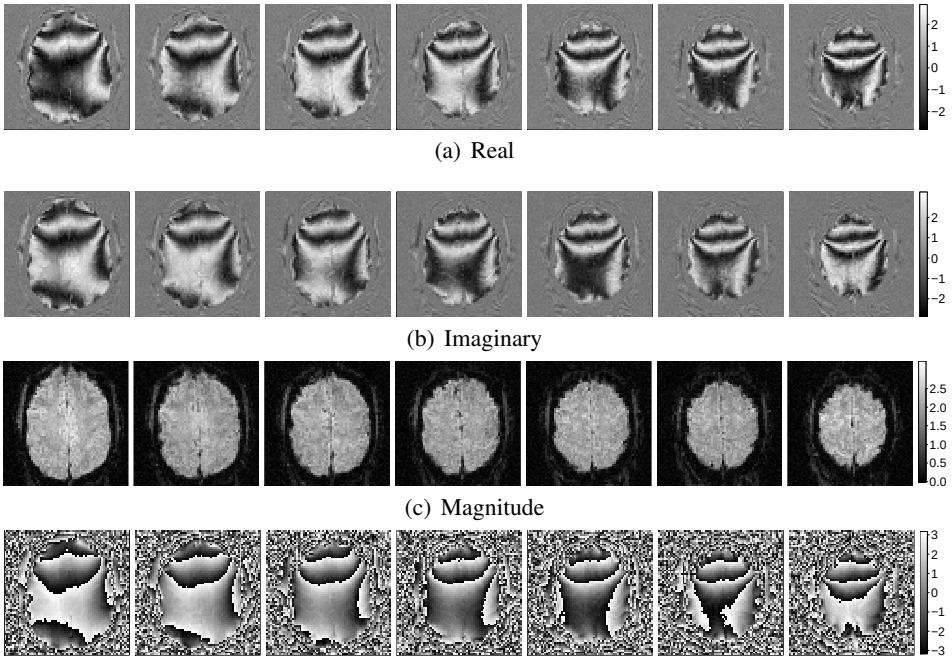
Submitted to the Annals of Applied Statistics

SUPPLEMENT TO "COMPLEX-VALUED TIME SERIES MODELING FOR IMPROVED ACTIVATION DETECTION IN FMRI STUDIES"

BY DANIEL W. ADRIAN[‡] AND RANJAN MAITRA[*,§] AND DANIEL B. ROWE[†,¶]

Grand Valley State University[‡] and Iowa State University[§] and Marquette University[¶]

S-1. Supplement to Section 2 – Further description of the finger-tapping dataset. Figure S-1 shows images of the real, imaginary, magnitude, and phase



(a) Real

(b) Imaginary

(c) Magnitude

(d) Phase

FIG S-1. *Images of the (a) real, (b) imaginary, (c) magnitude, and (d) phase components of the seven slices of the image at the first time point, ordered from superior to inferior when viewed left to right.*

components of the data for all slices at the first time point.

S-2. Supplement to Section 3 – Further Methodological Development.

[*]Research supported in part by the National Science Foundation CAREER Grant # DMS-0437555 and by the National Institutes of Health (NIH) award #R21EB016212.

[†]Research supported in part by the National Institutes of Health (NIH) award #R21NS087450.

S-2.1. *Supplement to Section 3.1 – Complex-valued running line derivations.* First, we note that (3.1) assumes that η_{Rt}, η_{It} are independent identically distributed as $N(0, \sigma^2)$ random variables and that the real and imaginary errors are temporally independent. Now, to evaluate the complex-valued running line at t^*, the MLEs of $\boldsymbol{\beta} = (\beta_0, \beta_1)'$, γ_0, and γ_1 can be obtained by iteratively updating the equations

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{r}^* \tag{S-1}$$

$$\hat{\gamma}_0 = \arctan_4(S, C) \tag{S-2}$$

$$\hat{\gamma}_1^* = \hat{\gamma}_1 - \ell'(\hat{\gamma}_1)/\ell''(\hat{\gamma}_1) \tag{S-3}$$

where, in (S-1), $\boldsymbol{X}$ is a matrix with two columns: an intercept and the arithmetic sequence with tth entry s_t, and $\boldsymbol{r}^*$ is a vector with tth entry $r_t \cos(\phi_t - \hat{\theta}_t)$. In (S-2), $S = \sum_{t \in N(t^*)} r_t \hat{\rho}_t \sin[\phi_t - g(s_t\hat{\gamma}_1)]$ and $C = \sum_{t \in N(t^*)} r_t \hat{\rho}_t \cos[\phi_t - g(s_t\hat{\gamma}_1)]$. In the Newton-Raphson update (S-3) which computes the new estimate $\hat{\gamma}_1^*$ from the old estimate $\hat{\gamma}_1$,

$$\ell'(\hat{\gamma}_1) = \sum_{t \in N(t^*)} s_t r_t \hat{\rho}_t g'(s_t\hat{\gamma}_1)\sin(\phi_t - \hat{\theta}_t) \tag{S-4}$$

$$\ell''(\hat{\gamma}_1) = \sum_{t \in N(t^*)} s_t^2 r_t \hat{\rho}_t \left\{ \begin{array}{l} g''(s_t\hat{\gamma}_1)\sin(\phi_t - \hat{\theta}_t) \\ -[g'(s_t\hat{\gamma}_1)]^2 \cos(\phi_t - \hat{\theta}_t) \end{array} \right\} \tag{S-5}$$

where $g'(x) = 2/(1+x^2)$ and $g''(x) = 2x/(1+x^2)^2$. We have applied the starting values $\hat{\gamma}_0^{(0)} = \arctan_4(\sum_{t \in N(t^*)} r_t \sin\phi_t, \sum_{t \in N(t^*)} r_t \cos\phi_t)$, $\hat{\beta}_1^{(0)} = (1/n)\sum_{t \in N(t^*)} r_t \cos(\phi_t - \hat{\gamma}_0^{(0)})$, and $\hat{\beta}_1^{(0)} = \hat{\gamma}_1^{(0)} = 0$. We use the log-likelihood function to assess convergence, which only takes a few iterations of (S-1)-(S-3).

S-2.2. *Supplement to Section 3.3 – Magnitude-only AR(p) model derivations.* The log-likelihood function is given by $\log L(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2|\boldsymbol{r}) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\log|\boldsymbol{R}_n| - \frac{1}{2\sigma^2}(\boldsymbol{r} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{R}_n^{-1}(\boldsymbol{r} - \boldsymbol{X}\boldsymbol{\beta})$, where $\boldsymbol{R}_n$ is such that $\sigma^2\boldsymbol{R}_n = \mathrm{Cov}(\boldsymbol{\epsilon})$. The unrestricted MLEs of $\boldsymbol{\beta}$ and σ^2 are given by $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\hat{\boldsymbol{R}}_n^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{R}}_n^{-1}\boldsymbol{r}$ and $\hat{\sigma}^2 = (\boldsymbol{r} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'\hat{\boldsymbol{R}}_n^{-1}(\boldsymbol{r} - \boldsymbol{X}\hat{\boldsymbol{\beta}})/n$, respectively. We obtain $\hat{\boldsymbol{\alpha}}$ by solving the system of equations (due to Miller, 1995): $\sum_{j=1}^{p}[\hat{d}_{jk} + (j/n)\hat{d}_{0,|j-k|}]\hat{\alpha}_j = \hat{d}_{0k}$, $k = 1, \ldots, p$, where $\hat{d}_{ij} = \sum_{t=1}^{n-i-j}\hat{\epsilon}_{t+i}\hat{\epsilon}_{t+j}$, for $0 \le i, j \le p$, and $\hat{\epsilon}_t = r_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}$, $t = 1, \ldots, n$. The estimation procedure, due to Cochrane and Orcutt (1949), begins with $\hat{\boldsymbol{R}}_n = \boldsymbol{I}_n$ and then iteratively updates $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\alpha}}$, and $\hat{\boldsymbol{R}}_n^{-1}$ until convergence. The

restricted MLEs under $H_0 : \boldsymbol{C\beta} = \mathbf{0}$ follow the equations

$$\tilde{\boldsymbol{\beta}} = \boldsymbol{\Psi}(\boldsymbol{X}'\tilde{\boldsymbol{R}}_n^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\tilde{\boldsymbol{R}}_n^{-1}\boldsymbol{r}, \tag{S-6}$$

$$\tilde{\sigma}^2 = (\boldsymbol{r} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})'\tilde{\boldsymbol{R}}_n^{-1}(\boldsymbol{r} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})/n, \tag{S-7}$$

$$\tilde{d}_{0k} = \sum_{j=1}^{p}[\tilde{d}_{jk} + (j/n)\tilde{d}_{0,|j-k|}]\tilde{\alpha}_j, \quad k = 1, \ldots, p, \tag{S-8}$$

where

$$\boldsymbol{\Psi} = \boldsymbol{I}_q - (\boldsymbol{X}'\tilde{\boldsymbol{R}}_n^{-1}\boldsymbol{X})^{-1}\boldsymbol{C}'\left[\boldsymbol{C}(\boldsymbol{X}'\tilde{\boldsymbol{R}}_n^{-1}\boldsymbol{X})^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C}, \tag{S-9}$$

q is the length of $\boldsymbol{\beta}$, and $\tilde{\boldsymbol{R}}_n^{-1}$ and $\tilde{d}_{ij}$ are the restricted estimation analogs of their counterparts above. The LRT statistic is given by

$$\Lambda_{M,p} = n\log(\tilde{\sigma}^2/\hat{\sigma}^2) - \log\left(\left|\tilde{\boldsymbol{R}}_p^{-1}\right| / \left|\hat{\boldsymbol{R}}_p^{-1}\right|\right), \tag{S-10}$$

where $\boldsymbol{R}_p$ is such that $\sigma^2\boldsymbol{R}_p = \mathrm{Cov}(\epsilon_1, \ldots, \epsilon_p)$.

