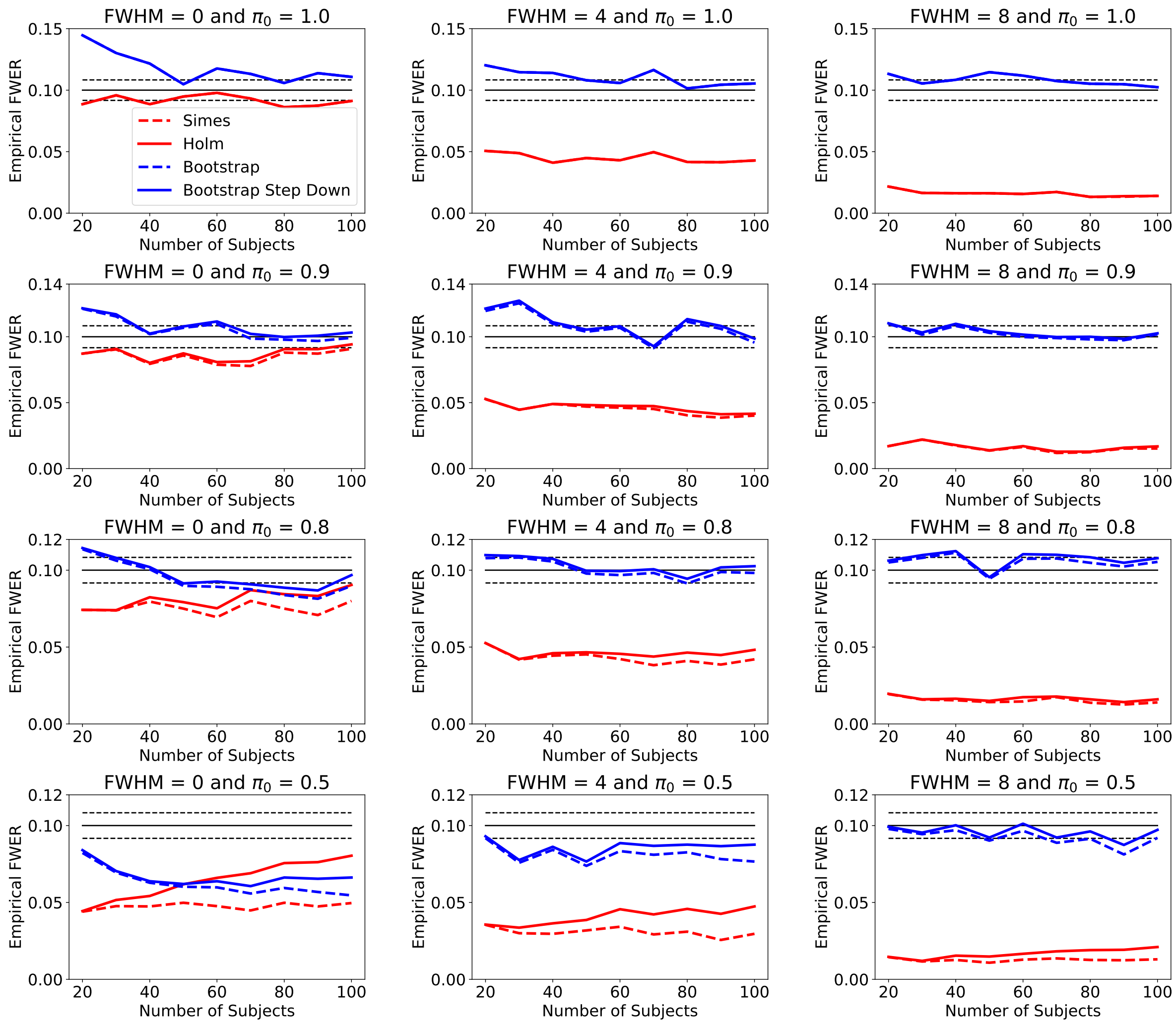

Figure S-1: Direct FWER control using the different methods. Here the parametric procedures are Bonferroni and its step-down: Holm (1979).

