Chapter 11: <u>Advanced Remedial Measures</u>

Weighted Least Squares (WLS)

• When the error variance appears nonconstant, a transformation (of *Y* and/or *X*) is a quick remedy.

• But it may not solve the problem, or it may create an inappropriate regression relationship.
• A more advanced approach is WLS regression.

• If $\mathrm{var}(Y_i) = \sigma_i^2$, then give observations with higher variance __________ weight in the regression fitting.

• For example, let and use

• But $\sigma_1^2, \ldots, \sigma_n^2$ are typically <u>unknown</u>.

Note σ_i^2 =

• Thus

• To estimate how σ_i varies a function of X_i (or of $\hat{Y}_i$), regress

Procedure for Determining Weights w_i:

(1) Regress Y against predictor variable(s) as usual (OLS).

(2) Regress absolute residuals $|e_i|$ against predictor X_j (if error variance is nonconstant as a function of X_j) or against fitted values $\hat{Y}$ (if error variance is nonconstant as a function of $\hat{Y}$).

(3)

(4)

(5)

• SAS or R will do WLS once we find the weights w_i.

<u>SAS Example</u> (Blood pressure data):

<u>Note</u>: R^2 does not have a standard interpretation with WLS.
<u>Note</u>: Standard error of b_1 decreases somewhat in the WLS regression →
<u>Note</u>: If the WLS estimates differ greatly from the OLS estimates, we may iterate this algorithm one or two times.
<u>Note</u>: In WLS, standard inferences about coefficients may not be valid for small sample sizes, when the weights are estimated from data.
<u>Note</u>: If the MSE of the WLS regression is near 1, then our estimation of the "error standard deviation" function is trustworthy.

Ridge Regression and LASSO Regression

• Ridge regression is an advanced remedy for multicollinearity.
• Idea: Instead of using our unbiased ordinary estimate ,
use a biased estimate, denoted

• Although $\underline{\mathbf{b}}_R$ is biased, it may have <u>less variance</u> (so that multicollinearity is reduced).

Procedure [typically these calculations are done on standardized (centered and scaled) regression coefficients]:
• Add a biasing constant *c* into the normal equations (see pg. 433).

• If *c* = 0, then

• As *c* increases:

• R provides automated choices for *c*, and it will perform ridge regression.

<u>R example</u> (body fat data):

Disadvantage to ridge regression: We <u>cannot</u> use ordinary inference procedures (bootstrapping can be used for inference)

• Ridge regression is an example of <u>shrinkage estimation</u>: The process will typically "shrink" the least-squares estimates toward <u>zero</u> because of the biasing constant.
• This "shrinkage" increases bias, but reduces variance.

• Ridge regression estimates may be obtained by minimizing the <u>penalized least-squares criterion</u>:

• The solution to this is the vector $\underline{b}_R$ that minimizes

The LASSO is a similar method which chooses $\underline{b}$ to minimize

• An advantage of LASSO regression (and ridge regression, to some degree) is that this constraint leads to some b_j's being set very close to zero, so LASSO can be viewed as a method of <u>variable selection</u> as well as coefficient estimation.

• Traditionally, ridge regression estimates have been easier to obtain computationally than the LASSO estimates.
• In 2000, an efficient algorithm was developed to solve for the LASSO estimates, making LASSO regression very popular.

<u>R example with LASSO regression</u> (body fat data):

<u>Robust Regression</u>

• If we have highly influential observations, we can reduce their impact on the regression equation (without discarding them entirely) using <u>robust</u> regression methods.

• Similarly, robust regression is effective when the error distribution is not normal, but rather <u>heavy-tailed</u>.

• M-estimation is a general class of estimation methods.
• We choose

<u>Note</u>: (1) If $p(u) = u^2$, then this is

(2) If $p(u) = |u|$, then the criterion is

• This method is called Least Absolute Residuals (LAR) regression, also called L_1 regression.
• It uses absolute residuals rather than squared residuals, so the effect of outliers is not as great.
<u>Note</u>: Residuals from LAR regression might not sum to zero.

(3) Huber's method uses a $p(\cdot)$ function that is a compromise between least-squares and LAR regression:

<u>R example</u> (math proficiency data):

<u>Note</u>: Inference on the regression coefficients is more complex for robust regression.
• For large samples, the robust estimators are approximately normal, so we can perform <u>approximate</u> CIs and tests about the coefficients.

<u>Inference Using the Bootstrap Method</u>

• We have seen how to perform inference (CIs, tests) with the general linear model with normal errors.

• Bootstrapping is a general method of inference that can often be used in nonstandard situations when our usual inferential methods are not valid.

<u>Examples</u>: We can evaluate the precision of estimates such as estimated coefficients and fitted values in:
• Weighted least squares
• Ridge and LASSO regression
• Robust regression

General Procedure:
(1) We select a random sample (of size *n*), <u>with replacement</u>, from the observations in the original sample.
• This is called a <u>bootstrap sample</u>.
• This bootstrap sample will likely contain some duplicate values from the original data, and some original data will be omitted in the bootstrap sample.
(2) We perform the original regression procedure on the bootstrap sample, and obtain the estimate of interest, say, $b_1^{*[1]}$**.**
(3) We repeat the sampling with replacement a large number (say, *B*) of times, and for each new bootstrap sample, we obtain the estimate of interest, so that we have a collection of bootstrap estimates, e.g., $b_1^{*[1]}, \ldots, b_1^{*[B]}$**.**
(4) The estimated standard deviation of these bootstrap estimates $b_1^{*[1]}, \ldots, b_1^{*[B]}$ **is an estimate of the standard error of the original estimator** b_1 **itself.**

Two Types of Bootstrap Sampling in Regression

• "<u>Fixed *X* resampling</u>" which is used when:
(1)

(2)

AND (3)

• With fixed *X* resampling, we fit the original regression and sample the residuals $e_1, \ldots, e_n$, with replacement, to obtain the bootstrap sample of *n* residuals $e_1^*, \ldots, e_n^*$.
• Then the bootstrap sample of response values is

• Then we regress the Y_i^* values against the original X_i values to obtain the bootstrap estimate, say, b_1^*.
• This is done *B* times, so that we obtain $b_1^{*[1]}, \ldots, b_1^{*[B]}$.

• "<u>Random *X* resampling</u>" which is used when:
(1)

(2)

OR (3)

• With random *X* resampling, we sample the data pairs (X_i, Y_i) with replacement, so that we obtain a bootstrap sample of *n* data pairs (X_i^*, Y_i^*).
• Then we regress the Y_i^* values against the X_i^* values to obtain the bootstrap estimate, say, b_1^*.
• This is done *B* times, so that we obtain $b_1^{*[1]}, \ldots, b_1^{*[B]}$.

Bootstrap Confidence Intervals

• Bootstrap CIs are based on the empirical distribution of b_1*.

• The percentile method to obtain a 100(1 – α)% bootstrap CI for, say, β_1 is to use the interval (L, U), where

• The reflection method to obtain a 100(1 – α)% bootstrap CI for, say, β_1 is

• The above methods tend to produce similar results.

• It is recommended to let B be at least 500 when constructing bootstrap CIs (often 1000 resamples are used).

Examples in R (Toluca data and blood pressure data):