Non-Constant Variance

Brenton Kenkel — PSCI 8357
February 11, 2016

From the amount of attention heteroskedasticity receives in graduate statistical modeling courses—including this one!—you would think it is a dire problem for statistical inference. It isn't.[1] It probably doesn't rank among the top 10. Nonetheless, your future advisors and reviewers will expect you to be familiar with heteroskedasticity and the methods for dealing with it, so today's goal is to make sure you're well equipped.

Note: Much of the material in these notes comes from Greene (2003, chaps. 10–11).

What Is Heteroskedasticity?

Heteroskedasticity is when the variance of the error term, or the residual variance, is not constant across observations. Graphically, it means the spread of points around the regression line is variable.

```
x <- rnorm(250)
y <- 1 + x + rnorm(250, sd = abs(x))
plot(x, y, main = "Heteroskedasticity")
```

---

[1] I think maybe people get fixated on it because it's got an imposing name. I mean, *heteroskedasticity*, right? You tell your drunk uncles you study public opinion, and they'll probably start talking about how Donald Trump is going to Make America Great Again. But you tell them you're working on overcoming heteroskedasticity and they'll shuffle off to another room real fast.

Heteroskedasticity
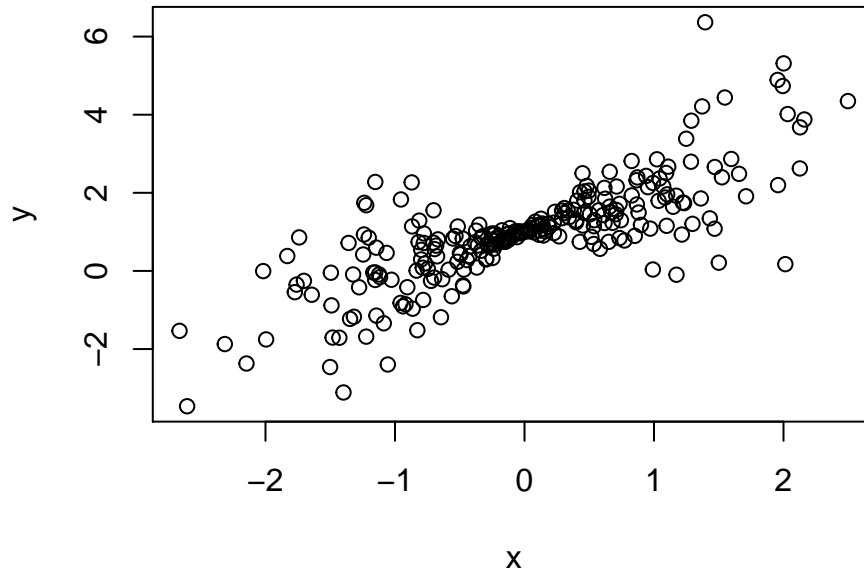
Under *homoskedasticity*, we have $V[\epsilon_i \mid x_i] = \sigma^2$, a constant, for all $i = 1, \ldots, N$. Under heteroskedasticity, this no longer holds; we have $V[\epsilon_i \mid x_i] \neq V[\epsilon_j \mid x_j]$ for some i, j. If we continue to assume that there is no autocorrelation—that the covariance of each pair of distrinct ϵ_i and ϵ_j is 0—then we can write the variance matrix of the vector ϵ as

$$
V[\epsilon \mid \mathbf{X}] = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega_N \end{bmatrix} = \sigma^2 \Omega.
$$

Why Do We Care?

You will sometimes hear it said, even by people who ought to know better, that "homoskedasticity is one of the OLS assumptions". This is nonsense on a stick. It's like saying "one of the assumptions of pants is that you have legs". Pants are pants regardless of your appendages, though it may or may not be a good idea for you to wear them. Analogously, OLS is just a statistic—a function of your

sample data. The only "assumption" is that $\mathbf{X}$ has rank p, because otherwise $\mathbf{X}^\top\mathbf{X}$ isn't invertible and the OLS estimate is ill-defined.

