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

Samuel Davenport, Bertrand Thirion, Pierre Neuvial

August 21, 2022

Contents

S-1 Consistency of the bootstrap in the linear model **S-2**
- S-1.1 Further theory for random fields S-2
- S-1.2 Lindeberg Central Limit Theorem S-2
- S-1.3 Convergence in the linear model S-3
- S-1.4 Proof of Theorem 3.1 . . . . . S-4

S-2 Proofs for the main text **S-6**
- S-2.1 Proofs for Section 2 . . . . . S-6
  - S-2.1.1 Proof of Claim 2.4 . . . . . S-6
- S-2.2 Proofs for Section 3 . . . . . S-7
  - S-2.2.1 Convergence in the Linear Model S-7
  - S-2.2.2 Proof of Theorem 3.1 . . . . . S-8
  - S-2.2.3 Proof of Theorem 3.2 . . . . . S-9
- S-2.3 Proof of Theorem 3.3 . . . . . S-9
- S-2.4 Proofs for Section 4 . . . . . S-10
  - S-2.4.1 Setup S-10
  - S-2.4.2 Proof of Theorem 4.1 . . . . . S-10
  - S-2.4.3 Proof of Corollary 4.2 . . . . . S-12
  - S-2.4.4 Proof of Theorem 4.3 . . . . . S-12

S-3 Further Theory **S-12**
- S-3.1 FWER inference S-12
- S-3.2 Permutation in the Linear Model S-13
- S-3.3 Additional Lemmas for the proofs S-14

S-4 fMRI data pre-processing **S-15**

S-5 Further figures **S-16**
- S-5.1 Simes vs ARI for the IQ contrast S-16
- S-5.2 The contrast for sex S-17
- S-5.3 Illustrating the simulation setup S-17
- S-5.4 Additional JER control plots S-18
- S-5.5 Additional power plots S-19

S-1 Consistency of the bootstrap in the linear model

S-1.1 Further theory for random fields

Given two random fields $f, g : \mathcal{V} \to \mathbb{R}^L$ operations of addition and subtraction can be performed pointwise and so $f + g$ and $f - g$ 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, fg and f/g are well-defined.

Given $D \in \mathbb{N}$, suppose that $\mathcal{V} = \{u_1, \dots, u_V\}$ for some $V \in \mathbb{N}$ and $u_1, \dots, u_V \in \mathbb{R}^D$. For $L \in \mathbb{N}$, let $f : \mathcal{V} \to \mathbb{R}^L$ be a random field. Then we define $\mathrm{vec}(f) \in \mathbb{R}^{LV}$ to be the vector whose $((i-1)L + j)$th element is $f_j(u_i)$ 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: $((f_n)_{n\in\mathbb{N}}, f)$ of random fields from $\mathcal{V}$ to $\mathbb{R}^L$ we say that f_n converges to f in distribution (resp. probability/almost surely) if $\mathrm{vec}(f_n)$ converges in distribution (resp. probability/almost surely) to $\mathrm{vec}(f)$. We will write this as $f_n \xrightarrow{d} f$ (resp. $f_n \xrightarrow{\mathbb{P}} f / f_n \xrightarrow{a.s.} f$). Given such a sequence we will write $f_{n,j}$ $(1 \leq j \leq L)$ to denote its components.

Definition S-1.1. For $L, L' \in \mathbb{N}$ let $f : \mathcal{V} \to \mathbb{R}^L$ be a random field then we define the random field Mf which sends $v \in \mathcal{V}$ to $Mf(v) \in \mathbb{R}^{L'}$.

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

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

Moreover if f is Gaussian then so is Mf.

S-1.2 Lindeberg Central Limit Theorem

In what follows we will require CLTs to hold to ensure that we have convergence in our setting. To do so we state and prove Proposition S-1.4 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-1.3. *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 $a \in \mathbb{R}$,*

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

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

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

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

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

Proposition S-1.4. *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 $u, v \in \mathcal{V}$,

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

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

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

We can expand the left hand side as

$$\begin{aligned}&\sum_{i=1}^{k_n}\mathbb{E}\left[\sum_{v\in\mathcal{V}}\|a_{ni}\xi_{n,i}(v)\|^2 1\left[\sum_{u\in\mathcal{V}}\|a_{ni}\xi_{n,i}(u)\|^2 > \gamma^2\right]\right] && (1)\\ &\le \sum_{i=1}^{k_n}\sum_{v\in\mathcal{V}}\|a_{ni}\|^2\mathbb{E}\left[\xi_{n,i}(v)^2\sum_{u\in\mathcal{V}}1\left[\|a_{ni}\||\xi_{n,i}(u)| > \gamma|\mathcal{V}|^{-1/2}\right]\right] && (2)\\ &= \sum_{i=1}^{k_n}\|a_{ni}\|^2\sum_{u,v\in\mathcal{V}}\mathbb{E}\left[\xi_{n,i}(v)^2 1\left[\|a_{ni}\||\xi_{n,i}(u)| > \gamma|\mathcal{V}|^{-1/2}\right]\right] && (3)\\ &\le C\sum_{i=1}^{k_n}\|a_{ni}\|^{2+K/q} && (4)\end{aligned}$$

for some fixed constant $C > 0$, chosen in accordance with Lemma S-1.3. This bound converges to zero as $n \to \infty$. □

S-1.3 Convergence in the linear model

In this section we demonstrate how to obtain conditional convergence of the least squares estimates in the linear model. We require the following assumption from the main text in what follows, which we have repeated here for reference.

