

Models, Testing, and Correction of Heteroskedasticity

JAMES L. POWELL
DEPARTMENT OF ECONOMICS
UNIVERSITY OF CALIFORNIA, BERKELEY

Aitken's GLS and Weighted LS

The Generalized Classical Regression Model (in Goldberger's (1990) terminology) has

$$E(\mathbf{y}|\mathbf{X}) = \mathbf{X}\boldsymbol{\beta}, \quad \mathbf{V}(\mathbf{y}|\mathbf{X}) = \boldsymbol{\Sigma},$$

where the matrix $\boldsymbol{\Sigma}$ is *not* proportional to an identity matrix. The special case of a heteroskedastic linear model assumes $\boldsymbol{\Sigma}$ is a diagonal matrix, i.e.,

$$\boldsymbol{\Sigma} = \text{diag}[\sigma_i^2]$$

for some variances $\{\sigma_i^2, i = 1, \dots, N\}$ which can vary across observations (usually as some functions of \mathbf{x}_i).

For example, in the *grouped-data regression model*

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \varepsilon_{ij}, \quad j = 1, \dots, M_i, \quad i = 1, \dots, N,$$

where only the group average values $\bar{y}_i \equiv \frac{1}{M_i} \sum_j y_{ij}$ and $\bar{\mathbf{x}}_i \equiv \frac{1}{M_i} \sum_j \mathbf{x}_{ij}$ are observed, the diagonal elements of the $\boldsymbol{\Sigma}$ matrix are of the form $\sigma_i^2 = \sigma^2/M_i$.

When the diagonal elements of $\boldsymbol{\Sigma}$ are known (as in the grouped-data regression model), we can transform the data to satisfy the conditions of the Classical Regression Model; the Classical Least Squares Estimator applied to the transformed data yields the Generalized Least Squares Estimator, which in this case reduces to Weighted Least Squares (WLS):

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{WLS} &= (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}) \\ &= \arg \min_{\mathbf{c}} (\mathbf{y} - \mathbf{X}\mathbf{c})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{c}) \\ &\equiv \arg \min_{\mathbf{c}} \sum_{i=1}^N w_i (\mathbf{y}_i - \mathbf{x}'_i \mathbf{c})^2, \end{aligned}$$

where $w_i \equiv 1/\sigma_i^2$. That is, each term in the sum of squares is weighted by the inverse of its error variance. If the covariance matrix $\boldsymbol{\Sigma}$ involves unknown parameters (aside from a constant of proportionality), then

this estimator isn't feasible; to construct a Feasible WLS estimator for β , replacing an estimator $\hat{\Sigma}$ for the unknown Σ , we need a model for the variance terms $Var(y_i) = \sigma_i^2$.

Multiplicative Heteroskedasticity Models

Virtually all of the applications of Feasible WLS assume a *multiplicative heteroskedasticity* model, in which the linear model $y_i = \mathbf{x}_i' \beta + u_i$ has error terms of the form

$$u_i \equiv c_i \varepsilon_i$$

for $\varepsilon_i \sim i.i.d.$, $E(\varepsilon_i) = 0$, $V(\varepsilon_i) = \sigma^2$. (If the errors are normally distributed given \mathbf{x}_i , then this representation is always available.) Also, it is almost always assumed that the heteroskedasticity function c_i^2 has an underlying linear (or "single index") form,

$$c_i^2 = h(\mathbf{z}_i' \theta),$$

The variables \mathbf{z}_i are some observable functions of the regressors \mathbf{x}_i (excluding a constant term); and the function $h(\cdot)$ is normalized so that $h(0) = 1$ with a derivative $h'(\cdot)$ assumed to be nonzero at zero, $h'(0) \neq 0$.

Here are some examples of models which fit into this framework.

