Gov 2000 - 10. Troubleshooting the Linear Model II: Heteroskedasticity

Matthew Blackwell

December 4, 2015

1. Heteroskedasticity
2. Clustering
3. Serial Correlation
4. What's next for you?

Where are we? Where are we going?

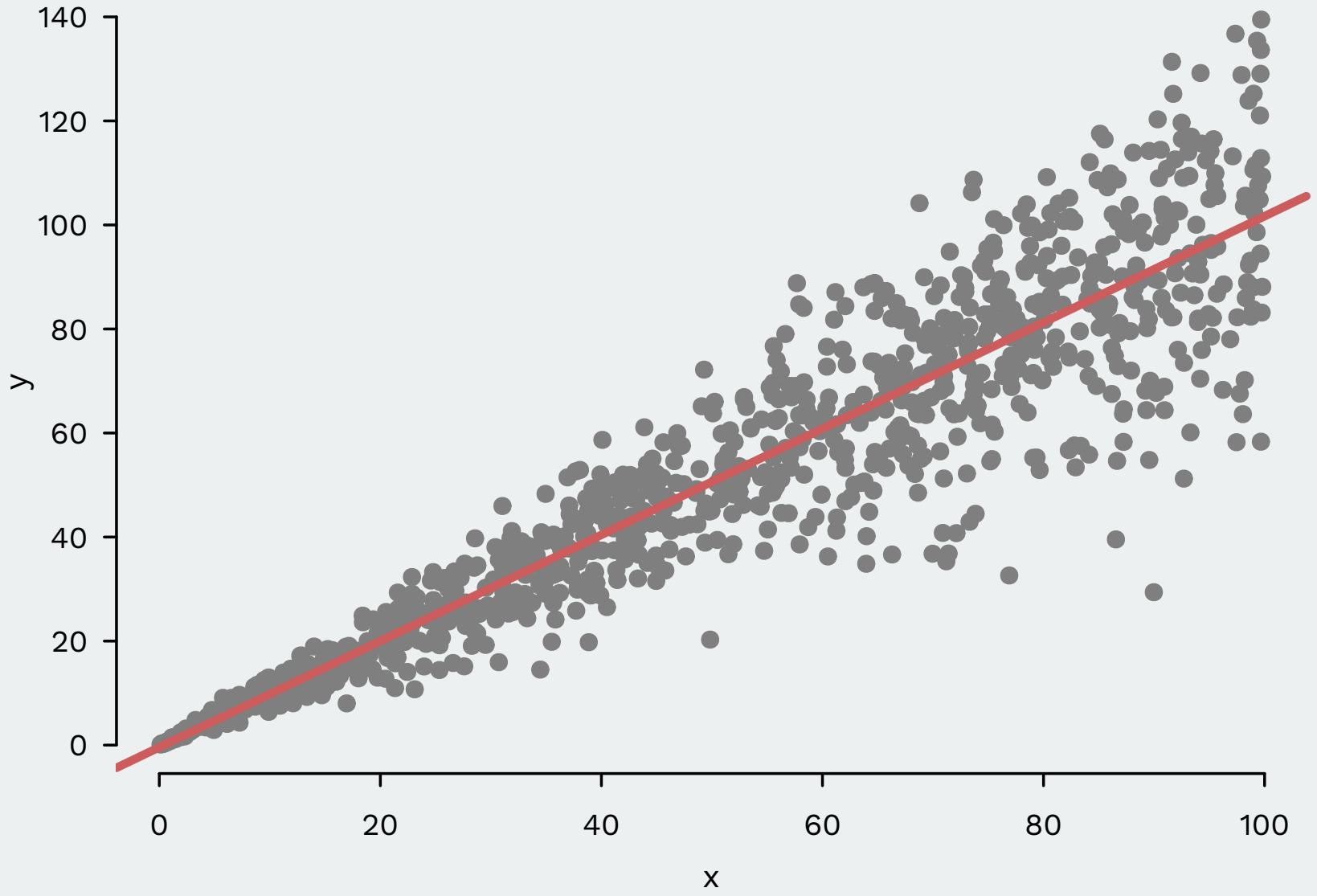
- Last week: finding and correcting violations of linearity and non-Normal errors
- This week: detecting and correcting violations of homoskedasticity

Review of the OLS assumptions

1. Linearity: $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$
2. Random/iid sample: $(y_i, \mathbf{x}_i')$ are a iid sample from the population.
3. No perfect collinearity: $\mathbf{X}$ is an $n \times (k+1)$ matrix with rank $k+1$
4. Zero conditional mean: $\mathbb{E}[\mathbf{u}|\mathbf{X}] = \mathbf{0}$
5. Homoskedasticity: $\mathbb{V}(\mathbf{u}|\mathbf{X}) = \sigma_u^2 \mathbf{I}_n$
6. Normality: $\mathbf{u}|\mathbf{X} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I}_n)$

- 1-4 give us unbiasedness/consistency
- 1-5 are the Gauss-Markov, allow for large-sample inference
- 1-6 allow for small-sample inference

How do we deal with this?



Plan for today

- Talk about different forms of error variance problems

1. Heteroskedasticity
2. Clustering
3. Serial correlation

- Each is a violation of heteroskedasticity, but each has its own diagnostics and corrections

1/ Heteroskedasticity

Review of homoskedasticity

- Remember:

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

- Let $\mathbb{V}[\mathbf{u}|\mathbf{X}] = \boldsymbol{\Sigma}$
- Using assumptions 1 and 4, we can show that we have the following:

$$\mathbb{V}[\hat{\boldsymbol{\beta}}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

- To derive this, use three facts about matrices:
  - If $\mathbf{A}$ is a constant matrix, then $\mathbb{V}[\mathbf{A}\mathbf{y}] = \mathbf{A}\mathbb{V}[\mathbf{y}]\mathbf{A}'$
  - $(\mathbf{A}\mathbf{B})' = \mathbf{B}'\mathbf{A}'$
  - $(\mathbf{X}'\mathbf{X})^{-1}$ is symmetric $\rightsquigarrow \left((\mathbf{X}'\mathbf{X})^{-1}\right)' = (\mathbf{X}'\mathbf{X})^{-1}$

Review of homoskedasticity

- With homoskedasticity, $\Sigma = \sigma^2 \mathbf{I}$

$$\begin{aligned}
\mathbb{V}[\hat{\boldsymbol{\beta}}|\mathbf{X}] &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \text{ (by homoskedasticity)} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}$$

