# AN EXAMINATION OF THE ROWE MODEL AND LEE MODEL FOR COMPLEX-VALUED FUNCTIONAL MAGNETIC RESONANCE IMAGING

by

Yuning Chen

A Thesis submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Master of Science

Milwaukee, Wisconsin

May 2013

ABSTRACT
AN EXAMINATION OF THE ROWE MODEL AND LEE MODEL
FOR COMPLEX-VALUED FUNCTIONAL MAGNETIC
RESONANCE IMAGING

Yuning Chen

Marquette University, 2013

The problem being addressed in this paper is the validity of two competing models for complex-valued fMRI activation. Most of the previous models only use the magnitude portion of the data and discard the phase portion of data which may contain useful biological information. These two models were designed to detect activation using both the magnitude and/or phase in fMRI data. Hence the results can be more accurate and reliable. The first model published by Rowe (2005) and the second model published by Lee (2007) both claim to perform the same job. After the two models have been published, Dr. Rowe pointed out four inaccurate items in Lee model. But Dr. Lee only agreed with Dr. Rowe's first item and he claimed that his model is correct and the data in polar coordinates misled Dr. Rowe's conclusion. In this paper, I will discuss the four items Dr. Rowe pointed out and use some examples and mathematical arguments to support Dr. Rowe's conclusions. The work is mainly about two aspects of these two models. The first aspect is the estimation of the parameter which would be used in the model. The second aspect is the distribution of the test statistic for activation. For the estimation of the parameters, I will analyze the conditions under which the two models can be used and restrictions on the design matrices. For the test statistic, I will present a detailed mathematical derivation of its distribution.

ACKNOWLEDGEMENTS

Yuning Chen

I would like to express my most sincere gratitude to my academic advisor Dr. Daniel B. Rowe for his guidance and support in this paper. I would like to thank Dr. Naveen Bansal for his help and suggestions in this work. I would also like to thank Mary Kociuba, Chaitan Parikh and William Castedo for their help in our research seminars.

# TABLE OF CONTENTS

# 1. BACKGROUND

In human brain research, functional magnetic resonance imaging (fMRI) is an important method to observe brain activity. It measures the blood oxygen level-dependent (BOLD) to detect activity in human brain. Due to the noise from various sources in this process, mathematical models and statistical methods are usually used to detect brain activity.

Statistical models have been used to compute brain activation for many years using magnitude data, more recently models have been published using magnitude and phase for activation from complex-valued data. The first model published by Rowe (2005) and the second model published by Lee (2007) both claim to perform the same job. In 2009, Rowe wrote a letter to the editor to point out four items which he claimed to be inaccurate in the Lee model. After Dr. Rowe published this letter, Dr. Lee responded to Dr. Rowe in another letter to the editor. In this second letter, Dr. Lee only agreed with Dr. Rowe's first statement and he gave another counter example and some explanation as response to Dr. Rowe's other three statements in order to prove that his model was correct. However, Dr. Rowe is not satisfied with Dr. Lee's explanations and he seeks a more detailed mathematical argument to show that the Lee model is not equivalent to the Rowe model and can not be used in practical situations.

In Dr. Rowe's letter, the first item is a mathematical proof by Dr. Lee that to show the equivalence of the Lee model to the Rowe model. Dr. Rowe pointed out that this proof is merely a derivation of the test statistic using a likelihood ratio test in the Lee model. It did not show the equivalence of the two models. In Dr. Lee's response, he agreed with Dr. Rowe's comment on this item.

The second and third items in Dr. Rowe's letter is regarding the validity of the Lee model. Dr. Rowe pointed out that the Lee model is only correct when we add some strong constraints on the design matrix. Dr. Rowe also gave an example to show that Lee model does not work under some conditions. In the response, Dr. Lee stated that the Lee model is in Cartesian coordinates and Rowe model is in polar coordinates, and that this difference misled Dr. Rowe to give a wrong conclusion. Dr. Lee also gave a simple example with the goal to show that the Lee model works while Rowe model does not work.

The fourth item is regarding the degrees of freedom of the test statistic using a likelihood ratio test in the Lee model. In Appendix B of his original paper, Dr. Lee directly gave a conclusion that the test statistic when using a likelihood ratio test is the same as the test statistic when using Hotelling's $T^2$ test, and hence the test statistic should be F distributed and its degrees of freedom should be (m, n-m). But Dr. Rowe stated in his letter that it is F distributed with 2 and 2n-4 degrees of freedom. Dr. Lee responded in his letter that Dr. Rowe assumed the residual errors are independent, hence the degrees of freedom are correct as stated in his original paper. In addition, when the independence of errors is assumed, the degrees of freedom should be $2n-2$ and $2n-4$. So Dr. Lee still does not agree with Dr. Rowe.

# 2. METHODS

## 2.1 The Rowe Model

In fMRI, the temporally changing intensity value in each voxel $y_t$ is complex-valued. The observation $y_t$ can be described with a nonlinear multiple regression model that includes both a temporally varying magnitude $\rho_t$ and phase $\theta_t$ given by

$$y_t = [\rho_t cos\theta_t + \eta_{Rt}] + [\rho_t sin\theta_t + \eta_t]$$
$$\rho_t = x_t'\beta = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_{q1} x_{q_1 t}$$
$$\theta_t = u_t'\gamma = \gamma_0 + \gamma_1 u_{1t} + \cdots + \gamma_{q2} u_{q_2 t}, \qquad t = 1, \cdots, n$$

where $(\eta_{Rt}, \eta_{It})' \sim N(0, \Sigma)$, $x_t'$ is the $t^{th}$ row of an $n \times (q_1 + 1)$ design matrix X for the magnitude, $u_t'$ is the $t^{th}$ row of an $n \times (q_2 + 1)$ design matrix U for the phase, and $\Sigma = \sigma^2 I_2$ while $\beta$ and $\gamma$ are magnitude and phase regression coefficient vectors respectively. Then activation can be determined with a generalized likelihood ratio statistic

$$\text{-2log}\lambda\text{=2nlog}(\tfrac{\tilde{\sigma}^2}{\hat{\sigma}^2}).$$

This statistic has an asymptotic $\chi^2_r$ dsitribution where $r$ is the difference in the number of constraints between the alternative and null hypotheses.

In the Rowe model, there are four potential hypotheses.

$H_a : C\beta \neq 0, D\gamma \neq 0$ does not constrain the coefficients of magnitude and phase,

$H_b : C\beta = 0, D\gamma \neq 0$ only constrains the coefficients of magnitude,

$H_c : C\beta \neq 0, D\gamma = 0$ only constrains the coefficients of phase,

$H_d : C\beta = 0, D\gamma = 0$ constrains the coefficients of both magnitude and phase, which can be combined in various ways to test different hypotheses.

## 2.2 The Lee Model

The complex-valued time series data can be decomposed into real and imaginary parts.

$$y_r = X\beta_r + \epsilon_r$$
$$y_i = X\beta_i + \epsilon_i$$

where $y = y_r + iy_i$ is the time-series data of an individual voxel, $X$ is the design matrix, $\beta_r$ and $\beta_i$ are the parameters of the model, $\epsilon_r$ and $\epsilon_i$ are residual errors. The least-squares estimates of $\hat{\beta}_r$ and $\hat{\beta}_i$ without any constraints are $(X^TX)^{-1}X^Ty_r$ and $(X^TX)^{-1}X^Ty_i$ respectively.

To detect activation, Lee used Hotelling's $T^2$-test. The model describes the null hypothesis is with the constraints $v^T\beta_r = 0$ and $v^T\beta_i = 0$, then the $T^2$ value can be obtained as follow:

$$T^2 = v^T[\hat{\beta}_r \;\; \hat{\beta}_i] \cdot [COV(\hat{\epsilon}_r, \hat{\epsilon}_i)v^T(X^TX)^{-1}v]^{-1} \cdot [\hat{\beta}_r \;\; \hat{\beta}_i]^T v.$$

If both the real and imaginary parts of the data are normally distributed, the $T^2$ statistic will follow an F-distribution with 2 and n-2 degrees of freedom. Thus for a given significance level $\alpha$, the null hypothesis is rejected when

$$T^2 > [\tfrac{2(n-1)}{n-2}] \; F_{2,n-2}(\alpha).$$

Similar to the Rowe model, there are also four hypotheses in the Lee model. But because the deign matrices for real and imaginary are the same in Lee model, there is only one contrast vector $v$ in each of the four hypotheses:

$H_a : v\beta_r \neq 0, v\beta_i \neq 0$ does not constrain the coefficients of real and imaginary.