Assumption 1.

a. For $n \in \mathbb{N}$, $X_n = [x_1, \dots, x_n]^T$ for a sequence of i.i.d vectors $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{R}^p$ such that $\mathbb{E}\left[\|x_1\|^{5/2}\right] < \infty$ and whose multivariate density is bounded above.

b. $(\epsilon_n)_{n\in\mathbb{N}}$ is an i.i.d sequence of 1-dimensional random fields on $\mathcal{V}$ which is independent of $(x_n)_{n\in\mathbb{N}}$ and such that $\max_{v\in\mathcal{V}} \mathbb{E}[\epsilon_1(v)^4] < \infty$ and $\min_{v\in\mathcal{V}} \mathrm{var}(\epsilon_1(v)) > 0$.

Lemma S-1.5. *Suppose that $(X_m)_{m\in\mathbb{N}}$ and $(\epsilon_m)_{m\in\mathbb{N}}$ satisfy Assumption 1. Then conditional on $(X_m)_{m\in\mathbb{N}}$, for almost all sequences $(X_m)_{m\in\mathbb{N}}$, as $n \to \infty$*

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

Proof.

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

Applying Proposition S-1.4 (with $\eta = 2$ and $K = 1$), conditional on $(X_m)_{m\in\mathbb{N}}$, for almost all sequences $(X_m)_{m\in\mathbb{N}}$, as $n \to \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n x_i \epsilon_i \xrightarrow{d} \mathcal{G}(0, \mathfrak{c}_\epsilon \Sigma_X)$$

since $\frac{X_n^T X_n}{n} \longrightarrow \Sigma_X$ almost surely as $n \to \infty$. $\left(\frac{X_n^T X_n}{n}\right)^{-1}$ converges almost surely to Σ_X^{-1} by Lemma S-2.2 and so the result follows by Slutsky. □

S-1.4 Proof of Theorem 3.1

Here we prove Theorem 3.1 from the main text.

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-2.2, $\left(\frac{X_n^T X_n}{n}\right)^{-1}$ converges a.s. to Σ_X^{-1}. Moreover, $(E_{n,i}^b)_{n\in\mathbb{N},1\leq i\leq n}$ is a triangular array which is mean-zero and i.i.d within rows so if we can show that its covariance converges the result will follow by applying Proposition S-1.4 with $\eta = 2$ and $K = 1$. To demonstrate this convergence, for each $u, v \in \mathcal{V}$, conditional on $(X_n, Y_n)_{n\in\mathbb{N}}$

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

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.

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

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

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

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

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^n \hat{E}_{n,j}(u) &= \frac{1}{n}1_n^T(I_n - P_n)E_n(u) = \frac{1}{n}1_n^T E_n(u) - \frac{1}{n}1_n^T X_n(X_n^T X_n)^{-1}X_n^T E_n(u) \\ &= \frac{1}{n}\sum_{i=1}^n \epsilon_i(u) - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)\left(\frac{X_n^T X_n}{n}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n x_i\epsilon_i(u)\right)\end{aligned}$$

which converges almost surely to 0 as $n \to \infty$. To show that the variance converges, note 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, $\frac{1}{\sqrt{n}}\sum_{i=1}^n E_{n,i}^b$ converges in distribution and, dividing by $\sqrt{n}$, it follows that $\frac{1}{n}\sum_{i=1}^n E_{n,i}^b$ converges almost surely to zero as $n \to \infty$. For the first term, note that

$$\mathbb{E}(E_{n,i}^b)^2 = \sum_{j=1}^n \frac{1}{n}\left(\hat{E}_{n,j} - \frac{1}{n}\sum_{l=1}^n \hat{E}_{n,l}\right)^2 \tag{5}$$

which converges to σ^2 almost surely as $n \to \infty$. As such the result follows by the triangular weak law of large numbers so long as we can demonstrate that $\sup_{n\in\mathbb{N},1\le i\le n}\mathbb{E}(E_{n,i}^b)^4 < \infty$. To show this note that for each $n \in \mathbb{N}$ and $1 \le i \le n$ and $1 \le b \le B$,

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

Now $\|P_nE_n\|$ converges in probability to 0 by Lemma S-1.6 (see below) and so

$$\max_{1\le l\le n}|(P_nE_n)_l| \xrightarrow{\mathbb{P}} 0,$$

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

$$\lim_{k\to\infty}\mathbb{P}\left(\max_{n\ge k}\max_{1\le l\le n}|(P_nE_n)_l| > M\right) \to 0 \tag{6}$$

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 (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

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

We can thus bound $\mathbb{E}(E^b_{n,i})^4$ 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 $\mathbb{E}(E^b_{n,i})^4$ is bounded, a fact that is true almost everywhere since $\mathbb{P}(A) = 1$. $\square$

Lemma S-1.6. *Under Assumption 1, 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. We have,

$$P_n E_n = X_n (X_n^T X_n)^{-1} X_n^T E_n = \frac{X_n}{n^{0.45}} \left(\frac{X_n^T X_n}{n} \right)^{-1} \left(\frac{X_n^T E_n}{n^{0.55}} \right).$$