Now

$$P\epsilon \sim \epsilon$$

by exchangability. However $PX\beta \neq 0$ so it is not true that $PY \sim \epsilon$ which is what we want (because we need to simulate under the null model, in order to provide strong control over contrasts). $PX\beta$ is a random variable (due to randomness in P) with a non-zero mean and variance. So regressing PY against X gives linear model coefficients of

$$\begin{aligned}\hat{\beta}^* &= (X^TX)^{-1}X^TPY = (X^TX)^{-1}X^TP(X\beta+\epsilon)\\ &= (X^TX)^{-1}X^TPX\beta + (X^TX)^{-1}X^TP\epsilon.\end{aligned}$$

Now, under exchangeability,

$$(X^TX)^{-1}X^TP\epsilon \sim (X^TX)^{-1}X^T\epsilon$$

which indeed is the distribution of the linear model estimates under the null, however

$$(X^TX)^{-1}X^TPX\beta \neq 0$$

which causes a problem. □

S-4.4 Additional Lemmas for the proofs

Lemma S-4.3. *Suppose that $(Z_n)_{n\in\mathbb{N}}, Z$ are $\mathbb{R}^M$ valued random variables, for some $M \in \mathbb{N}$. Let $(f_n)_{n\in\mathbb{N}}, f$ be functions from $\mathbb{R}^M \to I$ for some compact set $I \subset \mathbb{R}$. Suppose that f_n converges uniformly to f, that f is continuous and that $Z_n \xrightarrow{d} Z$, then*

$$f_n(Z_n) \xrightarrow{d} f(Z).$$

Proof. Given any continuous and bounded function $h : \mathbb{R} \to \mathbb{R}$,

$$|\mathbb{E}[h(f_n(Z_n))] - \mathbb{E}[h(f(Z))]| \leq |\mathbb{E}[h(f_n(Z_n))] - \mathbb{E}[h(f(Z_n))]| + |\mathbb{E}[h(f(Z_n))] - \mathbb{E}[h(f(Z))]|.$$

The set I is compact and so the restriction of h to I is uniformly continuous. So for any $\epsilon > 0$ there is some δ such that $|h(x) - h(y)| < \epsilon$ for all $x, y \in I$ such that $|x - y| < \delta$. By uniform convergence, there is some $N \in \mathbb{N}$ such that for all $n > N$, $|f_n(z) - f(z)| < \delta$ for all $z \in \mathbb{R}^L$. The functions $(f_n)_{n\in\mathbb{N}}, f$ take values within I so it follows that

$$|\mathbb{E}[h(f_n(Z_n))] - \mathbb{E}[h(f(Z_n))]| \leq \mathbb{E}[|h(f_n(Z_n)) - h(f(Z_n))|] < \mathbb{E}[\epsilon] = \epsilon.$$

So this term converges to zero as $n \to \infty$. The second term: $|\mathbb{E}[h(f(Z_n))] - \mathbb{E}[h(f(Z))]|$ also converges to zero as $h \circ f$ is a continuous bounded function and $Z_n \xrightarrow{d} Z$ as $n \to \infty$ (by applying the Portmanteau Lemma). Thus, as $n \to \infty$,

$$\mathbb{E}[h(f_n(Z_n))] \to \mathbb{E}[h(f(Z))].$$

Since this holds for any continuous bounded h the result follows by Portmanteau. □

Lemma S-4.4. *Let F_H be the CDF of $\min_{1\leq k\leq K\wedge|\mathcal{H}|} t_k^{-1}(p_{(k:H)}(T))$ where $T \sim \mathcal{G}(0, \mathfrak{c}')$ and $\mathfrak{c}'$ is defined as in Section 2.4, then F_H is continuous.*