$H_b : v\beta_r = 0, v\beta_i \neq 0$ only constrains the coefficients of real.

$H_c : v\beta_r \neq 0, v\beta_i = 0$ only constrains the coefficients of imaginary.

$H_d : v\beta_r = 0, v\beta_i = 0$ constrains the coefficients of both real and imaginary.

We can find that both of the two models can be divided into two parts. The first part is estimating the parameters used in the model. The second part is using the estimated parameters to determine activation. In the Rowe model, the data is in polar coordinates and uses a generalized likelihood ratio test to determine activation. The likelihood must be maximized under appropriate null and alternative hypotheses to estimate the parameters in the model. In this process, a multiple iterations method is used to estimate the parameters. In the Lee model, the data is in Cartesian coordinates and Hotelling's $T^2$-test is used to determine activation. Lee also used the maximum likelihood estimate under constrained null and unconstrained alternative hypotheses to obtain those parameters.

## 3. RESULTS

From last section, we know that the parameters in the Lee model without constraints can be estimated by

$$\hat{\beta}_r = (X'X)^{-1}X'y_r,\ \hat{\beta}_i = (X'X)^{-1}X'y_i$$

The fitted time series can be obtained by

$$\hat{y}_r = X(X'X)^{-1}X'y_r,\ \hat{y}_i = X(X'X)^{-1}X'y_i$$

In the Rowe model, the same design matrix $X$ can be used for both magnitude and phase which is useful in comparison to the Lee model which requires the same design matrix. To estimate the parameters $\hat{\beta}$ and $\hat{\gamma}$, the multiple iterations method is used. Then the estimated magnitude and phase at time $t$ are

$$\hat{\rho}_t = x_t'\hat{\beta},$$
$$\hat{\theta}_t = x_t'\hat{\gamma}$$

and the fitted real and imaginary parts of the data is

$$\hat{y}_r = \hat{\rho}_t \cos\hat{\theta}_t = x_t'\hat{\beta}cos(x_t'\hat{\gamma})$$
$$\hat{y}_i = \hat{\rho}_t \sin\hat{\theta}_t = x_t'\hat{\beta}sin(x_t'\hat{\gamma})$$

where $x_t'$ is the $t^{th}$ row of design matrix $X$.

### 3.1 Validity of Rowe Model and Lee Mode

To examine these two models, I assume that there are three columns in the design matrix. We can build the design matrix by a constant vector $[1\ 1 \ldots\ldots 1]'$ (a real n×1 vector), a counting number vector $[1\ 2\ 3\ \ldots \mathrm{n}]'$ (a real n×1 vector) and an on/off(0/1) reference vector. I started with different phase regression coefficient vectors $\gamma$. There are also three components in the $\gamma$ vector.

All possible $\gamma$'s are

$$\gamma = \begin{pmatrix} C \\ 0 \\ 0 \end{pmatrix}, \gamma = \begin{pmatrix} 0 \\ 0 \\ C \end{pmatrix}, \gamma = \begin{pmatrix} C \\ 0 \\ C \end{pmatrix}, \gamma = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\gamma = \begin{pmatrix} 0 \\ C \\ 0 \end{pmatrix}, \gamma = \begin{pmatrix} C \\ C \\ 0 \end{pmatrix}, \gamma = \begin{pmatrix} 0 \\ C \\ C \end{pmatrix}, \gamma = \begin{pmatrix} C \\ C \\ C \end{pmatrix},$$

where C represents an arbitrary non-zero positive number and 0 represents zero.

All possible $\beta$'s are

$$\beta = \begin{pmatrix} C \\ 0 \\ 0 \end{pmatrix}, \beta = \begin{pmatrix} 0 \\ 0 \\ C \end{pmatrix}, \beta = \begin{pmatrix} C \\ 0 \\ C \end{pmatrix}, \beta = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\beta = \begin{pmatrix} 0 \\ C \\ 0 \end{pmatrix}, \beta = \begin{pmatrix} C \\ C \\ 0 \end{pmatrix}, \beta = \begin{pmatrix} 0 \\ C \\ C \end{pmatrix}, \beta = \begin{pmatrix} C \\ C \\ C \end{pmatrix}.$$

We do not need to examine the two choices

$$\beta = \begin{pmatrix} 0 \\ 0 \\ C \end{pmatrix}, \beta = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

because the magnitude could be zero in these two cases.

I examined all the possible combinations of $\beta$'s and $\gamma$'s. The result is that the Rowe model works as described in all the cases, while Lee model does not work when

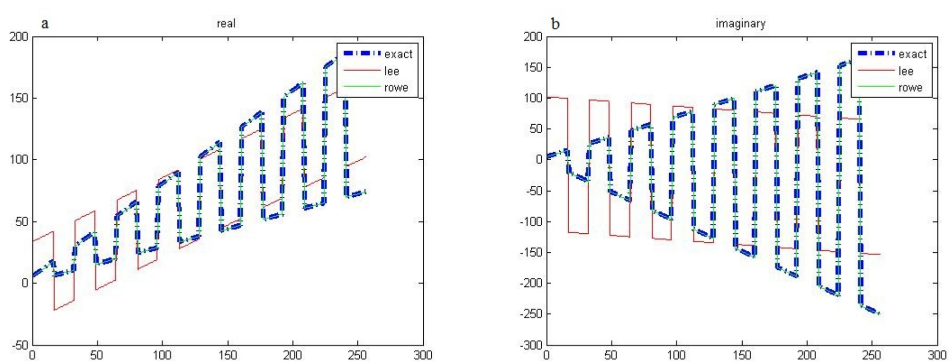
$$\gamma = \begin{pmatrix} 0 \\ 0 \\ C \end{pmatrix}, \gamma = \begin{pmatrix} C \\ 0 \\ C \end{pmatrix},$$

$$\text{and } \beta = \begin{pmatrix} 0 \\ C \\ 0 \end{pmatrix}, \beta = \begin{pmatrix} C \\ C \\ 0 \end{pmatrix}, \beta = \begin{pmatrix} C \\ 0 \\ C \end{pmatrix}, \beta = \begin{pmatrix} C \\ C \\ C \end{pmatrix}.$$

Graphs and differences between the Rowe model, the Lee model and the true values are shown below. In these examples, we assume they are noiseless. We first generate a data set from the Rowe model, then we use both the Rowe model and Lee model to estimate the data set. If the Lee model is correct, it should give us a very close estimation. I assumed that there are 256 time points and that the size of the design matrix is 256×3. We build the design matrix by a constant vector $[1\ 1 \ldots\ldots 1]'$ (a real 256×1 vector), a counting number vector $[1\ 2\ 3 \ldots 256]'$ (a real 256×1 vector) and an on/off (0/1) reference vector.

### 3.1.1 Example 1

When $\beta = \begin{pmatrix} C \\ C \\ C \end{pmatrix}, \gamma = \begin{pmatrix} C \\ 0 \\ C \end{pmatrix}$, we use $\beta$=[5 1 2]$'$,$\gamma$=[5 0 2]$'$ to generate data set from the Rowe model, then we use both of the two models to estimate data. Here are the results.



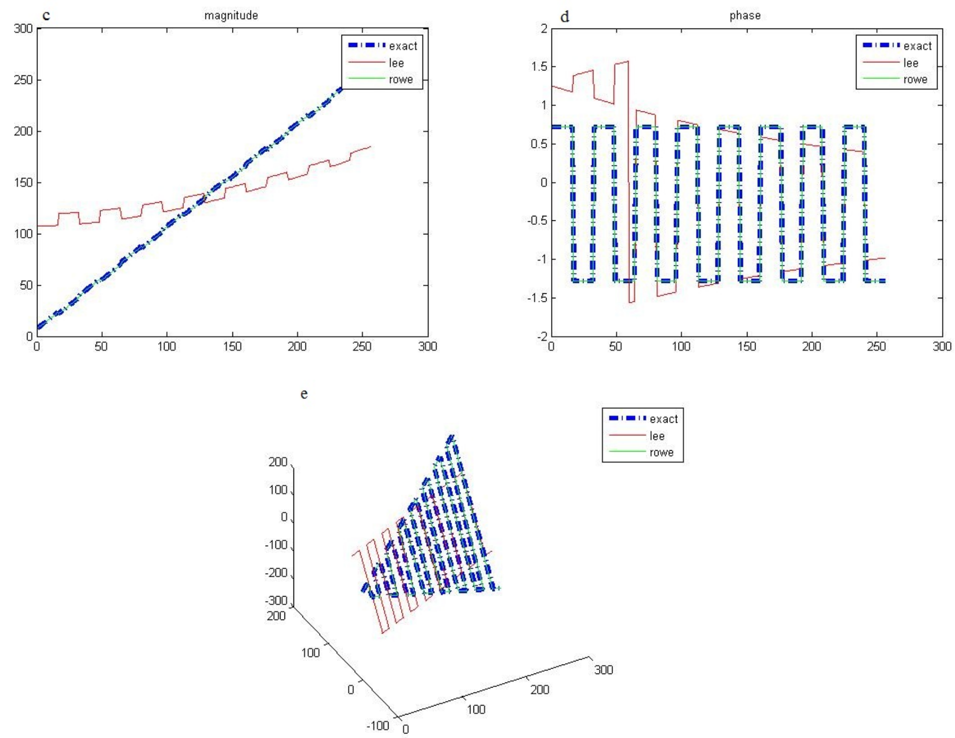



Figure 1

The norm of the difference between the fitted value from the two models and the true value are

lee_rowe$=1.2261 \times 10^3$

true_lee$=1.2261 \times 10^3$

true_rowe$=9.7479 \times 10^{-3}$.