Thus,

$$\|P_n E_n\| = \left\| \frac{X_n}{n^{0.45}} \left(\frac{X_n^T X_n}{n} \right)^{-1} \left(\frac{X_n^T E_n}{n^{0.55}} \right) \right\| \leq \left\| \frac{X_n}{n^{0.45}} \right\| \left\| \left(\frac{X_n^T X_n}{n} \right)^{-1} \right\| \left\| \left(\frac{X_n^T E_n}{n^{0.55}} \right) \right\|.$$

$\frac{X_n^T E_n}{\sqrt{n}}$ converges in distribution (see e.g. the proof of Lemma S-1.5) so $\left\| \left(\frac{X_n^T E_n}{n^{0.55}} \right) \right\| \xrightarrow{\mathbb{P}} 0$ and $\left(\frac{X_n^T X_n}{n} \right)^{-1}$ converges almost surely to Σ_X^{-1} by Lemma S-2.2. Applying the Gershgorin circle theorem and the AM-RM inequality, we have

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

$n^{-0.45} \max_{1 \leq i \leq n} \|x_i\| \xrightarrow{a.s.} 0$ since $\mathbb{E}(\|x_1\|^{5/2}) < \infty$, so in particular $\|n^{-0.45} X_n\| \xrightarrow{a.s.} 0$. Combining these results and using Slutsky, it follows that $\|P_n E_n\| \xrightarrow{\mathbb{P}} 0$. $\square$

S-2 Proofs for the main text

S-2.1 Proofs for Section 2

S-2.1.1 Proof of Claim 2.4

Proof. The event

$$\{|R_k(\lambda) \cap \mathcal{N}| > k - 1\} = \{|\{(l, v) \in \mathcal{N} : p_{n,l}(v) \leq t_k(\lambda)\}| > k - 1\}$$

$$= \left\{p^n_{(k:\mathcal{N})} \leq t_k(\lambda)\right\} = \left\{t_k^{-1}(p^n_{(k:\mathcal{N})}) \leq \lambda\right\}$$

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-2.1. *This claim can be generalized to arbitrary ζ_k. 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. As such the results in Section 4 can trivially be generalized to arbitrary ζ_k.

S-2.2 Proofs for Section 3

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

Lemma S-2.2. *Suppose that $(X_n)_{n\in\mathbb{N}}$ satisfies Assumption 1a and let $\Sigma_X = \mathbb{E}\big[x_1x_1^T\big]$, then Σ_X is invertible and*

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

S-2.2.1 Convergence in the Linear Model

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-2.3. *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

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

for $u, v \in \mathcal{V}$. $\left(\frac{X_n^TX_n}{n}\right)^{-1}$ converges almost surely to Σ_X^{-1} by Lemma S-2.2 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 $u, v \in \mathcal{V}$

$$\mathfrak{c}'(u,v) = \rho_\epsilon(u,v) AC\Sigma_X^{-1}C^T A^T \tag{7}$$

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.

Theorem S-2.4. *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-2.3, 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-2.2 and $\hat{\sigma}_n \xrightarrow{a.s.} \sigma$ as $n \to \infty$. So applying Lemmas S-2.3 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-2.2.2 Proof of Theorem 3.1

Proof. Our proof of this result is available, see Section S-1.4. What follows here is an alternative proof using Theorem 1 of Eck (2018).

Applying Eck (2018)'s Theorem 1 (conditioning on $(Y_n)_{n\in\mathbb{N}}$ and restricting to the probability 1 event that $(\frac{1}{n}X_n^T X_n)^{-1} \longrightarrow \Sigma_X^{-1}$), we see that

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

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

$$(\Sigma \otimes \Sigma_X^{-1}))_{L(l-1)+j, L(m-1)+k} = \mathfrak{c}_\epsilon(u_j, u_k)(\Sigma_X^{-1})_{lm}. \tag{8}$$

□

Remark S-2.5. *Eck (2018)'s theorem needs to be applied with care as they write the model $Y = \beta X + \epsilon$ rather than via the more standard formulation of $Y = X\beta + \epsilon$, i.e. they takes β to be a row vector rather than a column vector. Their vec operation is thus the result of stacking a transposed matrix the resulting distribution in the statement of their Theorem 1 is $N(0, \Sigma_X^{-1} \otimes \Sigma)$ rather than $N(0, \Sigma \otimes \Sigma_X^{-1})$.*

Remark S-2.6. *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).*

We prove this result an alternative, somewhat simpler way, using the Lindeberg Central Limit Theorem. See Section S-1 for details.

S-2.2.3 Proof of Theorem 3.2

Proof. We can write

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

where A_n is defined as in the proof of S-2.4. Applying Eck (2018)'s Theorem 1b it follows that, as $n \to \infty$, $\hat{\sigma}_n^b \xrightarrow{a.s.} \sigma$. Moreover $A_n \xrightarrow{a.s.} A$ so applying Theorem 3.1, Lemma S-1.2 and Slutsky, the first result holds. The second result immediately follows from the first. □

S-2.3 Proof of Theorem 3.3

Proof. Let $F_n : \mathbb{R} \to [0, 1]$ send $\lambda \in \mathbb{R}$ to $\mathbb{P}(f(T_n^1) \leq \lambda | (Y_n)_{n\in\mathbb{N}})$. Define a sequence $(\eta_n)_{n\in\mathbb{N}} \geq 0$ 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 $\eta_n = \frac{1}{2n^n}$ is α otherwise. Note that there are at most n^n distinct values that $f(T_n^1)$ can take, so 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,

