9.073/HST.460 Statistics for Neuroscience Research
Emery N. Brown

Lecture 15: Linear Models II: Multiple Linear Regression

I. Objectives

Understand the theory of multiple linear regression

Understand the geometry of multiple linear regression

Understand how to assess model goodness-of-fit and make model comparisons

Understand how to construct relevant hypothesis tests

Understand its application in fMRI

II. Multiple Linear Regression

A. Motivation

Because one of the widest uses of multiple linear regression in neuroscience is functional magnetic resonance imaging (fMRI) data analysis, we use an fMRI study to motivate its formulation.

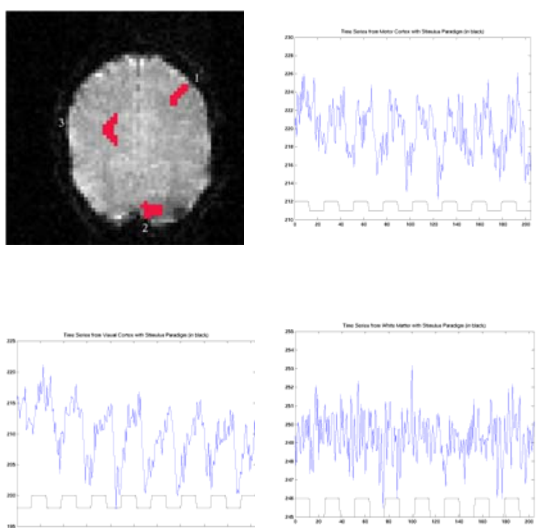

Figure 15.1 FMRI Experiment using an interleaved visual and motor task.

Example 15.3 fMRI Data Analysis. Figure 15.1 shows a slice taken from a combined visual and motor fMRI experiment. The subject was presented with a full-field flickering checkerboard, in a 12.8-s OFF, 12.8-s ON pattern, repeated 8 times. Alternating out of phase with the flickering checkerboard, the subject is instructed to finger tap. The slice shown was chosen to transect both the visual and motor cortices, and was imaged once every 800ms for the duration of the experiment. Three regions of interest have been selected, corresponding to B. motor cortex; C. visual cortex; and D. white matter. Panels B-D illustrate the raw time series taken from each of these regions, along with timing diagrams of the input stimuli.

A question which might be asked of this experiment is whether there significant activation in the visual and motor cortices during the respective phases of the motor and visual tasks? A second question might be whether the level of activation is greater in the visual area than in the motor area? To investigate these questions, we must develop a model for the time-series of fMRI responses on a each voxel. A plausible model for these data based on the physiology and biophysics fMRI should contain: 1) a hemodynamic response component, representing the blood oxygen level dependent (BOLD) response of the brain vasculature to the stimulus; 2) a drift term to account for subject motion artifacts and/or drift in the imaging parameters; 3) physiological noise due to low frequency vibrations in the cerebral vasculature; and 4) Johnson noise in the scanner (Purdon et al. 2001). We will construct an elementary multiple linear regression model based on these modeling considerations to analyze the time-series from this data set.

B. Multiple Linear Regression Model Assumptions

We can derive the multiple linear regression model by extending the sample linear regression model assumption that the number of regressors is $p > 1$. We assume a model of the form

$$y_i = \beta_0 + \beta_i x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \varepsilon_i \tag{15.1}$$

i) $E[y_i \mid x_i] = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}$ and the y_i's are sometimes referred to as the **independent variable**.

ii) The $x_i = (x_{i1}, ..., x_{ip})$ are fixed non-random constants termed **covariates**, **regressors, independent variables, explanatory variables or carriers**.

iii) The ε_i's are independent Gaussian random variables with mean zero and variance σ^2.

C. Model Parameter Estimation

If we let

$$Y = \begin{bmatrix} y_1 \\ . \\ . \\ . \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & ... & x_{1p} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ 1 & x_{1n} & ... & x_{np} \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix} \tag{15.2}$$

and $\beta = (\beta_0, ..., \beta_p)$, then the model in Eq. 15.1 becomes in matrix form

$$Y = X\beta + \varepsilon. \tag{15.3}$$

Our objective is to estimate β and σ^2. By our assumption, we have a Gaussian distribution for Y given X, and hence a Gaussian likelihood defined as

$$
\begin{aligned}
L(\beta,\sigma^2 \mid Y,X) &= f(Y \mid X,\beta,\sigma^2) \\
&= \prod_{i=1}^{n} f(y_i \mid x_i,\beta,\sigma^2) \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\{-\frac{1}{2}(\sum_{i=1}^{n} y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\{-\frac{1}{2\sigma^2}(Y-X\beta)^T(Y-X\beta)\}.
\end{aligned}
\tag{15.4}
$$

where X^T denotes the transpose of the matrix X. Proceeding as in the simple regression case, we have the log likelihood is

$$
\log L(\beta,\sigma^2 \mid Y,X) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y-X\beta)^T(Y-X\beta). \tag{15.5}
$$