In Figure 1a, it shows the real part of the fitted value from the Rowe model and the Lee model compared with the true value.

In Figure 1b, it shows the imaginary part of the fitted value from the Rowe model and the Lee model compared with the true value.

In Figure 1c, it shows the magnitude of the fitted value from the Rowe model and the Lee model compared with the true value.

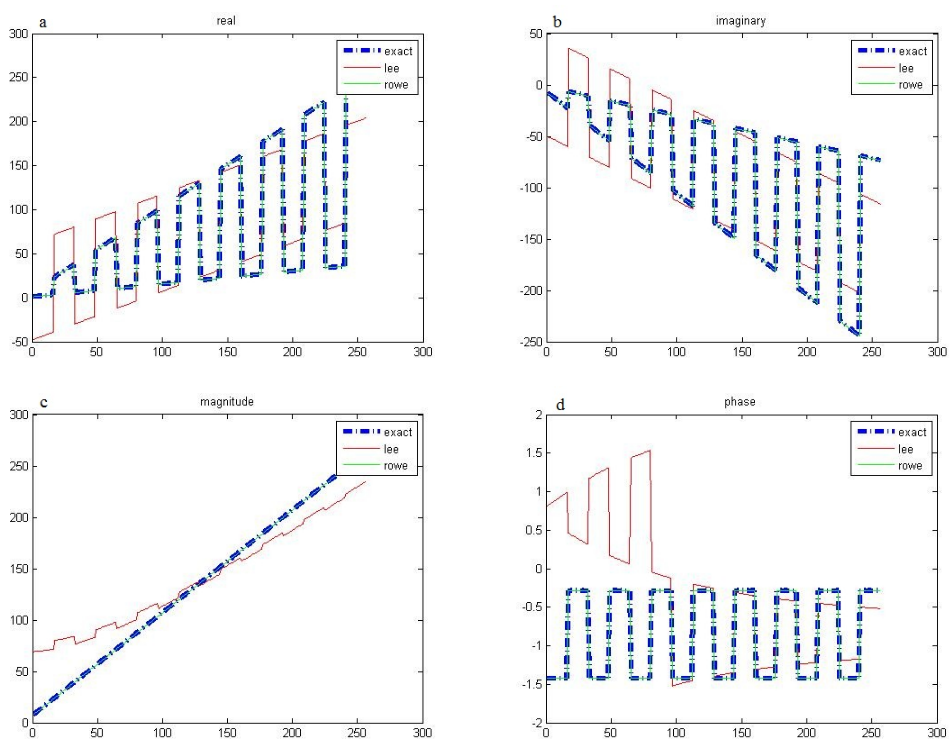
In Figure 1d, it shows the phase of the fitted value from the Rowe model and the Lee model compared with the true value.

In Figure 1e, it shows the fitted complexed-valued estimation from the Rowe model and the Lee model compared with the true value.

One can see that in all of these five figures, the fitted value from the Rowe model (the green one) is very close to the true value from the data (the blue one). They are almost the same. But the Lee model gave us a very bad fitted value (the red one).

### 3.1.2 Example 2

When $\beta = \begin{pmatrix} C \\ C \\ 0 \end{pmatrix}$, $\gamma = \begin{pmatrix} C \\ 0 \\ C \end{pmatrix}$, we use $\beta$=[2 10 0]$'$,$\gamma$=[1 0 10]$'$ to generate data set from the Rowe model, then we use both of the two models to estimate the data. Here are the results.



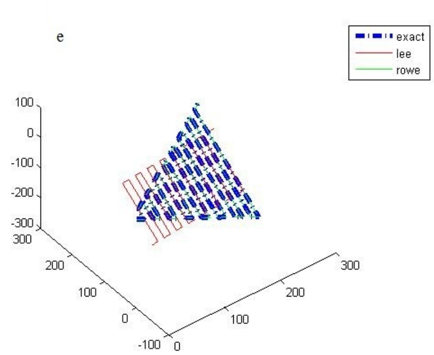



Figure 2

The norm of the difference between the fitted value from the two models and the true value are

lee_rowe=896.0575

true_lee=896.0575

true_rowe=$1.5787 \times 10^{-12}$.

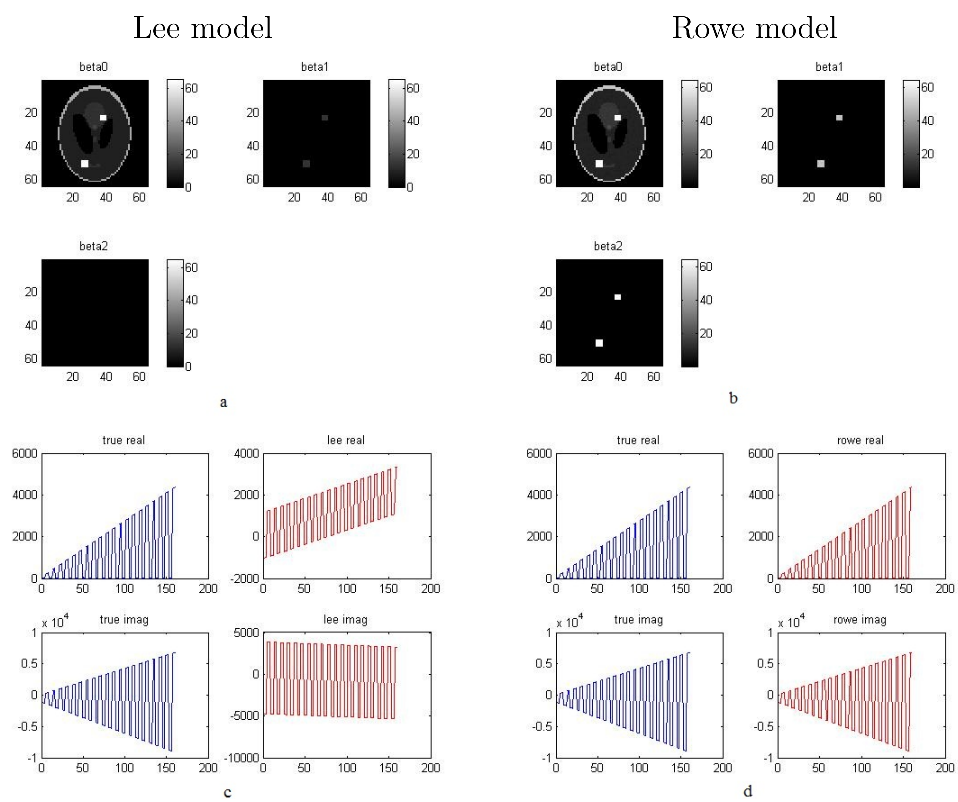
In Figure 2a, it shows the real part of the fitted value from the Rowe model and the Lee model compared with the true value.
In Figure 2b, it shows the imaginary part of the fitted value from the Rowe model and the Lee model compared with the true value.
In Figure 2c, it shows the magnitude of the fitted value from the Rowe model and the Lee model compared with the true value.
In Figure 2d, it shows the phase of the fitted value from the Rowe model and the Lee model compared with the true value.
In Figure 2e, it shows the fitted complexed-valued estimation from the Rowe model and the Lee model compared with the true value.