Proof. It is sufficient to show that for all $\lambda \in \mathbb{R}$, $\mathbb{P}\big(\min_{1\leq k\leq K\wedge|\mathcal{H}|} t_k^{-1}(p_{(k:|\mathcal{H}|)}(T)) = \lambda\big) = 0$. To show this, choose $\lambda \in \mathbb{R}$, then

$$\begin{aligned}
\mathbb{P}\left(\min_{1\leq k\leq K\wedge|\mathcal{H}|} t_k^{-1}(p_{(k:H)}(T)) = \lambda\right) &\leq \mathbb{P}\big(\exists 1 \leq k \leq |\mathcal{H}| \text{ s.t. } t_k^{-1}(p_{(k:H)}(T)) = \lambda\big)\\
&= \mathbb{P}\big(\exists 1 \leq k \leq m \text{ s.t. } p_{(k:H)}(T) = t_k(\lambda)\big)\\
&\leq \sum_{k=1}^m \mathbb{P}\big(p_{(k:H)}(T) = t_k(\lambda)\big)\\
&\leq \sum_{k=1}^m \mathbb{P}(\exists (l, v) \in \mathcal{H} : 2(1 - \Phi(|T_l(v)|)) = t_k(\lambda))\\
&\leq \sum_{k=1}^m \sum_{(l,v)\in\mathcal{H}} \mathbb{P}(2(1 - \Phi(|T_l(v)|)) = t_k(\lambda)).
\end{aligned}$$

Now given $(l, v) \in \mathcal{H}$ and $1 \leq k \leq m$,

$$\mathbb{P}(2(1 - \Phi(|T_l(v)|)) = t_k(\lambda)) = \mathbb{P}\big(|T_l(v)| = \Phi^{-1}(1 - t_k(\lambda)/2)\big) = 0$$

since $T_l(v)$ is a Gaussian random variable. The result follows. □

S-5 fMRI data pre-processing

Participants underwent a working memory task in which they were shown images asked to remember them. They were reshown them at a subsequent point. This is known as an m-back task when $m \in \mathbb{N}$ is number of intervals between when each image is shown and then repeated - see Barch et al. (2013) for further details. The data we have consists of images that give the difference between the brain scans of participants under the 2-back and 0-back conditions. The data was pre-processed at the first level using nilearn. the images were then smoothed using an isotropic Gaussian kernel with an FWHM of 4/3 voxels (4 mm).

