NBER TECHNICAL WORKING PAPERS SERIES

HETEROSCEDASTICITY DIAGNOSTICS BASED ON "CORRECTED" STANDARD ERRORS

Edward E. Leamer

Technical Working Paper No. 94

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 1991

NBER Technical Working Paper #94
January 1991

HETEROSCEDASTICITY DIAGNOSTICS BASED ON "CORRECTED" STANDARD ERRORS

ABSTRACT

Weights are found for weighted least squares estimates such that a selected coefficient (a) changes by one standard deviation or (b) changes in sign. The length of the vector of weight changes is equal to the usual OLS standard error divided by the White-corrected standard errors. Thus the White-corrected standard errors can help decide if it is necessary to adjust the location of the confidence sets to correct for heteroscedasticity. The vector of weight changes is similar to the effect of omitting observations, one at a time. The sensitivity diagnostics of Belsley, Kuh and Welsch are therefore linked with heteroscedasticity issues.

Edward E. Leamer
Department of Economics
University of California
405 Hilgard Avenue
Los Angeles, CA 90024

/DIAGNOSTICS/HETERO/HETERO.DOC November 6, 1990

Heteroscedasticity Diagnostics
Based on "Corrected" Standard Errors

by

Edward E. Leamer *
Department of Economics
UCLA
Los Angeles, CA 90024

Telephone (213) 825-3557
Messages: (213) 825-2443
FAX (213) 825-9528
BITNET: RELEAMR@UCLASSCF

1. Introduction

If the (n×1) data vector y is drawn from a normal distribution with mean $\mathbf{X}\beta$ and covariance $\mathbf{W}^{-1}$, where β is an unknown (k×1) parameter vector and $\mathbf{W}$ is a known (n×n) precision matrix, then the generalized least squares estimator is $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}y$ which has variance $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$. Rarely, however, will the precision matrix $\mathbf{W}$ be known. When $\mathbf{W}$ is unknown but probably close to an identity matrix, then it is common practice to use the least squares estimator $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y$. Though this estimator has variance $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$, most practitioners report instead the ordinary least squares estimated variance, $s^2(\mathbf{X}'\mathbf{X})^{-1}$, where s^2 is the sum of squared residuals divided by the degrees of freedom. But Eicker(1967) and White(1980) show that, if $\mathbf{W}$ is diagonal but unknown, and $\mathbf{E}$ is a diagonal matrix with the least squares residuals on the diagonal, then the matrix $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ is a consistent estimate of the least-squares covariance matrix.

White's article has had great impact and it is now quite common to report ordinary least squares estimates together with "White-corrected"

* Support from NSF grant SES 8910950 is gratefully acknowledged. Computations done by Kishore Gawande.

standard errors. But this practice is disturbing because it is based on the idea that heteroscedasticity requires an adjustment to the length but not the location of a confidence set. This is correct asymptotically, as White has shown, but in small samples the location of the confidence sets may sensibly require substantial adjustment.

Unfortunately, adjustment of the location of the confidence sets seems to require a traditional approach in which the form of the heteroscedasticity is modelled in terms of a fixed set of parameters. If one were to approach the choice of estimates without commitment to the form of heteroscedasticity, one might try iterating the equations, $\mathbf{X}'\mathbf{E}^{-2}\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{E}^{-2}\mathbf{y}$, $\mathbf{e} = \mathbf{y}-\mathbf{X}\mathbf{b}$, $\mathbf{E} = \mathrm{diag}(e_1, e_2, \ldots, e_n)$ that jointly determine the estimates and the weights. But the fixed points of this iteration occur when all the weight ($e_i=0$) is assigned to k arbitrarily selected observations, and the estimate b is selected to fit these k observations exactly. There are n-choose-k such solutions. And the likelihood function has essential singularities at each of these points. This unappealing treatment of the problem of heteroscedasticity arises precisely because the variances are allowed to be anything. A sensible treatment must therefore impose some limits on the weight matrix **W**. A stochastic or deterministic model would serve this function.

In my opinion, the availability of White-corrected standard errors has had an undesirable affect on real data analyses because ,perhaps unintentionally, the use of the corrected standard errors has discouraged treatments that alter the location as well as the size of the confidence sets. But I report in this paper that the corrected standard errors may be used to indicate when changing the weights can alter substantially the <u>location</u> of the confidence sets. Namely, if there is an <u>increase</u> in the standard error after correction, then we can expect that there would likewise be a <u>large</u> change in the location of the confidence sets. If, on the other hand, the White-correction leaves unchanged or <u>reduces</u> the standard errors, then the estimates are likely to be very <u>insensitive</u> to reweighting.