S-2.3. *Supplement to Section 3.4 – Restricted MLEs under a complex-valued AR(p) model.* Recall from Section 3.4 that restricted MLEs are denoted with tildes, in contrast to the unrestricted MLEs denoted with "hats". The restricted MLE of $\boldsymbol{\beta}$ under $H_0 : C\boldsymbol{\beta} = 0$ is given by

$$\tilde{\boldsymbol{\beta}} = \frac{\boldsymbol{\Psi}\tilde{\boldsymbol{\beta}}_R\left[\frac{\cos\tilde{\theta}}{\tilde{\sigma}_R^2} - \frac{\tilde{\rho}\sin\tilde{\theta}}{\tilde{\sigma}_R\tilde{\sigma}_I}\right] + \boldsymbol{\Psi}\tilde{\boldsymbol{\beta}}_I\left[\frac{\sin\tilde{\theta}}{\tilde{\sigma}_I^2} - \frac{\tilde{\rho}\cos\tilde{\theta}}{\tilde{\sigma}_R\tilde{\sigma}_I}\right]}{\frac{\cos^2\tilde{\theta}}{\tilde{\sigma}_R^2} + \frac{\sin^2\tilde{\theta}}{\tilde{\sigma}_I^2} - 2\frac{\tilde{\rho}}{\tilde{\sigma}_R\tilde{\sigma}_I}\sin\tilde{\theta}\cos\tilde{\theta}}, \tag{S-11}$$

which has a similar form to (3.5) except for premultiplication by the matrix $\boldsymbol{\Psi}$ defined in (S-9). In the above, we have $\tilde{\boldsymbol{\beta}}_R = (\boldsymbol{X}'\tilde{\boldsymbol{R}}_n^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\tilde{\boldsymbol{R}}_n^{-1}\boldsymbol{y}_R$ and $\tilde{\boldsymbol{\beta}}_I = (\boldsymbol{X}'\tilde{\boldsymbol{R}}_n^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\tilde{\boldsymbol{R}}_n^{-1}\boldsymbol{y}_I$. The restricted MLE for θ is given by

$$\tilde{\theta} = \left[\arcsin\left(\tilde{c}/\sqrt{\tilde{a}^2 + \tilde{b}^2}\right) - \arctan_4(\tilde{b}, \tilde{a})\right]/2, \tag{S-12}$$

in which $\tilde{a}$, $\tilde{b}$, and $\tilde{c}$ are defined by replacing hats by tildes in the definitions of a through f following (3.6). In addition, B_{RR}, B_{II}, and B_{RI} must be redefined as $\tilde{B}_{RR} = \tilde{\boldsymbol{\beta}}_R'(\boldsymbol{\Psi}'\boldsymbol{X}'\tilde{\boldsymbol{R}}_n^{-1}\boldsymbol{X})\tilde{\boldsymbol{\beta}}_R$, $\tilde{B}_{II} = \tilde{\boldsymbol{\beta}}_I'(\boldsymbol{\Psi}'\boldsymbol{X}'\hat{\boldsymbol{R}}_n^{-1}\boldsymbol{X})\tilde{\boldsymbol{\beta}}_I$, and $\tilde{B}_{RI} = \tilde{\boldsymbol{\beta}}_R'(\boldsymbol{\Psi}'\boldsymbol{X}'\hat{\boldsymbol{R}}_n^{-1}\boldsymbol{X})\tilde{\boldsymbol{\beta}}_I$, respectively. Equations for the restricted MLEs for $\sigma_R^2, \sigma_I^2, \rho$, and $\boldsymbol{\alpha}$ may be obtained replacing hats by tildes in equations (3.7)-(3.10).

S-2.4. *Supplement to Section 3.5 – Controlling order over-detection via the significance level.* We show that the significance level, say δ, controls order over-detection for a single voxel time series in the sense that $\delta = P(\hat{p} > p | \hat{p} \geq p)$. To show this, first recall that the detected order $\hat{p} = k' - 1$, where k' is the first k for which we are unable to reject $H_0 : p = k-1, k \geq 1$. Note two facts for any $k \geq 1$: first, rejecting $H_0 : \alpha_k = 0$ means that $k' > k \Rightarrow \hat{p} > k-1$. Second, simply testing $H_0 : p = k - 1$ in the context of the procedure implies that $k' \geq k \Rightarrow \hat{p} \geq k - 1$. From the definition of δ, for $k > p$, $\delta = P(\text{reject } H_0 | H_0 \text{is true}) = P(\hat{p} > k - 1 | \hat{p} \geq k - 1)$. Substituting $k = p + 1$ yields the above result.

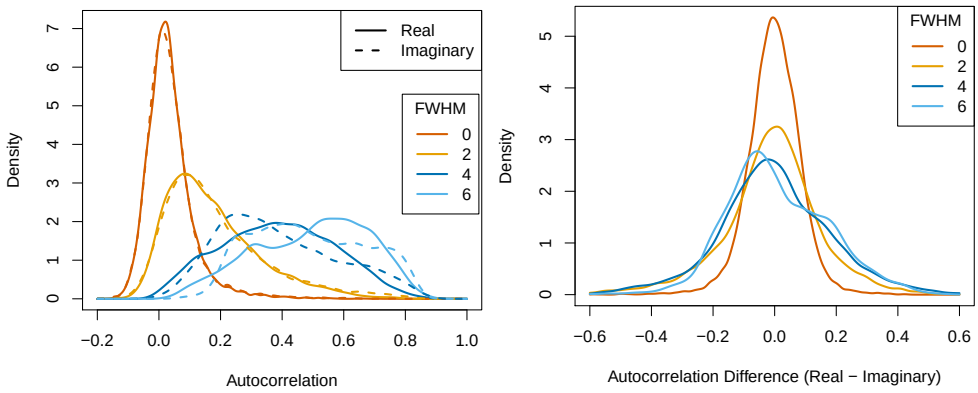
S-3. Supplement to Section 4 – Further Analysis of Finger-Tapping Dataset.

S-3.1. *Supplement to Section 4.2 – Addressing model fit and model assumptions.* Model (3.2) specifies that all effects in the $\boldsymbol{X}$ matrix are phase-coupled through θ. That is, as pointed out by a reviewer, (3.2) is equivalent to the two linear models

$$\boldsymbol{y}_R = \boldsymbol{X}\boldsymbol{\beta}_R + \boldsymbol{\eta}_R, \; \boldsymbol{y}_I = \boldsymbol{X}\boldsymbol{\beta}_I + \boldsymbol{\eta}_I, \tag{S-13}$$

where $\beta_{Ij}/\beta_{Rj} = \tan\theta$ for each element j of $\boldsymbol{\beta}_R$ and $\boldsymbol{\beta}_I$, where the last is the phase coupling assumption. To address this assumption, we performed an LRT comparing, for each voxel time series in the finger-tapping dataset model (S-13) without the phase coupling assumption with (3.2). This LRTS has an asymptotic χ^2_1 null distribution. After controlling for FDR (Benjamini and Hochberg, 1995) at the 5% level, the LRT statistic was significant for only one and two in-mask voxels for the cases with FWHM=0 and 2 voxels respectively: there were no significant voxels for FWHM=4 and 6 voxels. Thus, we conclude that deviations from the phase coupling assumption in model (3.2) are at best marginal.

Next, we focus on the assumption that the real and imaginary errors share the same AR coefficients. This assumption is supported by the similarity of the distributions of the lag-1 autocorrelations of the real and imaginary residuals shown in Figure S-2. Specifically, the real and imaginary residuals are given by $\hat{\boldsymbol{\eta}}_R = \boldsymbol{y}_R - \boldsymbol{X}\hat{\boldsymbol{\beta}}\cos\hat{\theta}$ and $\hat{\boldsymbol{\eta}}_I = \boldsymbol{y}_I - \boldsymbol{X}\hat{\boldsymbol{\beta}}\sin\hat{\theta}$, respectively, where $\hat{\boldsymbol{\beta}}$ and $\hat{\theta}$ are model (3.2) MLEs under the assumption that $p = 0$. Additionally, we can use an LRT to check this assumption. Because ML estimation is intractable when allowing for different AR coefficients of η_R and η_I, dependence between them, and phase coupling (which itself is an argument for the equality of the AR coefficients), we assume real/imaginary independence and drop the phase coupling assumption when calculating the LRTS. The full model is then given by (S-13) with separate AR coefficients $\boldsymbol{\alpha}_R$ and $\boldsymbol{\alpha}_I$ for the real and imaginary errors, respectively; the reduced model assumes that $\boldsymbol{\alpha}_R = \boldsymbol{\alpha}_I$. Under the assumption that $\boldsymbol{\alpha}_R$ and $\boldsymbol{\alpha}_I$ have the same order p, the LRTS follows an asymptotic χ^2_p distribution under $H_0 : \boldsymbol{\alpha}_R = \boldsymbol{\alpha}_I$. Table S-1



(a) (b)

FIG S-2. *Kernel density estimates of (a) the autocorrelation of the real and imaginary residuals and (b) their difference at different amounts of spatial smoothing (determined by FWHM).*

shows the proportions of voxels in which the LRT shows evidence of violation of the $\boldsymbol{\alpha}_R = \boldsymbol{\alpha}_I$ assumption at a per comparison error rate of 0.05 and a familywise error rate of 0.05 (using the Bonferroni correction). These proportions indicate that