S-6 Further figures

S-6.1 Simes vs ARI for the IQ contrast

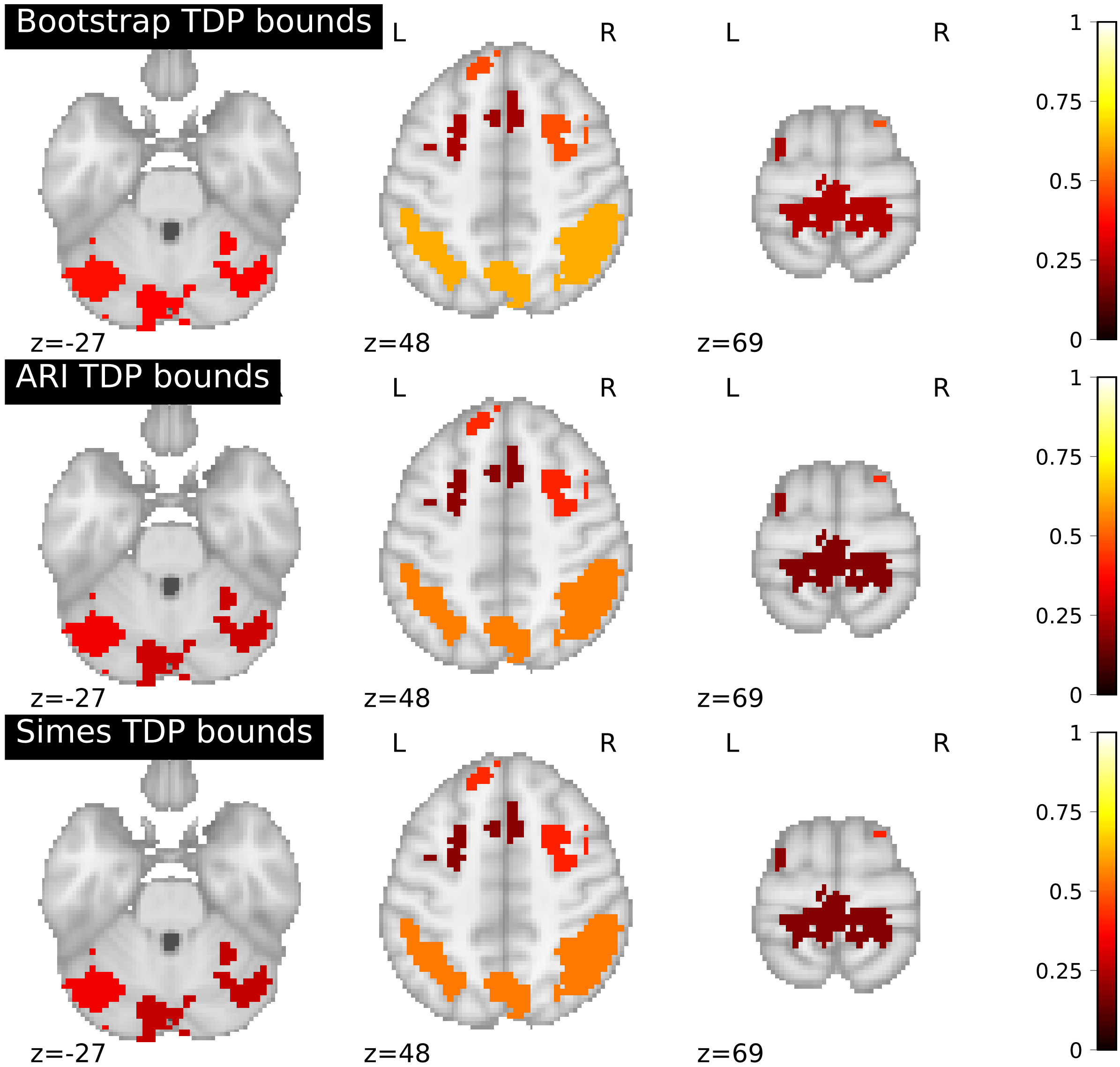

Figure S-2: TDP bounds within clusters for the contrast for IQ in the linear regression model fit to the HCP data. Each cluster is shaded a single colour which is the lower bound on the TDP. The upper panel gives the TDP bounds within each cluster provided by the bootstrap procedure. The lower panels gives the bounds provided by using ARI and the Simes procedure. The bounds given by the bootstrap are larger (as indicated by the light colours) indicating that the method is more powerful. (Note that the step-down bootstrap gave the same bounds as the bootstrap and so is not shown.) Note that these images are 2D slices through the 3D brain and so voxels that are part of the same cluster are not necessarily connected.

S-6.2 The contrast for sex

Less activation is found for the contrast of sex in the linear model fit to the HCP data. In this case only a single cluster above the cluster defining threshold has non-zero lower bound. The bound provided is the same for all the parametric and bootstrap methods that we consider, in particular they all conclude that at least one of the 17 voxels within this cluster has non-zero activation. The cluster (and its TDP) is illustrated in Figure S-3.

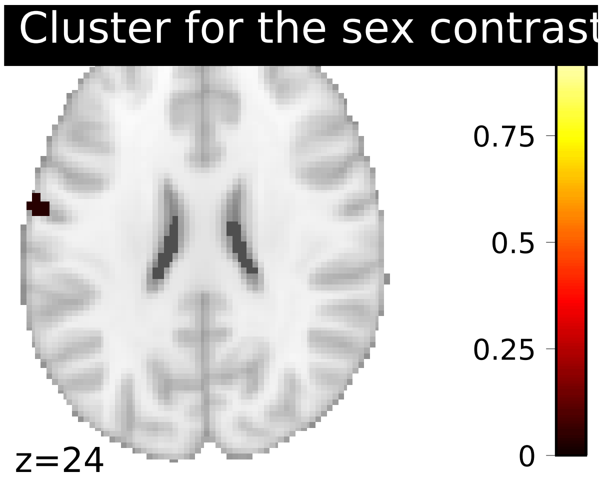

Figure S-3: Illustrating the cluster in the sex contrast with non-zero activation.

S-6.3 Illustrating the simulation setup

Figure S-4: Illustrating the simulation setup for a domain of size [25,25] and a smoothness of 4 pixels. Left: the signal for the first contrast. Right: a realisation of one of the subjects in G_2.