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

The values $\alpha \pm \eta_n$ are continuity points of F_n^- and for $\lambda \in \mathbb{R}$, conditional on $(Y_n)_{n\in\mathbb{N}}$, by the SLLN, $\frac{1}{B}\sum_{b=1}^{B} 1\left[f(T_n^b) \leq \lambda\right]$ converges almost surely to $F_n(\lambda)$ as $B \to \infty$. As such, applying Lemma 1.1.1 from De Haan and Ferreira (2006) it follows that $\lambda^*_{\alpha\pm\eta_n,n,B} \to F_n^-(\alpha \pm \eta_n)$ as $B \to \infty$. Moreover, as $n \to \infty$, F_n converges pointwise to F (as $f(T_n^1)|(Y_n)_{n\in\mathbb{N}} \xrightarrow{d} f(\mathcal{G}(0, \mathfrak{c}')))$ 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 $n \geq N, \eta_n < \delta$ and

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

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. F^{-1} is continuous and so $F^{-1}(\alpha+\delta) \to F^{-1}(\alpha)$ as $\delta \to 0$. Arguing similarly for the sequence $\alpha - \eta_n$ the result follows. Taking limits and using the bound in equation (9), it follows that

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

□

S-2.4 Proofs for Section 4

S-2.4.1 Setup

In what follows we will require the following Lemma.

Lemma S-2.7. *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 $n \to 0$ and $F(\lambda_n) - F(\lambda) \to 0$ because F is continuous. □

Moreover we will want to restrict random fields to subsets. This is defined formally as follows.

Definition S-2.8. Given a a set valued function: $\mathcal{N}$ on $\mathcal{V}$, such that for each $v \in \mathcal{V}, \mathcal{N}_v \subset \{1, \dots, L\}$, we define the **restriction** of f to $\mathcal{N}$ to be the map $f|_{\mathcal{N}} : \Omega \to \left\{g : \mathcal{V} \to \bigcup_{1\leq j\leq L} \mathbb{R}^j\right\}$ such that $f|_{\mathcal{N}}(\omega)(v)$ is the vector $(f_k(v) : k \in \mathcal{N}_v)^T \in \mathbb{R}^{|\mathcal{N}_v|}$.

Given a set function $\mathcal{N}$, defined as in Definition S-2.8, we can stack the entries of $f|_{\mathcal{N}}$ to create $\text{vec}(f|_{\mathcal{N}})$ and thus define $f_n|_{\mathcal{N}} \xrightarrow{d} f|_{\mathcal{N}}$,$f_n|_{\mathcal{N}} \xrightarrow{\mathbb{P}} f|_{\mathcal{N}}$ and $f_n|_{\mathcal{N}} \xrightarrow{a.s.} f|_{\mathcal{N}}$. Because of the Central Limit Theorem convergence will typically be to a Gaussian random field which is defined as follows.

Definition S-2.9. Moreover, for a set function $\mathcal{N}$ as defined above, we shall write $\mathcal{G}|_{\mathcal{N}}(\mu, \mathfrak{c})$ to denote the distribution of the restricted random field. Given

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

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

$$h : \left\{g : \mathcal{V} \to \bigcup_{1\leq j\leq L} \mathbb{R}^j\right\} \to \mathbb{R} \tag{10}$$

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

S-2.4.2 Proof of Theorem 4.1

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

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

where Φ is the CDF of a standard normal distribution. Secondly, given $H \subset \mathcal{H}$, let $f_H : \{g : \mathcal{V} \to \mathbb{R}^L\} \to \mathbb{R}$ send $T \in \{g : \mathcal{V} \to \mathbb{R}^L\}$ 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

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

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 $\phi_n : \left\{g : \mathcal{V} \to \bigcup_{0\le j\le L} \mathbb{R}^j\right\} \to \mathbb{R}$ which send

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

and $\phi : \left\{g : \mathcal{V} \to \bigcup_{0\le j\le L} \mathbb{R}^j\right\} \to \mathbb{R}$ 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 4.1.

Proof. Defining $\mathfrak{c}'$ as in Section 3.2, $T_n|_{\mathcal{N}}$ converges to $\mathcal{G}(0,\mathfrak{c}')|_{\mathcal{N}}$ in distribution by Theorem S-2.4. As such, using the fact that $f_{n,\mathcal{N}}$ is the composition of functions which are either continuous or converge uniformly with range $[0,1]$, by Lemma S-3.3 and the Continuous Mapping Theorem,

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

By the same logic, and applying Theorem 3.2, for $\mathcal{N} \subset H \subset \mathcal{H}$

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

This convergence occurs conditional on the data, a fact that we take as implicit in (13) and in the rest of the proof. As such, applying Theorem 3.3, it follows 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 and the fact that the density of the multivariate normal distribution is positive everywhere) and continuous, by Lemma S-3.4). 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 (12) and the fact that F_0 is continuous (which follows from Lemma S-3.4). As such, applying Lemma S-2.7 (since F_n and F_0 are CDFs and F_0 is continuous), it follows that

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

which proves the result. Note that the 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

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

□

S-2.4.3 Proof of Corollary 4.2

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),

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

The result follows by sending ϵ to zero. □

S-2.4.4 Proof of Theorem 4.3

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 4.1,

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

We claim that on the event Ω_n, $\mathcal{N} \subset \hat{H}_n$. We prove this inductively, using the notation from Algorithm 1. $\mathcal{N} \subset H^{(0)}$ trivially. Assuming that $\mathcal{N} \subset H^{(j-1)}$ 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),