- Replace σ^2 with estimate $\widehat{\sigma}^2$ will give us our estimate of the covariance matrix

Non-constant error variance

- Homoskedastic:

$$\mathbb{V}[\mathbf{u}|\mathbf{X}] = \sigma^2\mathbf{I} = \begin{bmatrix} \sigma^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma^2 & 0 & \dots & 0 \\ & & & \vdots & \\ 0 & 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

- Heteroskedastic:

$$\mathbb{V}[\mathbf{u}|\mathbf{X}] = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ & & & \vdots & \\ 0 & 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}$$

- Independent, not identical
- $\text{Cov}(u_i, u_j|\mathbf{X}) = 0$
- $\mathbb{V}(u_i|\mathbf{X}) = \sigma_i^2$

Consequences of Heteroskedasticity

- Standard error estimates biased, likely downward
- Test statistics won't have t or F distributions
- α-level tests, the probability of Type I error $\neq \alpha$
- Coverage of $1 - \alpha$ CIs $\neq 1 - \alpha$
- OLS is not BLUE
- $\widehat{\boldsymbol{\beta}}$ still unbiased and consistent for $\boldsymbol{\beta}$

Visual diagnostics

1. Plot of residuals versus fitted values
   - In R, `plot(mod, which = 1)`
   - Residuals should have the same variance across x-axis
2. Spread location plots
   - y-axis: Square-root of the absolute value of the residuals
   - x-axis: Fitted values
   - Usually has loess trend curve, should be flat
   - In R, `plot(mod, which = 3)`

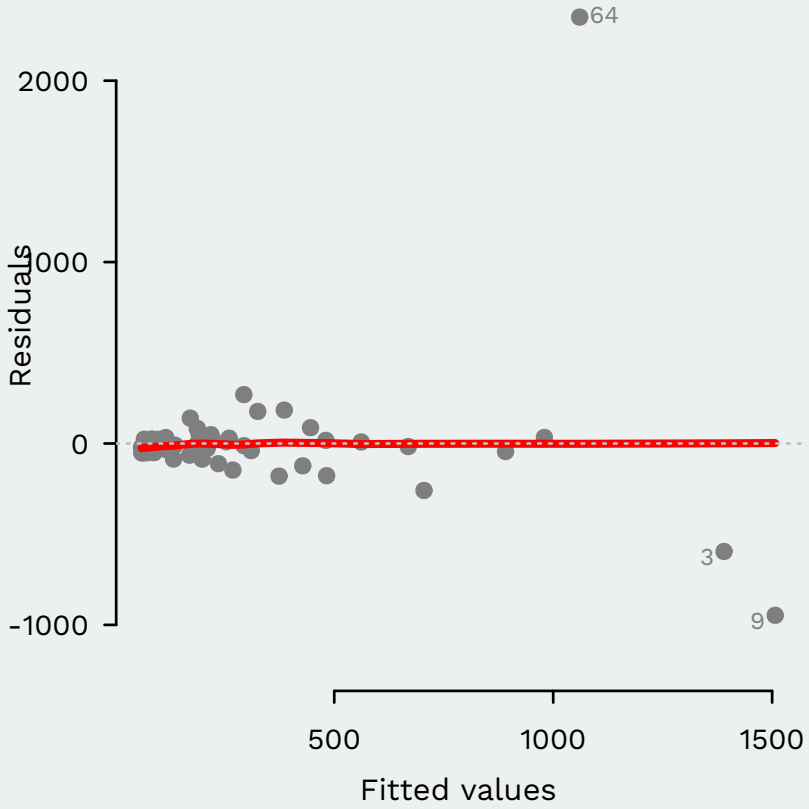
Example: Buchanan votes

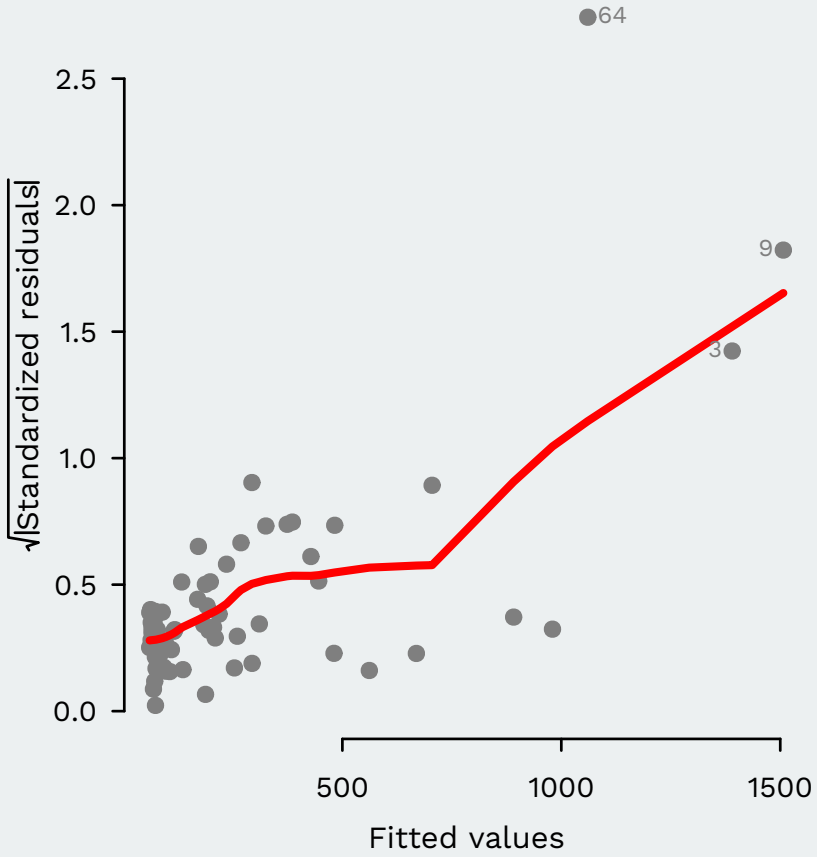
```r
flvote <- foreign::read.dta("flbuchan.dta")
mod <- lm(edaybuchanan ~ edaytotal, data = flvote)
summary(mod)
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 54.22945   49.14146    1.10     0.27
## edaytotal    0.00232    0.00031    7.48  2.4e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 333 on 65 degrees of freedom
## Multiple R-squared:  0.463,  Adjusted R-squared:  0.455
## F-statistic:   56 on 1 and 65 DF,  p-value: 2.42e-10
```