S-6.4 Additional JER control plots

In this section we present the results of the simulations to consider JER control where the domain of the data in the simulations is 25 by 25 or 100 by 100 rather than 50 by 50. The results for the 25 by 25 simulations are shown in Figure S-5 and those for the 100 by 100 simulations are shown in Figure S-6. The results are similar to those in the main text.

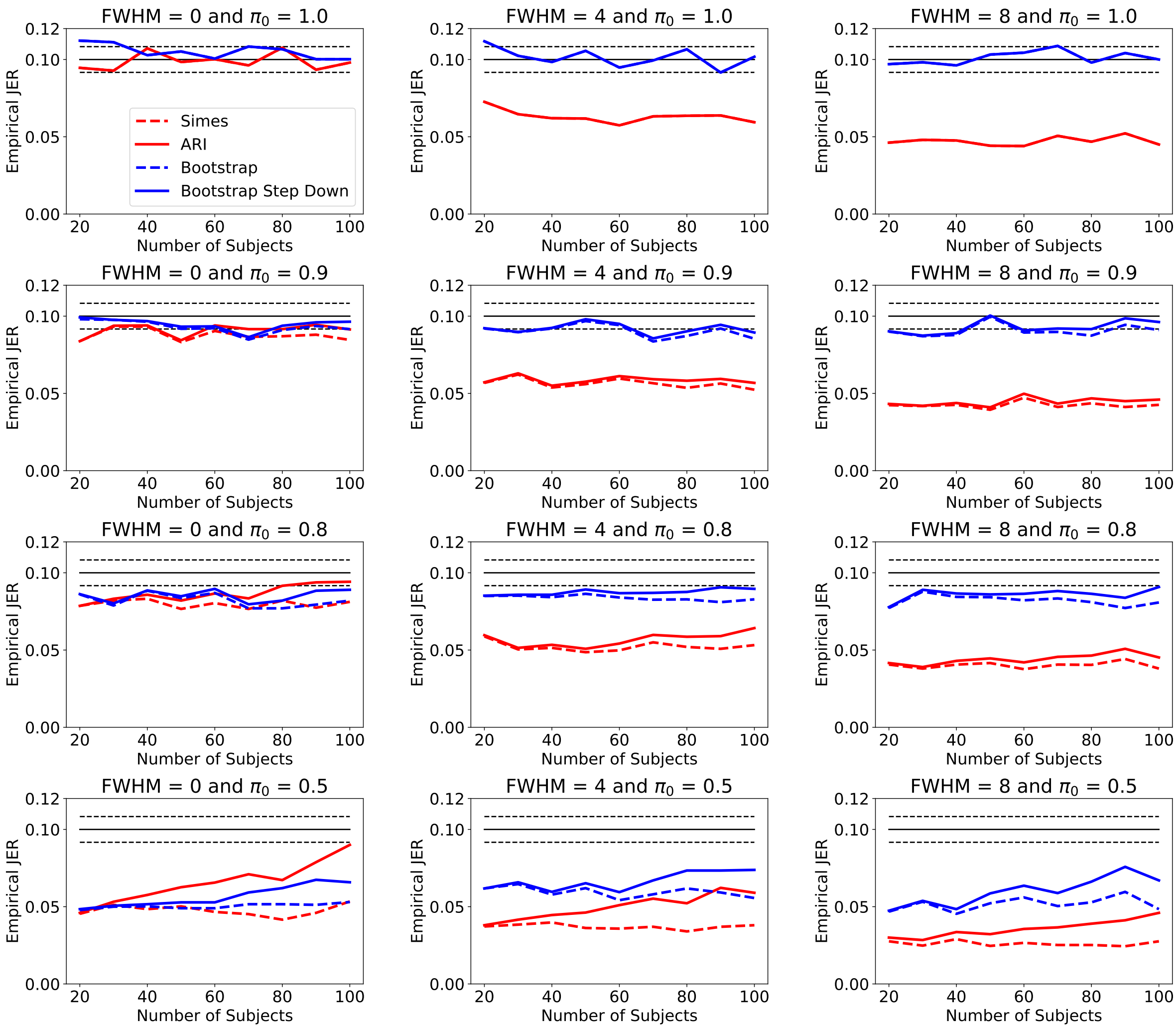

Figure S-5: Empirical joint error rate for the 25 by 25 simulations.