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

which implies that $\mathcal{N} \subset H^{(j)}$. Thus $\mathcal{N} \subset \hat{H}_n$ 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 result follows as in the proof of Corollary 4.2. □

S-3 Further Theory

S-3.1 FWER inference

FWER inference is commonly used in brain imaging in order to identify areas of activation in the brain. This corresponds to performing multiple testing inference on the data and returning a set of active hypotheses $R \subset \mathcal{H}$ such that the familywise error rate (FWER), defined as

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

When a single test is being used (for a single contrast or an F-test at each voxel), brain imaging studies have typically used 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 - see Section S-3.2. However the bootstrap approach can be applied. In particular we have the following theorem which follows as a corollary of Theorem 4.1 by taking $K = 1$ and using the linear template.

Theorem S-3.1. *For $0 \leq \alpha \leq 1$ 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

$$\lim_{n\to\infty} \lim_{B\to\infty} \mathbb{P}(R_{n,B} \cap \mathcal{N}) \leq \alpha.$$

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

This control of the FWER does not occur simultaneously with the control of the joint error rate. However let $\lambda^*_{\alpha,n,B}$ is the $\alpha-$quantile of the bootstrap distribution of $f_{n,\mathcal{N}}$ (as defined in the statement of Theorem 4.1). Then FWER 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-3.1 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-3.2 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), permuting the data does not necessarily generate data under the global null even when the noise is exchangeable under permutation.

Claim S-3.2. *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

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

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 apply the subset pivotality condition 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

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

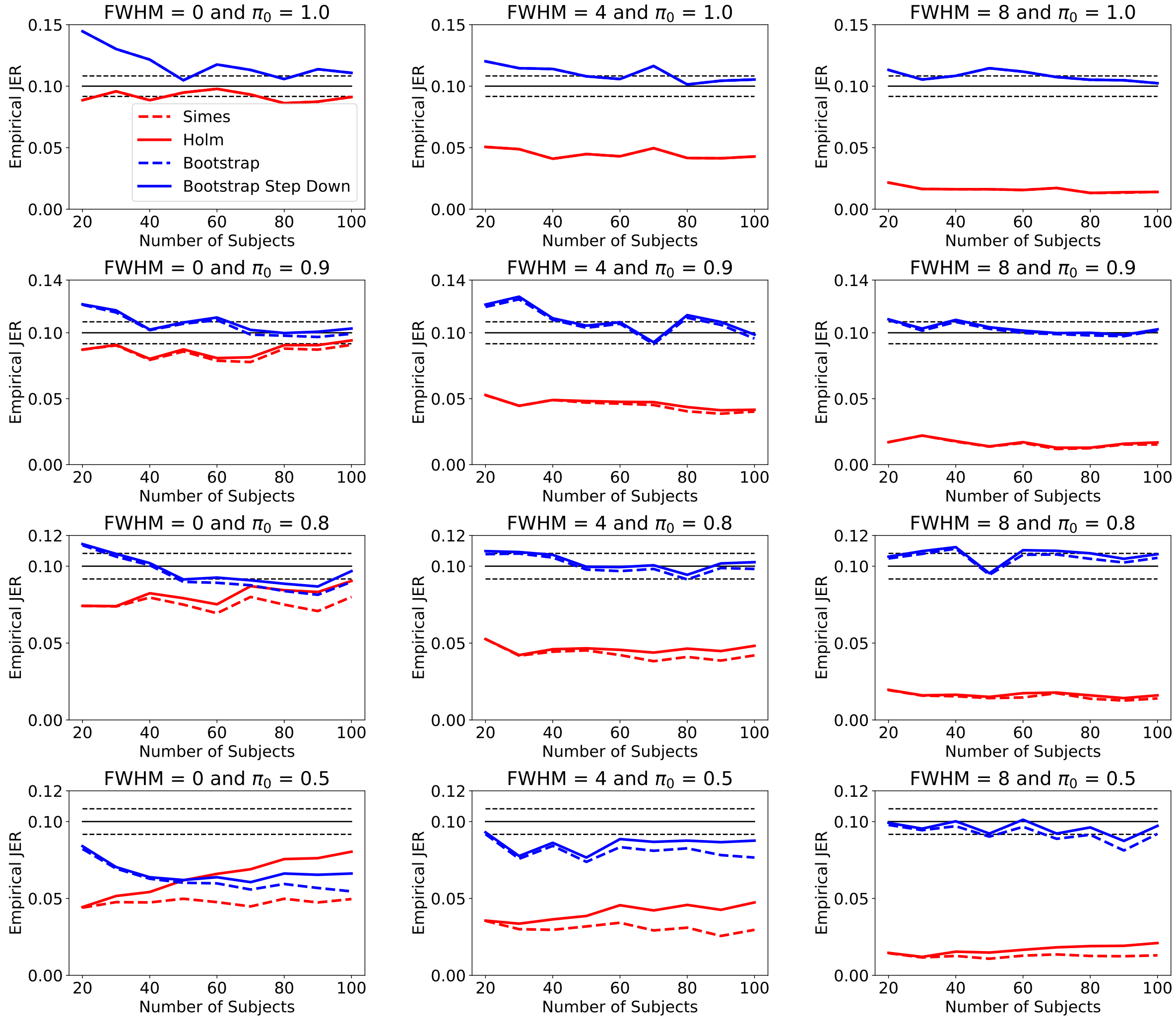

Figure S-1: Direct FWER control using the different methods. Here the parametric procedures are Bonferroni and its step-down: Holm (1979).