FWHM	0	2	4	6
PCER = 0.05	0.262	0.397	0.495	0.571
FWER = 0.05	0.030	0.074	0.105	0.105

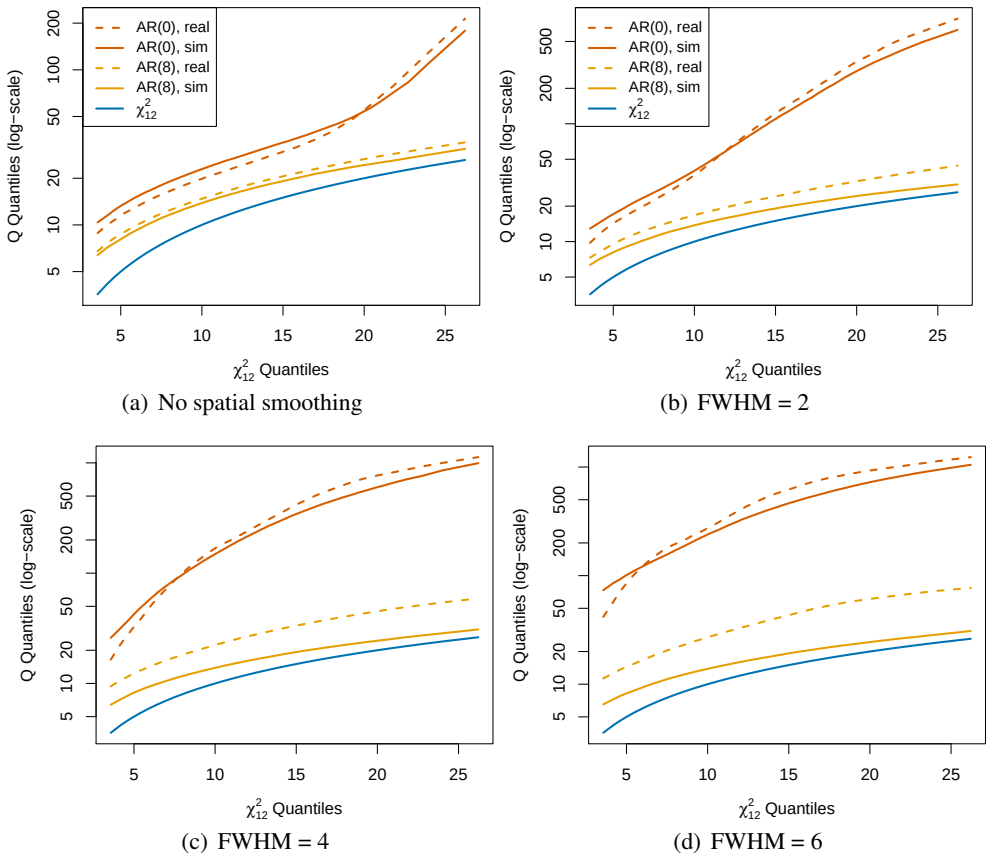
TABLE S-1

Proportions of voxel time series in the finger-tapping dataset showing a violation of the assumption of equality of the AR coefficients of the real and imaginary errors according to a LRT at a per comparison error rate (PCER) and familywise error rate (FWER) of 0.05.

a substantial proportion of voxels show statistically significant evidence that this assumption is violated. Whether this result is practically significant in terms of the difference in autocorrelations of real and imaginary residuals is a matter of debate.

In addition, to check the sufficiency of the AR(p) model, we computed Ljung-Box-Pierce Q-statistics (Box and Pierce, 1970; Ljung and Box, 1978). The Q-statistic for a (real-valued) time series of length n is defined as $Q = n(n+2)\sum_{k=1}^{K}(n-k)^{-1}\hat{\rho}^2(k)$, where $\hat{\rho}(k)$ is the lag-k sample autocorrelation of the model fit residuals, and K is typically chosen to be $K = 20$ (Shumway and Stoffer, 2006). When an AR(p) model is fit to an AR(p) time series, Q is asymptotically χ^2_{K-p}. Under the complex-valued AR(p) model, the real and imaginary model fit residuals are given by $\hat{e}_{R,t+p} = \hat{\eta}_{R,t+p} - \sum_{k=1}^{p}\hat{\alpha}_k\hat{\eta}_{R,t+p-k}$ and $\hat{e}_{I,t+p} = \hat{\eta}_{I,t+p} - \sum_{k=1}^{p}\hat{\alpha}_k\hat{\eta}_{I,t+p-k}$, respectively, $t = 1, \ldots, n-p$, where $\hat{\eta}_{Rt}$ and $\hat{\eta}_{It}$ are tth entries of the vectors $\hat{\boldsymbol{\eta}}_R = \boldsymbol{y}_R - \boldsymbol{X}\hat{\boldsymbol{\beta}}\cos\hat{\theta}$ and $\hat{\boldsymbol{\eta}}_I = \boldsymbol{y}_I - \boldsymbol{X}\hat{\boldsymbol{\beta}}\sin\hat{\theta}$, respectively. For each complex-valued time series, we computed separate Q-statistics based on the real and imaginary residuals. We computed Q-statistics based on AR(0) and

AR(8) complex-valued model fits – the latter because $\hat{p}_C \leq 8$ for all voxels – to the finger-tapping dataset, and also simulated time series based on complex-valued AR(8) model parameter estimates. Figure S-3 shows quantile-quantile plots com-



(a) No spatial smoothing (b) FWHM = 2

(c) FWHM = 4 (d) FWHM = 6

FIG S-3. *Quantile-quantile plots comparing the theoretical χ^2_{12} distribution of the Ljung-Box-Pierce Q-statistics to the distributions of the Q-statistics computed from the simulated ("sim") and real finger-tapping data under complex-valued AR(0) and AR(8) model fits for various amounts of spatial smoothing (FWHMs).*

paring the distribution of these four Q-statistics to the theoretical χ^2_{12} null distribution of Q – *i.e.* under $K = 20$ and $p = 8$. The AR(8)-model residual Q-statistics for the dataset are well-below the residual Q-statistics for the AR(0) model and relatively close to the AR(8)-model residual Q-statistics for the simulated data and the χ^2_{12} distribution. This indicates that the AR model fit removes the vast majority of autocorrelation from the dataset and gains from introducing more complex (and computationally intensive) methods, such as incorporating moving average, integrated, or seasonal components, would be small.

Last, in response to a reviewer's concern about the increased computational bur-

den imposed by analyzing the complex-valued data, we compared the computational times for the CV and MO data-based analyses of the finger-tapping dataset. These times are based on a CV running line that uses linear interpolation between every $j = 10$ time points, spatial smoothing with FWHM = 4 voxels, and an Intel Core i5-4300M CPU 64-bit processor running `C` code within `R` (R Core Team, 2016). The computational times are decomposed into the steps of detrending, spatial smoothing, order detection, and activation statistics in Table S-2. As should be

Data Analysis	Detrending	Spatial Smoothing	Order Detection	Activation Statistics	Other	Total
Magnitude-only	13.1	23.0	16.8	10.7	7.3	70.8
Complex-valued	95.6	45.2	55.4	22.8	7.8	226.8

TABLE S-2

Computational times (in seconds) for the various processing steps for the magnitude-only and complex-valued data-based analyses of the finger-tapping dataset.

expected (since there are twice as many observations), the complex-valued data-based analysis takes more time, roughly three times so, but the entire analysis takes less than four minutes. Further, the computational time is scalable, meaning that this ratio of three-to-one will hold regardless of the dataset size.

S-3.2. *Supplement to Section 4.3 – Additional activation maps.* Figure 5 showed activation images for the four activation statistics $\Lambda_{C,\hat{p}_C}$, $\Lambda_{M,\hat{p}_M}$, $\Lambda_{C,0}$, and $\Lambda_{M,0}$ for slice 5 of the dataset when a spatial smoothing was performed with Gaussian filter of FWHM = 4 voxels. To display more of the dataset and the effect of each level of spatial smoothing, we provide activation images for the full set of seven slices and at the four FWHMs of 0 (indicating no smoothing), 2, 4, and 6 voxels. Each of Figures S-4 through S-6 shows the activation images for slices 4, 5, and 6 of the dataset, with each figure showing the 16 combinations of the 4 FHWMs and the 4 activation statistics. Examining all the activation maps, we note that the complex-valued AR(p) model-based statistic best identifies the anatomical left central sulcus over each slice as long as some spatial smoothing is done. However, if the data are not smoothed in space, there is much less difference between the activation maps of the four activation statistics.

We also computed activation maps using two other methodologies. We display the maps for the fifth slice and smoothing FWHM = 4 voxels for direct comparison to Figure 5. Figure S-7 displays the activation maps thresholded at the 5% false discovery rate. Specifically, we applied the Benjamini and Yekutieli (2001) method to the voxel-wise p-values and cluster thresholded the resulting q-values with a third-order neighborhood structure, a size threshold of 11 voxels, and a significance level threshold of 0.05. These maps look quite similar to Figure 5. Second, we were motivated by a reviewer to examine maps based on assumed AR orders of

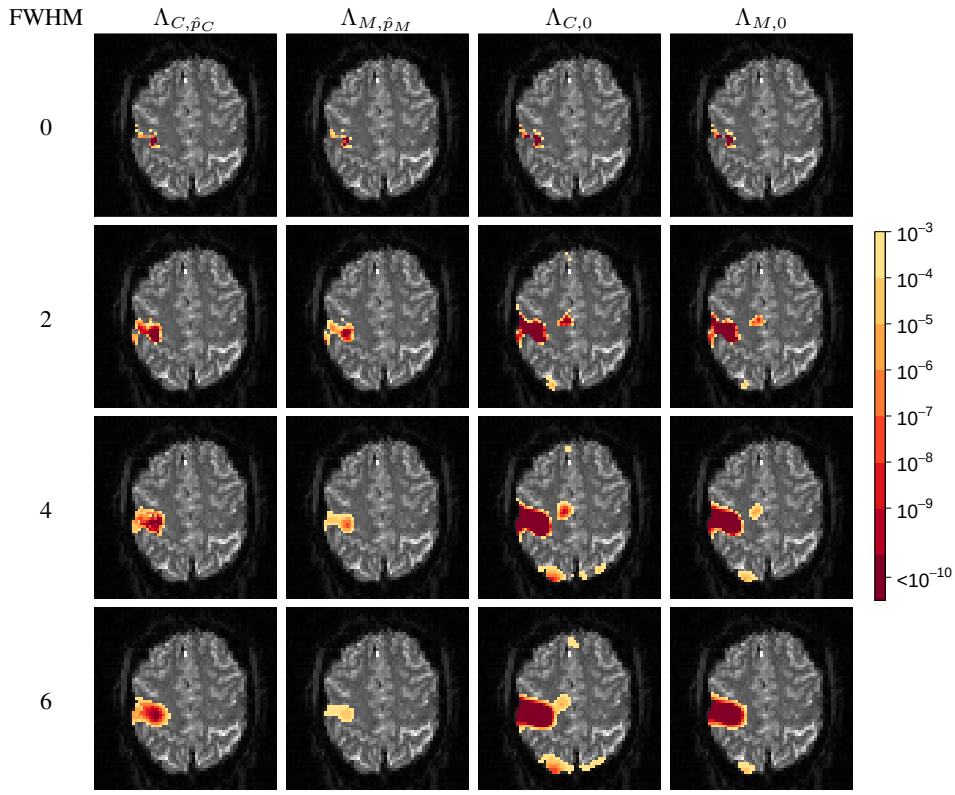


FIG S-4. *Activation images for slice 4 organized according to FWHM of kernel smoother (rows) and activation statistic (columns).*

1. He/she stated that from experience, the exact time series model does not matter very much in terms of activation maps, as long as some type of AR model is used instead of an independence assumption (and also pointed out mention of this point in the documentation on the `3dREMLfit` function in the Analysis of Functional Neuroimaging (AFNI) (Cox, 1996, 2012) software). Thus, Figure S-8 displays the complex-valued and magnitude-only model-based maps where $p = 1$ is assumed at each voxel. Comparing these maps to Figure 5, it could be argued that the $\Lambda_{M,1}$ map detects basically the same region of voxels as the $\Lambda_{M,\hat{p}_M}$ map. However, this does not seem to be true for CV data-based LRTSs. The $\Lambda_{C,\hat{p}_C}$ map has lower p-values inside the left functional central sulcus than the $\Lambda_{C,1}$ map, but the latter has lower p-values for voxels detected outside this region. This suggests the trouble of detecting the AR order for the complex-valued model is worth the effort.