So is it a bad idea to use OLS in the presence of heteroskedasticity? To put it more precisely, do the OLS properties we like break down in the presence of heteroskedasticity? Let's check in on them.

- **Unbiasedness:** Nope, OLS is unbiased regardless. You can verify this yourself: we proved its unbiasedness in class, and the only thing we assumed about the errors was $E[\epsilon_i\,|\,x_i] = 0$.
- **Consistency:** Under reasonable technical conditions I won't get into here (look at Greene or any other graduate-level econometrics textbook if you're curious), OLS is still consistent in the presence of heteroskedasticity.
- **Asymptotic normality:** Ditto.
- **Efficiency:** Aha! If you carefully read the Gauss-Markov Theorem, as I'm sure you did, you remember that it depends on the assumption that $V[\epsilon_i\,|\,x_i] = \sigma^2$, a constant, for all x_i. We didn't assume this just for fun. If the errors are heteroskedastic, then there is an unbiased linear estimator with a lower variance than OLS. The problem is, to use that estimator, we must know each individual error variance up to a multiplicative constant. In other words, we must know Ω. We usually don't. So there's a more efficient estimator out there, but we're unlikely to know what it is.

In sum, the OLS estimator is unbiased, consistent, and asymptotically normal despite heteroskedasticity. It is inefficient, but the optimal estimator may be unavailable to us. It sounds like heteroskedasticity is not such a big problem. It isn't—unless we want to draw inferences and test hypotheses.

Let us briefly remember some Important Things about Statistics. The standard error is the standard deviation of the sampling distribution of a statistic. We like estimators with low standard errors because it is easier to draw inferences from them—the estimates are less prone to fluctuate due to sampling variation. But we almost never know the true standard error of an estimator, so we have to estimate it. (Yo dawg, I heard you like estimators, so. . .)

Under homoskedasticity, the variance matrix of the OLS estimator (treating the covariates $\mathbf{X}$ as fixed) is

$$\Sigma = V[\hat{\beta}_{\text{OLS}}(Y, \mathbf{X})\,|\,\mathbf{X}] = \sigma^2(\mathbf{X}^\top\mathbf{X})^{-1}.$$

The typical estimate of this variance matrix is

$$\hat{\Sigma}_{\text{OLS}} = \frac{\sum_{i=1}^{N} \hat{e}_i^2}{N - p}(\mathbf{X}^\top \mathbf{X})^{-1},$$

where $\hat{e}_i$ is the residual of the i'th observation under the OLS estimate. This is what R spits out when you run `summary(lm(y ~ ...))`. Under homoskedasticity, this is an unbiased and consistent estimator of the true variance matrix. With heteroskedasticity, however, $\hat{\Sigma}_{\text{OLS}}$ is biased and inconsistent. If we go ahead and use it to calculate inferential statistics, our measures of uncertainty will be misleading. Typically, we will too readily reject the null hypothesis—our reported p-values will be understated.

To sum up, although heteroskedasticity doesn't cause much of a problem for the OLS estimate of β itself, it does throw a wrench into our efforts to draw inferences about β from the OLS estimate. We are left with two options:

1. Use an estimator other than OLS.
2. Make a correction to the estimated standard errors that accounts for the possibility of heteroskedasticity.

These correspond, respectively, to the cases when the heteroskedasticity is of *known* and *unknown* form. But first we should know how to infer whether there's heteroskedasticity at all.

Detecting Heteroskedasticity

In a bivariate regression model, you can usually detect heteroskedasticity via the eye test. Not so much when you have multiple covariates. In this case, you may want to formally test for heteroskedasticity.

There are a few such tests, but we will just look at the *Breusch-Pagan test*, which was developed by Breusch and Pagan (1980) and refined by Koenker and Bassett (1982). The null hypothesis of the test is that $\sigma_i^2 = \sigma^2$ for all $i = 1, \ldots, N$. The test procedure is as follows.