Diagnostics

```
plot(mod, which = 1, lwd = 3)
plot(mod, which = 3, lwd = 3)
```





Formal tests

- Plots are usually sufficient, but can use formal hypothesis test for heteroskedasticity:

$$H_0 : \mathbb{V}[u_i|\mathbf{X}] = \sigma^2$$

- Under zero conditional mean, this is equivalent to

$$H_0 : \mathbb{E}[u_i^2|\mathbf{X}] = \mathbb{E}[u_i^2] = \sigma^2$$

- Under the null, the squared residuals should be unrelated to the independent variables
- Breush-Pagan test:
  1. Regression y_i on $\mathbf{x}_i'$ and store residuals, $\hat{u}_i$
  2. Regress $\hat{u}_i^2$ on $\mathbf{x}_i'$
  3. Run F-test against null that all slope coefficients are 0
- In R, `bptest()` in the `lmtest` package

Breush-Pagan example

```
library(lmtest)
bptest(mod)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  mod
## BP = 13, df = 1, p-value = 0.0004
```

Dealing with non-constant error variance

1. Transform the dependent variable
2. Model the heteroskedasticity using Weighted Least Squares (WLS)
3. Use an estimator of $\mathbb{V}[\widehat{\boldsymbol{\beta}}]$ that is robust to heteroskedasticity
4. Admit we have the wrong model and use a different approach

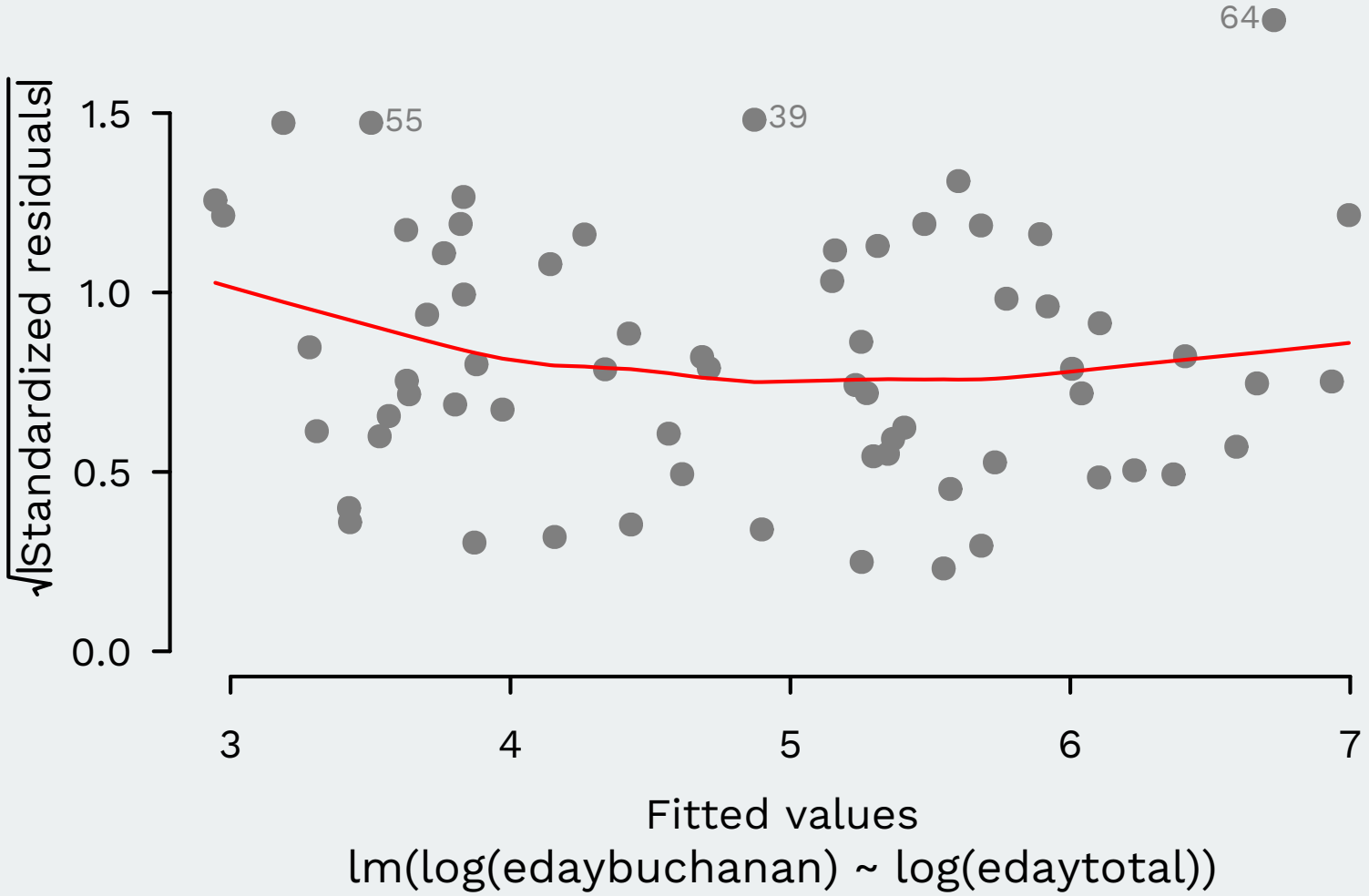
Example: Transforming Buchanan votes

```
mod2 <- lm(log(edaybuchanan) ~ log(edaytotal), data = flvote)
summary(mod2)
```

```
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -2.728      0.400   -6.83  3.5e-09 ***
## log(edaytotal)     0.729      0.038   19.15  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.469 on 65 degrees of freedom
## Multiple R-squared:  0.849,  Adjusted R-squared:  0.847
## F-statistic:  367 on 1 and 65 DF,  p-value: <2e-16
```

Example: Transformed scale-location plot

```
plot(mod2, which = 3)
```



Example: Transformed

```
bptest(mod, studentize = FALSE)
```

```
##
##  Breusch-Pagan test
##
## data:  mod
## BP = 250, df = 1, p-value <2e-16
```

```
bptest(mod2, studentize = FALSE)
```

```
##
##  Breusch-Pagan test
##
## data:  mod2
## BP = 0.011, df = 1, p-value = 0.9
```