The same as example 1, one can easily see that in all of these five figures, the fitted value from the Rowe model (the green one) is very close to the true value (the blue one). But the Lee model gave us a very bad fitted value (the red one). In addition, the difference between the true value and estimation from Lee model is very large. But the difference between the true value and estimation from Rowe model is very close to 0.

### 3.1.3 Example 3

To demonstrate the validity of the two models, I will choose one combination of $\beta$ and $\gamma$ from above, then use the Shepp-Logan Phantom and add activation in certain areas. I will plot the graphs of $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \theta, \hat{\theta}, \sigma^2$ and $\hat{\sigma}^2$ in the Shepp-Logan phantom. Assume there are 160 time points.

I chose to use $\beta$=[2 1 20]$'$, $\gamma$=[1 0 10]$'$ in the Rowe model to generate the activation in the specified areas. Then I use the Lee model and the Rowe model to estimate the parameters used in the models.



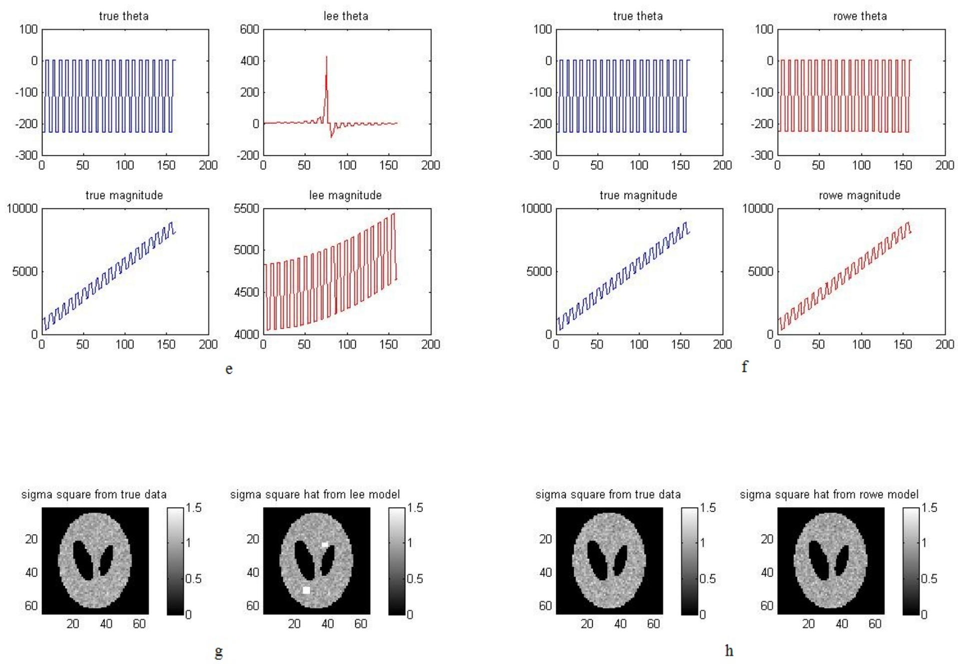



Figure 3

Figures 3a and 3b show $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\beta}_2$ from the Lee model and the Rowe model. Figures 3c, 3d, 3e and 3f show the real, imaginary, magnitude and phase of the data from the Rowe model and the Lee model compared with the true values. Figures 3g and 3h show the $\hat{\sigma}^2$ from the Rowe model and the Lee model.

In the graphs of $\beta_0$, although the areas of activation are both white, $\hat{\beta}_0 = 1160$ in the Lee model and $\hat{\beta}_0 = 100.2$ in the Rowe model. In figure 3g, the $\hat{\sigma}^2$ in the two white areas is $2.874 \times 10^6$, while in figure 3h, all the values of $\hat{\sigma}^2$ are close to 1. And from the graphs of real, imaginary, magnitude and phase in the activation area, one can convincingly conclude that the Lee model does not give us a good estimation.

We can see that the Lee model does not work under these conditions, while the Rowe model works as described. Because when we build the design matrix in this way, in fact the Lee model is a linear regression model. A linear regression model can only give us a "good"estimation when there is a constant slope of the upper side and a constant slope of the lower side in the graph. But when we generate data from the Rowe model using the parameter vectors we choose, although the graphs of the magnitude and phase have the constant slops of the upper and lower sides, the graphs of the real and imaginary parts do not satisfy this condition. Hence, the Lee model does not work.

### 3.2 Items 2 and 3 Design Matrix

In Lee's response letter to the editor (2009), he presented an example to demonstrate that the Rowe model does not work, but his example is very misleading.

Lee considered an example in which L=2, n=3 and the design matrix has the first column [1 1 1]′ and the second column [0 0.5 1]′. He assumed the observation made was $y_{Rt}$=[4 5 6]′ and $y_{It}$=[8 7 6]′. This is also a noiseless observation in Cartesian coordinates. In this case, the Lee model is also a linear regression model. Lee was using the linear function $\hat{y} = \hat{a} + \hat{b}x$ to estimate the data when he chose

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{pmatrix}, \beta = \begin{pmatrix} a \\ b \end{pmatrix}$$

We know the linear regression model works better when $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ almost lie in a line. So when we are estimating the real part of the data,

$$y_r = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}, X = \begin{pmatrix} 1 & 0 \\ 1 & 0.5 \\ 1 & 1 \end{pmatrix}$$

We can see $(4, 0), (5, 0.5), (6, 1)$ lie in a line. This is the same in the imaginary part, $(8, 0), (7, 0.5), (6, 1)$ lie in a line too. So in this misleading special example chosen by Lee, his model works. But when we make a very small change such that the three points do not lie perfectly in a line, the Lee model does not work. We assume the design matrix is the same, the observation was

$$y_{Rt}\text{=[4 5 6]}',\ y_{It}\text{=[8 6 7]}'$$

and the estimation is

$\hat{\beta}_r = (X'X)^{-1}X'y_r$ =[4 2]$'$

$\hat{\beta}_i = (X'X)^{-1}X'y_i$ =[7.5 -1]$'$

$\hat{y}_r = X\hat{\beta}_r$ =[4 5 6]$'$

$\hat{y}_i = X\hat{\beta}_i$ =**[7.5 7 6.5]**$'$

We can see that the Lee model does not work because in the imaginary part, $(8,0),(6,0.5),(7,1)$ do not lie in the same line. Furthermore, when there are two columns in the design matrix, one is a constant baseline, the other one is on/off reference vector, the Lee model works well only when the graph of the real and imaginary data set has a constant zero slope or the slopes of upper and lower sides are zero. That means the graph of real and imaginary parts consists of horizontal lines.

### 3.3 Item 4 Degrees of Freedom of Test Statistic

To find the test statistic of the Lee model when we use a likelihood ratio test, consider $H_d$ vs. $H_a$. Under the alternative hypothesis $H_a : v^T\beta_r \neq 0, v^T\beta_i \neq 0$, we have

$$\hat{\beta}_r = (X^TX)^{-1}X^Ty_r$$

$$\hat{\beta}_i = (X^TX)^{-1}X^Ty_i$$

$$\hat{\sigma}^2_{H_a} = \tfrac{1}{2n}[(y_r - X\hat{\beta}_r)^T(y_r - X\hat{\beta}_r) + (y_i - X\hat{\beta}_i)^T(y_i - X\hat{\beta}_i)]$$

Under the null hypothesis $H_d : v^T\beta_r = 0, v^T\beta_i = 0$, we have

$$\tilde{\beta}_r = \hat{\beta}_r - (X^TX)^{-1}v[v^T(X^TX)^{-1}v]^{-1}v^T\hat{\beta}_r,$$

$$\tilde{\beta}_i = \hat{\beta}_i - (X^TX)^{-1}v[v^T(X^TX)^{-1}v]^{-1}v^T\hat{\beta}_i,$$

$$\tilde{\sigma}^2_{H_d} = \frac{1}{2n}[(y_r - X\tilde{\beta}_r)^T(y_r - X\tilde{\beta}_r) + (y_i - X\tilde{\beta}_i)^T(y_i - X\tilde{\beta}_i)]$$

In the generalized likelihood ratio test, the likelihood ratio statistic is

$$\lambda = (\frac{\tilde{\sigma}^2_{H_d}}{\hat{\sigma}^2_{H_a}})^{-n}$$

The activation test statistic is

$$F = \lambda^{-\frac{1}{n}} - 1 = \frac{\tilde{\sigma}^2_{H_d} - \hat{\sigma}^2_{H_a}}{\hat{\sigma}^2_{H_a}}$$

The numerator of the activation statistic can be written as $[(y_r - X\tilde{\beta}_r)^T(y_r - X\tilde{\beta}_r) - (y_r - X\hat{\beta}_r)^T(y_r - X\hat{\beta}_r)] + [(y_i - X\tilde{\beta}_i)^T(y_i - X\tilde{\beta}_i) + (y_i - X\hat{\beta}_i)^T(y_i - X\hat{\beta}_i)]$.