1. Calculate the OLS estimate, $\hat{\beta}_{\text{OLS}}$.
2. Calculate the OLS residuals, $\hat{e} = Y - \mathbf{X}\hat{\beta}_{\text{OLS}}$. Let $\hat{u}$ be the vector of squared residuals, $\hat{u} = (\hat{e}_1^2, \ldots, \hat{e}_N^2)$.

3. Run a regression of $\hat{u}$ on $\mathbf{Z}$, an $N \times q$ matrix of covariates. Let $R^2_{\hat{u}}$ denote the R^2 of this regression.
4. Reject the null hypothesis if $NR^2_{\hat{u}}$ exceeds the critical value for a χ^2_{q-1} distribution.

In the canonical version of this test, $\mathbf{Z}$ is equal to $\mathbf{X}$. A more powerful version is the *White test* (White 1980), in which $\mathbf{Z}$ contains each variable in $\mathbf{X}$ as well as all second-order terms (squares and interactions).

To illustrate this test—and our upcoming solutions—we will use data from the **car** package on professors' salaries. This is a topic that, hopefully, you will come to care about rather much in 5ish years.

```
library("car")
data(Salaries)
head(Salaries)
```

```
##         rank discipline yrs.since.phd yrs.service  sex salary
## 1      Prof          B            19          18 Male 139750
## 2      Prof          B            20          16 Male 173200
## 3   AsstProf         B             4           3 Male  79750
## 4      Prof          B            45          39 Male 115000
## 5      Prof          B            40          41 Male 141500
## 6 AssocProf          B             6           6 Male  97000
```

There is an obvious source of heteroskedasticity here. We should expect much more variation in the salaries of faculty who've been around for a while than for those fresh out of grad school.

```
library("ggplot2")

ggplot(Salaries, aes(x = yrs.since.phd, y = salary)) +
    geom_point()
```
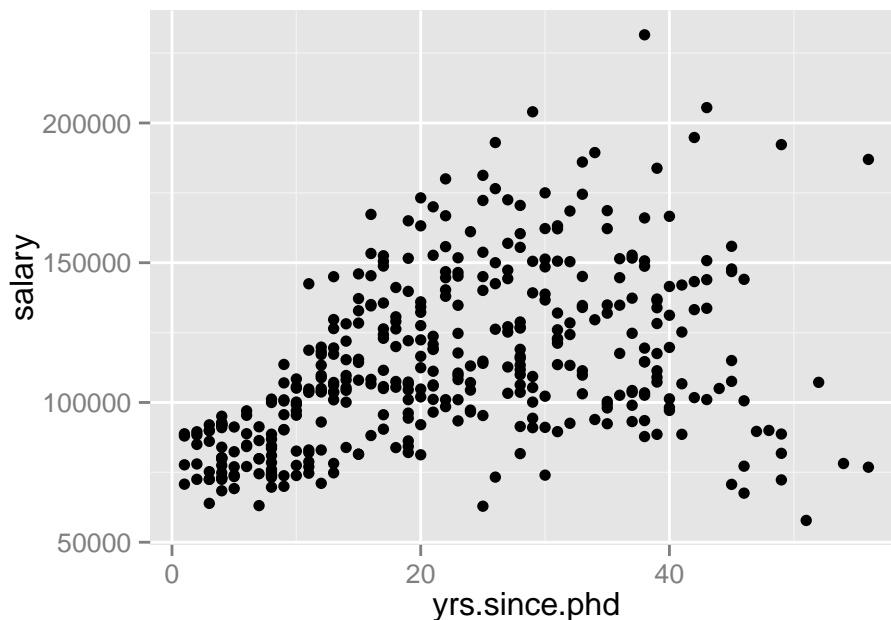
Let's regress salary on years since PhD and years of service, and use the Breusch-Pagan test to confirm what we already know about heteroskedasticity. We'll use the bptest() function from the **lmtest** package.

```
ols_salaries <- lm(salary ~ yrs.since.phd + yrs.service,
                   data = Salaries)
summary(ols_salaries)
```

```
##
## Call:
## lm(formula = salary ~ yrs.since.phd + yrs.service, data = Salaries)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -79735 -19823  -2617  15149 106149
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      89912       2844   31.62  < 2e-16
## yrs.since.phd     1563        257    6.09  2.8e-09
## yrs.service       -629        254   -2.47    0.014
##
## Residual standard error: 27400 on 394 degrees of freedom
```

```
## Multiple R-squared:  0.188,  Adjusted R-squared:  0.184
## F-statistic: 45.7 on 2 and 394 DF,  p-value: <2e-16
```

```r
library("lmtest")
bptest(ols_salaries)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  ols_salaries
## BP = 50, df = 2, p-value = 1e-11
```