- (1) *Random Coefficients Model*: The observable variables \mathbf{x}_i and y_i are assumed to satisfy

$$y_i = \alpha_i + \mathbf{x}_i' \beta_i,$$

where α_i and β_i are jointly *i.i.d.* and independent of x_i , with $E[\alpha_i] = \alpha$, $E[\beta_i] = \beta$, $Var(\alpha_i) = \sigma^2$, $\mathbf{V}(\beta_i) = \Gamma$, $\mathbf{C}(\alpha_i, \beta_i) = \gamma$. Defining

$$u_i = (\alpha_i - \alpha) + \mathbf{x}_i' (\beta_i - \beta),$$

the regression model can be rewritten in standard form as

$$y_i \equiv \alpha + \mathbf{x}_i' \beta + u_i,$$

with

$$\begin{aligned} E(u_i | x_i) &= 0, \\ V(u_i | x_i) &= \sigma^2 + 2\mathbf{x}_i' \gamma + \mathbf{x}_i' \Lambda \mathbf{x}_i \\ &\equiv \sigma^2 (1 + \mathbf{z}_i' \theta), \end{aligned}$$

where \mathbf{z}_i has the levels and cross-products of the components of the regression vector \mathbf{x}_i . When α_i and β_i are jointly normal, it is straightforward to write the error term u_i as a multiple of an i.i.d error term $\varepsilon_i \sim N(0, \sigma^2)$, as for the multiplicative heteroskedasticity model.

(2) *Exponential Heteroskedasticity*: Here it is simply assumed that $c_i = \exp\{\mathbf{x}'_i \boldsymbol{\theta}/2\}$, so that

$$\sigma_i^2 = \sigma^2 \exp\{\mathbf{x}'_i \boldsymbol{\theta}\},$$

with $\mathbf{z}_i \equiv \mathbf{x}_i$.

(3) *Variance Proportional to Square of Mean*: Assume

$$y_i = \alpha + \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i,$$

with $E(\varepsilon_i | \mathbf{x}_i) = 0$, $V(\varepsilon_i | \mathbf{x}_i) = \gamma^2 (\alpha + \mathbf{x}'_i \boldsymbol{\beta})^2$, so that

$$\sigma_i^2 = \sigma^2 (1 + \mathbf{x}'_i \boldsymbol{\theta})^2,$$

for $\sigma^2 \equiv \gamma^2 / \alpha^2$ and $\boldsymbol{\beta} \equiv \boldsymbol{\theta} / \alpha$ (assuming $\alpha \neq 0$).

“Squared Residual Regression” Tests for Heteroskedasticity

For these models, a diagnostic test of heteroskedasticity reduces to a test of the null hypothesis $H_0 : c_i \equiv 1 \iff H_0 : \boldsymbol{\theta} = \mathbf{0}$. Under H_0 , $E(\varepsilon_i^2) = E(u_i^2) = \sigma^2$ and $V(\varepsilon_i^2) = V(u_i^2) \equiv \tau$ for some $\tau > 0$. Thus, the null hypothesis generates a linear model for the squared error terms,

$$\varepsilon_i^2 \equiv \sigma^2 + \mathbf{z}'_i \boldsymbol{\delta} + r_i,$$

where $E(r_i | z_i) = 0$, $V(r_i | z_i) \equiv \tau$, and the true $\boldsymbol{\delta} = \mathbf{0}$ under $H_0 : \boldsymbol{\theta} = \mathbf{0}$. (A Taylor’s series expansion would suggest that $\boldsymbol{\delta} \cong h'(0) \cdot \boldsymbol{\theta}$ if $\boldsymbol{\theta} \cong \mathbf{0}$.) If ε_i^2 were observed, we could test $\boldsymbol{\delta} = \mathbf{0}$ in usual way; since it isn’t, we can use the squared values of the least squares residuals $e_i^2 \equiv (y_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}})^2$ in their place, since these are consistent estimators of the true squared errors. The resulting test, termed the “Studentized LM Test” by Koenker (1981), is a modification of the Score or Lagrange Multiplier (LM) test for heteroskedasticity proposed by Breusch and Pagan (1979). The steps to carry out this test of $H_0 : \boldsymbol{\theta} = \mathbf{0}$ are:

(1) Construct the squared least squares residuals

$$e_i^2 \equiv (y_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}})^2,$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$;

(2) Regress the squared residuals e_i^2 on 1 and \mathbf{z}_i , and obtain the R^2 from this “squared residual regression.”