In particular, the vector

$$\Delta_s\text{Weights}(c) = c'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}\ s(c'(\mathbf{X}'\mathbf{X})^{-1}c)^{1/2}/\ (c'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}^2\mathbf{X}\ (\mathbf{X}'\mathbf{X})^{-1}c),$$

is shown to be the change in the diagonal values of the weight matrix $\mathbf{W}$ from the initial values of $\mathbf{W} = I$ that is just enough to change the estimate of $c'\beta$ by one standard deviation. (The subscript s on Δ_s is a reference to the standard error.) The length of this vector is the ratio of the OLS standard error divided by the White corrected standard error:

$$||\Delta_s\text{Weights}(c)|| = s\ (c'(\mathbf{X}'\mathbf{X})^{-1}c)^{1/2}\ /\ (c'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}^2\mathbf{X}\ (\mathbf{X}'\mathbf{X})^{-1}c)^{1/2}$$

$$= \text{S.E.Ratio}(c)$$

where s^2 is the OLS residual sum of squares divided by the degrees of freedom. A "large" value of this diagnostic would indicate that the issue is very insensitive to choice of weights since "large" changes in the least-squares weights from their initial value of one are required to alter the estimate by one standard deviation. The meaning of "small" and "large" is further discussed below. Until then, quotations will surround these vague words.

Another diagnostic indicates the change in the vector of least-squares weights that is enough to change the <u>sign</u> of the estimate. This vector of weight changes is equal to the product of the t-statistic times the vector of weight changes necessary to alter the estimate by one standard deviation:

$$\Delta_b\text{Weights}(c) = |t(c)|\ \Delta_s\text{Weights}(c)$$

$$||\Delta_b\text{Weights}|| = |t(c)|\ \text{S.E.Ratio}(c)$$

where $t(c) = c'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y/s\ (c'(\mathbf{X}'\mathbf{X})^{-1}c)^{1/2}$ is the t-statistic for testing $c'\beta = 0$. (The subscript b on Δ_b is a reference to the least-squares estimate b, signalling that here we are perturbing the estimate by enough to change its sign.)

These diagnostics are different for each coefficient or linear combination, but the continuum of possible diagnostics can be summarized in terms of the minimum and maximum. Letting the eigenvalues λ_i be roots of the characteristic equation $|s^2\mathbf{X}'\mathbf{X} - \lambda\mathbf{X}'E^2\mathbf{X}|$ where s is the unbiased estimator of the residual variance, a pair of general diagnostics are the smallest and largest roots:

$$\inf_{c}\ \text{S.E.Ratio}(c) = (\lambda_{min})^{1/2} \qquad \sup_{c}\ \text{S.E.Ratio}(c) = (\lambda_{max})^{1/2}$$

The statistic $(\lambda_{min})^{1/2}$ indicates the change in the weights that is enough to change *some* coefficient by one standard deviation. The statistic $(\lambda_{max})^{1/2}$ indicates the change in the weights that is enough to change *any* linear combination of coefficients by one standard deviation. A problem is judged to be insensitive to heteroscedasticity problems if $(\lambda_{min})^{1/2}$ is "large". A problem is judged to be very sensitive to heteroscedasticity problems if $(\lambda_{max})^{1/2}$ is "small".

The extreme values of the second diagnostic indicating the change in the weights necessary to alter the sign of the coefficient are shown to be

$$\inf_{c}\ |t(c)|\ \text{S.E.Ratio}(c) = 0 \qquad \sup_{c}\ |t(c)|\ \text{S.E.Ratio}(c) = (\mathbf{y}'\mathbf{E}^{-2}\mathbf{y})^{1/2}$$

The infimum is zero since there is always some linear combination with an estimate $c'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y$ which is so close to zero that the slightest change in the weights will reverse its sign. The supremum is then the interesting result. A "small" value of the supremum indicates that all estimates are sensitive to the choice of weights, since "small" changes in the weights can be found to alter the signs of any linear combination. A "large" value for the supremum indicates that there are some linear combinations that are resistent to reweighting.

The vagueness in the terms "small" and "large" will have to be eliminated if these diagnostics are to have practical value. I regard changing each of the weights by ten percent to be a small change, one that seems easily justified by formal models of heteroscedasticity. If each weight changes by ten percent, the length of the weight change vector is $(\sum(.1)^2)^{1/2} = .1*n^{1/2}$. This measure of smallness increases with sample size, which may seem unnatural, but another way to think about it is to suppose that the weights come from a distribution with central tendency equal to one and second moment $E(w_i-1)^2 = v$. Then the expected sum is $E(\sum_i(w_i-1)^2) = nv$ with square root equal to $n^{1/2}\, v^{1/2}$, which also grows like $n^{1/2}$. A change in one of the weights as big as .9 seems to me to be large since then the relative weights can be as extreme as $1.9/.1 = 19$. For these reason, I will regard any vector of weight changes that is smaller than $.1\, n^{1/2}$ to be "small, and any vector larger than $.9\, n^{1/2}$ to be "large". Classification of numbers between these extremes is open for discussion.

