Computing FMRI Activations: Coefficients and t-Statistics by Detrending and Multiple Regression

Daniel B. Rowe and Steven W. Morgan

Division of Biostatistics
Medical College of Wisconsin

Technical Report 39

June 2002

Division of Biostatistics
Medical College of Wisconsin
8701 Watertown Plank Road
Milwaukee, WI 53226
Phone:(414) 456-8280

MEDICAL COLLEGE OF WISCONSIN

Computing FMRI Activations: Coefficients and t-Statistics by Detrending and Multiple Regression

Daniel B. Rowe[*,†,‡] and Steven W. Morgan[*]

Abstract

Detrending is often performed when analyzing FMRI data in order to remove any linear (or higher-order) drift and offset from the signal. When performing simple linear regression after detrending, the results obtained are not identical to those obtained by multiple linear regression. In fact, even if there is no error in the data, detrending and simple linear regression are unable to determine exactly the correct coefficients in simulations. In addition, the choice of reference function, whether it be a square wave of ones and zeros or negative ones and ones, also affects the obtained result. Particular care should be taken when detrending and using a zero-one square wave as a reference function. The t-statistics for such a case do not give an accurate estimation of the error in the data, yielding too many false negatives.

[*]Biophysics Research Institute, Medical College of Wisconsin
[†]Division of Biostatistics, Medical College of Wisconsin
[‡] Corresponding author:
Biophysics Research Institute
Medical College of Wisconsin
8701 Watertown Plank Rd.
Milwaukee, WI 53226
Phone: 414-456-4027
Email: dbrowe@mcw.edu

1 Introduction

When fitting the FMRI signal to an idealized reference function such as a square wave, the method of detrending followed by simple linear regression is often used where an estimated linear trend (and offset) is subtracted and the difference fit to the reference function. The motivation for this is that there is often linear drift that occurs throughout time in the FMRI signal (such as due to patient motion), in addition to an offset in the signal itself. The multiple linear regression method [6, 5, 8] takes no such step, instead dealing with any possible drift simultaneously with the reference function. Detrending followed by simple linear regression is in general not equivalent to multiple regression [3], so if the offset and trend are removed, care should be taken in how the data is analyzed. This paper will only look at the square wave reference function and its use in block design FMRI.

2 Mathematics

2.1 Model

The determination of whether functional activation has occurred in a block-design FMRI experiment is based upon the regression coefficients from the fit of the BOLD signal to an idealized reference function [1, 2] which is often taken to be a square wave. The equation to be fit for each (assumed to be independent) voxel is given by

$$\underset{n\times 1}{y} = \underset{n\times (q+1)}{X} \underset{(q+1)\times 1}{\beta} + \underset{n\times 1}{\epsilon}. \tag{1}$$

In Eq. 1, y is a vector containing the observed signal for each of n time points. The design matrix X has dimensions $n \times (q+1)$. An example in which $q = 2$ is given in Table 1, where it has columns which consist of an n dimensional column of ones (which corresponds to the intercept term), a column of the first n counting numbers which may or may not be centered about their mean (this vector is for the linear trend), and a column of the n-dimensional reference function. The matrix containing the first set of column vectors (in this case two) is denoted by x_1 and the matrix containing the last set of column vectors (the reference function) by x_2, so that $X = (x_1, x_2)$. The vectors β and ϵ have dimensions $(q+1) \times 1$ and $n \times 1$ respectively. They contain the coefficient and error vectors respectively.

x_1		x_2
1	1	0 or -1
1	2	0 or -1
$\vdots$	$\vdots$	$\vdots$
1	n-1	1
1	n	1

Table 1: The design matrix X, containing x_1 and x_2.