Now, under exchangeability,

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

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

$$(X^T X)^{-1} X^T P X\beta \neq 0$$

which causes a problem. □

S-3.3 Additional Lemmas for the proofs

Lemma S-3.3. *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 $g : \mathbb{R} \to \mathbb{R}$,

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

the functions $(f_n)_{n\in\mathbb{N}}, f$ range values within a compact set I and so without loss of generality we may assume that g is uniformly continuous. So for any $\epsilon > 0$ there is some δ such that $|g(x) - g(y)| < \epsilon$ for all $x, y \in \mathbb{R}$ 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 $z \in \mathbb{R}^L$. As such

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

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

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

Since this holds for any continuous bounded g the result follows by Portmanteau. $\square$

Lemma S-3.4. *Let F_0 be the CDF of $\min_{1\leq k\leq K\wedge|\mathcal{N}|} t_k^{-1}(p_{(k:H)}(T))$ where $T \sim \mathcal{G}(0, \mathfrak{c}')$ and $\mathfrak{c}'$ is defined as in Section 3.2, then F_0 is continuous.*

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

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

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. $\square$

S-4 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-5 Further figures

S-5.1 Simes vs ARI for the IQ contrast

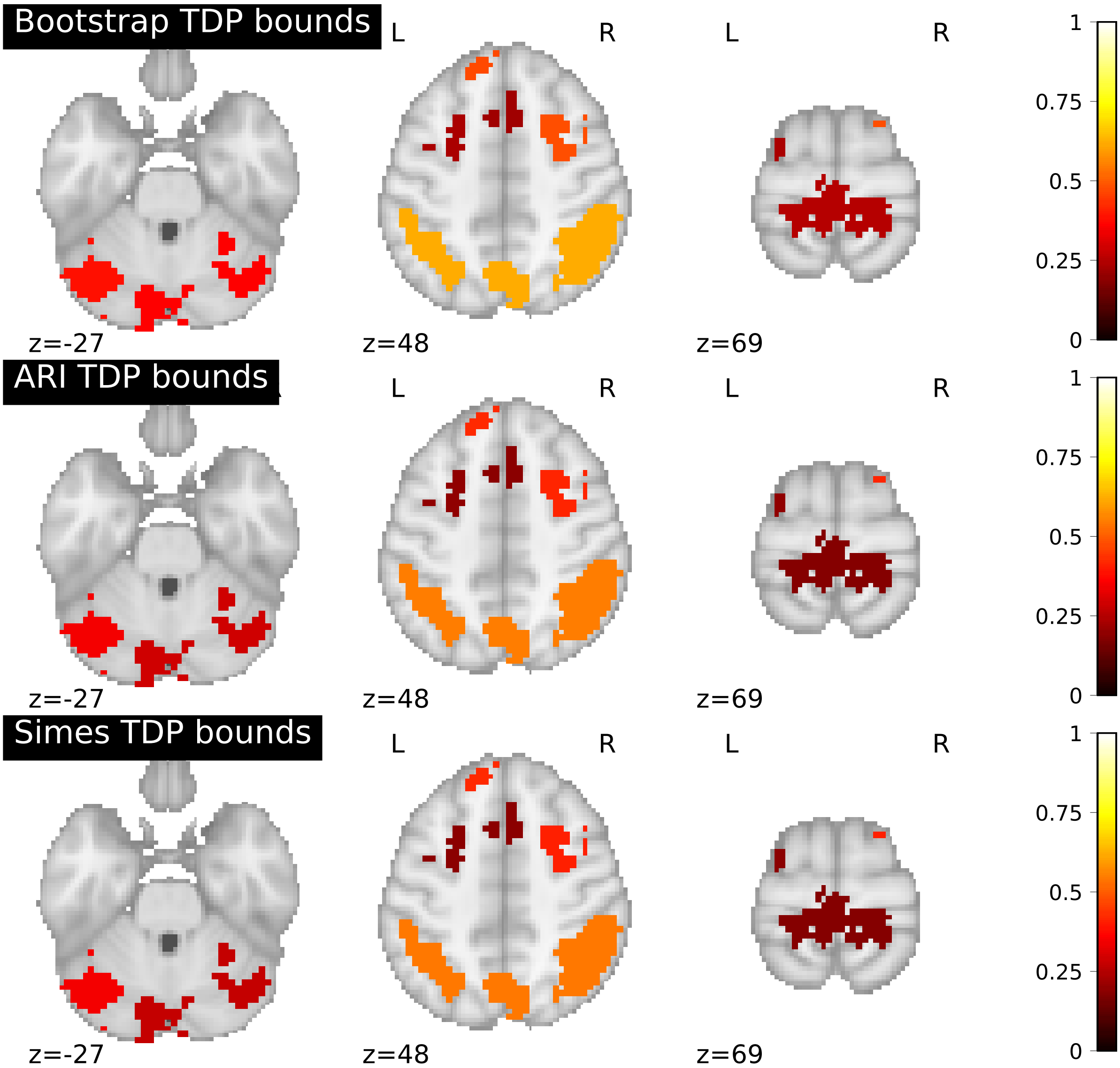

Figure S-2: TDP bounds within clusters for the contrast for IQ in the linear regression model fit to the HCP data. Each cluster is shaded a single colour which is the lower bound on the TDP. The upper panel gives the TDP bounds within each cluster provided by the bootstrap procedure. The lower panels 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-5.2 The contrast for sex

Much 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.