By default, bptest() uses the same variables as in the original regression in the regression of the squared residuals. To perform the White test, we can use an extra argument to bptest() to specify a different model formula.

```r
bptest(ols_salaries,
       ~ yrs.since.phd * yrs.service + I(yrs.since.phd^2) + I(yrs.service^2),
       data = Salaries)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  ols_salaries
## BP = 60, df = 5, p-value = 1e-11
```

In this case, regardless of which test we use, we reject the null hypothesis of homoskedasticity.

Heteroskedasticity of Known Form

Suppose we know the form of Ω up to a multiplicative constant. In other words, we know that the variance of each ϵ_i is proportional to some constant $\omega_i > 0$. In this lucky circumstance, there is an efficient estimator of β available to us—the *weighted least squares*, or WLS, estimator.

The principle behind WLS is that observations with lower residual variance should receive more weight in the estimation of β. The more that each response observation Y_i is expected to deviate from the true regression line, the less useful that observation is for estimating the slope of that line—the harder

7

it is to separate the signal from the noise.

Let us now formally define the WLS estimator. We will collect the proportional variance of each error term in an $N \times 1$ vector, $\omega = (\omega_1, \ldots, \omega_N)$. We will weight every observation by the inverse of the proportional standard deviation, dividing both the response and the covariates (including the constant term) by the corresponding $\sqrt{\omega_i}$. Then we will run OLS on the weighted data. Formally, the estimator is

$$\hat{\beta}_{\text{WLS}}(Y, \mathbf{X}, \omega) = \hat{\beta}_{\text{OLS}}\left(\begin{bmatrix} Y_1/\sqrt{\omega_1} \\ Y_2/\sqrt{\omega_2} \\ \vdots \\ Y_N/\sqrt{\omega_N} \end{bmatrix}, \begin{bmatrix} x_{11}/\sqrt{\omega_1} & x_{12}/\sqrt{\omega_1} & \cdots & x_{1p}/\sqrt{\omega_1} \\ x_{21}/\sqrt{\omega_2} & x_{22}/\sqrt{\omega_2} & \cdots & x_{2p}/\sqrt{\omega_2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1}/\sqrt{\omega_N} & x_{N2}/\sqrt{\omega_N} & \cdots & x_{Np}/\sqrt{\omega_N} \end{bmatrix}\right)$$

It turns out this is equivalent to

$$\hat{\beta}_{\text{WLS}}(Y, \mathbf{X}, \omega) = (\mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Omega^{-1} Y,$$

where Ω, as before, is the matrix with ω along the diagonal and 0 everywhere else. Notice that OLS is a special case of WLS, with $\omega = (1, 1, \ldots, 1)$. Just like OLS, WLS is unbiased and (under reasonable conditions) consistent, even if Ω is misspecified. But if we have Ω right—and *only* if we have Ω right—then WLS is efficient in the class of linear unbiased estimators. In addition, our estimated variance matrix,

$$\hat{\Sigma}_{\text{WLS}} = \frac{\sum_{i=1}^{N} \hat{e}_i^2/\omega_i}{N - p} (\mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1},$$

is unbiased and consistent.

