WEIGHTED LEAST SQUARES

Recall: We can fit least squares estimates just assuming a linear mean function. Without the constant variance assumption, we can still conclude that the coefficient estimators are unbiased, but we can't say anything about their variances; consequently, the inference procedures are not applicable.

Moreover, if we fit least squares with non-constant variance, the values with larger variance typically have more influence on the result; values with lower variance typically are fit poorly.

Example: Geese

Recall: Sometimes we can find transformations to achieve constant variance. But sometimes we can't do this without messing up the linear mean or normality assumptions. An alternative that works sometimes is *weighted least squares*.

Model assumptions for weighted least squares:

1. $E(Y|\underline{x}_i) = \underline{\eta}^T\underline{u}_i$ (linear mean function -- as for ordinary least squares)
2. $Var(Y|\underline{x}_i) = \sigma^2/w_i$, where the w_i's are known, positive constants (called *weights*) (Different from OLS!)

Observe:

- w_i is inversely proportional to $Var(Y|\underline{x}_i)$. This is sometimes helpful in getting suitable w_i's.
- the w_i's aren't unique – we could multiply all of them by a constant c, and divide σ by $\sqrt{c}$ to get an equivalent model.

Error: For WLS, the *error* is defined as

$$e_i = \sqrt{w_i}\,[\,Y|\underline{x}_i - \underline{\eta}^T\underline{u}_i]$$ (Different from OLS!)

Then (exercise)

$$E(e_i) = 0 \text{ and } Var(e_i) = \sigma^2$$

Reformulating (1) in terms of errors:

$$1\text{'}: Y|\underline{x}_i = \underline{\eta}^T\underline{u}_i + e_i/\sqrt{w_i}$$

Note: WLS is not a universal remedy for non-constant variance, since weights are needed. But it is useful in many types of situations. Examples:

A. If $Y|\underline{x}_i$ is the <u>sum</u> of m_i independent observations $v_1, v_2, \ldots, v_{mi}$, each with variance σ^2, then

$Var(Y|\underline{x}_i)$ = _____________ = _____, so we could take w_i = ___________.

B. If $Y|\underline{x}_i$ is the <u>average</u> of m_i independent observations $v_1, v_2, \ldots, v_{mi}$, each with variance σ^2, then

$\mathrm{Var}(Y|\underline{x}_i)$ = ____________ = _____ , so we could take w_i = ___________.

C. Sometimes visual or other evidence suggests a pattern of how $\mathrm{Var}(Y|\underline{x}_i)$ depends on x_i. For example, if it looks like $\sqrt{Var(Y \mid x_i)}$ is a linear function of x_i [Sketch a picture of this!], then we can fit a line to the data points (x_i, s_i), where s_i = sample standard deviation of observations with x value x_i. If we get

$\hat{s}_i = \hat{\gamma}_0 + \hat{\gamma}_1 x_i$, try w_i = __________.

Caution: This involves looking at the data to decide on the supposedly "known" weights, which is iffy. A slightly better approach is to use w_i's as above, then "iterate" by using the standard errors calculate from the first WLS regression.

D. Sometimes theoretical considerations may suggest a choice of weights. For example, if theoretical considerations suggest that the conditional distributions are Poisson, then the conditional variances are equal to the conditional means. This suggests taking w_i = ______________.

E. Weighted least squares is also useful for other purposes – e.g., in calculating the lowess estimate, lines are fit so that points at the ends of the range count less than points at the middle of the range.

Fitting WLS: A WLS model may be fit by least squares: Find $\hat{\underline{\eta}}$ to minimize the "weighted residual sum of squares"

$$\mathrm{RSS}(\underline{h}) = \sum w_i (y_i - \underline{h}^T \underline{u}_i)^2$$

$\hat{\underline{\eta}}$ is called the "WLS estimate" of the coefficients.

Comments:

a. If all $w_i = 1$, we get ______________________________.
b. The larger w_i is, the more the i^{th} observation "counts" (and the __________er the variance at x_i – think of the geese example.)
c. $\mathrm{RSS}(\underline{h}) = \sum [\sqrt{w_i}\, y_i - \underline{h}^T(\sqrt{w_i}\, \underline{u}_i)]^2$, so we could get $\hat{\underline{\eta}}$ by using OLS to regress the $\sqrt{w_i}\, y_i$'s on the $\sqrt{w_i}\, \underline{u}_i$'s, *but*, we would need to fit *without* an intercept, since the first component of $\sqrt{w_i}\, \underline{u}_i$ is not 1. However, most statistics software has a routine to fit WLS directly – it will ask for weights; typically you need to have stored them as a "variable" or column.

Example: Coin data.

Residuals in WLS: Recall that the errors in WLS are $e_i = \sqrt{w_i}\,[\, Y|\underline{x}_i - \underline{\eta}^T \underline{u}_i]$. Analogously, the *residuals* are defined as $\hat{e}_i = \sqrt{w_i}\,(y_i - \hat{y}_i)$

Caution: Some software provides only the *unweighted* residuals $y_i - \hat{y}_i$; you need to multiply by the factors $\sqrt{w_i}$ in order to make residual plots (to be discussed shortly)

RSS and σ^2 estimates: RSS = $\Sigma w_i(y_i - \hat{y}_i)^2 = \Sigma \hat{e}_i^{\,2}$

$\hat{\sigma}^2$ = RSS/(n-k) (estimate of σ^2, which is *not* the variance)

Example: With the coins data, does $\hat{\sigma}^2$ seem reasonable?

Inference for WLS: Proceeds similarly to inference for ordinary least squares -- Model assumptions for inference are (1) and (2) above, plus

3) Independence of observations, and
4) Normal conditional distributions.

Cautions in WLS inference:

- Estimating variances to get weights (as in coins example) introduces more uncertainty.
- The interpretation of R^2 is questionable – some software doesn't even give it.
- Inference for means and prediction requires a weight (see pp. 209 – 210 for details)
- Is prediction appropriate for the coins example?

Diagnostics with WLS:

As for OLS, *except* use the WLS (weighted) residuals $\hat{e}_i = \sqrt{w_i}\,(y_i - \hat{y}_i)$.