Dealing with Heteroskedasticity

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

Dealing with Heteroskedasticity

Introduction

- In this lecture, we review some new concepts involving weighted least squares, evaluation of model fit, sandwich estimators, and bootstrap confidence interval calculation introduced in ALR, Chapter 7.

Weighted Least Squares Estimation

- Suppose that the conditional mean follows the linear regression rule, but the conditional variance does not, i.e.,

$$
\begin{aligned}
E(Y|X = \mathbf{x}_i) &= \boldsymbol{\beta}'\mathbf{x}_i && (1)\\
\mathrm{Var}(Y|X = \mathbf{x}_i) &= \mathrm{Var}(e_i) = \sigma^2/w_i && (2)
\end{aligned}
$$

- The matrix equivalent of the above is

$$
\mathbf{y} = \mathbf{X}\beta + \mathbf{e} \qquad \mathrm{Var}(\mathbf{e}) = \sigma^2\mathbf{W}^{-1} \qquad (3)
$$

- *Weighted Least Squares* (WLS) estimation minimizes, under choice of β, the function

$$
\begin{aligned}
RSS(\beta) &= (\mathbf{y} - \mathbf{X}\beta)'\mathbf{W}(\mathbf{y} - \mathbf{X}\beta) && (4)\\
&= \mathbf{e}_\beta'\mathbf{W}\mathbf{e}_\beta && (5)\\
&= \sum_i w_i e_{\beta_i}^2 && (6)
\end{aligned}
$$

Weighted Least Squares Estimation

- The solution to the WLS estimation problem is well-known to be

$$\hat{\beta} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y} \tag{7}$$

- Since $\mathbf{W} = \mathbf{W}^{1/2}\mathbf{W}^{1/2}$, we can consider redefining the problem as follows. Let $\mathbf{z} = \mathbf{W}^{1/2}\mathbf{y}$, $\mathbf{M} = \mathbf{W}^{1/2}\mathbf{X}$, and $\mathbf{d} = \mathbf{W}^{1/2}\mathbf{e}$.
- What will the covariance matrix of $\mathbf{d}$ be?
- Since $\mathbf{d} = \mathbf{W}^{1/2}\mathbf{e}$, it follows that $\text{Var}(\mathbf{d}) = \mathbf{W}^{1/2}\,\text{Var}(\mathbf{e})\mathbf{W}^{1/2} = \mathbf{W}^{1/2}(\sigma^2\mathbf{W}^{-1})\mathbf{W}^{1/2} = \sigma^2\mathbf{W}^{1/2}\mathbf{W}^{-1}\mathbf{W}^{1/2} = \sigma^2\mathbf{I}$.
- Since $\mathbf{d}$ will have covariance matrix $\sigma^2\mathbf{I}$, all the previously developed mechanics of OLS regression apply to the model

$$\mathbf{z} = \mathbf{M}\beta + \mathbf{d} \tag{8}$$

In OLS regression, $\hat{\beta} = (\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{z}$ which is easily shown (C.P.) to be equal to $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y}$.

Weighted Least Squares Estimation

- The conclusion from the previous slide is that if we have a set of weights, we can simply rescale the criterion variable and predictors by these weights, then apply ordinary least squares regression to the transformed Y and X variables.
- Some key questions remain:
  1. How, in practice, do you get the weights?
  2. Can R simplify the calculations and do them automatically?

Getting the Weights

- Known weights w_i can occur in many ways. If the ith response is an average of n_i equally variable observations, then $\mathrm{Var}(y_i) = \sigma^2/n_i$, and $w_i = n_i$.
- If y_i is a sum of n_i observations, $\mathrm{Var}(y_i) = n_i\sigma^2$, and $w_i = 1/n_i$.
- If variance is proportional to some predictor x_i, $\mathrm{Var}(y_i) = x_i\sigma^2$, then $w_i = 1/x_i$.

An Example From Physics

- An experiment was carried out with a beam having various values of s, the square of the total energy in the center-of-mass frame of reference system. For each value of s, we observe the scattering cross-section y, measured in millibarns (μb).
- A theoretical model predicts

$$E(y|s) = \beta_0 + \beta_1 s^{-1/2} + \text{relatively small terms} \tag{9}$$

- The theory makes quantitative predictions about β_0 and β_1 and their dependence on particular input and output particle type.
- Data for the π^- meson are in the data file `physics` in ALR4.