2.2 Coefficients

The multiple regression method estimates the coefficients of the model given in Eq. 1 above. To estimate the coefficients $\hat{\beta}_m$, the error elements of ϵ are assumed to be independently normally distributed with common mean zero and variance σ^2, so that the likelihood of the observations is found to be the normal distribution

$$f(y) = (2\pi\sigma^2)^{-n/2} e^{-(y-X\beta)'(y-X\beta)/2\sigma^2}. \tag{2}$$

It is noted [7], that by performing some algebra in the exponent of the above likelihood,

$$(y - X\beta)'(y - X\beta) = (\beta - \hat{\beta}_m)'(X'X)(\beta - \hat{\beta}_m) + y'[I_n - X(X'X)^{-1}X']y, \tag{3}$$

where $\hat{\beta}_m$ is defined to be

$$\hat{\beta}_m = (X'X)^{-1}X'y. \tag{4}$$

Having done this, the likelihood becomes

$$f(y) = (2\pi\sigma^2)^{-n/2} e^{-[(\beta-\hat{\beta}_m)'(X'X)(\beta-\hat{\beta}_m)+C]/(2\sigma^2)}, \tag{5}$$

where

$$\begin{aligned} C &= y'[I_n - X(X'X)^{-1}X']y \\ &= (y - X\hat{\beta}_m)'(y - X\hat{\beta}_m) \end{aligned} \tag{6}$$

does not depend on β. It is seen that the value of β which maximizes the likelihood (or minimizes the exponent) in Eq. 5 is given by $\beta = \hat{\beta}_m$.

The matrix $(X'X)$ given by

$$A = (X'X) = \begin{pmatrix} x_1'x_1 & x_1'x_2 \\ x_2'x_1 & x_2'x_2 \end{pmatrix}, \tag{7}$$

is a partitioned matrix [6] whose inverse is given by

$$(X'X)^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \tag{8}$$

where the matrix elements for the inverse of $X'X$ are given by

$$\begin{aligned} B_{11} &= (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} &(9)\\ B_{22} &= (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} &(10)\\ B_{12} &= -A_{11}^{-1}A_{12}B_{22} &(11)\\ B_{21} &= -A_{22}^{-1}A_{21}B_{11}. &(12)\end{aligned}$$

This gives estimates for the last set of coefficients

$$\begin{aligned} \hat{\beta}_{2,m} &= [-(x_2'x_2)^{-1}x_2'x_1[x_1'x_1 - x_1'x_2(x_2'x_2)^{-1}x_2'x_1]^{-1}x_1' \\ &+ [x_2'x_2 - x_2'x_1(x_1'x_1)^{-1}x_1'x_2]^{-1}x_2']y. \qquad (13)\end{aligned}$$

where $\hat{\beta}_m = (\hat{\beta}_{1,m}, \hat{\beta}_{2,m})'$.

The estimated coefficients $\hat{\beta}_{2,d}$ for the last set of coefficients in the detrend followed by simple linear regression method (without an intercept) is obtained as follows. First subtract the estimated offset and linear trend then form the equation

$$z = y - x_1\hat{\beta}_{1,d} = x_2\beta_2 + \delta \qquad (14)$$

where δ is another $n \times 1$ error vector with elements which are independently normally distributed with common mean zero and variance γ^2. In the above equation, $\hat{\beta}_{1,d}$ is given by $(x_1'x_1)^{-1}x_1'y$. By the same steps as the previously described multiple regression method, the estimated regression coefficient for the detrend followed by simple linear regression method (without an intercept) is

$$\begin{aligned} \hat{\beta}_{2,d} &= (x_2'x_2)^{-1}x_2'z \\ &= (x_2'x_2)^{-1}x_2'(y - x_1\hat{\beta}_{1,d}) \\ &= (x_2'x_2)^{-1}x_2'(I - x_1(x_1'x_1)^{-1}x_1')y \\ &= [(x_2'x_2)^{-1}x_2' - (x_2'x_2)^{-1}x_2'x_1(x_1'x_1)^{-1}x_1']y. \qquad (15)\end{aligned}$$

It is easily seen that the coefficients estimates in Eqs. 13 and 15 are not identical. These two expressions are only equivalent if x_1 and x_2 are orthogonal (i.e. $x_1'x_2$ is the zero vector). This is not the case since an offset vector of 1's and a linear trend cannot both be orthogonal to the reference function. In fact, a relationship between these two coefficients has been presented [4].