These diagnostics indicate the local sensitivity of an estimate to changes in assumptions about the heteroscedasticity weights. Global sensitivity results for the same problem are reported in Gilstein and Leamer(1983) who provide algorithms for characterizing the full set of weighted regressions. Global sensitivity results are also reported in

Belsley, Kuh and Welsch(1980) who suggest studying the changes in the least-squares estimates induced by the omission of observations one-at-a-time. This seems like an unusual and unlikely form of reweighting since zero weight is put on one observation and equal weights on all the others. The BKW approach does usefully detect "lonely" outliers but not outliers in bunches. Multiple deletion is computationally burdensome, and Belsley, Kuh and Welsch(1980) suggest instead a stepwise approach which doesn't seem likely to deal fully with the problems of clusters of outliers. The local sensitivity analysis discussed here seems automatically to identify sets of influential observations since a vector of changes in the weights is selected that most affects the estimate. But this comparison with BKW is misleading since this local sensitivity analysis is built on the approximation that the weighted least squares estimate is linear in the weights. Part of the problem of multiple deletions is a form of nonlinearity. This point is further discussed in Section 3 which makes some additional comparisons of BKW and my results.

The final section contains an example built on a model and data taken from BKW(1980). The intent of the example is to demonstrate the value of these diagnostics and also suggest an informative and economical style of reporting that includes S.E.Ratio(c)/$n^{1/2}$ and Δ_sWeights(c) for selected linear combinations $c'\beta$, and also the extreme values of S.E.Ratio(c)/$n^{1/2}$ and S.E.Ratio(c)t(c)/$n^{1/2}$.

2. Derivations

The generalized least-squares estimator is a solution to the normal equations

$$X'WX\ b = X'Wy$$

where **W** is the inverse of the error covariance matrix which in the heteroscedastic case is assumed to be diagonal

$$W = \mathrm{diag}(w_1,\ w_2,\ \ldots,\ w_n).$$

Differentiation of this equation produces

$$X'WX\ db + X'(dW)X\ b - X'(dW)y = 0$$

or equivalently

$$\mathbf{X'WX}\ d\mathbf{b} = \mathbf{X'}(d\mathbf{W})\ \mathbf{e} \tag{1}$$

where **e** is the vector of residuals

$$\mathbf{e} = y - \mathbf{X}\ b$$

Let E be the diagonal matrix with the residuals on the diagonal and zeroes elsewhere:

$$E = \mathrm{diag}(e_1,\ e_2,\ \ldots,\ e_n).$$

and let

$$d\mathbf{w} = (dw_1,\ dw_2, \ldots\ dw_2)'.$$

Then $(d\mathbf{W})\ \mathbf{e} = \mathbf{E}\ d\mathbf{w}$ and if (1) is evaluated at the unweighted least squares estimate, $\mathbf{W} = \mathrm{I}$, it can be written as $d\mathbf{b} = (\mathbf{X'X})^{-1}\ \mathbf{X'E}\ d\mathbf{w}$. This matrix of derivatives can be collapsed into a vector if interest focuses on the linear combination **c'b**. Then the derivative of the estimate of this linear combination with respect to changes in the weights is

$$\mathbf{c'}d\mathbf{b} = \mathbf{c'}(\mathbf{X'X})^{-1}\ \mathbf{X'E}\ d\mathbf{w} \tag{2}$$

This vector of derivatives can be collapsed into a scalar by selecting a direction $d\mathbf{w}$ in which to perturb the weights. A worst-case direction can be found by maximizing $c'db$ subject to the constraint

$d\mathbf{w}'d\mathbf{w} \leq (d\omega)^2$ where $d\omega$ is an infinitesimal. The Lagrangian equation for this extreme value problem is $\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E} - \lambda d\mathbf{w}' = 0'$ where λ is the Lagrange multiplier. Thus $d\mathbf{w}' = \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}/\lambda$ and $\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}/\lambda^2 = (d\omega)^2$, from which we can solve for the Lagrange multipliers, $\lambda = \pm(\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c})^{1/2}/d\omega$. Using these values of the multiplier, we have the weight vector and derivative of the estimate as

$$d\mathbf{w}' = \pm\ \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}\ (d\omega)/\ (\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c})^{1/2} \tag{3}$$

$$\mathbf{c}'d\mathbf{b} = \pm(\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}^2\mathbf{X}\ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{c})^{1/2}\ (d\omega) \tag{4}$$