If we differentiate the log likelihood with respect to β and set the derivatives equal to zero we obtain the normal equations

$$
(X^T X)\beta = X^T Y \,. \tag{15.6}
$$

Solving for $\hat{\beta}$ yields the maximum likelihood estimate

$$
\hat{\beta} = (X^T X)^{-1} X^T Y \,. \tag{15.7}
$$

It follows that

$$
\frac{\partial \log L(\hat{\beta},\sigma^2 \mid Y,X)}{\partial(\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{(Y-X\hat{\beta})^T(y-X\hat{\beta})}{(\sigma^2)^2}, \tag{15.8}
$$

and solving for σ^2 after setting the derivatives equal to zero yields $\hat{\sigma}^2$, the maximum likelihood estimate of σ^2 is given as

$$
\begin{aligned}
\hat{\sigma}^2 &= n^{-1}(Y-X\hat{\beta})^T(Y-X\hat{\beta}) \\
&= n^{-1}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \sum_{j=1}^{p}\hat{\beta}_j x_{ij})^2.
\end{aligned}
\tag{15.9}
$$

Remark 15.1. Estimating β by maximum likelihood under a Gaussian error assumption is equivalent to the method of least-squares and

$$
\begin{aligned}
s_n(\beta) &= n^{-1}(Y-X\beta)^T(Y-X\beta) \\
&= n^{-1}\,\|Y-X\beta\|^2
\end{aligned}
\tag{15.10}
$$

where $\| x \|^2 = \sum_{i=1}^{n} x_i^2$ is the squared length of the vector x.

Remark 15.2. To invert the matrix $X^T X$ in Eq. 15.7, the columns of X must be **linearly independent** in the linear algebra sense. What this means practically is that no regressor can be expressed as a linear function of the other regressors in the model. If one regressor can be expressed as a linear function of any of the others, it means that the number of distinct explanatory variables is less than p. High dependence among the regressors can be detected by analyzing the correlation matrix of X before beginning the model fitting.

Remark 15.3. Computing the regression parameter estimate is equivalent by Eq. 15.7 to projecting the data Y onto the subspace (plane) spanned by the columns of X. This can be seen by noting that

$$\begin{aligned} \hat{Y} &= X\hat{\beta} \\ &= X(X^T X)^{-1} X^T Y \\ \hat{Y} &= HY. \end{aligned} \tag{15.11}$$

The matrix H is called the **"hat" matrix**, because it is the projection operator that projects Y onto $\hat{Y}$. It has the interesting property that it is **idempotent**. By idempotent we mean that $H \cdot H = H$. To see this, we simply note that

$$\begin{aligned} H \cdot H &= X(X^T X)^{-1} X^T X (X^T X)^{-1} X \\ &= X(X^T X) I X \\ &= X(X^T X) X \\ &= H. \end{aligned} \tag{15.12}$$

Remark 15.4. The Pythagorean relation among the sums of squares holds as in the case of simple linear regression. It is usually written in the form of an ANOVA table.

Source	Sum of Squares	Degrees of Freedom	Mean Square Error
Regression (explained)	$\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$	p	$\frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{p}$
Error (residual)	$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$	$n - p - 1$	s^2
Total	$\sum_{i=1}^{n} (y_i - \bar{y})^2$	$n - 1$	

Table 15.1 Analysis of Variance Summary for the Multiple Linear Regression Model.

where

$$s^2 = (n-p-1)^{-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{15.13}$$

and the $\hat{\varepsilon}_i = y_i - \hat{y}_i$ are the residuals. As discussed in the case of simple linear regression, s^2 provides an unbiased estimate of σ^2 whereas $\hat{\sigma}^2$ (Eq. 15.9) is the maximum likelihood estimate of σ^2. Again, by analogy with the simple linear regression case, it is s^2 that will be used to estimate σ^2 when we carry out our hypothesis tests and construct our confidence intervals.