An Example From Physics

- The data file entries represent *many* observations per cell. As a result, the relative conditional variances of $y|s = s_i$ are known to a high degree of accuracy. The conditional standard deviations are given in the data file in the variable `SD`.
- Using the approach in ALR section 5.1, let's assume that the conditional variances are of the form $\text{Var}(y|s = s_i) = \sigma^2/w_i$.
- However, before we can proceed, we need to recognize a subtle technical problem.

An Example From Physics

- Note that we have 10 "known" conditional standard deviations, but 11 parameters to estimate. This contrasts with the more common situation where the weights are known but the variance factor σ^2 is not. For identification, we arbitrarily set $\sigma = 1$, and treat the entries in the data file as $1/\sqrt{w_i}$.
- In effect, this allows us to use the model in a situation in which the conditional variances are known.
- x in the data file is $s^{-1/2}$, so predicting y from x is equivalent to predicting it from $s^{-1/2}$.
- If the model is correct, the standard error of estimate should have an estimated value of 1.

An Example From Physics

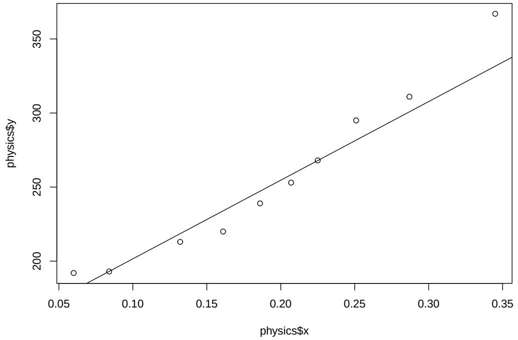
```
> m1 <- lm(y~x,weights=1/SD^2,data=physics)
> summary(m1)
Call:
lm(formula = y ~ x, data = physics, weights = 1/SD^2)

Weighted Residuals:
    Min      1Q  Median      3Q     Max
-2.3230 -0.8842  0.0000  1.3900  2.3353

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  148.473      8.079   18.38 7.91e-08 ***
x            530.835     47.550   11.16 3.71e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.657 on 8 degrees of freedom
Multiple R-squared:  0.9397,        Adjusted R-squared:  0.9321
F-statistic: 124.6 on 1 and 8 DF,  p-value: 3.71e-06
```

An Example From Physics

- We note that the R^2 value is quite high, but also that the standard error of estimate is 1.66, not the 1.0 that we expected.
- One reason that a standard error of estimate can be higher than a "known" value is if the form of the regression function (linear in this case) is wrong. Let's take a look at the regression of y on x.

An Example From Physics

```
> plot(physics$x,physics$y)
> abline(m1)
```



An Example From Physics

- We can see that there is a nonlinear trend in the data. In fact, there is a significant lack of fit in these data.
- In the next section, we examine a test of fit that can be applied in the case in which the standard error of estimate is assumed to be known.

Testing for Fit, Variance Known

- If $\hat{\sigma}^2$ is too large, we may have evidence that the mean function is wrong in our model.
- In the strong attraction data, we assumed $\sigma^2 = 1$, and that the stated model was linear. If the model is correct, the conditional variances around the regression line after using the weights should all be close to the actual conditional variances.
- If the estimated σ^2 exceeds its theoretical value, this can be evidence that the model is incorrect.
- From our previous output, we see an estimated residual standard error of 1.66, which translates into an estimated residual variance of about 2.74.
- We perform the classic χ^2 significance test discussed in Psychology 310,

$$\chi^2_{n-p'} = \frac{(n-p')\hat{\sigma}^2}{\sigma^2} = \frac{RSS}{\sigma^2} \tag{10}$$

Testing for Fit, Variance Known

- From the anova table we see that RSS is 21.953, which in this case (since we are testing that $\sigma^2 = 1$) is also the test statistic with 8 degrees of freedom.

```
> print(anova(m1),digits=6)
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value      Pr(>F)
x          1 341.991 341.991 124.629 3.7104e-06 ***
Residuals  8  21.953   2.744
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testing for Fit, Variance Known

- This is a 1-sided test. If the model is correct, the conditional variance should be 1.0. If it is incorrect, the conditional variance will be higher than 1.0.
- Consequently, we can calculate the *p*-value for the χ^2 test as

```
> 1 - pchisq(21.953,8)

[1] 0.005003683
```

- The test rejects beyond the 0.01 level, indicating that we have "significant lack of fit."
- We could try a transformation.
- An alternative approach is to try adding a quadratic term.

Testing for Fit, Variance Known

- The quadratic function fit is signficantly better than the lineara, as shown by the *p*-value of around 0.00038 in the ANOVA table.