The vector of weight-changes (3) can be used to point to specific observations that deserve closer scrutiny since changes in their weights can have a substantial affect of the estimate of the issue. The weight-change vector is defined up to the factor of proportionality $d\omega$ and the vector thus has to be scaled to make it comparable across issues and problems. One interesting scaling selects the minimal weight-change vector that is necessary to induce a one-standard deviation change in the issue. Using a linear approximation which is signaled by replacing the differential d with the discrete change notation Δ, the value of $\Delta\omega$ that is needed to induce a one-standard deviation change in the estimate is the solution to $\mathbf{c}'\Delta\mathbf{b} = (\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}^2\mathbf{X}\ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{c})^{1/2}\ (\Delta\omega)$ $= s\ (\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c})^{1/2}$. Using this value of $\Delta\omega$ we can solve for the vector of weight changes and its length:

$$\Delta_s\text{Weights}(\mathbf{c}) = \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}\ s(\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c})^{1/2}/\ (\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}^2\mathbf{X}\ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}) \tag{5}$$

$$\text{S.E.Ratio}(\mathbf{c}) = ||\Delta_s\text{Weights}(\mathbf{c})||$$

$$= s(\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c})^{1/2}/(\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}^2\mathbf{X}\ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{c})^{1/2} \tag{6}$$

The vector Δ_sWeights(c) is the approximate change in the weights from their initial unit values that is necessary to induce a one-standard deviation change in the estimate of the parameter $c'\beta$. The statistic S.E.Ratio(c) is the length of Δ_sWeights(c). The length S.E.Ratio(c) is equal to the ratio of the usual estimate of the standard error divided by the "White-corrected" estimate of the standard error. If the White corrected standard error is large compared with the uncorrected standard error then a small change in the weights can make a big difference in the estimate. But if the White corrected standard error is relatively small, then the estimate is insensitive to reweighting of observations.

An alternative choice of the weight change vector is a set of values for which the coefficient changes sign: $c'\Delta b = (c'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}^2\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}c)^{1/2}\ (\Delta\omega) = |c'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y|$. The corresponding diagnostics are

$$\Delta_b\text{Weights}(c) = \Delta_s\text{Weights}(c)\ |t(c)|$$

$$||\Delta_b\text{Weights}|| = \text{S.E.Ratio}(c)\ |t(c)|$$

where $t(c) = c'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y/s\ (c'(\mathbf{X}'\mathbf{X})^{-1}c)^{1/2}$ is the t-statistic for testing $c'\beta = 0$ and where the subscript b on Δ_b refers to the OLS estimate **b**. These statistics can be found from (5) and (6) simply by dividing by the t-values. An economical reporting scheme thus includes Δ_sWeights(c) and S.E.Ratio(c) together with the t-value t(c).

These diagnostics might be reported for selected issues c but they cannot be reported for all linear combinations. A way to deal with the dependence on c is to find the issue that is most sensitive to heteroscedasticity problems and the issue that is least sensitive. Note that by change of variable $c = \mathbf{X}'\mathbf{X}f$ the diagnostic is S.E.Ratio($\mathbf{X}'\mathbf{X}f$) = $s\ (f'\mathbf{X}'\mathbf{X}f)^{1/2}\ /\ (f'\mathbf{X}'\mathbf{E}^2\mathbf{X}\ f)^{1/2}$. Then letting the eigenvalues λ_i be roots of the characteristic equation $|s^2\mathbf{X}'\mathbf{X} - \lambda\mathbf{X}'\mathbf{E}^2\mathbf{X}|$ we have from Rao(1965, p. 50):

$$\inf_c \text{S.E.Ratio}(c) = (\lambda_{min})^{1/2} \qquad \sup_c \text{S.E.Ratio}(c) = (\lambda_{max})^{1/2}$$

One general diagnostic is thus $(\lambda_{min})^{1/2}$ which indicates the change in the weights that is enough to change <u>some</u> coefficient by one standard deviation. The statistic $(\lambda_{max})^{1/2}$ is also of interest since it indicates the change in the weights that is enough to change <u>any</u> linear combination of coefficients by one standard deviation. A problem is judged to be insensitive to heteroscedasticity problems if $(\lambda_{min})^{1/2}$ is large since this indicates that a large change in weights to change an estimate by one standard deviation. A problem is judged to be very sensitive to heteroscedasticity problems if $(\lambda_{max})^{1/2}$ is small since this indicates that only a small change in the weights is enough to change any estimate by one standard deviation.