Weighted least squares

- Suppose that the heteroskedasticity is known up to a multiplicative constant:

$$\mathbb{V}[u_i|\mathbf{X}] = a_i\sigma^2$$

  where $a_i = a_i(\mathbf{x}_i')$ is a positive and known function of $\mathbf{x}_i'$
- WLS: multiply y_i by $1/\sqrt{a_i}$:

$$y_i/\sqrt{a_i} = \beta_0/\sqrt{a_i} + \beta_1 x_{i1}/\sqrt{a_i} + \cdots + \beta_k x_{ik}/\sqrt{a_i} + u_i/\sqrt{a_i}$$

WLS intuition

- Rescales errors to $u_i/\sqrt{a_i}$, which maintains zero mean error
- But makes the error variance constant again:

$$\begin{aligned}\mathbb{V}\left[\frac{1}{\sqrt{a_i}}u_i|\mathbf{X}\right] &= \frac{1}{a_i}\mathbb{V}\left[u_i|\mathbf{X}\right] \\ &= \frac{1}{a_i}a_i\sigma^2 \\ &= \sigma^2\end{aligned}$$

- If you know a_i, then you can use this approach to makes the model homoeskedastic and, thus, BLUE again
- When do we know a_i?

WLS procedure

- Define the weighting matrix:

$$\mathbf{W} = \begin{bmatrix} 1/\sqrt{a_1} & 0 & 0 & 0 \\ 0 & 1/\sqrt{a_2} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1/\sqrt{a_n} \end{bmatrix}$$

- Run the following regression:

$$\begin{aligned} \mathbf{Wy} &= \mathbf{WX}\boldsymbol{\beta} + \mathbf{Wu} \\ \mathbf{y}^* &= \mathbf{X}^*\boldsymbol{\beta} + \mathbf{u}^* \end{aligned}$$

- Run regression of $\mathbf{y}^* = \mathbf{Wy}$ on $\mathbf{X}^* = \mathbf{WX}$ and all Gauss-Markov assumptions are satisfied
- Plugging into the usual formula for $\widehat{\boldsymbol{\beta}}$:

$$\widehat{\boldsymbol{\beta}}_W = (\mathbf{X}'\mathbf{W}'\mathbf{WX})^{-1}\mathbf{X}'\mathbf{W}'\mathbf{Wy}$$

WLS example

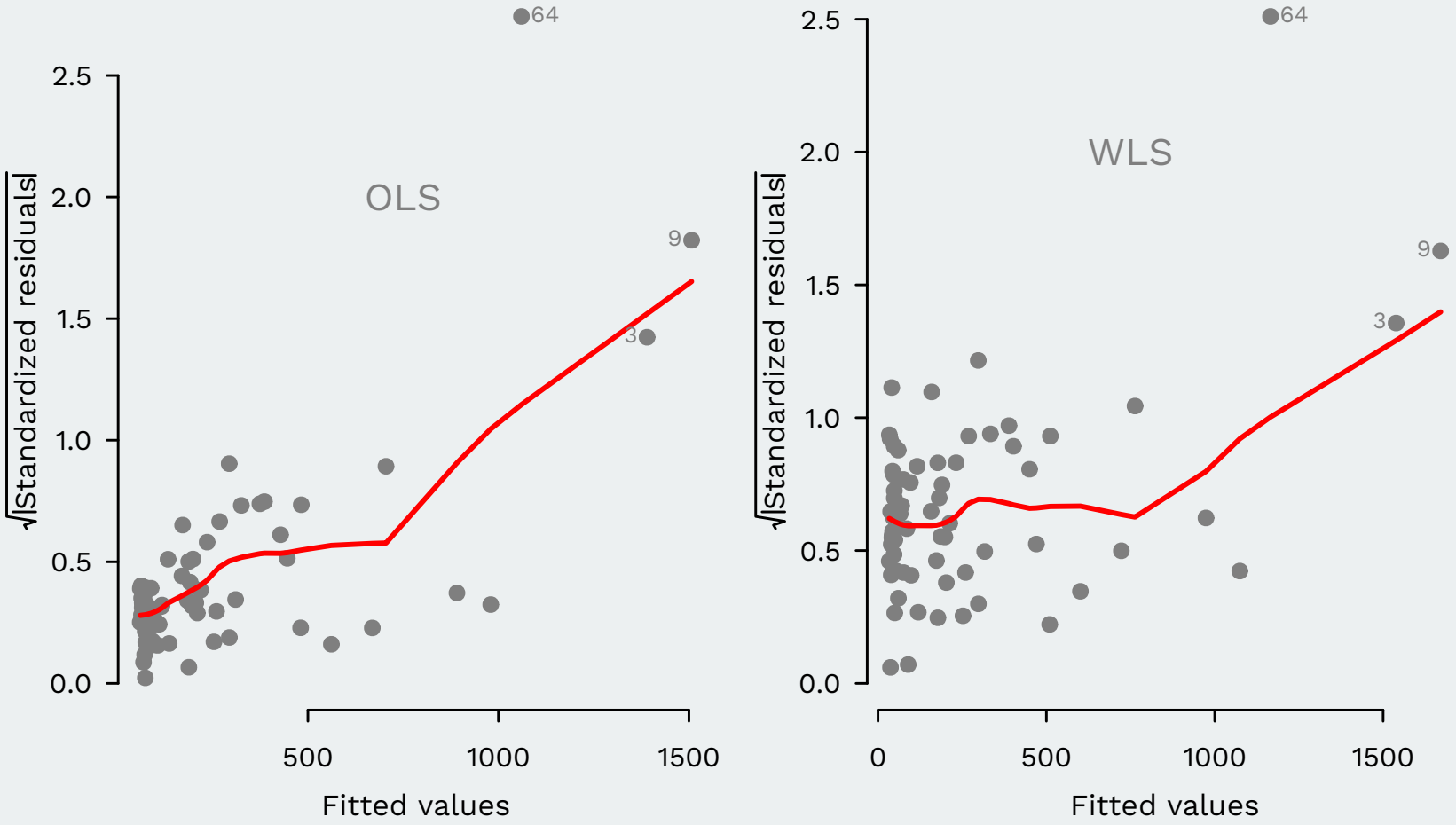
- In R, use `weights =` argument to `lm` and give the weights squared: $1/a_i$
- With the Buchanan data, maybe the variance is proportional to the total number of ballots cast:

```
mod.wls <- lm(edaybuchanan ~ edaytotal, weights = 1/edaytotal, data = flvot
summary(mod.wls)
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 27.06785    8.50723    3.18   0.0022 **
## edaytotal    0.00263    0.00025   10.50  1.2e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.565 on 65 degrees of freedom
## Multiple R-squared:  0.629,  Adjusted R-squared:  0.624
## F-statistic:  110 on 1 and 65 DF,  p-value: 1.22e-15
```