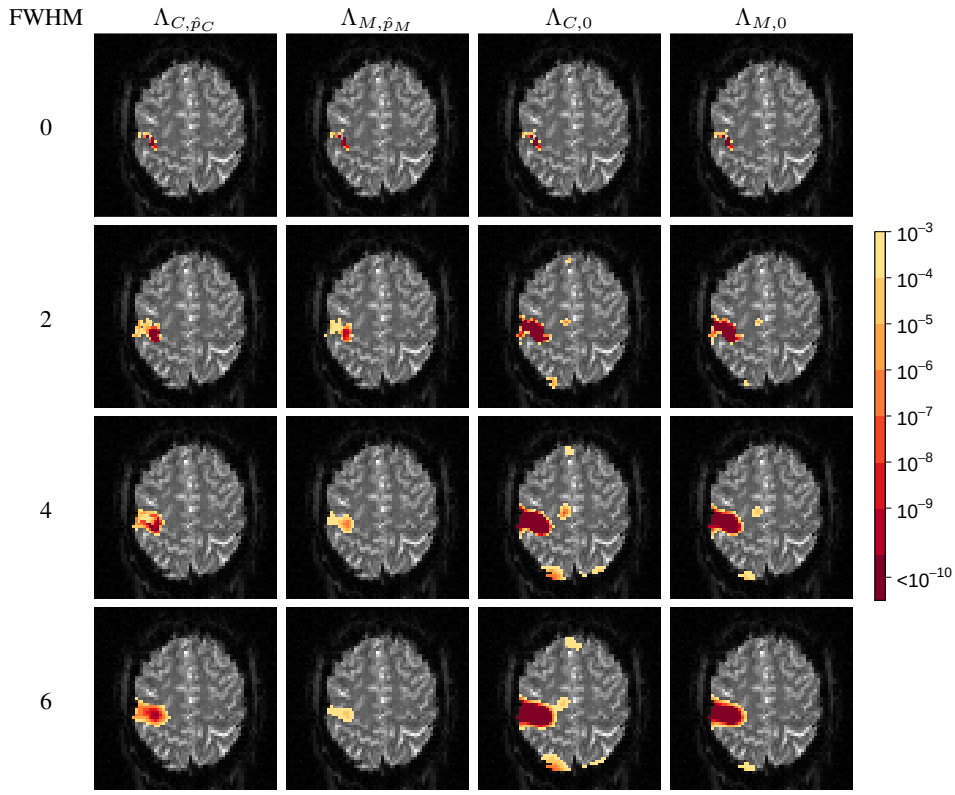


FIG S-5. *Activation images for slice 5 organized according to FWHM of kernel smoother (rows) and activation statistic (columns).*

S-3.3. *Supplement to Section 4.4 – Effect of spatial smoothing on parameter estimates.* Several supplementary figures are described in the following. Figures S-9 and S-10 are maps of the detected orders under the CV and MO models. respectively, for the different amounts of spatial smoothing, and Figure S-11 shows the corresponding frequency distributions. We note that $\hat{p} = 0$ (under both models) for the vast majority of voxels for the unsmoothed data, but the detected orders increase for smoothed data. Figure S-12 gives the images of the parameter estimates not shown in Figure 6. The effect of spatial smoothing on these parameter estimates is more predictable in that increased spatial smoothing decreases the resolution of $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\theta}$ and decreases the noise standard deviations $\hat{\sigma}_R$ and $\hat{\sigma}_I$ (as expected under the frameworks of Nencka, Hahn and Rowe, 2009; Rowe, 2016). Last, Figure S-13 shows, for different smoothing levels, the proportions of voxels in which the null hypotheses $H_0 : \sigma_R^2 = \sigma_I^2$ and $H_0 : \rho = 0$ are rejected after controlling the FDR at 0.05. For the unsmoothed data, it is evident that the assumptions of equal

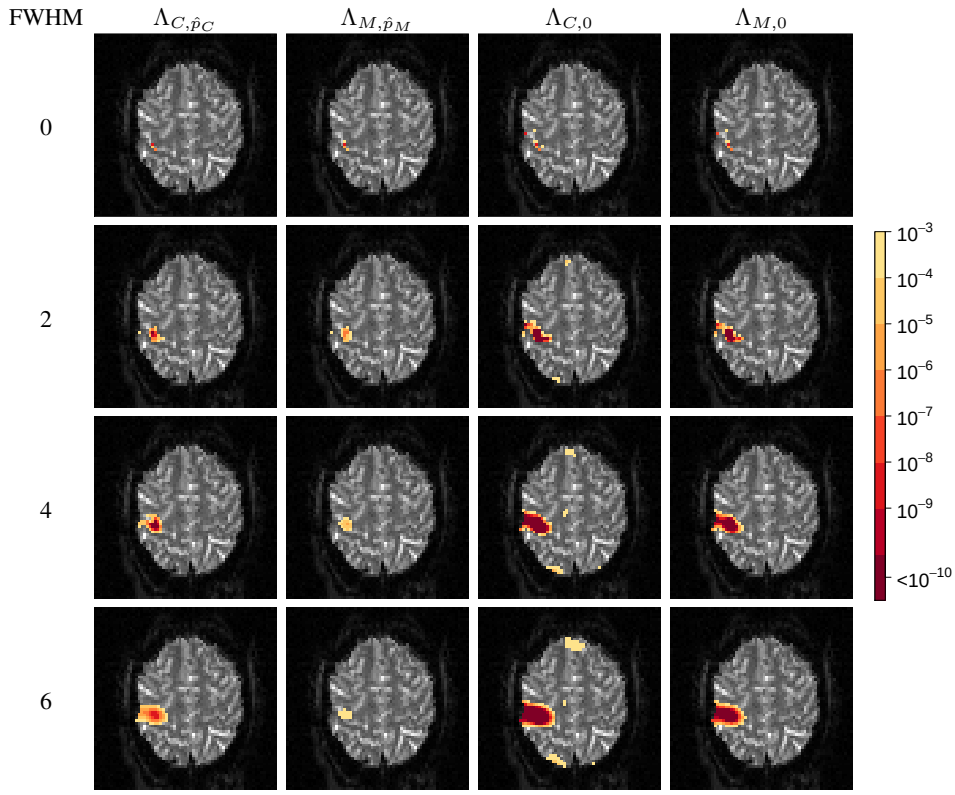
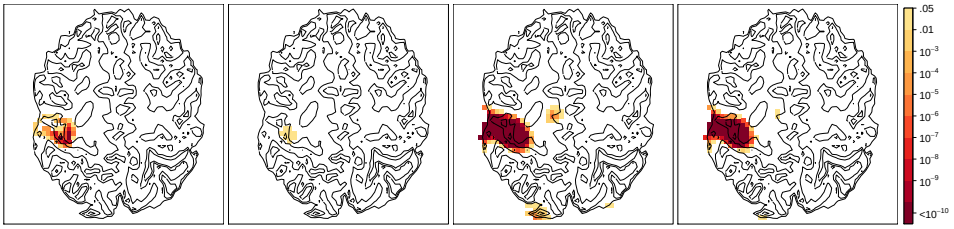


FIG S-6. *Activation images for slice 6 organized according to FWHM of kernel smoother (rows) and activation statistic (columns).*

variances and independence of the real and imaginary components are reasonable for the vast majority of voxels. However, these assumptions are rejected for sizable proportions of voxels for the smoothed data, and the proportions increase with the FWHM.

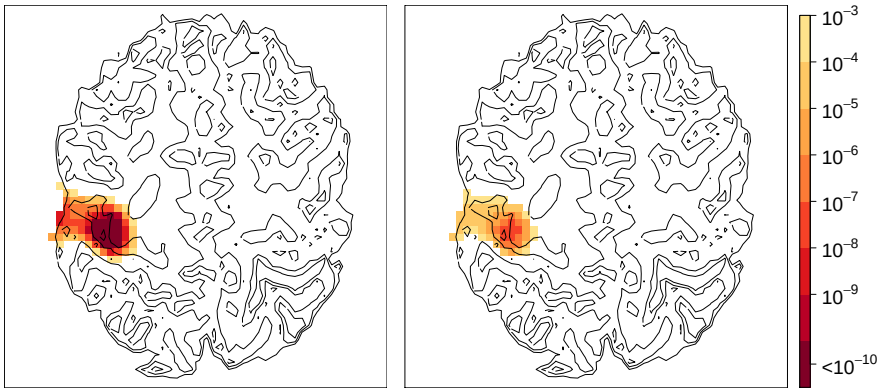
S-3.3.1. *Derivations under simplifying assumptions.* To clarify the discussion in Section 4.4 on how model (4.1) supports the observed dependence of the parameter estimates on spatial smoothing, we provide some derivations under simplifying assumptions. We will derive the real/imaginary correlation and the lag-1 temporal autocorrelation[1] within the real and imaginary time series and model (4.1) and the following assumptions. First, we assume that the spatial context is one-dimensional and that spatial smoothing is performed using a simple moving

[1] The lag-1 temporal autocorrelation is equal to the AR(1) coefficient for a truly AR(1) time series.