Remark 15.5. As was true for the simple linear regression model goodness-of-fit analyses, we have that the **square of the multiple correlation coefficient,** R^2, is

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{TSS - RSS}{TSS} = 1 - \frac{RSS}{TSS} \tag{15.14}$$

Here again, the R^2 measures the fraction of the variance in Y that is explained by the regressors. R^2 is also the cosine of the angle between Y and the plane (subspace) defined by the regressors, i.e. the columns of the X matrix. As $p \to n$, the R^2 will approach 1 because there is a parameter for every data point. This means that the regression line (plane) goes through every data point. The geometric interpretation is that when $p = n$, the data Y lie completely in the plane defined by the regressors. Therefore, instead of just reporting the R^2, it is customary to also report the **adjusted** R^2 defined as

$$R^2 adj = 1 - \frac{RSS/(n-p-1)}{TSS/(n-1)}, \tag{15.15}$$

because it adjusts the R^2 for the number of parameters in the model.

Remark 15.6. The adjusted R^2 generalizes approximately to certain non-linear Gaussian model problems but not at all to problems with non-Gaussian data. As a consequence, various criteria have been proposed to measure the parsimony of statistical models, i.e. the trade-off between of the model fit to the data and the number of parameters required to achieve that fit. The criterion we use to measure this trade-off is **Akaike's Information Criterion (AIC)**. It is defined as

$$\text{AIC} = -2\log L(\hat{\beta}, \hat{\sigma}^2) + 2p \tag{15.16}$$

where *p* is the number of the parameters in the model and the remaining expression on the rhs of Eq. 15.16 is the log likelihood function evaluated at the maximum likelihood estimates $\hat{\beta}$ and $\hat{\sigma}^2$. The log likelihood function, which decreases as more parameters are added to the model, is penalized by the quantity $2p$ to measure the marginal cost of improving the goodness-of-fit by increasing the number of parameters. AIC can be used to compare different models. The smaller the AIC, the more parsimonious is the model's description of the data. While the AIC is interpreted as measuring the trade-off between fit to the data and model complexity, this is not the logic Akaike used to derive it. What AIC provides is an approximately unbiased estimate of the average or expected Kullbach-Liebler distance between the model being fit to the data and the "true" model (Pawitan, 2001). Several extensions of AIC have been proposed. None of these model selection techniques should be viewed as an absolute criterion for choosing a model. Instead, they provide helpful guidelines that should be evaluated in the context of the problem being studied.

For the Gaussian multiple linear regression model, the AIC simplifies to

$$\begin{aligned} AIC &= -2\log L(\hat{\beta},\hat{\sigma}^2 \mid Y) + 2p \\ &= n\log(2\pi\hat{\sigma}^2) + \frac{\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p}\hat{\beta}_j x_{ij})^2}{\hat{\sigma}^2} + 2p \\ &= n\log(\hat{\sigma}^2) + n\log(2\pi) + n + 2p \end{aligned} \tag{15.17}$$

which, on removing the term $n\log(2\pi)+n,$ gives

$$AIC = n\log(\hat{\sigma}^2) + 2p. \tag{15.18}$$

We can omit this last term because for a given data set, it is a constant. Hence, we see that for the Gaussian multiple linear regression problem, the maximized log likelihood is simply n times the log of the estimate of the observation variance.

Remark 15.7. The residuals from the model fit should be analyzed as discussed in **Lecture 14** as part of the goodness-of-fit assessment.

Remark 15.8. To analyze the performance of our model , we must be able to test hypotheses about the model parameters. If the null hypothesis $H_0 : \beta_1 = \beta_2 = ... = \beta_p = 0$ is true, then

$$ESS = \sum_{i=1}^{n}(y_i - \bar{y})^2 \sim \chi_p^2 \tag{15.19}$$

and

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \sim \chi_{n-p-1}^2, \tag{15.20}$$

where χ_v^2 denotes a chi-squared random variable with v degrees of freedom. It also turns out that ESS and RSS are independent. We then have, using **Table 15.1** and Eq. 15.13, that the F- statistic for testing H_0 is

$$F_{p,n-p} = \frac{ESS/p}{RSS/(n-p-1)} \sim \frac{\chi_p^2/p}{\chi_{n-p-1}^2/(n-p-1)}. \tag{15.21}$$