(3) To test $H_0 : \boldsymbol{\theta} = \mathbf{0} = \boldsymbol{\delta}$, we can use the test statistic

$$T \equiv NR^2;$$

under H_0 , $T \xrightarrow{d} \chi^2(p)$, where $p = \dim(\boldsymbol{\delta}) = \dim(\mathbf{z}_i)$, so we would reject H_0 if T exceeds the upper critical value of a chi-squared variate with p degrees of freedom.

(3') A more familiar variant would use the “F” statistic

$$\mathcal{F} = (N - K) \frac{R^2/p}{1 - R^2} \overset{A}{\approx} F(p, N - K),$$

which should have $\mathcal{F} \cong T/p$ for large N (since $(N - p)/N \cong 1$ and, under H_0 , $1 - R^2 \cong 0$). Critical values from the F tables rather than the chi-squared tables would be appropriate here, though the results would converge to the corresponding chi-squared test results for large N .

(3'') When the error terms ε_i are assumed to be normally distributed, Breusch and Pagan (1979) showed that the Score test statistic for the null hypothesis that $\boldsymbol{\theta} = \mathbf{0}$ is of the form

$$S \equiv \frac{RSS}{2\hat{\sigma}^4},$$

where

$$RSS \equiv \sum_{i=1}^N [(\mathbf{z}_i - \bar{\mathbf{z}})' \hat{\boldsymbol{\delta}}]^2$$

is the “regression (or explained) sum of squares” from the squared residual regression and

$$\hat{\sigma}^2 \equiv \frac{1}{N} \sum_{i=1}^N e_i^2$$

is the ML estimator of σ^2 under the null of homoskedasticity. For the normal distribution, $\tau \equiv Var(\varepsilon_i^2) = 2[Var(\varepsilon_i)]^2 = 2\sigma^4$; more generally though, no such relation exists between the second moment and $\tau = E(\varepsilon_i^4) - \sigma^4$. It is straightforward to show that the Studentized LM test statistic can be written in the form

$$T \equiv \frac{RSS}{\hat{\tau}},$$

for

$$\hat{\tau} \equiv \frac{1}{N} \sum_{i=1}^N e_i^4 - \hat{\sigma}^4$$

which is the same form as the Score test with a more general estimator for $Var(\varepsilon_i^2)$.

Feasible WLS

If the null hypothesis $H_0 : \boldsymbol{\theta} = \mathbf{0}$ is rejected, a "correction for heteroskedasticity" (either by Feasible WLS or a heteroskedasticity-consistent covariance matrix estimator for LS) is needed. To do Feasible WLS, we can use fact that $E(\varepsilon_i^2 | \mathbf{x}_i) \equiv \sigma_i^2 = \sigma^2 \cdot h(\mathbf{z}'_i \boldsymbol{\theta})$, which is a nonlinear regression model for the squared error terms ε_i^2 . This proceeds in two steps:

(i) Replace ε_i^2 by $e_i^2 \cong \varepsilon_i^2$, and then estimate $\boldsymbol{\theta}$ (and σ^2) by nonlinear LS (which, in many cases, can be reduced to *linear* LS).

(ii) Do Feasible WLS using $\hat{\Omega} = \text{diag}[h(\mathbf{z}'_i \hat{\boldsymbol{\theta}})]$; that is, replace y_i and \mathbf{x}_i with

$$y_i^* \equiv y_i / \sqrt{h(\mathbf{z}'_i \hat{\boldsymbol{\theta}})}, \quad \mathbf{x}_i^* = \mathbf{x}_i \cdot [h(\mathbf{z}'_i \hat{\boldsymbol{\theta}})]^{-1/2},$$

and do LS using y_i^*, \mathbf{x}_i^* . If $\sigma_i^2 = \sigma^2 h(\mathbf{z}'_i \boldsymbol{\theta})$ is a correct specification of heteroskedasticity, the usual standard errors formulae for LS using the transformed data will be (asymptotically) correct.