As seen in Appendix A, we have

$(y_r - X\tilde{\beta}_r)^T(y_r - X\tilde{\beta}_r) = (\tilde{\beta}_r - \hat{\beta}_r)^T(X^TX)(\tilde{\beta}_r - \hat{\beta}_r) + y_r^T y_r - y_r^T X\hat{\beta}_r$

$(y_r - X\hat{\beta}_r)^T(y_r - X\hat{\beta}_r) = y_r^T y_r - \hat{\beta}_r^T X^T y_r$

Assume $y_r$ is $n \times 1$, design matrix $X$ is $n \times m$ and $\hat{\beta}_r$ is $m \times 1$, then the dimension of $y_r^T X\hat{\beta}_r$ is $1 \times 1$, hence $y_r^T X\hat{\beta}_r$ is a number and $y_r^T X\hat{\beta}_r = \hat{\beta}_r^T X^T y_r$.

Now

$(y_r - X\tilde{\beta}_r)^T(y_r - X\tilde{\beta}_r) - (y_r - X\hat{\beta}_r)^T(y_r - X\hat{\beta}_r) = (\tilde{\beta}_r - \hat{\beta}_r)^T(X^TX)(\tilde{\beta}_r - \hat{\beta}_r)$

We can do the same work to the imaginary part. We can get

$(y_i - X\tilde{\beta}_i)^T(y_i - X\tilde{\beta}_i) - (y_i - X\hat{\beta}_i)^T(y_i - X\hat{\beta}_i) = (\tilde{\beta}_i - \hat{\beta}_i)^T(X^TX)(\tilde{\beta}_i - \hat{\beta}_i)$

The activation test statistic becomes

$$F = \frac{[(\tilde{\beta}_r - \hat{\beta}_r)^T(X^TX)(\tilde{\beta}_r - \hat{\beta}_r) + (\tilde{\beta}_i - \hat{\beta}_i)^T(X^TX)(\tilde{\beta}_i - \hat{\beta}_i)]/\sigma^2}{[(y_r - X\hat{\beta}_r)^T(y_r - X\hat{\beta}_r) + (y_i - X\hat{\beta}_i)^T(y_i - X\hat{\beta}_i)]/\sigma^2}$$

We assume

$X_{1r} = (\tilde{\beta}_r - \hat{\beta}_r)^T(X^TX)(\tilde{\beta}_r - \hat{\beta}_r)$

$X_{1i} = (\tilde{\beta}_i - \hat{\beta}_i)^T(X^TX)(\tilde{\beta}_i - \hat{\beta}_i)$,

$X_{2r} = (y_r - X\hat{\beta}_r)^T(y_r - X\hat{\beta}_r)$,

$X_{2i} = (y_i - X\hat{\beta}_i)^T(y_i - X\hat{\beta}_i)$.

First of all, we will derive the distribution of $X_{2r}$ and $X_{2i}$. One can easily get $X_{2r} = y_r^T[I - X(X^TX)^{-1}X^T]y_r^T/\sigma^2$ (See Appendix).

We can show that $[I - X(X^TX)^{-1}X^T]$ is symmetric and idempotent. By Conclusion 2 (See Appendix), $X_{2r}/\sigma^2$ follows a $\chi^2$ distribution. The degrees of freedom come from the rank of $[I - X(X^TX)^{-1}X^T]$, which in turn is the trace by Conclusion 1 (See Appendix). That is $trace[I - X(X^TX)^{-1}X^T] = n - p$. So $X_{2r}/\sigma^2$ follows a $\chi^2_{n-p}$ distribution. We can do the same work to $X_{2i}/\sigma^2$ and we can show that it also follows a $\chi^2_{n-p}$ distribution.

Secondly, we will show the derivation of $X_{1r}$ and $X_{1i}$. Consider when the null hypothesis is $H_0 : C\beta = 0$ where $C$ is $r \times p$, full rank and $r < p$. Because matrix $C$ is of full rank, the row vectors are linearly independent. We can always add more linearly independent row vectors in $C$ to get a square matrix and the new square matrix is still full rank. We call the new square matrix $Q = \begin{pmatrix} C \\ C_1 \end{pmatrix}$. There are $(p - r)$ linearly independent row vectors in $C_1$. Q is $p \times p$, non-singular and invertible.

We have

$y = X\beta + \epsilon$

$y = XQ^{-1}Q\beta + \epsilon$

$y = XQ^{-1}\begin{pmatrix} C \\ C_1 \end{pmatrix}\beta + \epsilon$, where $Q = \begin{pmatrix} C \\ C_1 \end{pmatrix}$

$y = X_q\begin{pmatrix} C\beta \\ C_1\beta \end{pmatrix} + \epsilon$ , where $X_q = XQ^{-1}$

$y = \begin{pmatrix} X_q^{(1)}, X_q^{(2)} \end{pmatrix}\begin{pmatrix} r_1 \\ r_2 \end{pmatrix} + \epsilon,$

where $X_q = \begin{pmatrix} X_q^{(1)}, X_q^{(2)} \end{pmatrix}$, $\begin{pmatrix} C\beta \\ C_1\beta \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}$ .

$y = X_q^{(1)}r_1 + X_q^{(2)}r_2 + \epsilon.$

The null hypothesis $H_0 : C\beta = 0$ is the same as $H_0 : r_1 = 0$. Assume

$r_1 = \begin{pmatrix} r_{11} \\ r_{12} \\ \vdots \\ r_{1r} \end{pmatrix}$, the null hypothesis becomes $H_0 : r_{11} = 0, r_{12} = 0, \cdots, r_{1r} = 0$. Hence $X_{1r}/\sigma^2$ follows a $\chi^2$ distribution (See Appendix). The degrees of freedom is $r$ which is equal to the number of rows in $C$. We can also get $X_{1i}/\sigma^2 \sim \chi^2_r$.

Third, we will show that $X_{1r}$ and $X_{2r}$ are independent. By Conclusion 4 (See Appendix), this is the same as showing

$$\sigma^2 AB = 0,$$

where $A = X(X^TX)^{-1}X^T - X_2(X_2^TX_2)^{-1}X_2^T$ and $B = I - X(X^TX)^{-1}X^T$.

As proved in the Appendix,

$\sigma^2[X(X^TX)^{-1}X^T - X_2(X_2^TX_2)^{-1}X_2^T][I - X(X^TX)^{-1}X^T] = 0.$

So we proved that $X_{1r}$ and $X_{2r}$ are independent. Using the same method, we can also show that $X_{1i}$ and $X_{2i}$ are independent.

Hence under the null hypothesis $H_0 : C\beta = 0$ where C is $r \times p$, full rank and $r < p$,

$$\frac{X_{1r}}{\sigma^2} \sim \chi^2_r$$
$$\frac{X_{1i}}{\sigma^2} \sim \chi^2_r$$
$$\frac{X_{2r}}{\sigma^2} \sim \chi^2_{n-p}$$
$$\frac{X_{2i}}{\sigma^2} \sim \chi^2_{n-p}$$

$\frac{X_{1r}}{\sigma^2}$ and $\frac{X_{2r}}{\sigma^2}$ are independent. $\frac{X_{1i}}{\sigma^2}$ and $\frac{X_{2i}}{\sigma^2}$ are independent.

The test statistic is therefore

$$F = \frac{\frac{X_{1r}}{\sigma^2} + \frac{X_{1i}}{\sigma^2}}{\frac{X_{2r}}{\sigma^2} + \frac{X_{2i}}{\sigma^2}} \sim F_{2r,2(n-p)}$$

which has an F distribution with $2r$ numerator and $2(n-p)$ denominator degrees of freedom. This is exacely as described by Rowe in his letter where $r = 1$.