It is possible also to find extreme values of the diagnostic indicating the length of the weight-change vector that is necessary to change the sign of the coefficient: $||\Delta_b\text{Weights}(c)||$ $= ||\Delta_s\text{Weights}(c)||\ |t(c)| = c'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y \,/\, (c'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}^2\mathbf{X}\ (\mathbf{X}'\mathbf{X})^{-1}c)^{1/2} =$ $[f'\mathbf{X}'yy'\mathbf{X}f/f'\mathbf{X}'\mathbf{E}^2\mathbf{X}f]^{1/2}$ where $f' = c'(\mathbf{X}'\mathbf{X})^{-1}$. The extreme values of this ratio are the minimimum and maximum roots of the characteristic equation $|\mathbf{X}'yy'\mathbf{X} - \lambda\ \mathbf{X}'\mathbf{E}^2\mathbf{X}| = 0$. Using the formula $|\mathbf{A} - \mathbf{r}\mathbf{r}'| = |\mathbf{A}|(1-\mathbf{r}'\mathbf{A}^{-1}\mathbf{r})$ this characteristic equation can be expressed as $|\mathbf{X}'yy'\mathbf{X} - \lambda\ \mathbf{X}'\mathbf{E}^2\mathbf{X}|$ $= |\lambda\mathbf{X}'\mathbf{E}^2\mathbf{X}|(1- y'\mathbf{X}(\mathbf{X}'\mathbf{E}^2\mathbf{X})^{-1}\mathbf{X}'y/\lambda| = \lambda^{n-1}\ |\mathbf{X}'\mathbf{E}^2\mathbf{X}|(\lambda - y'\mathbf{X}(\mathbf{X}'\mathbf{E}^2\mathbf{X})^{-1}\mathbf{X}'y|$ which has minimum root 0 and maximum root $y'\mathbf{X}(\mathbf{X}'\mathbf{E}^2\mathbf{X})^{-1}\mathbf{X}'y$. Thus

$$\inf_c \text{S.E.Ratio}(c)t(c) = 0 \qquad \sup_c \text{S.E.Ratio}(c)\ t(c) = (y'\mathbf{X}(\mathbf{X}'\mathbf{E}^2\mathbf{X})^{-1}\mathbf{X}'y)^{1/2}$$

The infimum is zero since there is always some linear combination with an estimate $c'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y$ which is so close to zero that the slightest change in the weights will reverse its sign. The supremum is then the interesting result. It addresses the following problem: Find the linear combination that is most resistent to change as the weights are altered. By how much do the weights have to change in order for this linear combination to change in sign. It perhaps is wise to repeat again that this answer is only an approximation. As a matter of fact, the set of all weighted regressions as characterized by Gilstein and Leamer(1983) may not include the origin in which case there are linear combinations that cannot change their signs. In that event the correct solution should be infinity not $(y'\mathbf{X}(\mathbf{X}'E^2\mathbf{X})^{-1}\mathbf{X}'y)^{1/2}$. My recommendation therefore is not to take this number literally, especially when it is large. When $(y'\mathbf{X}(\mathbf{X}'E^2\mathbf{X})^{-1}\mathbf{X}'y)^{1/2}$ is small, all estimates are sensitive to the choice of weights, since fairly small changes in the weights can be found to alter the signs of any linear combination. When $(y'\mathbf{X}(\mathbf{X}'E^2\mathbf{X})^{-1}\mathbf{X}'y)^{1/2}$ is large there are some linear combinations that are resistent to reweighting.

Incidentally, the supremum $(y'\mathbf{X}(\mathbf{X}'E^2\mathbf{X})^{-1}\mathbf{X}'y)^{1/2}$ will necessarily grow with sample size which can be interpreted to mean that as sample size grows, there will always be some estimates that are difficult to change in sign by reweighting the observations, but the standard against which this is compared increases with $n^{1/2}$. Thus the sample-size corrected diagnostic is $(y'\mathbf{X}(\mathbf{X}'E^2\mathbf{X})^{-1}\mathbf{X}'y)^{1/2}/n^{1/2}$ which does not diverge.

Last in this section I make some remarks about the linear approximation that is implicit in many of these equations and I show how some aspects of the approximation can be completely overcome. The weighted least squares estimate that is considered here takes the form $\mathbf{b} = (\mathbf{X}'(I+\rho\Delta)^{-1}\mathbf{X})^{-1}\mathbf{X}'(I+\rho\Delta)^{-1}y$ where $\Delta = \text{diag}(\Delta w_1, \Delta w_2, \Delta w_n)$ and the scalar ρ is selected to make $c'\mathbf{b}$ approximately zero or make $c'\mathbf{b}$ and

$c'(X'X)^{-1}X'y$ differ approximately by one standard deviation. It is possible without undue effort to make these statments precise. Leamer(1984) has shown that this generalized least squares estimate can be written as

$$b = (X'(I+\rho\Delta)^{-1}X)^{-1}X'(I+\rho\Delta)^{-1}y \quad = (X'X)^{-1}X'(y - \mathbf{g})$$

where $\mathbf{g} = (\mathbf{M} + \rho^{-1}\Delta^{-1})^{-1}\mathbf{M}y$ where $\mathbf{M}$ is the usual idempotent matrix $\mathbf{M} = I - X(X'X)^{-1}X'$ and $\mathbf{M}y$ is the vector of least squares residuals. In words, b is an OLS estimate using "corrected data" $(y-\mathbf{g})$, where the correction $\mathbf{g}$ is the product of a matrix that depends on the matrix of weight changes $\rho\Delta$ times the OLS residuals $\mathbf{My}$. Using this result we have

$$c'b - c'(X'X)^{-1}X'y = -c'(X'X)^{-1}X'(M + \rho^{-1}\Delta^{-1})^{-1}\mathbf{My}$$