```
> m2 <- lm(y ~ x + I(x^2), weights=1/SD^2, data=physics)
> anova(m2)
```

```
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
x          1 341.99  341.99 742.185 2.303e-08 ***
I(x^2)     1  18.73   18.73  40.641 0.0003761 ***
Residuals  7   3.23    0.46
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Testing for Fit, Variance Known

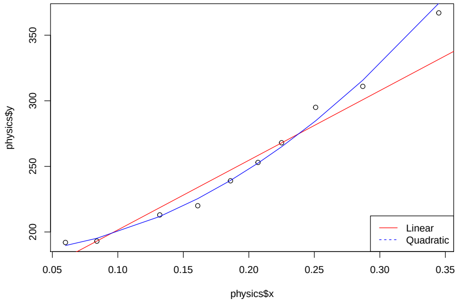
- From the table, we see that the residual sum of squares is now only 3.23, so the *p*-value for the χ^2 test of fit is now

```
> 1 - pchisq(3.23,7)

[1] 0.8629415
```

- Let's replot the linear function along with the quadratic on the next slide.
- We can see a big improvement.

Testing for Fit, Variance Known

```
> plot(physics$x,physics$y)
> abline(m1,col="red")
> lines(physics$x,predict(m2),type="l",col="blue")
> legend("bottomright",legend=c("Linear","Quadratic"),
+        col=c("red","blue"),lty=c(1,2))
```



Testing for Fit, Variance Known

- In fact, from a model summary, we see that R^2 has improved to 0.99!

```
> summary(m2)

Call:
lm(formula = y ~ x + I(x^2), data = physics, weights = 1/SD^2)

Weighted Residuals:
     Min       1Q   Median       3Q      Max
-0.89928 -0.43508  0.01374  0.37999  1.14238

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  183.8305     6.4591  28.461  1.7e-08 ***
x              0.9709    85.3688   0.011 0.991243
I(x^2)      1597.5047   250.5869   6.375 0.000376 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6788 on 7 degrees of freedom
Multiple R-squared:  0.9911,	Adjusted R-squared:  0.9886
F-statistic: 391.4 on 2 and 7 DF,  p-value: 6.554e-08
```

The Sandwich Estimator

- One can attempt to construct a model for the unequal variances, by constructing a regression function to predict the unequal variances, and using the predicted values as weights.
- However, an alternate approach to dealing with heteroskedasticity is based on the fact that if $\boldsymbol{\Omega}$ is the covariance matrix of the errors (and hence of the Y_i), then the OLS estimate of β, $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, has a covariance matrix that can be calculated from the fundamental theorem of multivariate analysis

$$\text{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} \tag{11}$$

- One can estimate the matrix $\boldsymbol{\Omega}$ in a variety of ways. A method called HC3 estimates it as a diagonal matrix with diagonal entry $\hat{e}^2/(1-h_{ii})^2$, where h_{ii} is the leverage of the ith observation.
- One can then use the diagonal elements of the estimated covariance matrix as estimates of the sampling variances of the $\hat{\beta}_i$, and their square roots as estimated standard errors.
- The ALR4 primer for Chapter 7 shows how do do this with some “prepackaged” R functions.