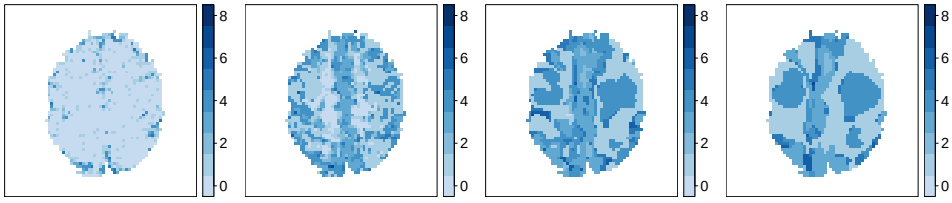
(a) $\Lambda_{C,\hat{p}_C}$ (b) $\Lambda_{M,\hat{p}_M}$ (c) $\Lambda_{C,0}$ (d) $\Lambda_{M,0}$

FIG S-7. *False discovery rate (FDR) thresholded activation maps corresponding to those in Figure 5.*



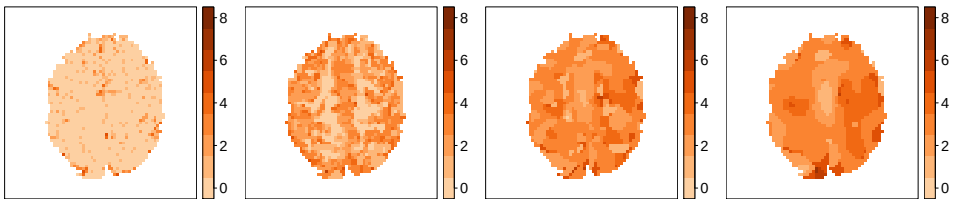
(a) Complex-valued AR(1) model (b) Magnitude-only AR(1) model

FIG S-8. *Activation maps for the finger-tapping experiment corresponding to Figure 5 where an AR order of 1 is assumed for every voxel time series.*



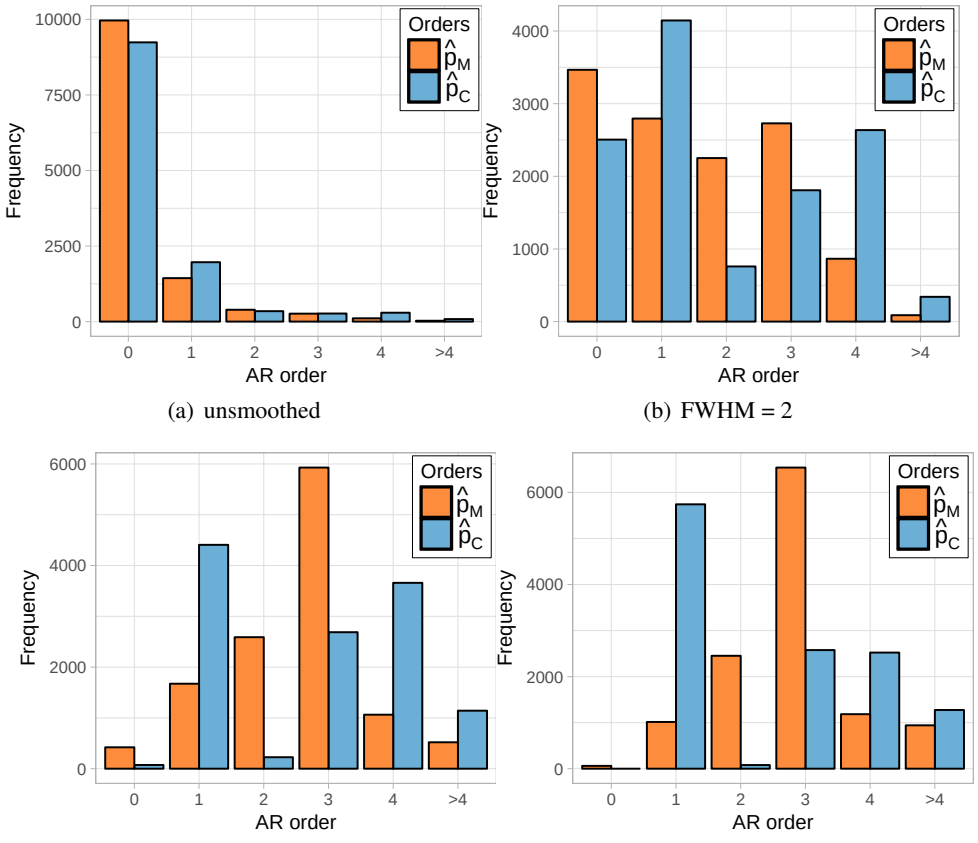
(a) $\hat{p}_C$, FWHM = 0 (b) $\hat{p}_C$, FWHM = 2 (c) $\hat{p}_C$, FWHM = 4 (d) $\hat{p}_C$, FWHM = 6

FIG S-9. *Images of the detected AR orders $\hat{p}_C$ under the complex-valued model for slice 5, under (a) the absence of spatial smoothing and (b-d) smoothing where the kernel has FWHMs = 2, 4, and 6 voxels.*



(a) $\hat{p}_M$, FWHM = 0 (b) $\hat{p}_M$, FWHM = 2 (c) $\hat{p}_M$, FWHM = 4 (d) $\hat{p}_M$, FWHM = 6

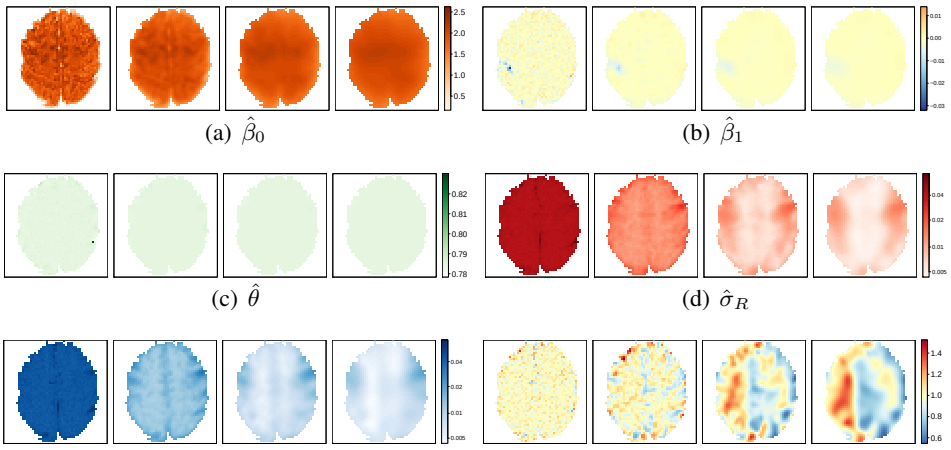
FIG S-10. *Images of the detected AR orders $\hat{p}_M$ under the magnitude-valued model for slice 5, under (a) the absence of spatial smoothing and (b-d) smoothing where the kernel has FWHMs = 2, 4, and 6 voxels.*



(a) unsmoothed (b) FWHM = 2

(c) FWHM = 4 (d) FWHM = 6

FIG S-11. *Frequency distributions of the complex-valued and magnitude-only fitted AR orders $\hat{p}_C$ and $\hat{p}_M$ with (a) no smoothing and (b-d) when spatial smoothing is first performed with different FWHMs.*



(a) $\hat{\beta}_0$ (b) $\hat{\beta}_1$

(c) $\hat{\theta}$ (d) $\hat{\sigma}_R$

(e) $\hat{\sigma}_I$ (f) $\hat{\sigma}_R/\hat{\sigma}_I$

FIG S-12. *Parameter estimates (a) $\hat{\beta}_0$, (b) $\hat{\beta}_1$, (c) $\hat{\theta}$, (d) $\hat{\sigma}_R$, (e) $\hat{\sigma}_I$, and (f) $\hat{\sigma}_R/\hat{\sigma}_I$ for (moving left to right within each subfigure) spatial smoothing with FWHM = 0, 2, 4, and 6 voxels. (FWHM = 0 means no spatial smoothing.)*

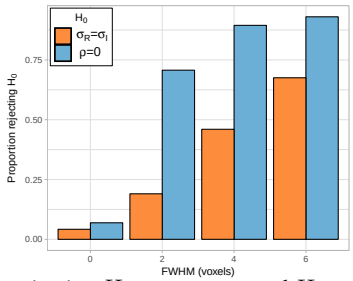
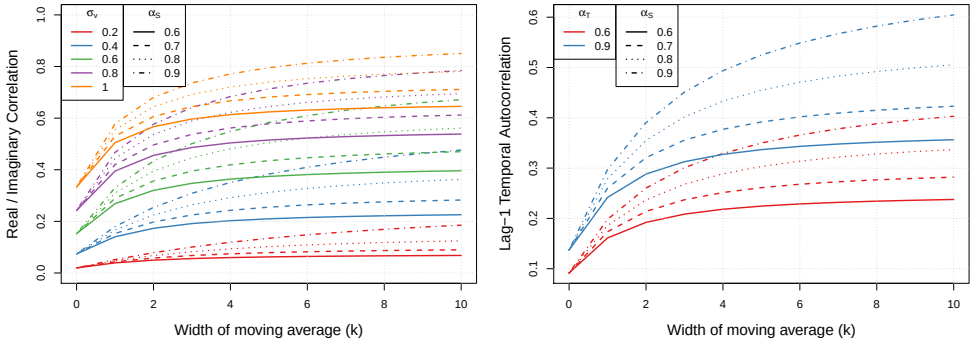


FIG S-13. *Proportion of voxels rejecting $H_0 : \sigma_R = \sigma_I$ and $H_0 : \rho = 0$ after controlling the FDR at 0.05 for different levels of spatial smoothing (FWHMs).*

average of length N (assumed to be odd such that $N = 2k + 1$ for $k \in \mathbb{N}$) such that spatial smoothing at time t results in the real/imaginary observations at voxel s given by $\tilde{y}_{Rst} = (1/N)\sum_{i=-k}^{k} y_{R,s+i,t}$ and $\tilde{y}_{Ist} = (1/N)\sum_{i=-k}^{k} y_{I,s+i,t}$. Next, we assume that $\boldsymbol{v} = (\boldsymbol{v}_1', \boldsymbol{v}_2', \ldots, \boldsymbol{v}_S')'$, where $\boldsymbol{v}_s = (v_{s1}, \ldots, v_{sn})'$, is multivariate normal with mean zero and a covariance matrix given as a Kronecker product of spatial and temporal AR(1) covariance matrices. That is, we assume that $\text{Cov}(\boldsymbol{v}) = \sigma_v^2 \boldsymbol{R}_S \otimes \boldsymbol{R}_T$, where $\boldsymbol{R}_S$ and $\boldsymbol{R}_T$ are AR(1) correlation matrices of orders S and n, respectively, based on the AR parameters α_S and α_T. Last, we assume that w_{Rst} and w_{Ist} are i.i.d. $N(0, \sigma_w^2)$, and $\theta = \pi/4$ as specified by the CV running line. Under this framework, the real/imaginary cross-correlation for the unsmoothed data is $\rho = \text{Cor}(y_{Rst}, y_{Ist}) = 0.5\sigma_v^2/(0.5\sigma_v^2 + \sigma_w^2)$ and the lag-1 autocorrelation is $\alpha_1 = \text{Cor}(y_{Rst}, y_{Rs,t-1}) = \text{Cor}(y_{Ist}, y_{Is,t-1}) = \alpha_T \rho$. Note that these expressions support the statement that the fact that $\hat{\rho}$ and $\hat{\alpha}_1$ are close to zero for the unsmoothed data suggests σ_w^2 is larger than σ_v^2. When smooth-