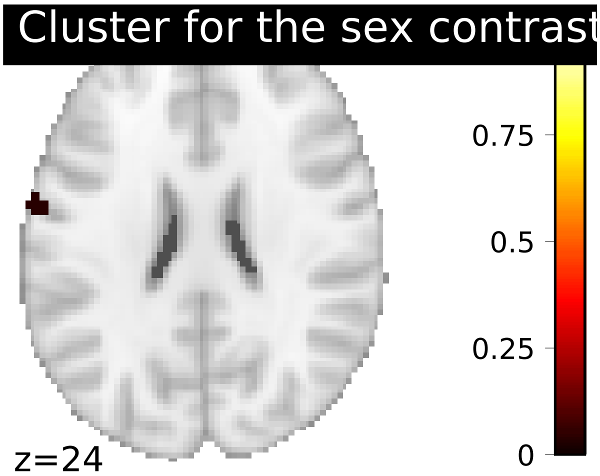

Figure S-3: Illustrating the cluster in the sex contrast with non-zero activation.

S-5.3 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-5.4 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.

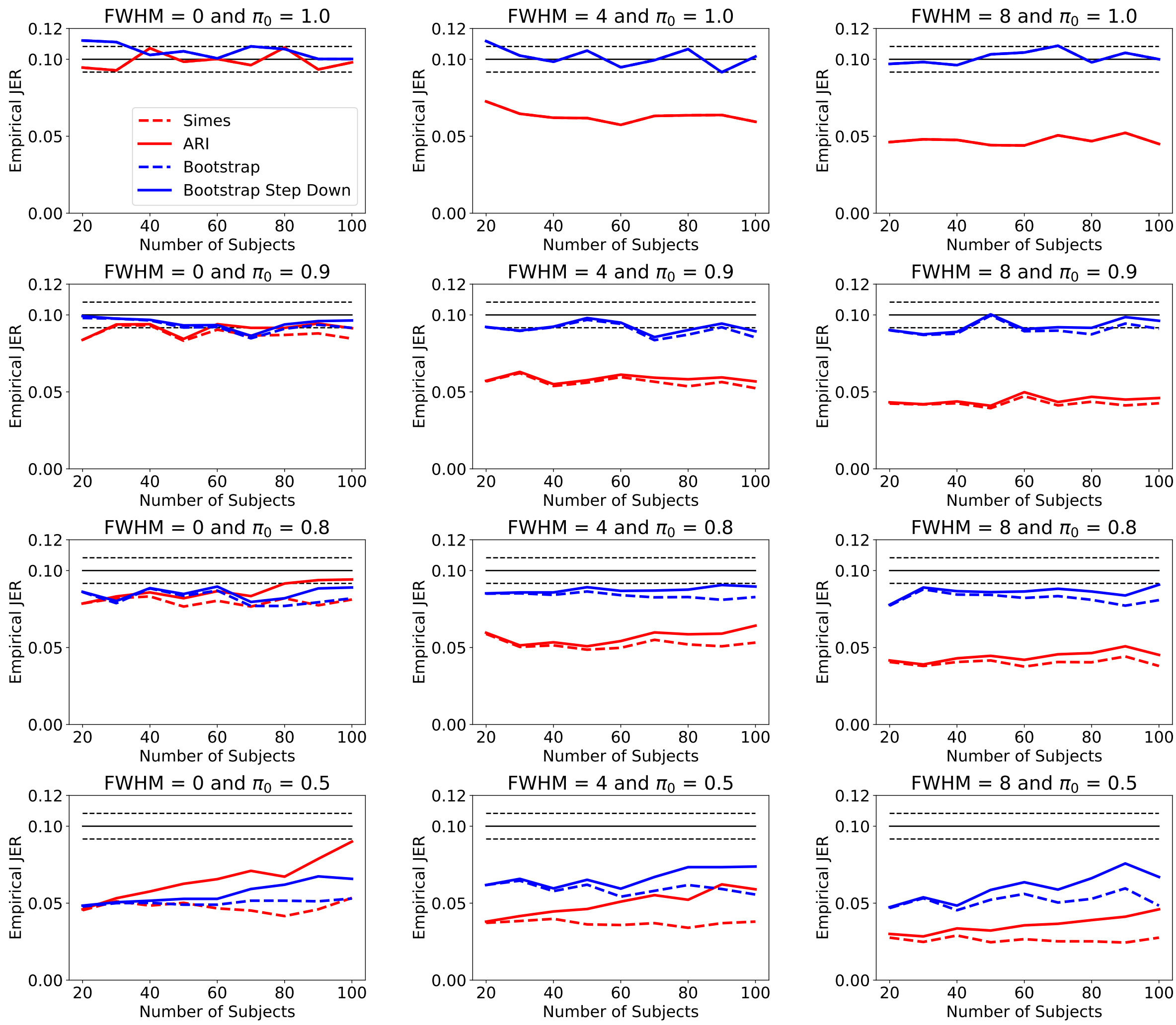

Figure S-5: Empirical joint error rate for the 25 by 25 simulations.