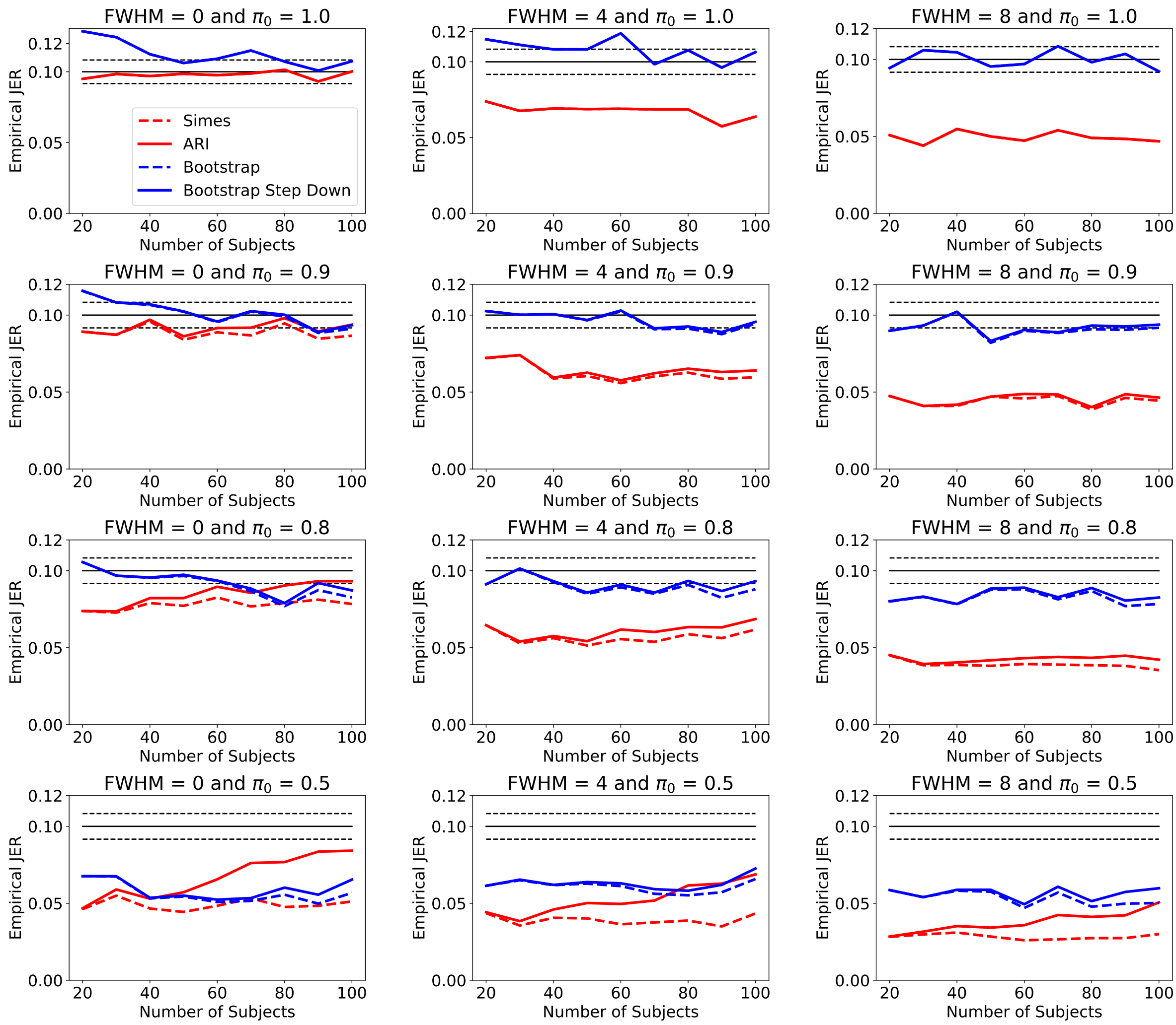

Figure S-6: Empirical joint error rate for the 100 by 100 simulations.