## 4. Conclusion

Including both the magnitude and phase activations, these two models do not discard the biological information that might be in the phase-signal changes. Hence we can hopefully obtain more reliable results. However, as shown above, we can conclude that when we build the design matrix by a constant vector [1 $1 \ldots \ldots 1]'$ (a real n$\times$1 vector), a counting number vector $[1\ 2\ 3 \ldots n]'$ (a real n$\times$1 vector) and an on/off (0/1) reference vector, the Lee model can not give us a good estimation under some conditions, while the Rowe model works all the time. In the Lee model , Dr. Lee can not conclude that his model using Hotelling's $T^2$ test is equivalent to the model using likelihood ratio test for the reason that the degrees of freedom of the two statistics are not the same. Different coordinate systems used is not the reason for poor parameter estimation. The poor estimation comes from the inaccuracy of the model. Despite these four critical items Dr. Rowe made, the Lee model is more computationally efficient and also works well when the phase of the data is a constant. Even though under some conditions as listed above in this paper, the Lee model works, but since the Rowe model works all the time, one should use the Rowe model to detect the magnitude and phase activation in fMRI.

## Appendix. Detailed Proof of Degrees of Freedom of Test Statistic

To find the test statistic of Lee model when we use likelihood ratio test, consider $H_d$ vs. $H_a$. Under alternative hypothesis $H_a : v^T\beta_r \neq 0, v^T\beta_i \neq 0$, we have

LLa=$\log P(y|H_a) = -n \log 2\pi\sigma^2_{H_a} - \frac{1}{2\sigma^2_{H_a}}[(y_r - X\beta_r)^T(y_r - X\beta_r) + (y_i - X\beta_i)^T(y_i - X\beta_i)]$

When $\frac{\partial LLa}{\partial \beta_r} = 0$ and $\frac{\partial LLa}{\partial \beta_i} = 0$,

$$\hat{\beta}_r = (X^TX)^{-1}X^Ty_r$$
$$\hat{\beta}_i = (X^TX)^{-1}X^Ty_i.$$

When $\frac{\partial LLa}{\partial \sigma^2_{H_a}} = 0$, we have

$$\hat{\sigma}^2_{H_a} = \frac{1}{2n}[(y_r - X\hat{\beta}_r)^T(y_r - X\hat{\beta}_r) + (y_i - X\hat{\beta}_i)^T(y_i - X\hat{\beta}_i)].$$

Under null hypothesis $H_d : v^T\beta_r = 0, v^T\beta_i = 0$, we have

LLd=$\log P(y|H_d) =$

$-n \log 2\pi\sigma^2_{H_d} - \frac{1}{2\sigma^2_{H_d}}[(y_r - X\beta_r)^T(y_r - X\beta_r) + (y_i - X\beta_i)^T(y_i - X\beta_i)].$

When $\frac{\partial LLd}{\partial \beta_r} = 0$ and $\frac{\partial LLd}{\partial \beta_i} = 0$

$$\tilde{\beta}_r = \hat{\beta}_r - (X^TX)^{-1}v[v^T(X^TX)^{-1}v]^{-1}v^T\hat{\beta}_r,$$
$$\tilde{\beta}_i = \hat{\beta}_i - (X^TX)^{-1}v[v^T(X^TX)^{-1}v]^{-1}v^T\hat{\beta}_i.$$

When $\frac{\partial LLd}{\partial \sigma^2_{H_d}} = 0$, we have

$$\tilde{\sigma}^2_{H_d} = \frac{1}{2n}[(y_r - X\tilde{\beta}_r)^T(y_r - X\tilde{\beta}_r) + (y_i - X\tilde{\beta}_i)^T(y_i - X\tilde{\beta}_i)].$$

In the generalized likelihood ratio test, the likelihood ratio is

$$\lambda = \left(\frac{\tilde{\sigma}^2_{H_d}}{\hat{\sigma}^2_{H_a}}\right)^{-n}.$$

The test statistic

$$\begin{aligned} F &= \lambda^{-\frac{1}{n}} - 1 \\ &= \frac{\tilde{\sigma}^2_{H_d}}{\hat{\sigma}^2_{H_a}} - 1 \\ &= \frac{\tilde{\sigma}^2_{H_d} - \hat{\sigma}^2_{H_a}}{\hat{\sigma}^2_{H_a}}. \end{aligned}$$

The numerator can be written as $[(y_r - X\tilde{\beta}_r)^T(y_r - X\tilde{\beta}_r) - (y_r - X\hat{\beta}_r)^T(y_r - X\hat{\beta}_r)] + [(y_i - X\tilde{\beta}_i)^T(y_i - X\tilde{\beta}_i) + (y_i - X\hat{\beta}_i)^T(y_i - X\hat{\beta}_i)]$.

We have

$$
\begin{aligned}
&(y_r - X\tilde{\beta}_r)^T(y_r - X\tilde{\beta}_r)\\
&= y_r^T y_r - y_r^T X\tilde{\beta}_r - \tilde{\beta}_r^T X^T y_r + \tilde{\beta}_r^T X^T X\tilde{\beta}_r\\
&= \tilde{\beta}_r^T[(X^T X)\tilde{\beta}_r - X^T y_r] + y_r^T y_r - y_r^T X\tilde{\beta}_r\\
&= \tilde{\beta}_r^T(X^T X)[\tilde{\beta}_r - (X^T X)^{-1}X^T y_r] + y_r^T y_r - y_r^T X\tilde{\beta}_r\\
&= \tilde{\beta}_r^T(X^T X)[\tilde{\beta}_r - \hat{\beta}_r] + y_r^T y_r - y_r^T X\tilde{\beta}_r\\
&= (\tilde{\beta}_r - \hat{\beta}_r)^T(X^T X)(\tilde{\beta}_r - \hat{\beta}_r) + \hat{\beta}_r^T(X^T X)(\tilde{\beta}_r - \hat{\beta}_r) + y_r^T y_r - y_r^T X\tilde{\beta}_r\\
&= (\tilde{\beta}_r - \hat{\beta}_r)^T(X^T X)(\tilde{\beta}_r - \hat{\beta}_r) + y_r^T X(X^T X)^{-1}(X^T X)(\tilde{\beta}_r - \hat{\beta}_r) + y_r^T y_r - y_r^T X\tilde{\beta}_r\\
&= (\tilde{\beta}_r - \hat{\beta}_r)^T(X^T X)(\tilde{\beta}_r - \hat{\beta}_r) + y_r^T X(\tilde{\beta}_r - \hat{\beta}_r) + y_r^T y_r - y_r^T X\tilde{\beta}_r\\
&= (\tilde{\beta}_r - \hat{\beta}_r)^T(X^T X)(\tilde{\beta}_r - \hat{\beta}_r) + y_r^T X\tilde{\beta}_r - y_r^T X\hat{\beta}_r + y_r^T y_r - y_r^T X\tilde{\beta}_r\\
&= (\tilde{\beta}_r - \hat{\beta}_r)^T(X^T X)(\tilde{\beta}_r - \hat{\beta}_r) + y_r^T y_r - y_r^T X\hat{\beta}_r.
\end{aligned}
$$

We also have

$$
\begin{aligned}
&(y_r - X\hat{\beta}_r)^T(y_r - X\hat{\beta}_r)\\
&= y_r^T y_r - y_r^T X\hat{\beta}_r - \hat{\beta}_r^T X^T y_r + \hat{\beta}_r^T X^T X\hat{\beta}_r\\
&= y_r^T y_r - y_r^T X\hat{\beta}_r - \hat{\beta}_r^T X^T y_r + y_r^T X(X^T X)^{-1}(X^T X)(X^T X)^{-1}X^T y_r\\
&= y_r^T y_r - y_r^T X\hat{\beta}_r - \hat{\beta}_r^T X^T y_r + y_r^T X(X^T X)^{-1}X^T y_r\\
&= y_r^T y_r - y_r^T X\hat{\beta}_r - \hat{\beta}_r^T X^T y_r + y_r^T X\hat{\beta}_r\\
&= y_r^T y_r - \hat{\beta}_r^T X^T y_r.
\end{aligned}
$$

Assume $y_r$ is $n \times 1$, design matrix $X$ is $n \times m$ and $\hat{\beta}_r$ is $m \times 1$, then the dimension of $y_r^T X\hat{\beta}_r$ is $1 \times 1$, hence $y_r^T X\hat{\beta}_r$ is a number and $y_r^T X\hat{\beta}_r = \hat{\beta}_r^T X^T y_r$.