As an example, let's run WLS on the professor salary data, under the assumption that the residual variance is proportional to years since PhD:

$$V[\epsilon_i \,|\, x_i] = \sigma^2 \times (\text{Years Since PhD})_i.$$

We can accomplish this using the `weights` argument of `lm()`.

```
wls_salaries <- lm(salary ~ yrs.since.phd + yrs.service,
                   weights = 1 / yrs.since.phd,
                   data = Salaries)
summary(wls_salaries)
```

```
##
## Call:
## lm(formula = salary ~ yrs.since.phd + yrs.service, data = Salaries,
##     weights = 1/yrs.since.phd)
##
## Weighted Residuals:
##    Min     1Q Median     3Q    Max
## -13520  -4386    -91   4101  16046
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      79672       1460   54.56  < 2e-16
## yrs.since.phd     1753        242    7.25  2.3e-12
## yrs.service       -289        265   -1.09     0.28
##
## Residual standard error: 5760 on 394 degrees of freedom
## Multiple R-squared:  0.427,  Adjusted R-squared:  0.425
## F-statistic:  147 on 2 and 394 DF,  p-value: <2e-16
```

Compared to the OLS estimates, we now estimate a stronger relationship between years since the PhD and the expected value of a professor's salary. In addition, we estimate a smaller coefficient on years of service, and would no longer reject the null hypothesis of no relationship.

Are the WLS results better than the OLS results? It depends—namely on whether our assumption that the residual variance is proportional to the number of years since PhD. If our assumption is correct, then the WLS estimates are efficient and our estimated standard errors are valid. But if we assumed incorrectly, then the coefficient estimates might be even less efficient than OLS, and the estimated standard errors might be even more biased.

Heteroskedasticity of Unknown Form

Suppose we don't trust ourselves to conjure up the right proportional residual variances, the ω_is.[2] Then we do not have an efficient estimator of β. We might

---

[2]There is an in-between solution known as *feasible generalized least squares*, whereby we estimate Ω from the data. We won't discuss FGLS in this course, but you can read about it in Wooldridge or virtually any other econometrics textbook.

be all right with that, but we would really like to have a good estimator for the standard errors of the OLS estimator, so that we can test hypotheses about the coefficients. Happily, we can estimate the variance matrix of the OLS estimator consistently even in the presence of heteroskedasticity.

White's heteroskedasticity-consistent estimator (White 1980) of the variance matrix starts by forming a diagonal matrix out of the squared residuals,

$$\hat{\mathbf{U}} = \begin{bmatrix} \hat{u}_1 & 0 & \cdots & 0 \\ 0 & \hat{u}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{u}_N \end{bmatrix} = \begin{bmatrix} \hat{e}_1^2 & 0 & \cdots & 0 \\ 0 & \hat{e}_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{e}_N^2 \end{bmatrix}$$

This lets us form the "meat" of the "sandwich" that is White's estimator of $\Sigma = V[\hat{\beta}_{\text{OLS}} | \mathbf{X}]$:

$$\hat{\Sigma}_{\text{HC}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \hat{\mathbf{U}} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1}.$$

You know I love proofs, but I am not even going to attempt to prove that this consistently estimates the (asymptotic) variance matrix of $\hat{\beta}_{\text{OLS}}$. See Greene (2003, 198–99) for a sketch of the proof.

White's estimator is consistent but not unbiased, so we may want to apply a sort of bias correction in small samples. A popular choice is the so-called "HC1" estimator, which corrects for the number of parameters estimated the same way the usual OLS variance estimator does:

$$\hat{\Sigma}_{\text{HC1}} = \frac{N}{N-p} \hat{\Sigma}_{\text{HC}}$$

In this scheme, the standard White estimator is called the "HC" or "HC0" estimator. There are many other consistent estimators that apply some or other finite-sample correction; see MacKinnon and White (1985) for the gory details.

Because of its association with the `, robust` option in Stata, people sometimes call the White estimator of the standard errors "robust standard errors". Don't do that. In your own work, if you estimate and report heteroskedasticity-consistent standard errors, report that you use the White (1980) estimator of the standard errors, and specify which variant (HC0, HC1, and so on). Remember that your goal is to give others enough information to replicate your analysis even if they don't have your code—"robust standard errors" has too many interpretations to accomplish that.