This formula is made more clear if the matrices $\mathbf{M}$ and Δ^{-1} are jointly diagonalized: $G'MG = D$ = diagonal matrix and $G'\Delta^{-1}G = I$. Then

$$c'b - c'(X'X)^{-1}X'y = -c'(X'X)^{-1}X'G(G'MG + \rho^{-1}G'\Delta^{-1}G)^{-1}G'\mathbf{My}$$

$$= c^{*\prime}(D + \rho^{-1}I)^{-1}y^{*} = \sum c^{*}_{i}y^{*}_{i}/(d_{i}+\rho^{-1})$$

which is a ratio of polynomialsin ρ. It is not difficult to find zeroes of the polynomial in the denominator and in that way to find exactly the matrices $\rho\Delta$ that are required to change an estimate by one standard deviation or to set an estimate to zero. For the example presented in section 4 the approximation is quite good and this somewhat more burdensome calculation seems unnecessary.

3. Relationship with Belsley, Kuh and Welsch

There are several vectors that could select observations for special scrutiny. One is the derivative (2) of the estimate with respect to the weights. Another is the vector of weight-changes (5) that are enough to alter the estimate by one standard deviation. These two vectors are proportional to each other if, as is assumed above, the vector of weight-changes is constrained to lie within a circle. A third, and different formula proposed by Belsley, Kuh and Welsch(1980) is the change in the coefficient induced by the omission of a single observation, BKW(1980, p.13). This differs from (2) and (5) by the multiplication of a positive diagonal matrix. In my notation this coefficient change is

$$\mathrm{DFBETA}(c) = c'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}\mathbf{D}^{-1} \qquad (7)$$

where $\mathbf{D} = \mathrm{diagonal}(\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$. BKW(1980, p.24) also report the derivative (2) and make the following comment:

> "[The derivative (2)]...is often viewed as the influence of the ith observation on the estimated coefficients. Its relationship to the formula (7) for DFBETA is obvious and it could be used as an alternative to that statistic."

In their example which is reanalyzed in the next section, BKW report the values of DFBETA scaled by the standard error of the estimate based on the data set with one observation omitted. This scaling is computationally burdensome and in the next section I will report instead the value of DFBETA scaled by the standard error estimated with all the data:

$$\mathrm{DFBETAS}^{*}(c) = \mathrm{DFBETA}(c) \,/\, s\,(c'(X'X)^{-1}c)^{1/2}$$

The vectors Δ_sWeights(c) and DFBETAS*(c) are different both in length and in direction. The directions in which these two vectors point may not be very different since the diagonal matrix $D = \text{diagonal}(I-X(X'X)^{-1}X')$ which links the two formula may be pretty close to the identity matrix. When these two vectors point toward different influential observations, I tend to prefer the question that underlies Δ_sWeights(c): "If you reweight the observations to alter the estimate of $c'\beta$ by as much as possible, which weights change the most?" The question implicit in DFBETAS* is: "Which observations, if deleted from the sample, would have the greatest affect on the estimate of $c'\beta$?" This question seems somewhat less interesting since I am unlikely to omit completely an observation. But the one clear advantage of DFBETAS* is that it is an exact answer to its question, whereas Δ_sWeights(c) is only an approximate answer to its question. Furthermore, the question that Δ_sWeights(c) answers has to be made more specific by defining the length of a vector which is here taken to be Euclidean. This choice of a circle within which to vary the weights makes an implicit reference to the spherical symmetry of the prior distribution for the inverses of the residual variances. The traditional choice of prior would be a gamma distribution, which has level curves $a = \sum_i \log w_i - \gamma\sum w_i$. The set $\sum_i \log w_i - \gamma\sum w_i \geq r^2$ can be far from spherical. Furthermore, the perturbation of weights within a sphere makes no reference to data information which might suggest that some perturbations are more likely than others. Weights that are selected when heteroscedasticity is modelled will necessarily be influenced by the data.

4.0 An Example

Belsley, Kuh and Welsch(1980, pp 40-63) report a data analysis that explains the variation in the personal savings rates for 50 countries in terms of differences in demographics and income. The equation that they estimate is

$$SR = \beta_1 + \beta_2 POP15_i + \beta_3 POP75_i + \beta_4 DPI_i + \beta_5 \Delta DPI_i + \epsilon_i, \quad i = 1,\ldots,50$$

where

SR = the average aggregate personal savings rate over the period 1960-1970 (per cent).

POP15 = the average percentage of the population under 15 years of age over the period 1960-1970.

POP75 = the average percentage of the population over 75 years of age over the period 1960-1970.

DPI = the average level of real per-capita disposable income over the period 1960-1970 (thousands of U.S. dollars).[2]

ΔDPI = the average percentage growth rate of DPI over the period 1960-1970. (percentage)

Estimates of this model and associated diagnostics are reported in Table 1. The least-squares estimates are somewhat different from BKW, presumably due to slight differences in the data sets caused by misreporting by BKW or misrecording by my research assistant. These differences are highly unlikely to affect the following conclusions substantially.