S-6.5 Additional power plots

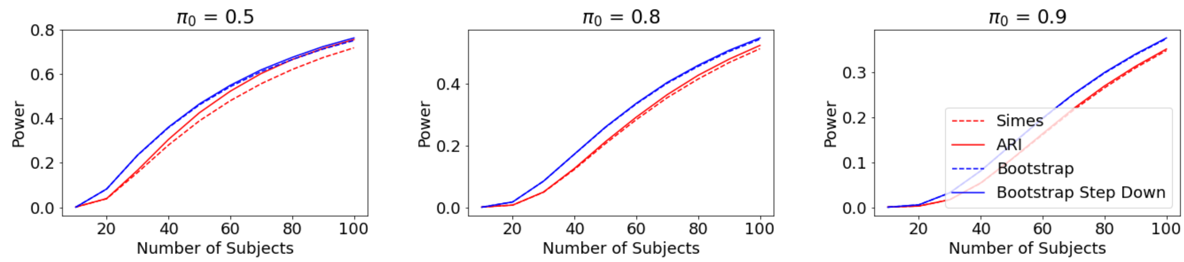

Figure S-7: Plotting the power of the different methods against the numbers of a subjects for setting 3, i.e. taking $R = \{(l, v) : p_{n,l}(v) \leq 0.05\}$ in (19).

References

Deanna M Barch, Gregory C Burgess, Michael P Harms, Steven E Petersen, Bradley L Schlaggar, Maurizio Corbetta, Matthew F Glasser, Sandra Curtiss, Sachin Dixit, Cindy Feldt, et al. Function in the human connectome: task-fmri and individual differences in behavior. *Neuroimage*, 80:169–189, 2013.

Gilles Blanchard, Pierre Neuvial, Etienne Roquain, et al. Post hoc confidence bounds on false positives using reference families. *Annals of Statistics*, 48(3):1281–1303, 2020.

Samuel Davenport, Fabian Telschow, Thomas E. Nichols, and Armin Schwarzman. Confidence regions for the location of peaks of a smooth random field. 2021.

Laurens De Haan and Ana Ferreira. *Extreme value theory: an introduction*, volume 21. Springer, 2006.

Daniel J Eck. Bootstrapping for multivariate linear regression models. *Statistics & Probability Letters*, 134:141–149, 2018.

David A Freedman. Bootstrapping regression models. *The Annals of Statistics*, 9(6): 1218–1228, 1981.

Jelle J. Goeman and Aldo Solari. Multiple Testing for Exploratory Research. *Statistical Science*, 26(4):584–597, 2011. ISSN 0883-4237. doi: 10.1214/11-STS356.

Jelle J Goeman, Rosa J Meijer, Thijmen JP Krebs, and Aldo Solari. Simultaneous control of all false discovery proportions in large-scale multiple hypothesis testing. *Biometrika*, 106(4):841–856, 2019.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scand. J. Statist.*, 6(2):65–70, 1979. ISSN 0303-6898.

Gerhard Hommel. A stagewise rejective multiple test procedure based on a modified bonferroni test. *Biometrika*, 75(2):383–386, 1988.

Michael R Kosorok. Bootstraps of sums of independent but not identically distributed stochastic processes. *Journal of Multivariate Analysis*, 84(2):299–318, 2003.

Bryan FJ Manly. Randomization and regression methods for testing for associations with geographical, environmental and biological distances between populations. *Researches on Population Ecology*, 28(2):201–218, 1986.

Jonathan D Rosenblatt, Livio Finos, Wouter D Weeda, Aldo Solari, and Jelle J Goeman. All-resolutions inference for brain imaging. *Neuroimage*, 181:786–796, 2018.

Sanat K Sarkar et al. On the simes inequality and its generalization. In *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen*, pages 231–242. Institute of Mathematical Statistics, 2008.

Fabian Telschow, Samuel Davenport, and Armin Schwartzman. Functional delta residuals. 2020.

Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

A.W. van der Vaart. *Asymptotic Statistics*. 1998.

Anderson M. Winkler, Gerard R. Ridgway, Matthew A. Webster, Stephen M. Smith, and Thomas E. Nichols. Permutation inference for the general linear model. *NeuroImage*, 92:381–397, 2014. ISSN 10959572. doi: 10.1016/j.neuroimage.2014.01.060.