Comparing WLS to OLS

```
plot(mod, which = 3, lwd = 2, sub = "")
plot(mod.wls, which = 3, lwd = 2, sub = "")
```



Heteroskedasticity consistent estimator

- Under non-constant error variance:

$$\mathbb{V}[\mathbf{u}|\mathbf{X}] = \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ & & & \vdots & \\ 0 & 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}$$

- When $\Sigma \neq \sigma^2 \mathbf{I}$, we are stuck with this expression:

$$\mathbb{V}[\widehat{\boldsymbol{\beta}}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\Sigma\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}$$

- White (1980) shows that we can consistently estimate this if we have an estimate of Σ:

$$\widehat{\mathbb{V}}[\widehat{\boldsymbol{\beta}}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\widehat{\Sigma}\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}$$

- Sandwich estimator with bread $(\mathbf{X}'\mathbf{X})^{-1}$ and meat $\mathbf{X}'\widehat{\Sigma}\mathbf{X}$

Computing HC/robust standard errors

1. Fit regression and obtain residuals $\hat{\mathbf{u}}$
2. Construct the "meat" matrix $\widehat{\Sigma}$ with squared residuals in diagonal:

$$\widehat{\Sigma} = \begin{bmatrix} \hat{u}_1^2 & 0 & 0 & \dots & 0 \\ 0 & \hat{u}_2^2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \hat{u}_n^2 \end{bmatrix}$$

3. Plug $\widehat{\Sigma}$ into sandwich formula to obtain HC/robust estimator of the covariance matrix:

$$\widehat{\mathbb{V}}[\hat{\boldsymbol{\beta}}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widehat{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

- Small-sample corrections (called 'HC1'):

$$\widehat{\mathbb{V}}[\hat{\boldsymbol{\beta}}|\mathbf{X}] = \frac{n}{n-k-1} \cdot (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widehat{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Robust SEs in Florida data

```
library(sandwich)
coeftest(mod)
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 54.22945   49.14146    1.10     0.27
## edaytotal    0.00232    0.00031    7.48  2.4e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
coeftest(mod, vcovHC(mod, type = "HC0"))
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 54.22945   40.61283    1.34   0.1864
## edaytotal    0.00232    0.00087    2.67   0.0096 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Robust SEs with correction

```
coeftest(mod, vcovHC(mod, type = "HC0"))
```

```
##
## t test of coefficients:
##
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 54.22945   40.61283    1.34   0.1864
## edaytotal    0.00232    0.00087    2.67   0.0096 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
coeftest(mod, vcovHC(mod, type = "HC1"))
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 54.229453  41.232904    1.32    0.193
## edaytotal    0.002323   0.000884    2.63    0.011 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

WLS vs. White's Estimator

- WLS:
  - With known weights, WLS is efficient
  - and $\widehat{SE}[\widehat{\boldsymbol{\beta}}_{WLS}]$ is consistent
  - but weights usually aren't known
- White's Estimator:
  - Doesn't change estimate $\widehat{\boldsymbol{\beta}}$
  - Consistent for $\mathbb{V}[\widehat{\boldsymbol{\beta}}]$ under any form of heteroskedasticity
  - Because it relies on consistency, it is a large sample result, best with large n
  - For small n, performance might be poor

2/ Clustering

Clustered dependence: intuition

- Think back to the Gerber, Green, and Larimer (2008) social pressure mailer example.
- Their design: randomly sample households and randomly assign them to different treatment conditions
- But the measurement of turnout is at the individual level
- Violation of iid/random sampling:
  - errors of individuals within the same household are correlated
  - $\rightsquigarrow$ violation of homoskedasticity
- Called clustering or clustered dependence

Clustered dependence: notation

- Clusters: $j = 1, \ldots, m$
- Units: $i = 1, \ldots, n_j$
- n_j is the number of units in cluster j
- $n = \sum_j n_j$ is the total number of units
- Units are (usually) belong to a single cluster:
  - voters in households
  - individuals in states
  - students in classes
  - rulings in judges
- Especially important when outcome varies at the unit-level, y_{ij} and the main independent variable varies at the cluster level, x_j.
- Ignoring clustering is "cheating": units not independent

Clustered dependence: example model

$$\begin{aligned} y_{ij} &= \beta_0 + \beta_1 x_{ij} + \varepsilon_{ij} \\ &= \beta_0 + \beta_1 x_{ij} + v_j + u_{ij} \end{aligned}$$

- $v_j \overset{iid}{\sim} N(0, \rho\sigma^2)$ cluster error component
- $u_{ij} \overset{iid}{\sim} N(0, (1-\rho)\sigma^2)$ unit error component
- v_{ij} and u_{ij} are assumed to be independent of each other
- $\rho \in (0, 1)$ is called the within-cluster correlation.
- What if we ignore this structure and just use ε_{ij} as the error?
- Variance of the composite error is σ^2:

$$\begin{aligned} \mathbb{V}[\varepsilon_{ij}] &= \mathbb{V}[v_j + u_{ij}] \\ &= \mathbb{V}[v_j] + \mathbb{V}[u_{ij}] \\ &= \rho\sigma^2 + (1-\rho)\sigma^2 = \sigma^2 \end{aligned}$$

Lack of independence

- Covariance between two units i and s in the same cluster is $\rho\sigma^2$:

$$\text{Cov}[\varepsilon_{ij}, \varepsilon_{sj}] = \rho\sigma^2$$

- Correlation between units in the same group is just ρ:

$$\text{Cor}[\varepsilon_{ij}, \varepsilon_{sj}] = \rho$$

- Zero covariance of two units i and s in different clusters j and k:

$$\text{Cov}[\varepsilon_{ij}, \varepsilon_{sk}] = 0$$

Example covariance matrix

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{1,1} & \varepsilon_{2,1} & \varepsilon_{3,1} & \varepsilon_{4,2} & \varepsilon_{5,2} & \varepsilon_{6,2} \end{bmatrix}'$$

$$\mathbb{V}[\boldsymbol{\varepsilon}|\mathbf{X}] = \boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 & 0 & 0 & 0 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & 0 & 0 & 0 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ 0 & 0 & 0 & \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ 0 & 0 & 0 & \rho\sigma^2 & \rho\sigma^2 & \sigma^2 \end{bmatrix}$$

Error covariance matrix with clustering

- In general, we can write the covariance matrix as a block diagonal
- By independence, the errors are uncorrelated across clusters:

$$\mathbb{V}[\boldsymbol{\varepsilon}|\mathbf{X}] = \Sigma = \left[\begin{array}{c|c|c|c} \Sigma_1 & \mathbf{0} & \dots & \mathbf{0} \\ \hline \mathbf{0} & \Sigma_2 & \dots & \mathbf{0} \\ \hline & & \ddots & \\ \hline \mathbf{0} & \mathbf{0} & \dots & \Sigma_m \end{array}\right]$$

- Here $\mathbf{0}$ is a matrix of 0's of the appropriate size.