ing is applied, $\tilde{\sigma}_w^2 \equiv \text{Var}(\tilde{w}_{Rst}) = \text{Var}(\tilde{w}_{Ist}) = \sigma_w^2/N$ and $\tilde{\sigma}_v^2 \equiv \text{Var}(\tilde{v}_{st}) = (\sigma_v^2/N^2)\sum_{i=1}^N \sum_{j=1}^N \alpha_S^{|i-j|}$, where the terms $\tilde{w}_{Rst}$, $\tilde{w}_{Ist}$, and $\tilde{v}_{st}$ are defined similarly to $\tilde{y}_{Rst}$ and $\tilde{y}_{Ist}$. When α_S is positive, $N < \sum_{i=1}^N \sum_{j=1}^N \alpha_S^{|i-j|}$, which supports the statement that the variance of the v_{st}s is less affected by spatial averaging than the (w_{Rst}, w_{Ist})s. It can be shown that for the smoothed data, the real/imaginary cross-correlation is $\tilde{\rho} = 0.5\tilde{\sigma}_v^2/(0.5\tilde{\sigma}_v^2 + \tilde{\sigma}_w^2)$ and the lag-1 temporal autocorrelation is $\tilde{\alpha}_1 = \alpha_T\tilde{\rho}$. It then follows that $\tilde{\rho}$ and $\tilde{\alpha}_1$ increase with the amount of spatial smoothing – given by the value of k. Figure S-14 displays values of $\tilde{\rho}$ and $\tilde{\alpha}_1$ based on the above derivations for different k at $\sigma_w^2 = 1$ and different values of σ_v^2, α_S, and α_T.

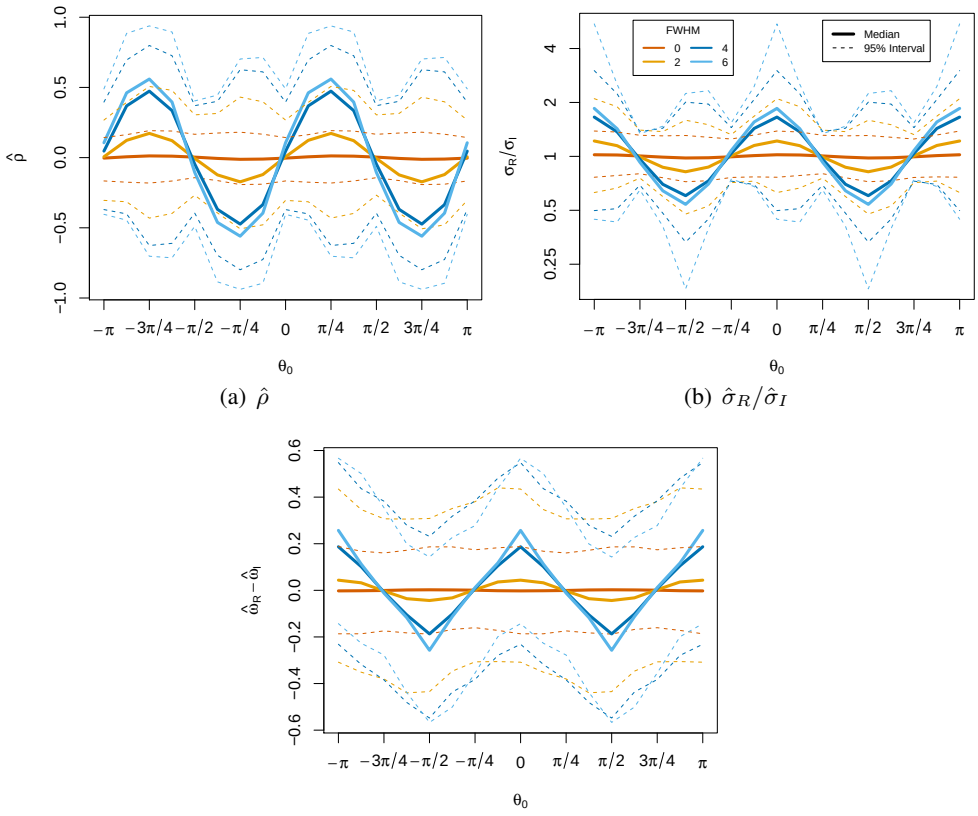


(a) Real/imaginary correlation $\tilde{\rho}$ (b) Lag-1 Temporal Autocorrelation $\tilde{\alpha}_1$

FIG S-14. *Values of $\tilde{\rho}$ and $\tilde{\alpha}_1$ for different k at $\sigma_w^2 = 1$ and different values of σ_v^2, α_S, and α_T, based on the derivations in Section S-3.3.1.*

S-3.3.2. *Effect of phase on parameter estimates.* Parameter estimates including $\hat{\rho}$, $\hat{\sigma}_R/\hat{\sigma}_I$, and the difference between the first-order real and imaginary residual autocorrelations (denoted by $\hat{\omega}_R - \hat{\omega}_I$) are dependent on the choice of θ_0. Figure S-15 shows that the medians of these parameter estimates are constant at zero for the unsmoothed data but are periodic in θ_0 for the smoothed data, with an amplitude that increases with the FWHM. Note that the periodicity of $\hat{\rho}$ shares the phase of $\sin 2\theta$ while the periodicity of $\hat{\sigma}_R/\hat{\sigma}_I$ and $\hat{\omega}_R - \hat{\omega}_I$ shares the phase of $\cos 2\theta$. In fact, we chose $\theta_0 = \pi/4$ because at this value the distribution of $\hat{\omega}_R - \hat{\omega}_I$ is centered around zero, which is necessary under the model assumption that the real and imaginary errors share the same AR coefficients.

In the following, we show that the periodicity discussed above is expected under model (4.1). Letting σ_v^2 = Var(v_{st}) and w_{Rst} and w_{Ist} be independent with equal



(a) $\hat{\rho}$ (b) $\hat{\sigma}_R/\hat{\sigma}_I$

(c) $\hat{\omega}_R - \hat{\omega}_I$

FIG S-15. *Plots showing the dependence of the distribution of (a) $\hat{\rho}$, (b) $\hat{\sigma}_R/\hat{\sigma}_I$, and (c) $\hat{\omega}_R - \hat{\omega}_I$ on the central phase θ_0 assigned to each voxel time series for different amounts of spatial smoothing.*

variance σ_w^2 and $\alpha_1 = \text{Cor}(v_{st}, v_{s,t-1})$, it can be shown that

$$\text{(S-14)} \qquad \rho = \text{Cor}(y_{Rst}, y_{Ist}) = \frac{\sigma_v^2 \sin 2\theta}{[\sigma_v^4 \sin^2 2\theta + 4\sigma_w^2(\sigma_v^2 + \sigma_w^2)]^{1/2}},$$

$$\text{(S-15)} \qquad \frac{\sigma_R^2}{\sigma_I^2} = \frac{\text{Var}(y_{Rst})}{\text{Var}(y_{Ist})} = 1 + \frac{2\sigma_v^2 \cos 2\theta}{\sigma_v^2(1 - \cos 2\theta) + 2\sigma_w^2},$$

$$\text{(S-16)} \quad \omega_R - \omega_I = \frac{4\alpha_1 \sigma_v^2 \sigma_w^2 \cos 2\theta}{\sigma_v^4(1 - \cos^2 2\theta) + 4\sigma_w^2(\sigma_v^2 + \sigma_w^2)},$$

and hence ρ shares the phase of $\sin 2\theta$. and σ_R^2/σ_I^2 and $\omega_R - \omega_I$ share the phase of $\cos 2\theta$.

S-4. Supplement to Section 5 – Further Simulation-based Analyses.

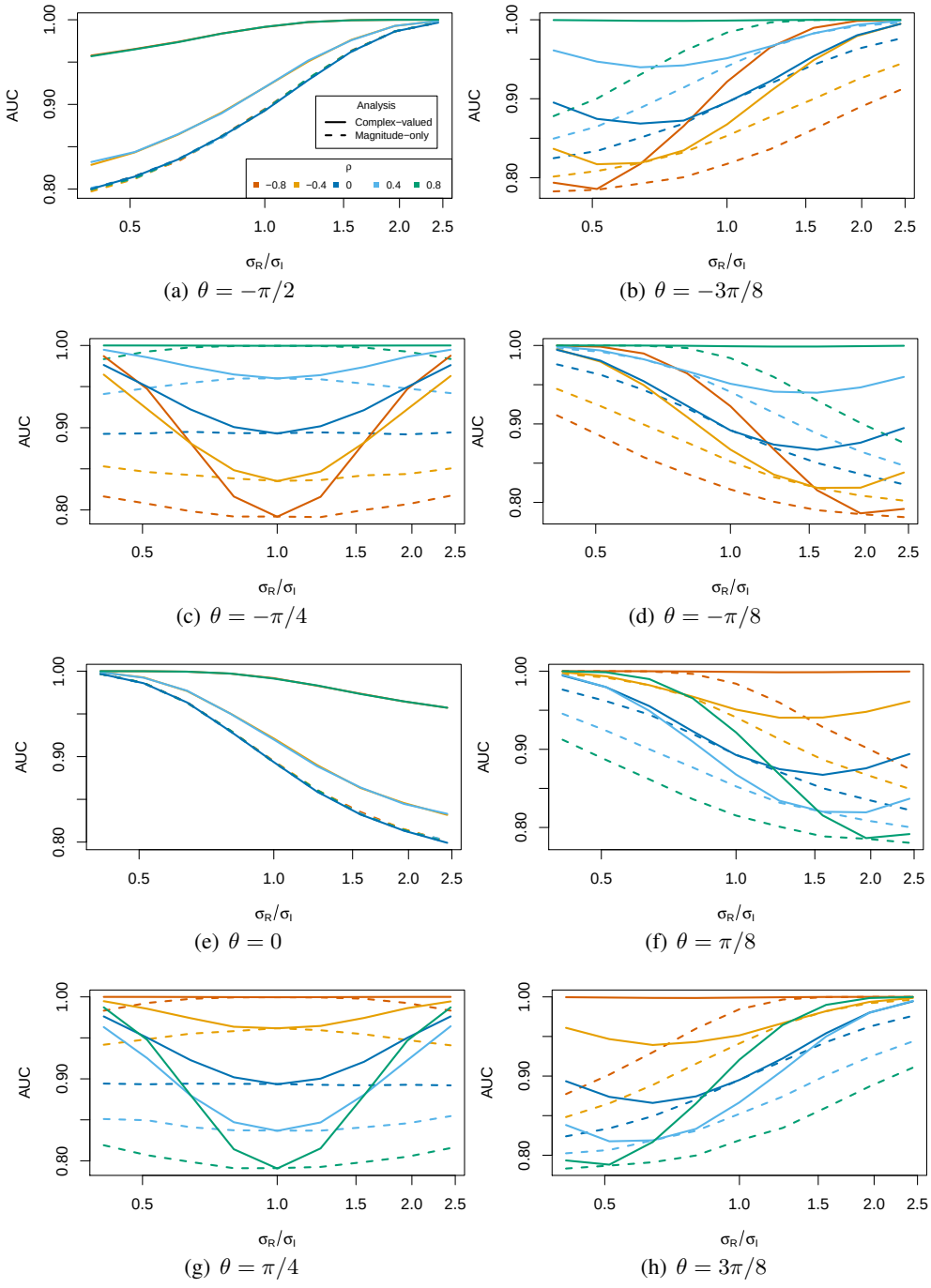
S-4.1. *Supplement to Section 5.1 – Calculation of the AUC.* Denoting the test statistics computed under H_0 and H_a by $\{T_{0i}\}_{i=1}^{n_0}$ and $\{T_{aj}\}_{j=1}^{n_a}$, respectively,