The F-statistic is the ratio of two independent chi-squared random variables. Both the R^2 and the F-statistic have straightforward geometric interpretations. However, F-statistic is easier to work with for hypothesis testing because, under the Gaussian assumptions about the data, the orthogonality of ESS and RSS (Pythagorean relation) means that these two quantities are independent. Therefore, it is easier to work out distribution of the F-statistic rather than that of the R^2. In simple linear regression model in which there is only one explanatory variable, we had $p=1$ and, as we pointed out $F = (t\text{-statistic})^2$. In the multiple linear regression this is no longer true as the F-statistic tests the composite null hypothesis $H_0 : \beta_1 = \beta_2 = ...\beta_p = 0.$ Therefore, the way to use the F-statistic is to first test the hypothesis that all the regression coefficients are zero. If this hypothesis is rejected, then there are individual t-statistics for each regression coefficient which may be used to determine which if any of the individual regression coefficients may be zero. We discuss this test in the next section.

D. Properties of the Model Parameters

The properties of the model parameter estimate $\hat{\beta}$ are easy to derive using a little matrix algebra. The estimate $\hat{\beta}$ is by Eq. 15.7, a linear function of Y so therefore, it has a Gaussian distribution. Because it is Gaussian, it suffices therefore, to define specify only its mean and variance. For the mean we have

$$\begin{aligned} E(\hat{\beta}) &= E[(X^T X)^{-1} X^T Y] \\ &= (X^T X)^{-1} X^T E(Y) \\ &= (X^T X)^{-1} X^T [X\beta + E(\varepsilon)] \\ &= I\beta + 0 \\ &= \beta \end{aligned} \tag{15.22}$$

Hence, $\hat{\beta}$ has mean β and hence, it is an unbiased estimate of $\beta.$ To compute the variance of $\hat{\beta},$ we need to use the fact that if X is a random vector with covariance matrix $\Sigma,$ then the covariance matrix of $Y = CX$ is $C\Sigma C^T,$ where C is a matrix. This is the matrix analog of the univariate result that if x has variance σ^2 then $y = cx$ has variance $c^2\sigma^2.$ If we use this result to compute the variance of $\hat{\beta},$ we find that

$$\begin{aligned} Var(\hat{\beta}) &= Var[X^T X)^{-1} X^T Y] \\ &= (X^T X)^{-1} X^T Var(Y) X (X^T X)^{-1} \\ &= (X^T X)^{-1} X^T (I\sigma^2) X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1} (X^T X)(X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1}. \end{aligned} \tag{15.23}$$

Therefore, $\hat{\beta}$ has a Gaussian distribution defined as

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1}), \tag{15.24}$$

which is the same result we stated in the simple linear regression case. Because we estimate σ^2 by s^2 in Eq. 15.13, we estimate the variance of $\hat{\beta}$ as

$$\mathrm{Var}(\hat{\beta}) = \mathrm{s}^2 (X^T X)^{-1}, \tag{15.25}$$

and the standard error of $\hat{\beta}_k$ for $k = 1, ..., p+1$ is

$$\mathrm{se}(\hat{\beta}_k) = [s^2 (X^T X)_{kk}^{-1}]^{\frac{1}{2}}, \tag{15.26}$$

where the $(X^T X)_{kk}^{-1}$ is the k^{th} diagonal element of this $p+1 \times p+1$ matrix for $k = 1, ..., p+1$. Therefore, we define the t-statistic for each individual coefficient as

$$t_{\hat{\beta}_k} = \frac{\hat{\beta}_k}{\mathrm{se}(\beta_k)} \sim t_{n-p-1}, \tag{15.27}$$

where t_{n-p-1} is a random variable with a t-distribution having $n-p-1$ degrees of freedom. Hence, we can construct a $100\%(1-\delta)$ confidence interval for each individual parameter as

$$\hat{\beta} \pm t_{n-p-1, 1-\delta/2} \mathrm{se}(\hat{\beta}_k). \tag{15.28}$$

