**CAS MA 575 – Linear Models.**

Boston University, Fall 2013

HOMEWORK 6 (CORRECTION)

1. Recall that in Lecture 6.2, we have defined a weighted least squares (WLS) regression problem, with the following mean and variance functions:

$$\mathbb{E}[Y|X = \mathbf{x}_i] = \mathbf{x}_i^T\boldsymbol{\beta}, \qquad \text{and} \qquad \mathbb{V}\text{ar}[Y|X = \mathbf{x}_i] = \frac{\sigma^2}{w_i},$$

where the $w_i$'s are known positive numbers, such that $w_i > 0$, for every $i = 1, \ldots, n$. Alternatively, this can be expressed in matrix notation, $\mathbb{V}\text{ar}[\mathbf{y}|\mathbf{X}] = \sigma^2\mathbf{W}^{-1}$, where $\mathbf{W}$ is a diagonal matrix, whose $\mathbf{W}_{ii}$ element is $w_i$.

Compute the following two quantities:

(a) The *variance/covariance matrix* of the WLS estimator

$$\widehat{\boldsymbol{\beta}} := (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y},$$

using the fact that $\mathbf{X}^T\mathbf{W}\mathbf{X}$ is symmetric.

*This can be done as follows, invoking the now familiar* $\mathbb{V}\text{ar}[\mathbf{A}\mathbf{y}] = \mathbf{A}\,\mathbb{V}\text{ar}[\mathbf{y}]\mathbf{A}^T$,

$$\begin{aligned}\mathbb{V}\text{ar}[\widehat{\boldsymbol{\beta}}|\mathbf{X}] &= \mathbb{V}\text{ar}[(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y}|\mathbf{X}] \\ &= (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\,\mathbb{V}\text{ar}[\mathbf{y}|\mathbf{X}]\mathbf{W}^T\mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}.\end{aligned}$$

*using the fact that* $\mathbf{X}^T\mathbf{W}\mathbf{X}$ *is symmetric. Moreover,* $\mathbb{V}\text{ar}[\mathbf{y}|\mathbf{X}] = \sigma^2\mathbf{W}^{-1}$ *and thus*

$$\begin{aligned}\mathbb{V}\text{ar}[\widehat{\boldsymbol{\beta}}|\mathbf{X}] &= \sigma^2(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{W}^{-1}\mathbf{W}^T\mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{W}\mathbf{X})(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1},\end{aligned}$$

*using the fact that* $\mathbf{W}$ *is also symmetric.*

(b) The *hat matrix* for this model can be defined through the transformations that we have previously seen with $\mathbf{W} = \mathbf{W}^{1/2}\mathbf{W}^{1/2}$,

$$\mathbf{z} := \mathbf{W}^{1/2}\mathbf{y}, \qquad \mathbf{M} := \mathbf{W}^{1/2}\mathbf{X}, \qquad \text{and} \qquad \mathbf{d} := \mathbf{W}^{1/2}\mathbf{e}.$$

Note that, in this case, the hat matrix projects $\mathbf{z}$ to

$$\widehat{\mathbf{z}} := \mathbf{M}\widehat{\boldsymbol{\beta}}.$$

Therefore, you should find an expression of the form $\widehat{\mathbf{z}} = \mathbf{Hz}$.
*Here, it suffices to observe that*

$$\begin{aligned}\widehat{\mathbf{z}} &= \mathrm{M}\widehat{\boldsymbol{\beta}} \\ &= \mathbf{W}^{1/2}\mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y} \\ &= \mathbf{W}^{1/2}\mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{1/2}\mathbf{W}^{1/2}\mathbf{y},\end{aligned}$$

*using the decomposition* $\mathbf{W} = \mathbf{W}^{1/2}\mathbf{W}^{1/2}$. *Therefore, we can define the hat matrix as,*

$$\mathbf{H} := \mathbf{W}^{1/2}\mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{1/2},$$

*since* $\mathbf{z} := \mathbf{W}^{1/2}\mathbf{y}$.
*In fact, this could be further simplified using the definition of* $\mathbf{M}$,

$$\mathbf{H} := \mathbf{M}(\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T.$$

*However, observe that pre-multiplying and post-multiplying by a diagonal matrix does not yield the same quantity. That is, one can easily verify that* $\mathbf{AD} \neq \mathbf{DA}$, *for any conformal matrices* $\mathbf{A}$ *and diagonal matrix* $\mathbf{D}$. *In general,*

$$\mathbf{DAD} \neq \mathbf{ADD} \neq \mathbf{DDA};$$

*and therefore, in the present case*

$$\mathbf{W}^{1/2}\mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{1/2} \neq \mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}.$$

*This type of cycling argument only works, when we take the trace of a matrix product, such that* $tr(\mathbf{AB}) = tr(\mathbf{BA})$. *Consequently, although we can talk about the projection of* $\mathbf{z}$ *as* $\widehat{\mathbf{z}}$ *with respect to* $\mathbf{H}$, *we cannot talk about the projection between* $\mathbf{y}$ *and* $\widehat{\mathbf{y}} := \mathbf{X}\widehat{\boldsymbol{\beta}}$, *because the hat matrix that we may define as* $\mathbf{H}_{\mathbf{y}}$ *in this context,*

$$\begin{aligned}\widehat{\mathbf{y}} &= \mathbf{X}(\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}\mathbf{y} \\ &=: \mathbf{H}_{\mathbf{y}}\mathbf{y},\end{aligned}$$

*can be shown to be not symmetric. However, this matrix is still idempotent. Therefore,* $\mathbf{H}_{\mathbf{y}}$ *is an oblique projection. The geometrical framework that justifies WLS only yields to an orthogonal projection, when considering* $\mathbf{z}$. *This relationship is best understood by noting that the RSS for WLS necessarily includes* $\mathbf{W}$. *That is, we are projecting a transformed version of* $\mathbf{y}$ *–i.e.* $\mathbf{z} = \mathbf{W}^{1/2}\mathbf{y}$*– and not* $\mathbf{y}$ *itself.*