Bamber (1975) computed the AUC as $[\sum_{i=1}^{n_0}\sum_{j=1}^{n_a}\mathbf{1}(T_{0i} < T_{aj})]/n_0 n_a$, where $\mathbf{1}(A)$ is the indicator function taking the value 1 if A is true and 0 otherwise. A test with higher AUC has greater ability to discriminate between statistics computed under H_0 and H_a, as the above formula computes the proportion of null-alternative statistic pairs in which the rule $\mathbf{1}(T_{0i} < T_{aj})$ discriminates the null and alternative statistics correctly.

S-4.2. *Supplement to Section 5.2 – Influence of central phase on AUC functions.* In the context of the simulation study in Section 5.2, the form of the power/AUC functions in terms of σ_R/σ_I and ρ depends on the value of θ. This is evident from Figure S-16, which shows the AUC functions corresponding to Figure 9(c) for various values of θ (Note that Figure 9(c) was calculated with $\theta = \pi/4$; further, the corresponding power functions are not shown for the sake of brevity because they show very similar patterns to the AUCs functions.) Two examples of θ-dependence are as follows: first, while the AUCs are greater for negative ρ than positive ρ when $\theta = \pi/4$, the opposite is true when $\theta = -\pi/4$. Second, when the value of σ_R/σ_I increases, AUCs increase for $\theta = -\pi/2$ but decrease for $\theta = 0$. Before we analyze these patterns, we emphasize that the AUCs for the CV model-based LRTs are greater than or equal to their MO counterparts in all cases. Further, only in "special cases" are the AUCs equal; some of these include when $\theta = \pm\pi/4$ and $\sigma_R = \sigma_I$ and when $\theta = -\pi/2$ or 0 and $\rho = 0$.[2]

We use two methods to explain the dependence of the power/AUC functions on σ_R/σ_I, ρ, and θ. The first involves visualizing the BOLD response and the real/imaginary error distribution of model (3.2) in the complex plane. As shown in Figure S-17, the BOLD response can be visualized as a periodic oscillation around the baseline signal in the θ direction. Specifically, in each subfigure, the large dashed circle centered at the origin has a radius that represents the magnitude of the baseline signal. The point on this circle (in the θ direction from the origin) represents the complex-valued baseline signal, and the double-headed arrow represents the oscillation of the BOLD response. Now, because the real/imaginary error distribution is bivariate normal with covariance matrix $\mathbf{\Sigma}$ (3.3), constant probability density contours are given by ellipses with major and minor axes in the direction of the eigenvectors of $\mathbf{\Sigma}$ (Johnson and Wichern, 2007). These ellipses are shown in Figure S-17 for different values of ρ assuming that $\sigma_R = \sigma_I$. When $\rho > 0$, the major (*i.e.* longer) axis of the elliptical contours is in the $\pi/4$ direction and the minor axis is in the $-\pi/4$ direction, and the direction of the major/minor axes is reversed when $\rho < 0$. Heuristically, a larger "effective CNR" (and therefore a

[2]The latter case is not obvious because many of the lines in Figures S-16(a),(e) coincide: the top line contains the AUCs for $\rho = \pm 0.8$ and the complex-valued model (only), the next line are for $\rho = \pm 0.4$ and the complex-valued model (only), and the bottom line contains the AUCs for $\rho = 0$ and the complex-valued model and all values of ρ for the magnitude-only model.



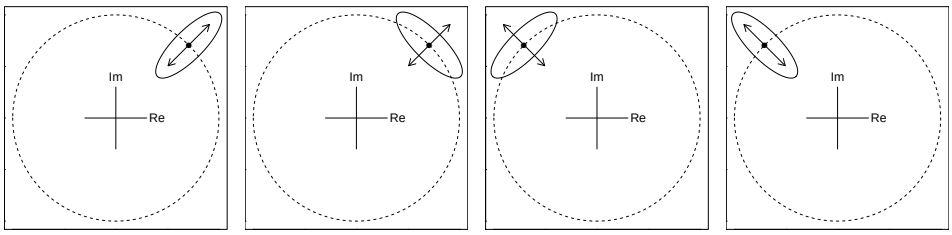
(a) $\theta = -\pi/2$ (b) $\theta = -3\pi/8$

(c) $\theta = -\pi/4$ (d) $\theta = -\pi/8$

(e) $\theta = 0$ (f) $\theta = \pi/8$

(g) $\theta = \pi/4$ (h) $\theta = 3\pi/8$

FIG S-16. *Simulation-based plots showing the AUC for the complex-valued and magnitude-only activation statistics in terms of σ_R/σ_I and ρ for different values of θ.*



(a) $\theta{=}\pi/4, \rho{=}0.8$ (b) $\theta{=}\pi/4, \rho{=}-0.8$ (c) $\theta{=}-\pi/4, \rho{=}0.8$ (d) $\theta{=}-\pi/4, \rho{=}-0.8$

FIG S-17. *Pictorial representation of the effect of θ and ρ on the "effective CNR".*

larger power and AUC as well) results when the direction of the oscillation of the BOLD response is aligned with the narrow (*i.e.* minor) axis than when it is aligned with the major axis of the elliptical contours of the error distribution. This explains why the power/AUC are greater for negative ρ than positive ρ when $\theta = \pi/4$, as in Figure 9(c) (see Figures S-17ab), but the reverse occurs when $\theta = -\pi/4$ (see Figures S-17cd).

The other explanation involves the (observed) Fisher information matrix, whose block corresponding to $\boldsymbol{\beta}$ is given by

$$\text{(S-17)} \quad \boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \frac{1}{2(1-\rho^2)} \boldsymbol{X}' \boldsymbol{R}_n^{-1} \boldsymbol{X} \left[\frac{1}{\sigma_R^2} \cos^2\theta + \frac{1}{\sigma_I^2} \sin^2\theta - \frac{2\rho}{\sigma_R \sigma_I} \sin 2\theta \right].$$

For example, considering $\theta = 0$, we see that $\boldsymbol{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}$ decreases when σ_R^2 increases. It then follows that the standard error of $\hat{\beta}_1$ increases, that the significance of β_1 decreases, and therefore that the power/AUC decreases as σ_R^2 increases, as shown in Figure S-16(e). In addition, the pairing and relative position of the $\rho = \pm 0.4$ and ± 0.8 lines in Figure S-16(e) can be explained by the $(1-\rho^2)$ term in (S-17).

S-4.3. *Supplement to Section 5.3 – Simulation study with higher AR order.* To examine the effect of temporal dependence of a higher AR order than one, we simulated complex-valued time series with AR(3) errors with coefficients $\boldsymbol{\alpha} = (0.2, 0.2, 0.2)$. We computed the complex-valued AR(p) model-based LRT statistics for activation based on orders of $p = 0, 1, 2, 3, 4, 5$ to examine the effect of under- and over-specification of the order. The false positive rates using a significance level of 0.05 are given in Table S-3. As demonstrated in Section 5.3, the

AR order p	0	1	2	3	4	5
False detection rate	0.255	0.117	0.061	0.053	0.053	0.053

TABLE S-3

False detection rates of complex-valued model-based activation statistics based on AR orders of $p = 0, 1, 2, 3, 4, 5$ for simulated complex-valued time series with a true order of 3.