In the middle of Table 1 are reported the values of BKW's DFBETAS*, indicating the change in the estimate if an observation is omitted, scaled by the standard error of the estimate. For example, the estimate

[2] BKW use dollars, not thousands of dollars. The units here make the tables a bit more readable.

of the intercept is reduced by 1.18 standard errors if Korea is omitted. The panel above indicates Δ_sWeight, the change in the weights that is necessary to change the estimates by one standard error. For example, the estimate of the intercept will change by one standard error if the Korean weight is reduced by .61 at the same time that the Costa Rican weight is reduced by .11 and the Japanese weight is increased by .34, and.... The intent of these two arrays of numbers is to point to observations that are especially influential in determining the estimates. For that purpose, the two arrays of numbers are virtually indistinguishable.

Below the DFBETAS are reported the length of the ΔWeights vector divided by the square root of the sample size. As we have indicated above, the length of the ΔWeights vector is the uncorrected standard error divided by the White-corrected standard error. The reported numbers range from .12 up to .24. These numbers are small enough to make one concerned about heteroscedasticity since changes in the weights by between 12 and 15 per cent on the average are enough to alter the estimates by one standard deviation. What I mean by this statement is that it seem probable to me that I could come up with a sensible model of heteroscedasticity that would lead me to changing the weights by .12 on the average. The row below indicates the average change in the weights that is necessary to change the sign of a coefficient. These are just equal to the preceeding numbers times the t-values. It appears rather difficult to change the sign of the intercept and also the coefficients on ΔDPI and POP15. But a little bit of change in the weights can change the sign of DPI.

Finally, in the last panel of the table are reported the extreme values of these statistics over all possible parameters or combinations thereof. The minimum of the first diagnostic referrring to changing coefficients by one standard deviation is .10 and the maximum is .268. The maximum is small enough to suggest that there is hardly any coefficient that is insensitive to heteroscedasticity if we are worried by changes in the coefficients equal to the standard error. But the minimum of .1 does indicate that it would take at least a ten per cent change in the weights on the average to change any coefficient by one standard deviation. If you regard that much change to be implausible, then you are free to conclude that heteroscecasticity is unimportant. The range of the other statistic is much wider, from zero to 3.25. The upper number indicates that there is some linear combination of coefficients that is very difficult to change in sign, requiring average changes in the weights equal to 325 per cent.

Finally, it seems wise to check the assumption that the weighted least squares estimates are approximately linear in the weights. In order to check the accuracy of this approximation, we first compute the weighted least squares estimates using the weights in Table 1 and verify that the selected coefficient changes by one standard error, as advertised. This calculation is reported in Table 2. The second column contains the unweighted OLS estimates. The third column contains weighted least squares estimates of the five coefficients using the five different sets of weights defined by the ΔWeights vectors that are partially reported in Table 1. The fourth column reports the difference between the OLS and the weighted least squares estimates, and the last column reports the OLS standard errors. If there were no approximation error, the last two columns would be identical. From my perspective, these columns are gratifyingly close. The approximation seems a little less accurate for the last two coefficients, which according to Table 1 do require relatively large changes in the weights. Of course, the linear approximation cannot possibly hold if one of the weights is required to take on a negative value. There are no such weights implied by the values of ΔWeights in Table 1.

More extreme weights are required to change the sign of some of the coefficients and then the linear approximation may not be as accurate. To check this I have computed the weighted least squares estimates generated when the weight vector is multiplied by the product of the absolute t-value and the numbers 1, 1.25 and 2. Note that in Table 2 the WLS estimate is always larger than the OLS estimate. This direction of change makes the coefficient smaller in absolute value only if the OLS estimate is negative. Thus the weight change vector has to be reversed in sign for the two coefficients with positive OLS estimates: Intercept and CH_DPI. When a weight change would make a weight negative, its value is set to zero, thereby obviously affecting the linear approximation. The OLS and the new weighted least-squares estimate are reported in Table 3. If the linear approximation were perfect the estimates in the column labelled WLS(t) would all be zero. These numbers are very small compared with the OLS estimates, but in four of the cases there have been no sign changes. If these weight changes are increased by the multiple 1.25, the sign changes do occur, except for the last coefficient. I find the results in this table to be further confirmation that the linear approximation is adequate.

References

Belsley, D.A., Kuh, E. and Welsch, R. (1980), Regression Diagnostics, New York: John Wiley and Sons, Inc..

Eicker, F. (1967), "Limit Theorems for Regressions with Unequal and Dependent Errors," in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, eds. L. Le Cam and J. Neyman, Berkeley: University of California Press, 59-82.

Gilstein, Z. and Leamer, E.E. (1983), "The Set of Weighted Regression Estimates," Journal of the American Statistical Association, 78, 942-948.