Example 15.3 (continued). To analyze fMRI data we consider two models. The first model is a hemodynamic response model defined for a given voxel as

$$y_t = \beta_0 + \beta_1 t + \beta_2 h(t) + \varepsilon_t, \tag{15.29}$$

where $h(t)$ is the physiological response function defined by

$$h(t) = \int_0^t g(u) c(t-u) du, \tag{15.30}$$

$c(t)$ is the ON-OFF light or motor stimulus which has the square wave shape in **Figure 15.1** and $g(t)$ is the hemodynamic response function defined by the gamma function

$$g(t) = t^{\alpha_1 - 1} e^{-\alpha_2 t}, \tag{15.31}$$

where for this problem, we chose $\alpha_1 = 7.5$ and $\alpha_2 = 1$. The first term in Eq. 15.29 represents a possible linear drift which may be due to subject movement and/or drift in the scanning parameters. This model is an elementary version of the types of regression models used to analyze fMRI time series (Purdon et al., 2001). It contains two of the four important components we stated in **IIA (Motivation)** as required to formulate a model of an fMRI time-series that can be used across whole brain regions. An fMRI image is made by plotting the estimate of β_2 or a normalized version of it for each voxel.

The second model we consider is a harmonic regression model defined as

$$y_t = \beta_0 + \beta_1 t + \beta_2 \cos(\omega t) + \beta_3 \sin(\omega t) + \varepsilon_t \tag{15.32}$$

where $\omega = \frac{2\pi}{\tau}$ and τ is the period of either the visual or motor stimulus. This model also includes a linear drift term. We propose this model as an alternative to the hemodynamic response model because of the obvious harmonic structure in the fMRI time-courses from the visual and motor areas (Fig. 15.1). The fMRI image for this model is constructed by plotting the estimate of

$\text{Amplitude} = (\beta_2{}^2 + \beta_3^2)^{\frac{1}{2}}$ on each voxel.

Harmonic Regression + Linear Drift

Source	**DF**	**SS**	**MS**	**F**	**p**
Regress	**3**	**15,071**	**5023.67**	**52.0**	**<1e-7**
Error	**252**	**9,321**	**36.99**		
Total	**255**	**24,392**			

RSQ = 0.618 adj_RSQ = 0.621 AIC = 926.28

	param	**se**	**t**	**p**
const	**243**	**0.48**	**504**	**<1e-10**
drift	**-13.4**	**1.68**	**-7.9**	**<1e-10**
sine	**2.3**	**0.68**	**3.4**	**0.008**
cosine	**-5.9**	**0.68**	**-8.6**	**<1e-10**

Table 15.2. Harmonic Regression + Linear Drift Model Fit

We use the two models to analyze the time course of one of the voxels from primary motor cortex (**Figs. 15.2A** and **15.3A**). **Table 15.2** shows the analysis of the fit of the harmonic regression model with linear drift. The F-statistic with 3 and 252 degrees of freedom is 52.0 with $p < 10^{-7}$. Hence we reject the null hypothesis of al the coefficients being zero. Each t-statistic for an individual coefficient is greater than 2 and has a $p < 10^{-10}$, suggesting that none of these coefficients is zero. The R^2 and adjusted R^2 show that this model with three regressors explains 62% of the variance in the time-series from this voxel.

Hemodynamic Response + Linear Drift

Source	DF	SS	MS	F	p
Regress	2	15,123	7,561.5	77.5	<1e-7
Error	253	9,269	36.4		
Total	255	24,392			

RSQ = 0.620 adjRSQ = 61.7 AIC = 922

	param	se	t	p
const	249	0.75	331	<1e-10
drift	-12.0	1.69	-7.1	<1e-10
hemod	11.2	1.22	9.2	<1e-10

Table 15.3. Harmonic Regression + Linear Drift Model Fit

Table 15.3 shows the analysis of the fit of the hemodynamic response model with a linear drift. The F-statistic with 2 and 253 degrees of freedom is 77.5 with $p < 10^{-7}$. Hence, we reject the null hypothesis of all the coefficients being zero. Each of the two t-statistics is greater than 2 and has a $p < 10^{-10}$, suggesting that none of these coefficients is zero. The R^2 and adjusted R^2 show that this model with 2 regressors explains 62% of the variance in the hemodynamic response time-series on this voxel.

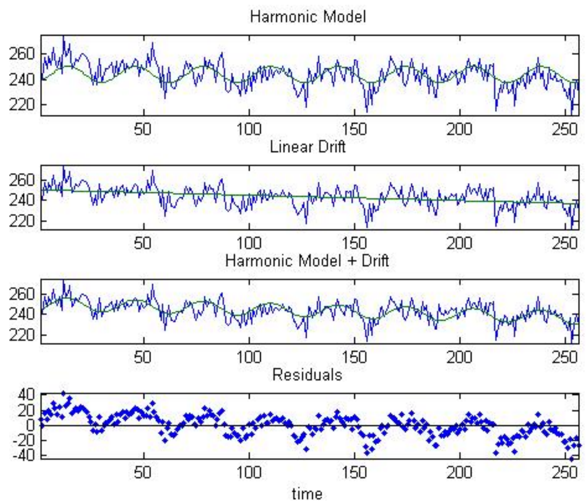

Figure 15.2 Harmonic Regression Model Analysis. A. fMRI time-series from a single voxel in motor cortex with the estimated harmonic response function; B. Linear drift; C. Estimated harmonic response + linear drift; D. Residuals.