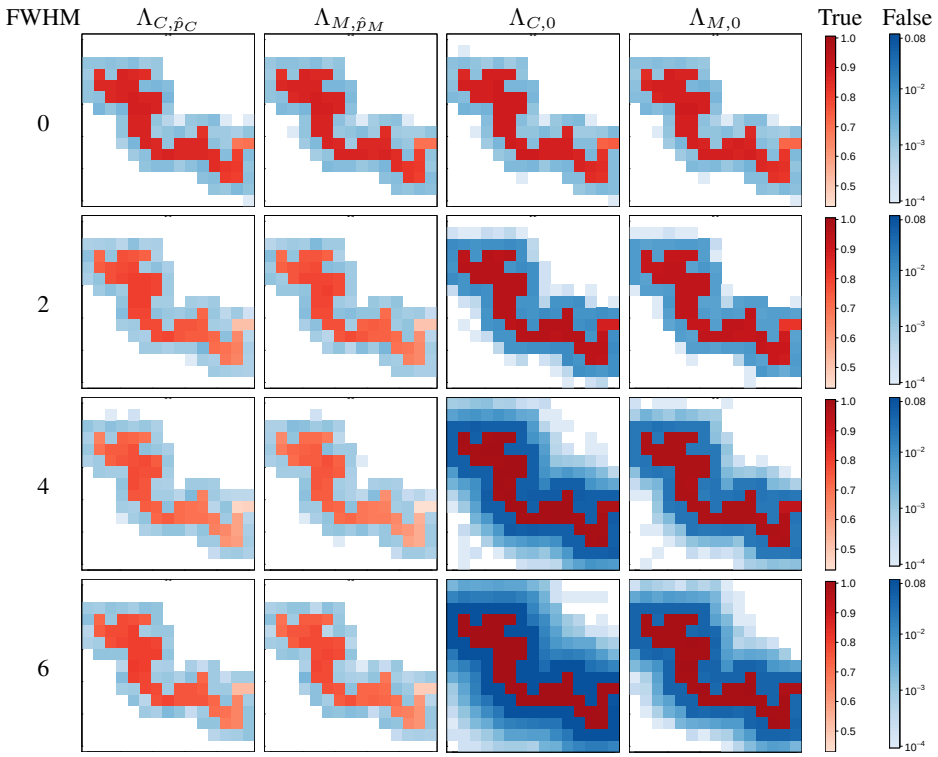
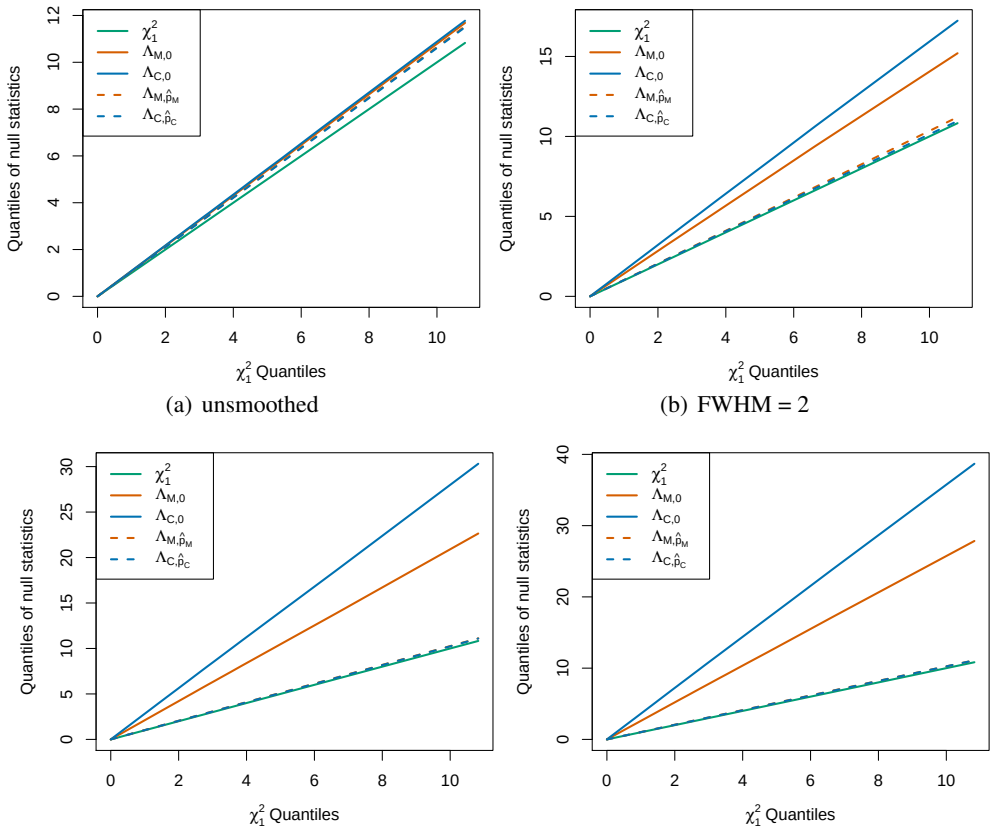


FIG S-18. *True and false detection rates (corresponding to the legends at right) for the simulated ROIs with parameter values corresponding to FWHM values at right as in Table 1.*

false positive rates of the LRTSs assuming independence are inflated when there is temporal dependence in the data. Further, we see in this study that even if AR dependence is modeled, the false positive rate is also inflated if the order is under-specified. However, the Type I Error rate is controlled appropriately the AR order is specified correctly or over-specified.

S-4.4. *Supplement to Section 5.4 – Examining a region of interest (ROI): more details.* We provide three sets of supplementary plots. First, Figure S-18 contains maps of the true and false activation detection rates for all four of the parameter sets in Table 1. These maps correspond to Figure 11, which shows the results for the FWHM = 4 parameter set. Second, Figure S-19 gives quantile-quantile plots comparing the bootstrap-based null distributions of the four statistics against the theoretical χ^2_1 distribution. These plots demonstrate that the independent-model-based activation statistics do not follow the χ^2_1 null distribution under H_0 for the parameter sets corresponding to the smoothed data. Last, Figure S-20 shows kernel density estimates of the distributions of the Jaccard indices over ROI replications for each statistic, which demonstrate the wide variability of the indices over repli-



(a) unsmoothed (b) FWHM = 2

(c) FWHM = 4 (d) FWHM = 6

FIG S-19. *Quantile-quantile plots comparing the null distributions of the two activation statistics against the χ^2_1 distribution for the four parameter sets (according to FWHMs) in Table 1.*

cations.

References.

BAMBER, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology* **12** 387-415.

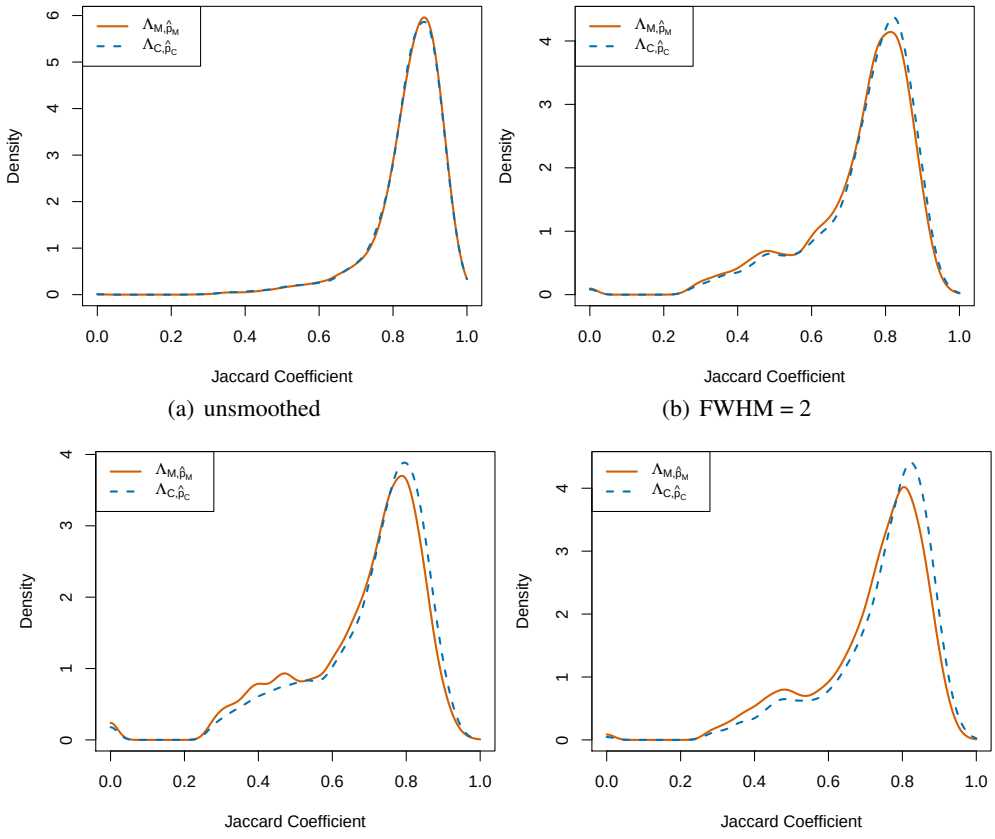
BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodology)* **57** 289-300.

BENJAMINI, Y. and YEKUTIELI, Y. (2001). The Control of the False Discovery Rate in Multiple Testing under Dependency. *The Annals of Statistics* **29** 1165-1188.

BOX, G. and PIERCE, D. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* **65** 1509-1526.

COCHRANE, D. and ORCUTT, G. (1949). Applications of least squares regression to relationships containing autocorrelated errors. *Journal of the American Statistical Association* **44** 32-61.

COX, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and biomedical research, an international journal* **29** 162-173.



(a) unsmoothed (b) FWHM = 2

(c) FWHM = 4 (d) FWHM = 6

FIG S-20. *Kernel density estimates of the distributions of the Jaccard indices over ROI replications for each statistic.*

COX, R. W. (2012). AFNI: What a long strange trip it has been. *NeuroImage* **62** 743-747.

JOHNSON, R. A. and WICHERN, D. W. (2007). *Applied Multivariate Statistical Analysis, 6th Edition*. Pearson Prentice Hall.

LJUNG, G. and BOX, G. (1978). On a measure of lack of fit in time series models. *Biometrika* **65** 297-303.

MILLER, J. W. (1995). Exact Maximum Likelihood Estimation in Autoregressive Processes. *Journal of Time Series Analysis* **16** 607-615.

NENCKA, A. S., HAHN, A. D. and ROWE, D. B. (2009). A Mathematical Model for Understanding the STatistical effects of k-space (AMMUST-k) preprocessing on observed voxel measurements in fcMRI and fMRI. *Journal of Neuroscience Methods* **181** 268-282.

R CORE TEAM, (2016). R: A Language and Environment for Statistical Computing, Vienna, Austria.

ROWE, D. B. (2016). *Handbook of Neuroimaging Data Analysis* Image Reconstruction in Functional MRI 205-232. Chapman & Hall/CRC.

SHUMWAY, R. H. and STOFFER, D. S. (2006). *Time Series Analysis and Its Applications*, Second Edition ed. Springer.

DEPARTMENT OF STATISTICS
GRAND VALLEY STATE UNIVERSITY
ALLENDALE, MI 49401-9403
USA
E-MAIL: adriand1@gvsu.edu

DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IA 50011-1090
USA
E-MAIL: maitra@iastate.edu

DEPARTMENT OF MATHEMATICS, STATISTICS AND COMPUTER SCIENCE
MARQUETTE UNIVERSITY
MILWAUKEE, WI 53233
USA
E-MAIL: daniel.rowe@marquette.edu