Correcting for clustering

1. Including a dummy variable for each cluster
2. "Random effects" models (take above model as true and estimate ρ and σ^2)
3. Cluster-robust ("clustered") standard errors
4. Aggregate data to the cluster-level and use OLS $\bar{y}_j = \frac{1}{n_j} \sum_i y_{ij}$
    - If n_j varies by cluster, then cluster-level errors will have heteroskedasticity
    - Can use WLS with cluster size as the weights

Cluster-robust SEs

- First, let's write the within-cluster regressions like so:

$$\mathbf{y}_j = \mathbf{X}_j \boldsymbol{\beta} + \boldsymbol{\varepsilon}_j$$

- $\mathbf{y}_j$ is the vector of responses for cluster j, and so on
- We assume that respondents are independent across clusters, but possibly dependent within clusters. Thus, we have

$$\mathbb{V}[\boldsymbol{\varepsilon}_j | \mathbf{X}_j] = \Sigma_j$$

- Remember our sandwich expression:

$$\mathbb{V}[\widehat{\boldsymbol{\beta}} | \mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\Sigma\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}$$

- Under this clustered dependence, we can write this as:

$$\mathbb{V}[\widehat{\boldsymbol{\beta}} | \mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{j=1}^{m} \mathbf{X}_j' \Sigma_j \mathbf{X}_j \right) (\mathbf{X}'\mathbf{X})^{-1}$$

Estimating CRSEs

- Way to estimate this matrix: replace Σ_j with an estimate based on the within-cluster residuals, $\widehat{\boldsymbol{\varepsilon}}_j$:

$$\widehat{\Sigma}_j = \widehat{\boldsymbol{\varepsilon}}_j \widehat{\boldsymbol{\varepsilon}}_j'$$

- Final expression for our cluster-robust covariance matrix estimate:

$$\widehat{\mathbb{V}}[\widehat{\boldsymbol{\beta}}|\mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{j=1}^{m} \mathbf{X}_j' \widehat{\boldsymbol{\varepsilon}}_j \widehat{\boldsymbol{\varepsilon}}_j' \mathbf{X}_j \right) (\mathbf{X}'\mathbf{X})^{-1}$$

- With small-sample adjustment (which is what most software packages report):

$$\widehat{\mathbb{V}}_a[\widehat{\boldsymbol{\beta}}|\mathbf{X}] = \frac{m}{m-1} \frac{n-1}{n-k-1} (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{j=1}^{m} \mathbf{X}_j' \widehat{\boldsymbol{\varepsilon}}_j \widehat{\boldsymbol{\varepsilon}}_j' \mathbf{X}_j \right) (\mathbf{X}'\mathbf{X})^{-1}$$

Example: Gerber, Green, Larimer

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY — VOTE!

MAPLE DR	Aug 04	Nov 04	Aug 06
9995 JOSEPH JAMES SMITH	Voted	Voted	_____
9995 JENNIFER KAY SMITH		Voted	_____
9997 RICHARD B JACKSON		Voted	_____
9999 KATHY MARIE JACKSON		Voted	_____

Social pressure model

```
load("gerber_green_larimer.RData")
social$voted <- 1 * (social$voted == "Yes")
social$treatment <- factor(social$treatment, levels = c("Control",
    "Hawthorne", "Civic Duty", "Neighbors", "Self"))
mod1 <- lm(voted ~ treatment, data = social)
coeftest(mod1)
```

```
##
## t test of coefficients:
##
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)          0.29664    0.00106  279.53  < 2e-16 ***
## treatmentHawthorne   0.02574    0.00260    9.90  < 2e-16 ***
## treatmentCivic Duty  0.01790    0.00260    6.88  5.8e-12 ***
## treatmentNeighbors   0.08131    0.00260   31.26  < 2e-16 ***
## treatmentSelf        0.04851    0.00260   18.66  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Social pressure model, CRSEs

- No canned CRSE in R, we posted some code on Canvas:

```
source("vcovCluster.R")
coeftest(mod1, vcov = vcovCluster(mod1, "hh_id"))
```

```
##
## t test of coefficients:
##
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           0.29664    0.00131  226.52  < 2e-16 ***
## treatmentHawthorne    0.02574    0.00326    7.90  2.8e-15 ***
## treatmentCivic Duty   0.01790    0.00324    5.53  3.2e-08 ***
## treatmentNeighbors    0.08131    0.00337   24.13  < 2e-16 ***
## treatmentSelf         0.04851    0.00330   14.70  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Cluster-robust standard errors

- CRSE do not change our estimates $\widehat{\boldsymbol{\beta}}$, cannot fix bias
- CRSE is consistent estimator of $\mathbb{V}[\widehat{\boldsymbol{\beta}}]$ given clustered dependence
  - Relies on independence between clusters, dependence within clusters
  - Doesn't depend on the model we present
  - CRSEs usually > conventional SEs—use when you suspect clustering
- Consistency of the CRSE are in the number of groups, not the number of individuals
  - CRSEs can be incorrect with a small (< 50 maybe) number of clusters
  - Block bootstrap can be a useful alternative (see Gov 2002)

3/ Serial Correlation

Time dependence: serial correlation

- Sometimes we deal with data that is measured over time, $t = 1, \ldots, T$
- Examples: a country over several years or a person over weeks/months
- Often have serially correlated: errors in one time period are correlated with errors in other time periods
- Many different ways for this to happen, but we often assume a very limited type of dependence called AR(1).