2.3 T-Statistics

The distribution for the vector of estimated regression coefficient from multiple regression which is a multivariate-t distribution were found by making a change in variable and marginalizing in the likelihood, and finally making another change of variable [7]. The resulting marginal distribution is

$$f(\hat{\beta}_m) \quad = \quad \frac{\gamma}{[(X'X)^{-1} + (\hat{\beta}_m - \beta)C^{-1}(\hat{\beta}_m - \beta)']^{n/2}} \tag{16}$$

where γ is a normalizing constant. Using the property that the marginal distribution of any element of $\hat{\beta}_m$ is student-t distributed, the t-statistic for the estimated coefficients from multiple linear regression are

$$t_k = \frac{\hat{\beta}_{k,m} - \beta_k}{[(n-q-1)^{-1}V_{kk}(y - X\hat{\beta}_{k,m})'(y - X\hat{\beta}_{k,m})]^{1/2}}, \tag{17}$$

where k denotes the coefficient number $(0, ..., q)$. The terms n and q refer to the $n \times (q+1)$ dimensions of X (for this papers example, $q = 2$). The matrix V is given by $V = (X'X)^{-1}$.

For the detrend followed by simple linear regression method (without an intercept), the t-statistics for x_2 are similarly found as above and given by

$$v_\ell = \frac{\hat{\beta}_{\ell,d} - \beta_\ell}{[(n-q_2)^{-1}P_{\ell\ell}(z - x_2\hat{\beta}_{2,d})'(z - x_2\hat{\beta}_{2,d})]^{1/2}} \tag{18}$$

for $\ell = q_1 + 1, \ldots, q$.

In this equation, P is given by $(x_2'x_2)^{-1}$ (note that for this papers example, P is a scalar). The parameter $q_2 = q - q_1$ (which is one for this papers example). Under the null hypothesis, $\beta_k = 0$ and $\beta_\ell = 0$,

$$t_k = \frac{\hat{\beta}_{k,m}}{[(n-q-1)^{-1}V_{kk}(y - x_1\hat{\beta}_{1,m} - x_2\hat{\beta}_{2,m})'(y - x_1\hat{\beta}_{1,m} - x_2\hat{\beta}_{2,m})]^{1/2}} \tag{19}$$

and

$$v_\ell = \frac{\hat{\beta}_{\ell,d}}{[(n-q_2)^{-1}P_{\ell\ell}(y - x_1\hat{\beta}_{1,d} - x_2\hat{\beta}_{2,d})'(y - x_1\hat{\beta}_{1,d} - x_2\hat{\beta}_{2,d})]^{1/2}}. \tag{20}$$

It is useful to recognize that the t-statistics for this detrend method gives more degrees of freedom, and that neither the estimated coefficients (the numerators) nor their standard errors (the denominators) are equivalent.

Detrending and Regression with Intercept

It is quite common both to detrend and to use a $0, 1$ square wave reference function, so it is important to investigate why the results it gives will be different than using a $-1, 1$ square wave reference function or using multiple regression.

The choice of a $0, 1$ reference function can cause problems, because when detrending, the offset has been subtracted, but the reference function the data is fit to has values of 0 and 1, where the data actually has some negative values. To overcome this particular point, it is possible to include an additional intercept term when performing regression after detrending. If the vector of ones for the inclusion of an additional intercept term is denoted by x_*, then the estimated last set of coefficients (containing the reference function) are given by

$$\hat{\beta}_{2,d*} = (x_3'x_3)^{-1}x_3'z = [(x_3'x_3)^{-1}x_3' - (x_3'x_3)^{-1}x_3'x_1(x_1'x_1)^{-1}x_1']y \tag{21}$$

where $x_3 = (x_*, x_2)$.