Now

$$
\begin{aligned}
&(y_r - X\tilde{\beta}_r)^T(y_r - X\tilde{\beta}_r) - (y_r - X\hat{\beta}_r)^T(y_r - X\hat{\beta}_r)\\
&= (\tilde{\beta}_r - \hat{\beta}_r)^T(X^T X)(\tilde{\beta}_r - \hat{\beta}_r) + y_r^T y_r - y_r^T X\hat{\beta}_r - (y_r^T y_r - \hat{\beta}_r^T X^T y_r)\\
&= (\tilde{\beta}_r - \hat{\beta}_r)^T(X^T X)(\tilde{\beta}_r - \hat{\beta}_r).
\end{aligned}
$$

We can do the same work to the imaginary part too. Then we have

$$(y_i - X\tilde{\beta}_i)^T(y_i - X\tilde{\beta}_i) - (y_i - X\hat{\beta}_i)^T(y_i - X\hat{\beta}_i) = (\tilde{\beta}_i - \hat{\beta}_i)^T(X^TX)(\tilde{\beta}_i - \hat{\beta}_i).$$

The test statistic becomes

$F = \frac{[(\tilde{\beta}_r-\hat{\beta}_r)^T(X^TX)(\tilde{\beta}_r-\hat{\beta}_r)+(\tilde{\beta}_i-\hat{\beta}_i)^T(X^TX)(\tilde{\beta}_i-\hat{\beta}_i)]}{[(y_r-X\hat{\beta}_r)^T(y_r-X\hat{\beta}_r)+(y_i-X\hat{\beta}_i)^T(y_i-X\hat{\beta}_i)]}$

$F = \frac{[(\tilde{\beta}_r-\hat{\beta}_r)^T(X^TX)(\tilde{\beta}_r-\hat{\beta}_r)+(\tilde{\beta}_i-\hat{\beta}_i)^T(X^TX)(\tilde{\beta}_i-\hat{\beta}_i)]/\sigma^2}{[(y_r-X\hat{\beta}_r)^T(y_r-X\hat{\beta}_r)+(y_i-X\hat{\beta}_i)^T(y_i-X\hat{\beta}_i)]/\sigma^2}$

We assume

$X_{1r} = (\tilde{\beta}_r - \hat{\beta}_r)^T(X^TX)(\tilde{\beta}_r - \hat{\beta}_r)$

$X_{1i} = (\tilde{\beta}_i - \hat{\beta}_i)^T(X^TX)(\tilde{\beta}_i - \hat{\beta}_i),$

$X_{2r} = (y_r - X\hat{\beta}_r)^T(y_r - X\hat{\beta}_r),$

$X_{2i} = (y_i - X\hat{\beta}_i)^T(y_i - X\hat{\beta}_i).$

To find the distribution of $X_{1r}$, $X_{1i}$, $X_{2r}$ and $X_{2i}$, we need the following conclusions.

Conclusion 1: Let A be an idempotent matrix. The rank of A is its trace.

Conclusion 2: Let A be a $k \times k$ matrix of constants and $y$ be a $k \times 1$ multivariate normal random vector with mean $\mu$ and variance matrix $\sigma^2 I$; thus, $y \sim N(\mu, \sigma^2 I)$. If A is idempotent with rank $p$, then

$$\frac{y'Ay}{\sigma^2} \sim \chi^2_{p,\lambda}$$

where $\lambda = \mu' A\mu/\sigma^2$.

Conclusion 3: Let X be an $n \times p$ matrix partitioned such that $X = [X_1, X_2]$, note that

$$X(X'X)^{-1}X'X = X$$

$$X(X'X)^{-1}X'[X_1X_2] = X$$

$$X(X'X)^{-1}X'[X_1X_2] = [X_1X_2]$$

Consequently,

$$X(X'X)^{-1}X'X_1 = X_1 \text{ and } X(X'X)^{-1}X'X_2 = X_2.$$

Similarly,

$$X_1'X(X'X)^{-1}X' = X_1' \text{ and } X_2'X(X'X)^{-1}X' = X_2'.$$

Conclusion 4: Let A and B be symmetric and idemponent matrices, $y = X\beta + \epsilon$, where $\epsilon \sim N(\mu, \sigma^2 I)$. Assume $U = y^T Ay$, $V = y^T By$, U and V are independent if

$$\sigma^2 AB = 0.$$

A simple proof is given as below.

Proof: Because A and B be symmetric and idemponent,

$y^T Ay = y^T A^T Ay = f_1(Ay)$

$y^T By = y^T B^T By = f_2(By)$

$y^T Ay$ can be seen as a function of Ay and $y^T By$ can be seen as a function of $By$.

Hence showing $y^T Ay$ and $y^T By$ are independent is the same as showing $Ay$ and $By$ are independent. Under the assumption that $Var(\epsilon) = \sigma^2 I$, $Ay$ and $By$ are independent when

$$cov(Ay, By) = Acov(y, y)B^T = A\sigma^2 IB^T = \sigma^2 AB = 0.$$

We will derive the distribution of $X_{1r}, X_{1i}, X_{2r}$ and $X_{2i}$ respectively. Firstly, we will show the derivation of $X_{2r}$ and $X_{2i}$.

$$\begin{aligned} X_{2r} &= (y_r - X\hat{\beta}_r)^T(y_r - X\hat{\beta}_r)/\sigma^2, \\ &= (y_r - X(X^TX)^{-1}X^Ty_r)^T(y_r - X(X^TX)^{-1}X^Ty_r)/\sigma^2 \\ &= y_r^T[I - X(X^TX)^{-1}X^T]y_r^T/\sigma^2. \end{aligned}$$

We can show that $[I - X(X^TX)^{-1}X^T]$ is symmetric idempotent. By Conclusion 2, $X_{2r}/\sigma^2$ follows a $\chi^2$ distribution. The degrees of freedom come from the rank of $[I - X(X^TX)^{-1}X^T]$, which in turn is the trace by Conclusion 1.

trace$[I - X(X^TX)^{-1}X^T]$

=trace$[I]$ - trace$[X(X^TX)^{-1}X^T]$

=trace$[I]$ - trace$[X^T X(X^T X)^{-1}]$

=trace$[I]$ - trace$[I_p]$

$=n-p.$

So $X_{2r}/\sigma^2$ follows a $\chi^2_{n-p}$ distribution. We can do the same work to $X_{2i}/\sigma^2$ and we can show it also follows a $\chi^2_{n-p}$ distribution.

Secondly, we will show the derivation of $X_{1r}$ and $X_{1i}$.

The design matrix is $X = \begin{pmatrix} 1 & x_{11} & x_{12} & ... & x_{1k} \\ 1 & x_{21} & x_{22} & ... & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & ... & x_{nk} \end{pmatrix} = \begin{pmatrix} x_{(1)} \\ x_{(2)} \\ \vdots \\ x_{(n)} \end{pmatrix}.$

Here $x_{(t)}$ represents the $t^{th}$ row in the design matrix and $x_{(t)} = (1, x_{(t)1}, x_{(t)2}, \cdots, x_{(t)k})$. $y_t = 1 \cdot \beta_0 + x_{(t)1} \cdot \beta_1 + x_{(t)2} \cdot \beta_2 + \cdots + x_{(t)k} \cdot \beta_k + \epsilon_t$.

If the null hypothesis is $H_0 : \beta_1 = 0, \beta_2 = 0, \cdots, \beta_m = 0$, under the null hypothesis, we can change the order of columns and write the design matrix as

$$X = \begin{pmatrix} x_{11} & x_{12} & ... & x_{1m} & 1 & x_{1,m+1} & x_{1,m+2} & ... & x_{1k} \\ x_{21} & x_{22} & ... & x_{2m} & 1 & x_{2,m+1} & x_{2,m+2} & ... & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & ... & x_{nm} & 1 & x_{n,m+1} & x_{n,m+2} & ... & x_{nk} \end{pmatrix} = [X_1, X_2]$$

where $X_1 = \begin{pmatrix} x_{11} & x_{12} & ... & x_{1m} \\ x_{21} & x_{22} & ... & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & ... & x_{nm} \end{pmatrix}$, $X_2 = \begin{pmatrix} 1 & x_{1,m+1} & x_{1,m+2} & ... & x_{1k} \\ 1 & x_{2,m+1} & x_{2,m+2} & ... & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,m+1} & x_{n,m+2} & ... & x_{nk} \end{pmatrix}$.

Note $\hat{y}_t = 1 \cdot \hat{\beta}_0 + x_{(t)1} \cdot 0 + x_{(t)2} \cdot 0 + \cdots + x_{(t)m} \cdot 0 + x_{(t)m+1} \cdot \hat{\beta}_{m+1} + \cdots + x_{(t)k} \cdot \hat{\beta}_k + \epsilon_t$.