Some examples of the first step for particular models are as follows:

(1) *Random Coefficients Model*: Since

$$\begin{aligned} V(\varepsilon_i | x_i) &= \sigma^2 + 2\mathbf{x}'_i \boldsymbol{\gamma} + \mathbf{x}'_i \boldsymbol{\Lambda} \mathbf{x}_i \\ &= \sigma^2 (1 + \mathbf{z}'_i \boldsymbol{\theta}) \\ &\equiv \sigma^2 + \mathbf{z}'_i \boldsymbol{\delta}, \end{aligned}$$

yields the linear regression model

$$\varepsilon_i = \sigma^2 + \mathbf{z}'_i \boldsymbol{\delta} + v_i$$

with $E[v_i | z_i] = 0$, regress e_i^2 on a constant and \mathbf{z}_i to estimate σ^2 and $\boldsymbol{\delta} = \boldsymbol{\theta} \cdot \sigma^2$.

(2) *Exponential Heteroskedasticity*: When

$$\varepsilon_i^2 = u_i^2 \exp\{\mathbf{x}'_i \boldsymbol{\theta}\},$$

we can take logarithms of both sides to obtain

$$\log(\varepsilon_i^2) = \log(u_i^2) + \mathbf{x}_i' \boldsymbol{\theta},$$

and, making the additional assumption that u_i is independent of \mathbf{x}_i (not just zero mean and constant variance), we get that $E[\log(\varepsilon_i^2)] = \alpha + \mathbf{x}_i' \boldsymbol{\theta}$, where $\alpha \equiv E[\log(u_i)]$, i.e.,

$$\log(\varepsilon_i^2) = \alpha + \mathbf{x}_i' \boldsymbol{\theta} + v_i$$

with $E(v_i | \mathbf{x}_i) \equiv 0$, so we would regress $\ln(e_i)$ on a constant and \mathbf{x}_i to get the preliminary estimator $\tilde{\boldsymbol{\theta}}$.

(3) *Variance Proportional to Square of Mean*: Here

$$\sigma_i^2 = E(\varepsilon_i^2 | \mathbf{x}_i) = \gamma^2 (\alpha + \mathbf{x}_i' \boldsymbol{\beta})^2,$$

where α and $\boldsymbol{\beta}$ are already estimated by the intercept term $\hat{\alpha}$ and slope coefficients $\hat{\boldsymbol{\beta}}$ of the classical LS estimator \mathbf{b} . Since γ is just a scaling factor common to all observations, we can take $h(\mathbf{z}_i' \hat{\boldsymbol{\theta}}) = (\hat{\alpha} + \mathbf{x}_i' \hat{\boldsymbol{\beta}})^2$.

Consistent Asymptotic Covariance Matrix Estimation

A possible problem with the use of Feasible WLS is that the form of the true heteroskedasticity may be misspecified, i.e., $\sigma_i^2 \neq h(\mathbf{z}_i' \boldsymbol{\theta})$. For example, it could be true that the original model is homoskedastic, $\sigma^2 = 1$, but a diagnostic test may have falsely rejected this null hypothesis, leading to a mistaken use of Feasible GLS. In a general heteroskedastic setting, we can calculate the covariance matrix of the WLS estimator $\hat{\boldsymbol{\beta}}_{GLS} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y$ to be

$$\mathbf{V} \equiv (\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Gamma} \mathbf{X} (\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1},$$

where

$$\boldsymbol{\Omega} \equiv \text{diag}[h(\mathbf{z}_i' \boldsymbol{\theta})]$$

$$\boldsymbol{\Gamma} \equiv \text{diag}[\sigma_i^2 / (h(\mathbf{z}_i' \boldsymbol{\theta}))^2].$$