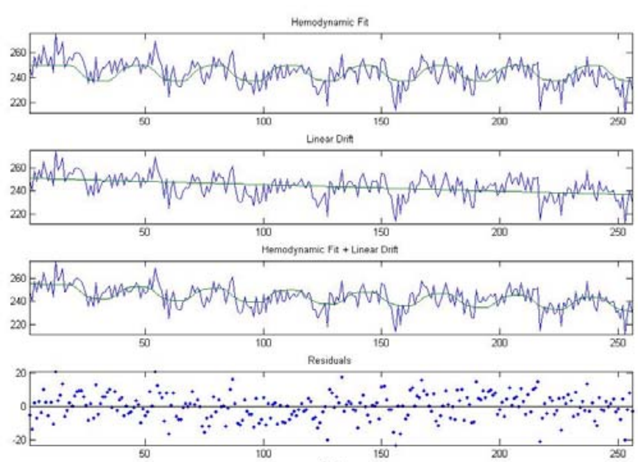

Figure 15.3 Hemodynamic Regression Model Analysis. A. fMRI time-series from a single voxel in motor cortex with the estimated hemodynamic response function; B. Linear drift; C. Estimated hemodynamic response + linear drift; D. Residuals.

Figures 15.2 and **15.3** show respectively the model fits of the harmonic regression and the hemodynamic response models. Both models identify a strong stimulus response and an appreciable negative linear drift. Two important differences are worth noting. First, the amplitude of the hemodynamic response is $\beta_2 = 11.2$ which is larger than the amplitude of the harmonic regression model which is approximately $6.6 = [(2.3)^2 + (5.9)^2]^{\frac{1}{2}}$. Second, the residuals from the harmonic regression model appear highly periodic, ranging from –40 to +40. The residuals from the hemodynamic response model appear much less structured and range from –20 to 20. The Q-Q plot analysis of the residuals from the harmonic regression model fit shows that these residuals are more non-Gaussian (**Figure 15.4**). In contrast, the Q-Q plot of the residuals from the hemodynamic model fit suggests that these residuals are consistent with them being zero mean Gaussian observations (**Figure 15.5**). Finally, the AIC of 922 (**Table 15.3**) for the hemodynamic response model and that of 926 (**Table 15.2**) for the harmonic regression model suggest that the former provides the more parsimonious description of the data.

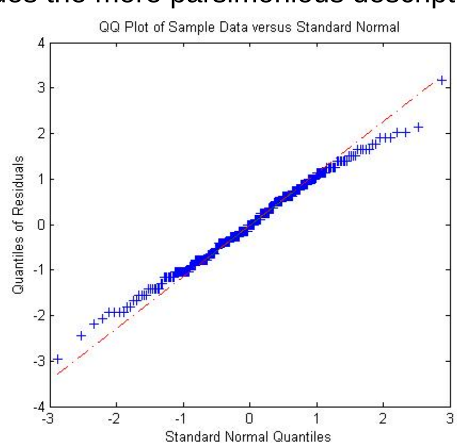

Figure 15.4 Q-Q plot analysis of the residuals from the Harmonic Regression + Linear Drift Model.

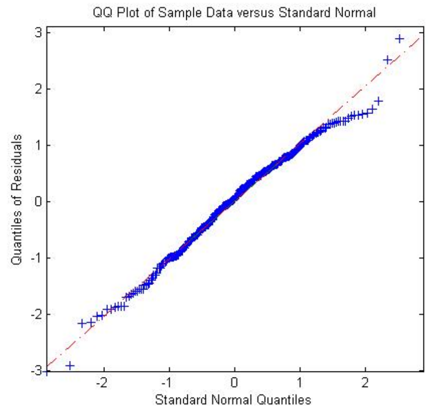

Figure 15.5 Q-Q plot analysis of the residuals from the Hemodynamic Response + Linear Drift Model

These analyses suggests that the hemodynamic response model would be the preferred model for characterizing these data because this model 1) is physiologically based; 2) is more parsimonious in that it achieves goodness-of-fit statistics, comparable to (R^2 and adjusted R^2) and better than (AIC) those of the harmonic regressions model with one fewer regressors; 3) identifies a stronger hemodynamic (stimulus) response component; and 4) has residuals that are smaller in range, more Gaussian like and less structured than those from the harmonic regression model.