The t-statistics for this method are given by

$$\begin{aligned} v_{m*} &= \frac{\hat{\beta}_{m,d*}}{[(n-q_2-1)^{-1}P_{mm}(z-x_3\hat{\beta}_{2,d*})'(z-x_3\hat{\beta}_{2,d*})]^{1/2}} \\ &= \frac{\hat{\beta}_{m,d*}}{[(n-q_2-1)^{-1}P_{mm}(y-x_1\hat{\beta}_{1,d}-x_3\hat{\beta}_{2,d*})'(y-x_1\hat{\beta}_{1,d}-x_3\hat{\beta}_{2,d*})]^{1/2}} \end{aligned} \tag{22}$$

where m denotes the coefficient number, $m = *, q_1+1, \ldots, q$ and $P = (x_3'x_3)^{-1}$. Note that there is an extra degree of freedom is subtracted off in the denominator due to the addition of another intercept.

3 Methods

3.1 Monte Carlo Simulation

Before performing the regression with data from an FMRI experiment, these three methods of estimating the coefficients and t-statistics were investigated by performing simulations. The simulation formulated the design matrix $X = (x_1, x_2)$, where the matrix containing the first set of regressors x_1 was a 128×2 vector containing a column of ones and a column of the counting numbers from 1 to 128 (which may or may not be centered about zero, this was seen to have no effect upon the resulting coefficients for β_2). The matrix containing the second set of regressors x_2, was a single column vector which consisted of a $-1, 1$ square wave of period 16 time points. The coefficient vector was arbitrarily chosen to be $\beta' = [3, 3, 3]$. Independent random error terms from a normal distribution with mean zero and variance three were generated. The values for the simulated data y were constructed from Eq. 1 since X, β, and ϵ were known. Coefficients are estimated using detrending followed by simple linear regression and multiple linear regression.

3.2 FMRI Experiment

Estimates of regression coefficients and t-statistics are computed for real FMRI data. Both detrending followed by simple regression and multiple regression were implemented. The FMRI task used was bilateral finger tapping at 3 T, using gradient echo EPI. Five axial slices were imaged, with $3.4 \times 1.7 \times 1.7$ mm voxels, and a 96×96 matrix with a 21.76 cm field-of-view. The TE was 41.6 ms and TR was 1000 ms. The finger tapping task was 20 s on and 20 s off per block. The total time of the task was 340 seconds, so finger tapping was off first and off last.

4 Results

4.1 Monte Carlo simulation

Two different sets of Monte Carlo simulations were conducted. One with the error variance $\sigma^2 = 0$ and the other with $\sigma^2 = 3$. For both simulations, a $-1, 1$ reference function was used to form the data.

The first simulation was with the error variance chosen to be $\sigma^2 = 0$. As shown in Table 2, if a $-1, 1$ reference function used to estimate the coefficients, then the estimated coefficients were similar. If a $0, 1$ reference function used to estimate the coefficients, then the detrend followed by simple linear regression method without an intercept yielded the same estimates as those with the $-1, 1$ reference function. With a $0, 1$ reference function, the multiple linear regression and detrend followed by simple linear regression with an intercept methods yielded estimates which were approximately twice their previous value with the $-1, 1$ reference function.

Also shown in Table 2, if a $-1, 1$ reference function were used to compute the t-statistics, the two detrending methods produced identical results as did using a $0, 1$ reference function for the detrending followed by simple linear regression with an intercept term. Using either a $-1, 1$ or a $0, 1$ reference function and multiple linear regression produced the same result of very large t-statistics. Using a $0, 1$ reference function for detrending followed by simple linear regression without an intercept term produced a t-statistic which is strikingly smaller than the other methods.

	-1,1		0,1	
	coef	t-stat	coef	t-stat
Detrend+slope regression	2.9648	103.4875	2.9648	**11.1381**
Multiple linear regression	3	$\approx 10^{14}$	6	$\approx 10^{14}$
Detrend+intercept&slope regression	2.9648	103.0793	5.9297	103.0793

Table 2: Coefficient estimates and t-statistics for the simulation with the three methods and two reference functions. No error was used ($\sigma^2 = 0$). Note that when detrending and using a $0, 1$ reference function without intercept, the t-statistics are lower than the other cases.