Leamer, Edward E. (1978), Specification Searches, New York: John Wiley and Sons, Inc..

Leamer, Edward E. (1984), "Global Sensitivity Results for Generalized Least Squares Estimates," Journal of the American Statistical Association, 79, 867-870.

Leamer, Edward E. and Leonard, H. B. (1983), "Reporting the Fragility of Regression Estimates," The Review of Economics and Statistics, LXV, 306-317.

Rao, C. Radhakrishna (1965), Linear Statistical Inference and Its Applications, New York: John Wiley and Sons, Inc..

White, Halbert (1980), "A Heteroskedasticity-consistent Covariance Matrix Estimator and A Direct Test for Heteroskedasticity," Econometrica, 48, 817-838.

Table 1

Estimates and Diagnostics
Dependent Variable: Savings Proportion
n = 50, R^2 = 0.2768

		Intercept		POP15		POP75		DPI		ΔDPI
Estimate		21.53		-0.32		-0.84		-0.19		0.42
Stnd. Error		7.00		0.14		1.06		0.97		0.20
t-value		3.08		-2.34		-0.79		-0.19		2.00
ΔWeights	Kor	-0.61	Jap	-0.36	Jap	-0.41	US	-0.68	Lib	-0.68
	CosR	-0.11	Lib	-0.19	Zam	-0.19	Swe	-0.66	Jam	-0.30
(Extremes)	Par	-0.11	Ice	-0.16	Lib	-0.15	Ire	-0.42	Per	-0.27
	Ire	-0.10	Rho	-0.13	Ice	-0.14	Ice	-0.33	Phi	-0.15
	Chl	-0.09	Tun	-0.07	Bra	-0.12	Rho	-0.31	Kor	-0.13
	Jam	0.07	Per	0.08	Fra	0.09	UK	0.35	Mal	0.17
	Ice	0.14	Ire	0.09	Phi	0.09	Bra	0.36	Ice	0.19
	Rho	0.14	Phi	0.13	Chl	0.12	Kor	0.37	Zam	0.23
	Lib	0.21	CosR	0.14	Ire	0.18	Den	0.39	Par	0.25
	Jap	0.34	Kor	0.60	Kor	0.59	Jap	0.47	Jap	0.40
DFBETAS	Kor	-1.18	Jap	-0.65	Jap	-0.69	US	-0.34	Lib	-1.28
	CosR	-0.19	Lib	-0.61	Lib	-0.45	Swe	-0.26	Jam	-0.31
(Extremes)	Ire	-0.18	Ice	-0.25	Zam	-0.28	Ire	-0.18	Per	-0.25
	Par	-0.18	Rho	-0.23	Ice	-0.21	Ice	-0.12	Kor	-0.15
	Chl	-0.15	Tun	-0.11	Bra	-0.18	Rho	-0.12	Phi	-0.14
	Jam	0.12	Per	0.13	Phi	0.13	Bra	0.13	Mal	0.16
	Ice	0.22	Ire	0.17	Fra	0.14	UK	0.13	Ice	0.18
	Rho	0.24	Phi	0.20	Chl	0.17	Den	0.14	Zam	0.21
	Jap	0.63	CosR	0.23	Ire	0.31	Kor	0.16	Par	0.23
	Lib	0.70	Kor	1.14	Kor	1.03	Jap	0.20	Jap	0.44
$(\Sigma\Delta W_i^2/n)^{1/2}$		0.12		0.12		0.12		0.24		0.15
$(t^2\Sigma\Delta W_i^2/n)^{1/2}$		0.37		0.28		0.09		0.05		0.30

$\min(\Sigma\Delta W_i^2/n)^{1/2} = 0.100$ $\max(\Sigma\Delta W_i^2/n)^{1/2} = 0.268$

$\min(t^2\Sigma\Delta W_i^2/n)^{1/2} = 0.000$ $\max(t^2\Sigma\Delta W_i^2/n)^{1/2} = 3.250$

Table 2
Weighted Least Squares Estimates

	OLS	WLS[1]	DIFFERENCE	ST.ERR.(OLS)
INTERCEP	21.531	29.00	7.469	7.000
POP15	-0.323	-0.19	0.133	0.138
POP75	-0.839	.15	0.989	1.058
DPI	-0.186	1.11	1.296	0.972
CH_DPI	0.411	.70	0.289	0.206

[1] Weighted least-squares estimates using weights defined in Table 1; different weights for each row.

Table 3
Weighted Least Squares Estimates

	OLS	WLS(t)	WLS(1.25t)	WLS(2t)
INTERCEP	21.531	3.45	-.37	-4.12
POP15	-0.323	-.04	.02	.08
POP75	-0.839	-.04	.15	.33
DPI	-0.186	.01	.06	.11
CH_DPI	0.411	.11	.05	-.01

Note:Weighted least-squares estimates using weights defined in Table 1 times the a multiple of the t-value; different weights for each row.