As an attempt at further improving the fit of the hemodynamic response model we also fit this model to the data using the first-order autoregressive hemodynamic response model

$$\varepsilon_t = \rho \varepsilon_{t-1} v_t \tag{15.33}$$

where ρ is the autocorrelation coefficient and the v_t's are assumed to be Gaussian noise with mean zero and variance σ^2. The fit of this model is shown in **Figure 15.6**. The third panel in this figure shows the estimate autoregressive process. We see that its effect is small, which is consistent with our finding that the residuals from the hemodynamic model had little to no structure.

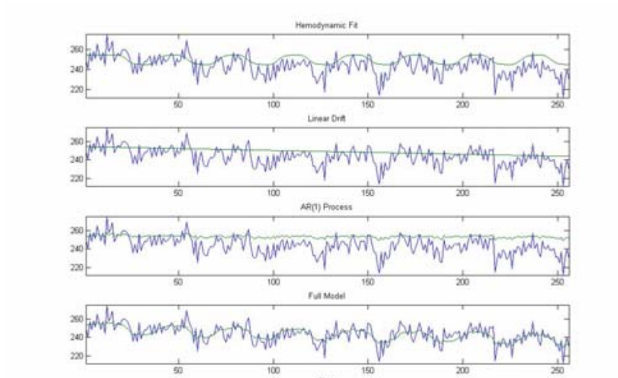

Figure 15.6 Hemodynamic + Linear Drift + First-Order Autoregressive Model.

Figure 15.7 shows the image constructed from analyzing all of the voxel time-series using the hemodynamic +linear drift+ first-order autoregressive model.

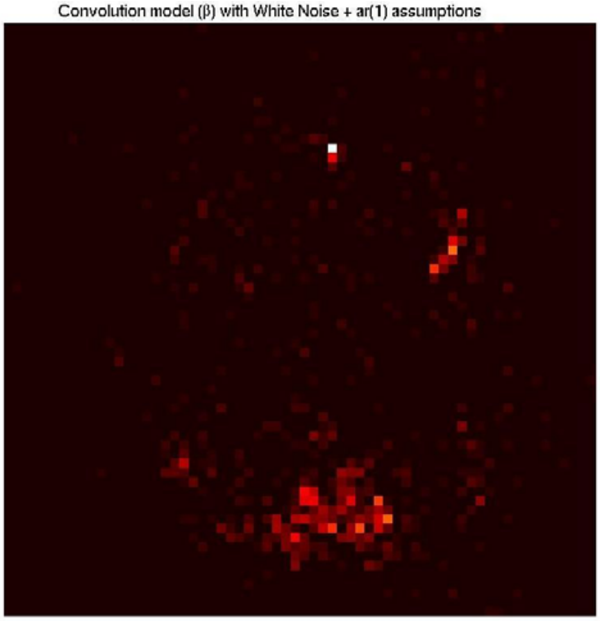

Figure 15.8. fMRI Images from the hemodynamic + linear drift + first-order autoregressive model applied to all of the voxels.

III. Summary

The concepts of the linear model will be crucial for our other model building techniques and paradigms.

Acknowledgments

I am grateful to Uri Eden for conducting all the data analyses and for making the figures. I am grateful to Julie Scott for technical assistance.

Textbook References

Draper NR, Smith H. *Applied Regression Analysis*, 3rd edition. New York: John Wiley & Sons, Inc., 1998.

Pawitan Y. *In All Likelihood: Statistical Modeling and Inference Using Likelihood*. London: Oxford, 2001.

Rice JA. *Mathematical Statistics and Data Analysis*, 3rd edition. Boston, MA, 2007.

Ryan TP. *Modern Regression Methods*. New York: John Wiley & Sons, Inc., 1997.

Literature Reference

Purdon PL, Solo V, Weisskoff RM, Brown EN. Locally regularized spatiotemporal modeling and model comparison for functional MRI. *NeuroImage* 2001; 14, 912-923.