AR(1) model

- Model for the mean:

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

- Autoregressive error:

$$u_t = \rho u_{t-1} + e_t \quad \text{where} \quad |\rho| < 1$$

- $e_t \sim N(0, \sigma_e^2)$
- ρ is an unknown autoregressive coefficient and measures the dependence/correlation between the errors and lagged errors
- Just one of many possible time-series models: AR(2) has $u_t = \rho u_{t-1} + \delta u_{t-2} + e_t$
- Model could be wrong!

Error structure of the AR(1) model

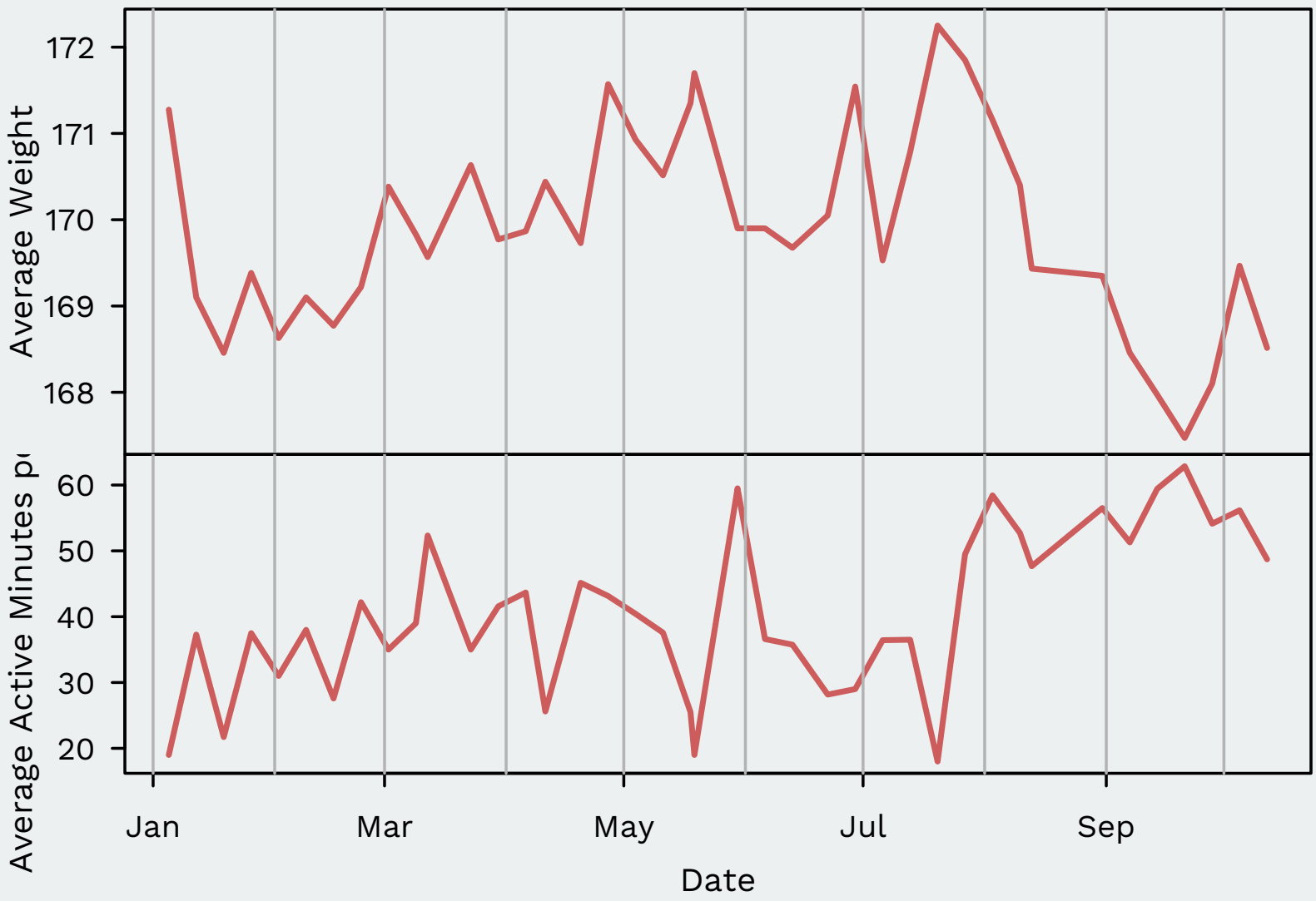
$$\mathbb{V}[\mathbf{u}] = \Sigma = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{bmatrix}$$

- What is this saying?
  - $\text{Cov}[u_1, u_2] = \sigma^2 \rho$
  - $\text{Cov}[u_1, u_3] = \sigma^2 \rho^2$
  - $\text{Cov}[u_1, u_4] = \sigma^2 \rho^3$
  - Covariance/correlation decreases as time between errors grows (because $|\rho| < 1$)
- ρ is usually positive, which means we underestimate the variance

Detecting and fixing serial correlation

- Detection:
  - Plot residuals over time
  - Formal tests (Durbin-Watson statistics)
- Correction:
  - Use SEs that are robust to serial correlation
  - AR corrections (e.g. Prais-Winston, Cochrane-Orcutt, etc)
  - Lagged dependent variables or other dynamic models
  - Transformations via first-differencing methods