The second simulation was with an error variance arbitrarily selected to be $\sigma^2 = 3$. As shown in Table 3, if a $-1, 1$ reference function used to estimate the coefficients, then the estimated coefficients were similar. If a $0, 1$ reference function used to estimate the coefficients, then the detrend followed by simple linear regression method without an intercept yielded the same estimates as those with the $-1, 1$ reference function. With a $0, 1$ reference function, the multiple linear regression and detrend followed by simple linear regression with an intercept methods yielded estimates which were approximately twice their previous values with the $-1, 1$ reference function.

Also shown in Table 2, if a $-1, 1$ reference function were used to compute the t-statistics, all three methods produced identical results as did using a $0, 1$ reference function for multiple regression and detrending followed by simple linear regression with an intercept term. Using a $0, 1$ reference function for detrending followed by simple linear regression without an intercept term produced a t-statistic which again is strikingly smaller than the other methods.

	-1,1		0,1	
	coef	t-stat	coef	t-stat
Detrend+slope regression	1.6837	11.8634	1.6837	**6.7291**
Multiple linear regression	1.7037	11.9178	3.4074	11.9178
Detrend+intercept&slope regression	1.6837	11.8166	3.3674	11.8166

Table 3: Multiple linear regression, detrending plus regression with intercept, and detrending plus simple linear regression coefficients for a simulation which computed y using a $0, 1$ square wave reference function whose true coefficient value was 3. The variance was $\sigma^2 = 3$.

4.2 FMRI Experiment

Estimates of coefficients and t-statistics for the FMRI data were found and displayed using 3dDeconvolve [8] and AFNI [2].

Figure 1 contains the statistical map of estimated coefficients using a $-1, 1$ reference function and multiple linear regression. Coefficient maps for the two detrend followed by simple linear regression methods using a $-1, 1$ reference function and the detrend followed by simple linear regression without an intercept using a $0, 1$ reference function produced nearly identical results. For brevity, these other statistical maps were omitted.

Figure 2 contains the statistical map of estimated coefficients using a $0, 1$ reference function and multiple linear regression. The coefficient map using a $0, 1$ reference function for detrending followed by simple linear regression produced a nearly identical map. This additional map was omitted for brevity.

Figure 3 contains a statistical map of t-statistics using a $-1, 1$ reference function and multiple linear regression. Maps of t-statistics for the two detrend followed by simple linear regression methods using a $-1, 1$ reference function, the multiple linear regression using a $0, 1$ reference function, and the detrend followed by simple linear regression without an intercept using a $0, 1$ reference function produced similar results. Recall that it was shown that the t-statistics for five of the six cases would be very similar in Table 3. For brevity, these other statistical maps were omitted.

Figure 4 contains a statistical map of t-statistics using a $0, 1$ reference function and detrending followed by simple linear regression without an intercept. Note that for the detrend followed by simple linear regression case without an intercept, the t-statistics for the $0, 1$ reference function using the same threshold values and same data, show fewer activations than the other cases. The $0, 1$ reference function, when used after detrending, gives false negatives.

5 Discussion

The method of detrending followed by simple linear regression will not make a noticeable difference in the estimated coefficients or t-statistics when using a $-1, 1$ reference function. However, the estimated coefficients and t-statistics from detrending followed by simple linear regression without an intercept will produce results different from the other two methods when using a $0, 1$ reference function.

The multiple linear regression and the detrend followed by simple linear regression with an intercept methods and a $0, 1$ reference function produced similar coefficient estimates and t-statistics. The detrend followed by simple linear regression without an intercept and a $0, 1$ reference produced different coefficient estimates and t-statistics than the other two methods using the same reference function. The coefficient is only estimated correctly with multiple regression only if the same reference function is used as that which produced the data. The true t-statistic is estimated correctly irrespective of the reference function only if the multiple regression method is used. The t-statistic for the detrend followed by simple linear regression with an intercept produced results very similar to multiple regression regardless of the reference function.