If $\sigma_i^2 \neq \sigma^2 h(\mathbf{z}_i' \boldsymbol{\theta})$, then the feasible WLS estimator

$$\hat{\boldsymbol{\beta}}_{FWLS} = (\mathbf{X}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{y} \stackrel{A}{\approx} \mathcal{N}(\boldsymbol{\beta}, \mathbf{V})$$

in general if the original linear model for y_i is correctly specified (where $\hat{\boldsymbol{\Omega}} = \boldsymbol{\Omega}(\hat{\boldsymbol{\theta}})$), but the usual estimator of the covariance matrix \mathbf{V} (assuming a correct specification of the form of heteroskedasticity) will be inconsistent. A consistent estimator of \mathbf{V} (properly normalized) would use

$$\begin{aligned}\hat{\mathbf{\Omega}} &= \text{diag}[h(\mathbf{z}'_i\tilde{\boldsymbol{\theta}})], \\ \hat{\mathbf{\Gamma}} &= \text{diag}[\tilde{e}_i^2/(h(\mathbf{z}'_i\tilde{\boldsymbol{\theta}}))^2]\end{aligned}$$

in place of $\mathbf{\Omega}$ and $\mathbf{\Gamma}$ in the expression for \mathbf{V} above, where $\tilde{e}_i = y_i - \mathbf{x}'_i\hat{\boldsymbol{\beta}}_{FWLS}$ are the residuals from the Feasible WLS fit. The resulting estimator $\hat{\mathbf{V}}$ is known as the *Eicker-White covariance matrix estimator* (after Eicker (1967) and White (1980)), and is usually applied in the special case of no heteroskedasticity correction - that is, with $\hat{\mathbf{\Omega}} = \mathbf{I}$, so that the heteroskedasticity-consistent covariance matrix estimator for least squares is

$$\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_{LS}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\{\text{diag}[(y_i - \mathbf{x}'_i\hat{\boldsymbol{\beta}}_{LS})^2]\}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.$$

As a side note, White proposed this estimator in the context of a test for consistency of the traditional estimator $\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$ of the covariance matrix of the classical LS estimator, and showed how such a test was equivalent to the test of the null hypothesis of homoskedasticity against the alternative of a random coefficients model, as described above.

Goldfeld-Quandt Test

When the error terms can be assumed to be normally distributed, and the regressor matrix \mathbf{X} can be taken to be fixed (as in the Classical or Neoclassical Normal Regression Model), Goldfeld and Quandt (1965) proposed an *exact* test of the null hypothesis of homoskedasticity. Their tests presumes that the possible form of heteroskedasticity permits division of the sample into "high" and "low" heteroskedasticity groups (without preliminary estimates of heteroskedasticity parameters). Separate least squares fits are obtained for the two groups, and the ratio of the residual variances will have an F distribution under the null hypothesis. A one-sided test would reject if the ratio of the "high" to "low heteroskedasticity" residual variances exceeds the upper α critical value from an F table, while a two-sided test would reject if either the ratio or its inverse exceeded the upper $\alpha/2$ cutoff point (assuming an equal number of observations in each subsample). Goldfeld and Quandt suggest that the power of the test can be improved by dropping 10% to 20% of the observations with "intermediate" magnitudes of the conditional variances under the alternative.

REFERENCES

- [1] Breusch, T. and A. Pagan(1979), “A Simple Test for Heteroscedasticity and Random Coefficient Variation,” *Econometrica*, 47,1287-1294.
- [2] Eicker, F. (1967), “Limit Theorems for Regression with Unequal and Dependent Errors,” *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press.
- [3] Goldberger, A.S. (1990), *A Course In Econometrics*. Cambridge: Harvard University Press.
- [4] Goldfeld, S. and R. Quandt (1965), “Some Tests for Heteroskedasticity,” *Journal of the American Statistical Association*, 60,539-547.
- [5] Koenker, R. (1981), “A Note on Studentizing a Test for Heteroskedasticity,” *Journal of Econometrics*, 17,107-112.
- [6] White, H. (1980), “A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity,” *Econometrica*, 48,817-838.