Example: weight and activity



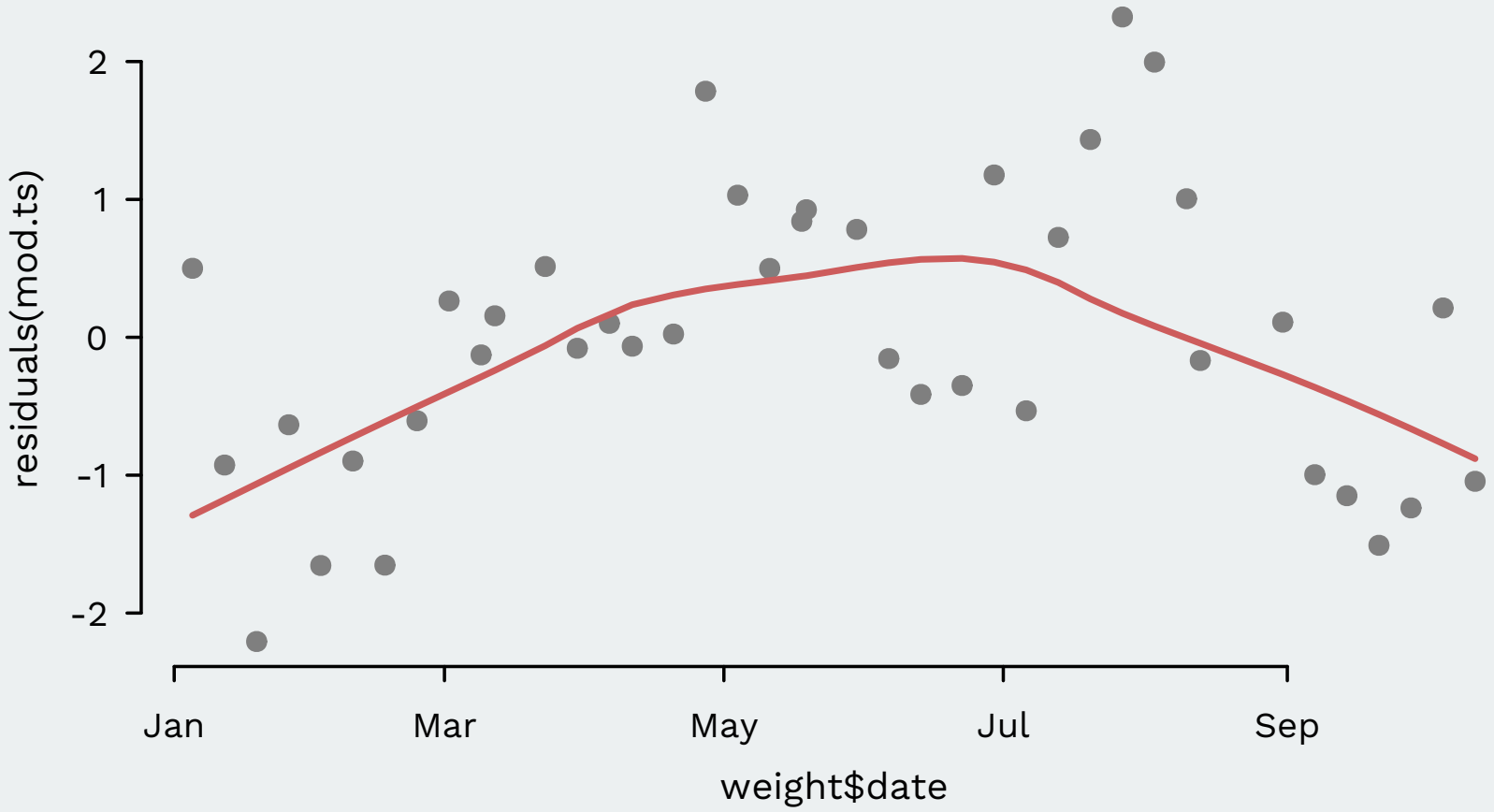
Regression ignoring serial dependence

```
mod.ts <- lm(weight ~ active.mins, data = weight)
summary(mod.ts)
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 171.5521     0.5832  294.16   <2e-16 ***
## active.mins  -0.0409     0.0138   -2.96   0.0053 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.06 on 38 degrees of freedom
## Multiple R-squared:  0.187,  Adjusted R-squared:  0.166
## F-statistic: 8.74 on 1 and 38 DF,  p-value: 0.00534
```

Residuals over time

```
plot(x=weight$date, y=residuals(mod.ts))
lines(lowess(x=weight$date, y=residuals(mod.ts)),
      col = 'indianred', lwd = 2)
```



A formal test: Durbin-Watson

- Null, $H_0 : \rho = 0$
- Alternative, $H_a : \rho \neq 0$
- Durbin-Watson statistic:

$$DW = \frac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n} \hat{u}_t^2} \quad \text{where} \quad DW \approx 2(1 - \widehat{\rho})$$

- If $DW \approx 2$ then $\widehat{\rho} \approx 0$
- $DW < 1$: strong evidence of positive serial correlation
- $DW > 3$: strong evidence of negative serial correlation

Durbin-Watson on weight

```
dwtest(mod.ts)
```

```
##
##  Durbin-Watson test
##
## data:  mod.ts
## DW = 0.75, p-value = 0.000002
## alternative hypothesis: true autocorrelation is greater than 0
```

Corrections: HAC standard errors

- We can generalize the HC/robust standard errors to be heteroskedastic and autocorrelation consistent (HAC) standard errors.
- Autocorrelation is just another term for serial correlation
- Very similar to HC/robust:
  - $\widehat{\boldsymbol{\beta}}$ remain as our esitmates
  - HAC SEs are consistent for $\mathbb{V}[\widehat{\boldsymbol{\beta}}]$ in the presence of heteroskedasticity and/or serial correlation
  - Can use the `sandwich` package in R, with covariance function `NeweyWest`

Example: Newey-West standard errors

```
coeftest(mod.ts, vcov = NeweyWest)
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 171.5521     0.8186  209.55   <2e-16 ***
## active.mins  -0.0409     0.0212   -1.93    0.061 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Summary

- Violations of homoskedasticity can come in many forms
  - Non-constant error variance
  - Clustered dependence
  - Serial dependence
- Use plots or formal tests to detect heteroskedasticity
- "Robust SEs" of various forms are consistent even when these problems are present
  - White HC standard errors
  - Cluster-robust standard errors
  - Newey-West HAC standard errors

4/ What's next for you?

Where are you?

- You've been given a powerful set of tools

Your new weapons

- **Probability**: if we knew the true parameters (means, variances, coefficients), what kind of data would we see?
- **Inference**: what can we learn about the truth from the data we have?
- **Regression**: how can we learn about relationships between variables?

You need more training!

- We got through a ton of solid foundation material, but to be honest, we have basically got you to the state of the art in political science in the 1970s

What else to learn?

- Non-linear models (Gov 2001/Stat E-200)
  - what if y_i is not continuous?
- Maximum likelihood (Gov 2001/Stat E-200)
  - a general way to do inference and derive estimators for almost any model
- Causal inference (Gov 2002, Stat 186)
  - how do we make more plausible causal inferences?
  - what happens when treatment effects are not constant?
- Bayesian statistics (Stat 120/220)
  - an alternative approach to inference based on treating parameters as random variables
- Machine Learning (Stat 121/CS 109)
  - how to handle massive data?
  - how can we use text as data?

Glutton for punishment?

- Stat 110/111: rigorous introduction to probability and inference
- Stat 210/211: Stats PhD level introduction to probability and inference (measure theory)
- Stat 221: statistical computing

Thanks!

Fill out your evaluations!