The detrend followed by simple linear regression method without an intercept will consistently give an approximately correct estimate of the coefficient regardless of the reference function. The detrend followed by simple linear regression method without an intercept will only give an approximately correct t-statistics if the data (i.e. the BOLD signal) originates from a $-1, 1$ reference function. To overcome this, the second regression should include an intercept term. It is not possible to determine whether the reference function centered about zero or not centered around zero should be used. Both procedures should be investigated.

Acknowledgments

The authors would like to acknowledge the assistance of Doug Ward, Yousef Mazaheri, Kevin Bennett, and Peter Kufahl.

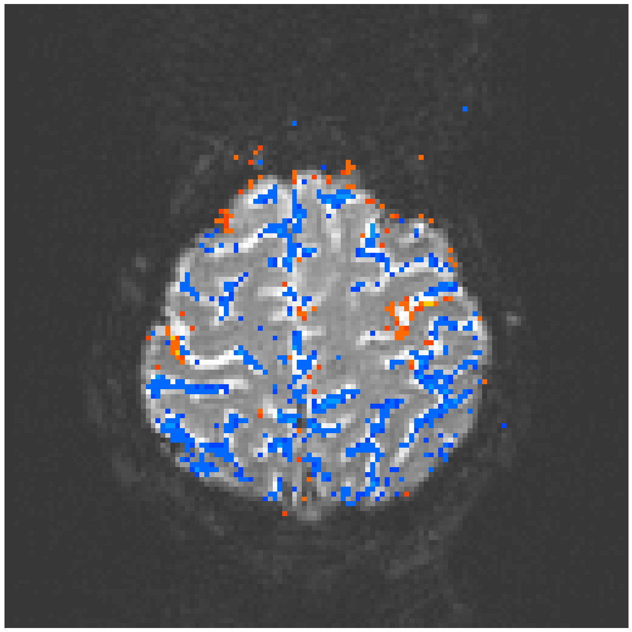

Figure 1: Estimated coefficients for the multiple regression case, $-1, 1$ reference function. Nearly identical to all but the cases in Fig. 2.

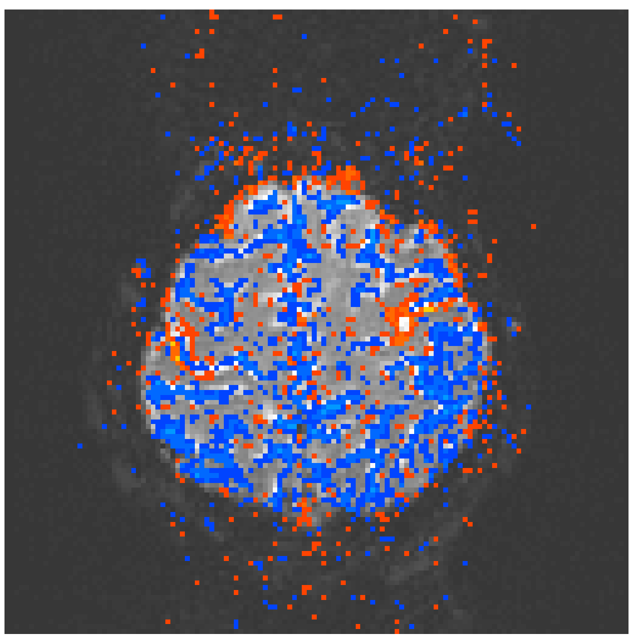

Figure 2: Estimated coefficients for the multiple regression case, $0, 1$ reference function. Nearly identical to detrend followed by simple linear regression with an intercept and $0, 1$ reference function.