The Sandwich Estimator

- Suppose we return to the sniffer data, and estimate Y from all 4 predictors.
- The fit object and summary are

```
> s1 <- lm(Y ~ ., data=sniffer)
> summary(s1)
Call:
lm(formula = Y ~ ., data = sniffer)

Residuals:
    Min      1Q  Median      3Q     Max
-6.5425 -1.2938  0.0495  1.2259  7.0413

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.15391    1.03489   0.149   0.8820
TankTemp    -0.08269    0.04857  -1.703   0.0912 .
GasTemp      0.18971    0.04118   4.606 1.03e-05 ***
TankPres    -4.05962    1.58000  -2.569   0.0114 *
GasPres      9.85744    1.62515   6.066 1.57e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.758 on 120 degrees of freedom
Multiple R-squared:  0.8933,        Adjusted R-squared:  0.8897
F-statistic: 251.1 on 4 and 120 DF,  p-value: < 2.2e-16
```

The Sandwich Estimator

- We can get R to output the entire estimated covariance matrix of the $\hat{\beta}$ as

```
> vcov(s1)
```

```
              (Intercept)     TankTemp      GasTemp   TankPres    GasPres
(Intercept)   1.070996244  0.008471429 -0.017735339 -0.20656649  0.09308897
TankTemp      0.008471429  0.002358852 -0.001002625 -0.04777401  0.02791454
GasTemp      -0.017735339 -0.001002625  0.001696097  0.04149452 -0.04686422
TankPres     -0.206566487 -0.047774008  0.041494518  2.49641354 -2.38665198
GasPres       0.093088965  0.027914540 -0.046864224 -2.38665198  2.64111761
```

- We can also see that the square roots of the diagonal elements of this matrix match the standard errors in the summary output on the preceding slide.

```
> sqrt(diag(vcov(s1)))
```

```
(Intercept)    TankTemp     GasTemp    TankPres     GasPres
 1.03488948  0.04856801  0.04118370  1.58000429  1.62515157
```

The Sandwich Estimator

- R will directly compute the sandwich estimator for the covariance matrix of $\hat{\beta}$ as

```
> m2 <- hccm(s1, type="hc3")
> m2
```

```
            (Intercept)     TankTemp      GasTemp    TankPres     GasPres
(Intercept)  1.09693263  0.0156217962 -0.0128307109 -0.26718035 -0.03244478
TankTemp     0.01562180  0.0019751133 -0.0006409877 -0.03915604  0.01822761
GasTemp     -0.01283071 -0.0006409877  0.0011424888  0.03985576 -0.04344723
TankPres    -0.26718035 -0.0391560420  0.0398557591  3.89032393 -3.86151048
GasPres     -0.03244478  0.0182276131 -0.0434472283 -3.86151048  4.22652409
```

-
- We can extract the diagonal elements and take their square roots to yield robust standard errors.

```
> sqrt(diag(m2))
```

```
(Intercept)    TankTemp     GasTemp    TankPres     GasPres
 1.04734551  0.04444225  0.03380072  1.97239041  2.05585118
```

The Sandwich Estimator

- We can then use them to construct confidence intervals or statistical tests.
- R has a function to do this in the `lmtest` library.

```
> library(lmtest)
> coeftest(s1, vcov=hccm)

t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)
(Intercept)  0.153908   1.047346  0.1470   0.88342
TankTemp    -0.082695   0.044442 -1.8607   0.06523 .
GasTemp      0.189707   0.033801  5.6125 1.306e-07 ***
TankPres    -4.059617   1.972390 -2.0582   0.04173 *
GasPres      9.857441   2.055851  4.7948 4.719e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Sandwich Estimator

- Let's compare these estimates and standard errors with a WLS analysis using 1/TankTemp as weights.

```
> m3 <- lm(Y~.,data=sniffer,weights=1/TankTemp)
> summary(m3)
```

```
Call:
lm(formula = Y ~ ., data = sniffer, weights = 1/TankTemp)

Weighted Residuals:
     Min       1Q   Median       3Q      Max
-0.85390 -0.20990  0.01953  0.17359  0.91192

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.17198    0.99087   0.174   0.8625
TankTemp    -0.06090    0.04256  -1.431   0.1551
GasTemp      0.18971    0.03732   5.083 1.38e-06 ***
TankPres    -3.18716    1.48894  -2.141   0.0343 *
GasPres      8.67599    1.50856   5.751 6.90e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3557 on 120 degrees of freedom
Multiple R-squared:  0.8886,	Adjusted R-squared:  0.8848
F-statistic: 239.2 on 4 and 120 DF,  p-value: < 2.2e-16
```