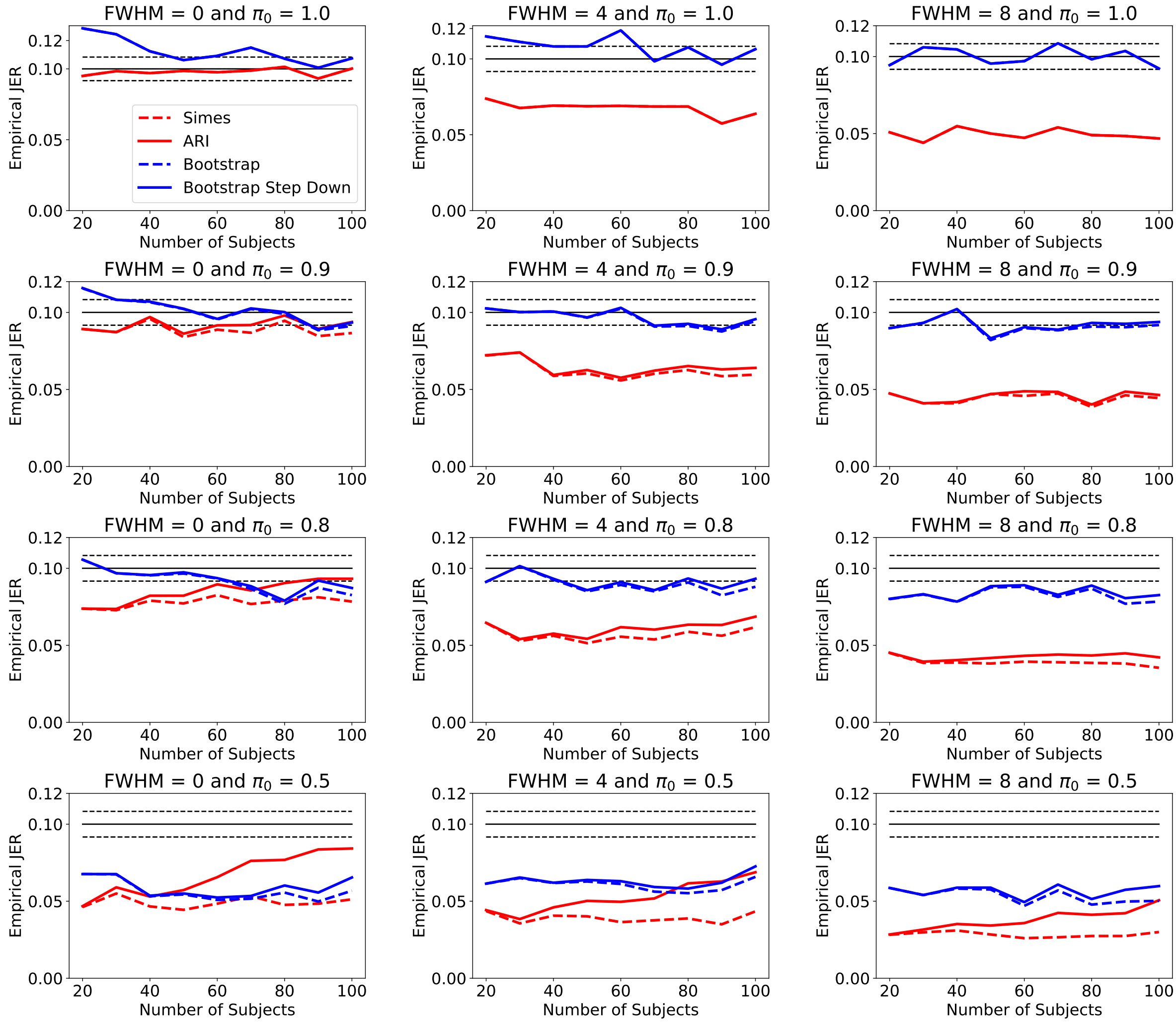

Figure S-6: Empirical joint error rate for the 100 by 100 simulations.

S-5.5 Additional power plots

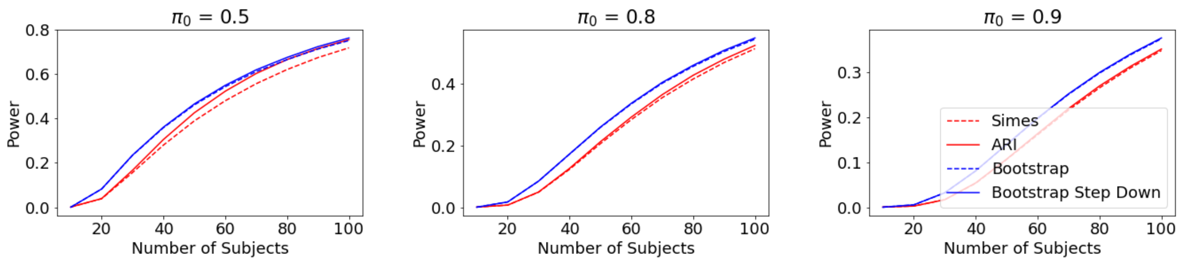

Figure S-7: Plotting the power of the different methods against the numbers of a subjects for setting 3, i.e. taking $R = \{(l, v) : p_{n,l}(v) \leq 0.05\}$ in (14).

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.

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

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.