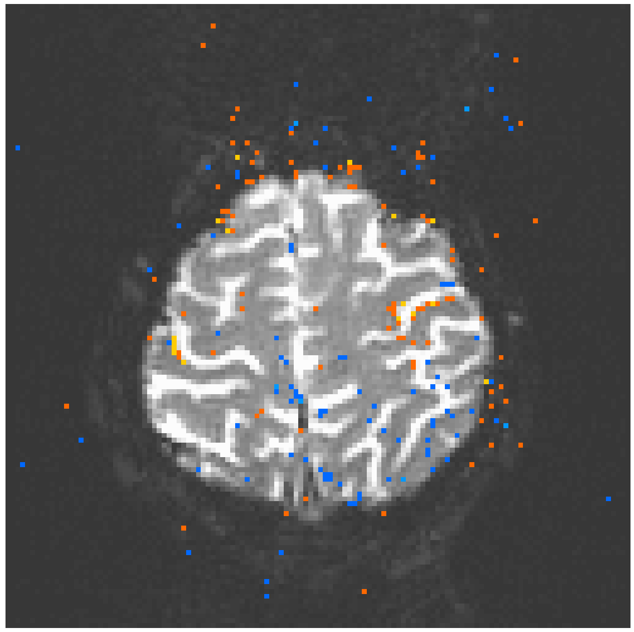

Figure 3: The t-statistics for the multiple linear regression case, 0, 1 reference function.

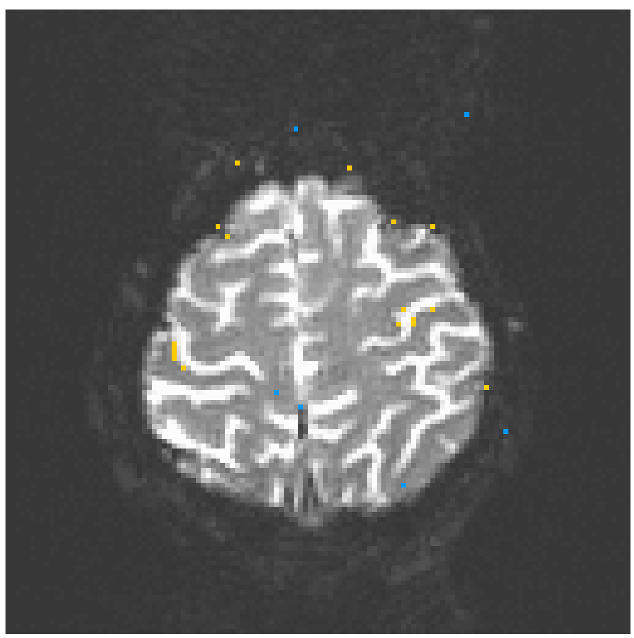

Figure 4: The t-statistics for the detrend followed by simple linear regression method, 0, 1 reference function.

References

[1] P.A. Bandettini, A. Jesmanowicz, E.C. Wong, and J.S. Hyde. Processing strategies for time-course data sets in functional MRI of the human brain. *Magnetic Resonance in Medicine*, 30, 1993.

[2] R.W. Cox, A. Jesmanowicz, and J.S. Hyde. Real time functional magnetic resonance imaging. *Magnetic Resonance in Medicine*, 33, 1995.

[3] M.J. Lowe and D.P. Russell. Treatment of baseline drifts in fMRI time series analysis. *Journal of Computer Assisted Tomography*, 23, 1999.

[4] S.W. Morgan and D.B. Rowe. The effect of detrending when computing regression coefficients in block design fMRI. In *Proceedings of the ISMRM*, 2002.

[5] R.H. Myers. *Classical and Modern Regression with Applications*. Duxbury Press, Pacific Grove, CA, second edition, 1990.

[6] S.J. Press. *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*. Robert E. Krieger Publishing, Malabar, FL, second edition, 1982.

[7] D.B. Rowe. *Multivariate Bayesian Statistics: Models for Source Separation and Signal Unmixing*. CRC Press, Boca Raton, FL, 2002. In Press.

[8] B.D. Ward. *Deconvolution Analysis of FMRI Time Series*. Medical College of Wisconsin, June 2001. AFNI Documentaion.