To calculate the White estimator and its friends in R, we use the `hccm()` function from the **car** package.[3]

```
vcv0 <- hccm(ols_salaries, type = "hc0")
vcv0
```

```
##              (Intercept) yrs.since.phd yrs.service
## (Intercept)     5809137       -340724      111808
## yrs.since.phd   -340724         77168      -75508
## yrs.service      111808        -75508       91091
```

To create a "regression table" using our new "robust" standard errors, we can use the `coeftest()` function from the **lmtest** package.

```
coeftest(ols_salaries)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)     89912       2844   31.62  < 2e-16
## yrs.since.phd    1563        257    6.09  2.8e-09
## yrs.service      -629        254   -2.47    0.014
```

```
coeftest(ols_salaries, vcov = vcv0)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)     89912       2410   37.30  < 2e-16
## yrs.since.phd    1563        278    5.63  3.5e-08
## yrs.service      -629        302   -2.08    0.038
```

Just like ordinary regression tables, the ones made by `coeftest()` can be "swept" into data frames using the tools in **broom**:

```
library("broom")
tidy(coeftest(ols_salaries, vcov = vcv0))
```

```
##            term estimate std.error statistic   p.value
```

---

[3]The `vcovHC()` function in the **sandwich** package produces the same results but is a bit more flexible. It can also be used for generalized linear models (e.g., logistic regression).

```
## 1   (Intercept)      89912       2410      37.30 2.33e-131
## 2 yrs.since.phd       1563        278       5.63  3.50e-08
## 3    yrs.service      -629        302      -2.08  3.78e-02
```

You may also want to use the White-estimated standard errors to conduct Wald tests of linear hypotheses. You can do that by supplying the relevant estimated variance matrix to the `vcov` argument of `linearHypothesis()`:

```
linearHypothesis(ols_salaries,
                 c("yrs.since.phd = 1500"),
                 vcov = vcv0,
                 test = "Chisq")
```

```
## Linear hypothesis test
##
## Hypothesis:
## yrs.since.phd = 1500
##
## Model 1: restricted model
## Model 2: salary ~ yrs.since.phd + yrs.service
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df Chisq Pr(>Chisq)
## 1    395
## 2    394  1  0.05        0.82
```

Finally, remember how earlier we talked about how the WLS estimates are only as good as the weights you choose. If they're not the true weights, then WLS is not efficient and the standard error estimator is inconsistent. We can't fix the first problem, but we can fix the second. To wit, you can estimate heteroskedasticity-consistent standard errors for WLS models too.

```
vcv0_wls <- hccm(wls_salaries, type = "hc0")
coeftest(wls_salaries, vcov = vcv0_wls)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)     79672       1474   54.06  < 2e-16
```

```
## yrs.since.phd     1753        245    7.16  3.9e-12
## yrs.service       -289        272   -1.06    0.29
```

So if you have a good idea about the residual variances but aren't sure you've nailed it down, you can have the best of both worlds—at least in terms of large-sample hypothesis testing.

Summary

- Heteroskedasticity doesn't affect the properties of the OLS estimator of β very much.
- But it does cause inefficiency and, even worse, break the ordinary estimator of the standard errors of $\hat{\beta}_{\text{OLS}}$.
- If we know the form of the heteroskedasticity, we can fix both of these problems with WLS.
- If we don't know the form of the heteroskedasticity, we can at least consistently estimate the standard errors of $\hat{\beta}_{\text{OLS}}$ via White's estimator.

References

Breusch, T.S., and A.R. Pagan. 1980. "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics." *Review of Economic Studies* 47 (1): 239–53. http://www.jstor.org/stable/2297111.

Greene, William H. 2003. *Econometric Analysis*. 5th ed. Prentice Hall.

Koenker, Roger, and Gilbert Bassett Jr. 1982. "Robust Tests for Heteroscedasticity Based on Regression Quantiles." *Econometrica* 50 (1): 43–61. http://www.jstor.org/stable/1912528.

MacKinnon, James G., and Halbert White. 1985. "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties." *Journal of Econometrics* 29 (3): 305–25. http://dx.doi.org/10.1016/0304-4076(85)90158-7.

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48 (4): 817–38. http://www.jstor.org/stable/1912934.