We can write the design matrix under the null hypothesis as

$$X_2 = \begin{pmatrix} 1 & x_{1,m+1} & x_{1,m+2} & \dots & x_{1k} \\ 1 & x_{2,m+1} & x_{2,m+2} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,m+1} & x_{n,m+2} & \dots & x_{nk} \end{pmatrix}.$$

Note $\tilde{\beta}_r = (X_2^T X_2)^{-1} X_2^T y_r$.

$$\begin{aligned} X_{1r} &= (\tilde{\beta}_r - \hat{\beta}_r)^T (X^T X)(\tilde{\beta}_r - \hat{\beta}_r) \\ &= (y_r - X\tilde{\beta})^T (y_r - X\tilde{\beta}) - (y_r - X\hat{\beta})^T (y_r - X\hat{\beta}) \\ &= y_r^T [I - X_2 (X_2^T X_2)^{-1} X_2^T] y_r - y_r^T [I - X(X^T X)^{-1} X^T] y_r \\ &= y_r^T [X(X^T X)^{-1} X^T - X_2 (X_2^T X_2)^{-1} X_2^T] y_r. \end{aligned}$$

By Conclusion 2, $X_{1r}/\sigma^2$ follows a $\chi^2$ distribution. The degrees of freedom come from the rank of $[X(X^T X)^{-1} X^T - X_2 (X_2^T X_2)^{-1} X_2^T]$, which in turn is the trace by Conclusion 1.

$$\begin{aligned} &\text{trace}[X(X^T X)^{-1} X^T] \text{ - trace}[X_2 (X_2^T X_2)^{-1} X_2^T] \\ =&\text{trace}[X^T X (X^T X)^{-1}] \text{ - trace}[X_2^T X_2 (X_2^T X_2)^{-1}] \\ =&\text{trace}[I_p] \text{ - trace}[I_{p-m}] \\ =&m. \end{aligned}$$

So the degrees of freedom is $m$.

Now consider when the null hypothesis is $H_0 : C\beta = 0$ where $C$ is $r \times p$, full rank and $r < p$.

In matrix $C$, because it is full rank, the row vectors are linearly independent. We can always add more linearly independent row vectors in $C$ to get a square matrix and the new square matrix is still full rank.

We call the new square matrix $Q = \begin{pmatrix} C \\ C_1 \end{pmatrix}$. There are $p - r$ linearly independent row vectors in $C_1$. Q is $p \times p$, non-singular and invertible.

We have

$$y = X\beta + \epsilon$$
$$y = XQ^{-1}Q\beta + \epsilon$$
$$y = XQ^{-1}\begin{pmatrix} C \\ C_1 \end{pmatrix}\beta + \epsilon, \text{ where } Q = \begin{pmatrix} C \\ C_1 \end{pmatrix}$$
$$y = X_q\begin{pmatrix} C\beta \\ C_1\beta \end{pmatrix} + \epsilon \text{ , where } X_q = XQ^{-1}$$
$$y = \begin{pmatrix} X_q^{(1)}, X_q^{(2)} \end{pmatrix}\begin{pmatrix} r_1 \\ r_2 \end{pmatrix} + \epsilon,$$
$$\text{where } X_q = \begin{pmatrix} X_q^{(1)}, X_q^{(2)} \end{pmatrix}, \begin{pmatrix} C\beta \\ C_1\beta \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}.$$
$$y = X_q^{(1)}r_1 + X_q^{(2)}r_2 + \epsilon.$$

The null hypothesis $H_0 : C\beta = 0$ is the same as $H_0 : r_1 = 0$. Assume $r_1 = \begin{pmatrix} r_{11} \\ r_{12} \\ \vdots \\ r_{1r} \end{pmatrix}$, the null hypothesis becomes $H_0 : r_{11} = 0, r_{12} = 0, \cdots, r_{1r} = 0$. This is the same situation as we showed above. Hence $X_{1r}/\sigma^2$ follows a $\chi^2$ distribution. The degrees of freedom is $r$ which is equal to the number of rows in $C$. We can also get $X_{1i}/\sigma^2 \sim \chi^2_r$.

Thirdly, we have shown that both $X_{1r}/\sigma^2$ and $X_{2r}/\sigma^2$ follow $\chi^2$ distribution. Now we will show that $X_{1r}$ and $X_{2r}$ are independent.

$$X_{1r} = (\tilde{\beta}_r - \hat{\beta}_r)^T(X^TX)(\tilde{\beta}_r - \hat{\beta}_r) = y_r^T[X(X^TX)^{-1}X^T - X_2(X_2^TX_2)^{-1}X_2^T]y_r$$
$$X_{2r} = (y_r - X\hat{\beta}_r)^T(y_r - X\hat{\beta}_r) = y_r^T[I - X(X^TX)^{-1}X^T]y_r^T.$$

By conclusion 4, this is the same as showing

$$\sigma^2 AB = 0,$$

where $A = X(X^TX)^{-1}X^T - X_2(X_2^TX_2)^{-1}X_2^T$ and $B = I - X(X^TX)^{-1}X^T$.

Note that

$$
\begin{aligned}
&\sigma^2[X(X^TX)^{-1}X^T - X_2(X_2^TX_2)^{-1}X_2^T][I - X(X^TX)^{-1}X^T]\\
=&\sigma^2[X(X^TX)^{-1}X^T - X_2(X_2^TX_2)^{-1}X_2^T - X(X^TX)^{-1}X^TX(X^TX)^{-1}X^T +\\
&X_2(X_2^TX_2)^{-1}X_2^TX(X^TX)^{-1}X^T]\\
=&\sigma^2[X(X^TX)^{-1}X^T - X_2(X_2^TX_2)^{-1}X_2^T - X(X^TX)^{-1}X^T + X_2(X_2^TX_2)^{-1}X_2^T]\\
=&0.
\end{aligned}
$$

So we proved that $X_{1r}$ and $X_{2r}$ are independent. Using the same method, we can also show that $X_{1i}$ and $X_{2i}$ are independent.

Hence we have proved that under the null hypothesis $H_0 : C\beta = 0$ where C is $r \times p$, full rank and $r < p$,

$$
\begin{gathered}
\frac{X_{1r}}{\sigma^2} \sim \chi^2_r\\
\frac{X_{1i}}{\sigma^2} \sim \chi^2_r\\
\frac{X_{2r}}{\sigma^2} \sim \chi^2_{n-p}\\
\frac{X_{2i}}{\sigma^2} \sim \chi^2_{n-p}
\end{gathered}
$$

$\frac{X_{1r}}{\sigma^2}$ and $\frac{X_{2r}}{\sigma^2}$ are independent. $\frac{X_{1i}}{\sigma^2}$ and $\frac{X_{2i}}{\sigma^2}$ are independent.

The test statistics

$$F = \frac{\frac{X_{1r}}{\sigma^2} + \frac{X_{1i}}{\sigma^2}}{\frac{X_{2r}}{\sigma^2} + \frac{X_{2i}}{\sigma^2}} \sim F_{2r,2(n-p)}.$$

# Bibliography

1. Rowe DB. Modeling both the magnitude and phase of complex-valued fMRI data. Neuroimage 2005;25:1310-1324.
2. Lee J, Sharam M, Schwartzman A, Pauly JM. Complex Data Analysis in High-Resolution SSFP fMRI. Magnetic Resonance in Medicine 2007;57:905-917.
3. Rowe DB. Magnitude and Phase Signal Detection in Complex-Valued fMRI Data. Magnetic Resonance in Medicine 2009;62:1356-1357.
4. Lee J, Sharam M and Pauly JM. Combining Complex Signal Change in Functional MRI. Magnetic Resonance in Medicine 2009;62:1358-1360.
5. Sruvastava MS. Methods of Multivariate Statistics. New York: John Wiley & Sons; 2002.P.89-151.
6. Rowe DB, Logan BR. A complex way to compute fMRI activation. Neuroimage 2004;23:1078-1092.
7. Rowe DB, Logan BR. Complex fMRI analysis with unrestricted phase is equivalent to a magnitude-only model. Neuroimage 2005;24:603-606.
8. Rowe DB. Parameter estimation in the magnitude-only and complex-valued fMRI data models. Neuroimage 2005;25:1124-1132.
9. Montgomery DC, Peck EA, Vining GG. Introduction to Linear Regression Analysis. New York: John Wiley & Sons; 2007.P.547-556.
10. Ogawa S, Lee TM, Kay AR and Tank DW. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci 1990;87:9868-9872.

Marquette University

This is to certify that we have examined this copy of the thesis by

Yuning Chen

and have found that it is complete and satisfactory in all respects.

The thesis has been approved by:

Dr. Daniel B. Rowe
Thesis Director, Department of Mathematics, Statistics and Computer Science

Approved on