Contents

References

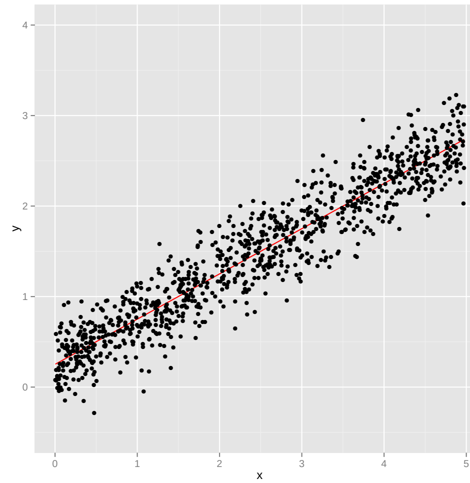
- Wooldridge (2013) chapter 8
- Stock and Watson (2009) chapter 10.5
- Angrist and Pischke (2014) chapter 5

1 Introduction

Introduction

Homoskedastic

MLR.5 (homoskedasticity) $\text{Var}(\epsilon_i|X) = \sigma_\epsilon^2$ is often an implausible assumption



Code

- *Heteroskedasticity* is when $\text{Var}(\epsilon|X)$ varies with X

Introduction

- Role of homoskedasticity:
  - Was *not* needed to show OLS unbiased and consistent
  - Was needed to calculate $\text{Var}(\hat{\beta}|X)$
- If errors are heteroskedastic then still have $\text{E}[\hat{\beta}_j] = \beta_j$ and $\hat{\beta}_j \xrightarrow{p} \beta_j$ (assuming MLR.1-4), but usual standard error is invalid
- If there is heteroskedasticity, the variance of $\hat{\beta}_1$ is different, but we can fix it
- "heteroskedasticity robust standard errors" / "heteroscedasticity consistent standard errors" / "Eicker–Huber–White standard errors"

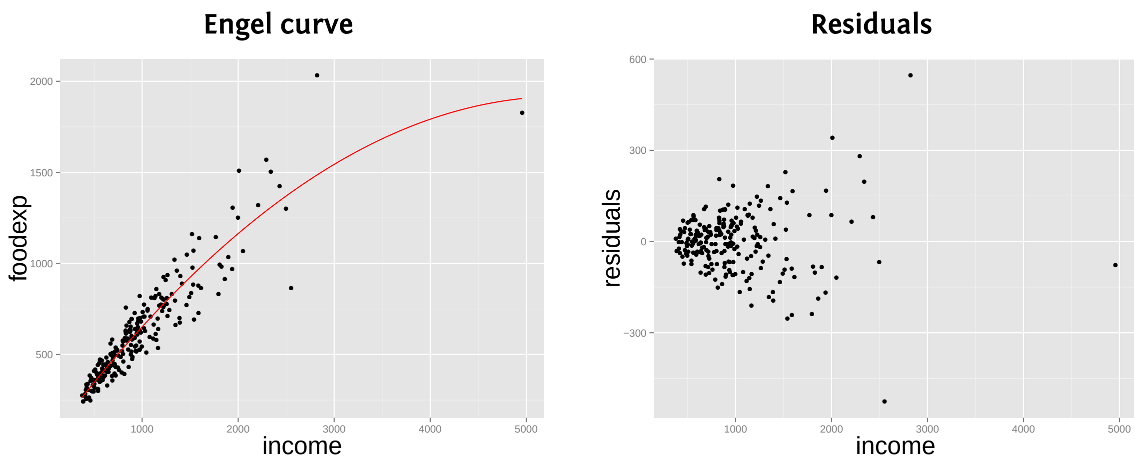
Example: Engel curves

- An *Engel curve* describes how a consumer's purchases of a good (like food) varies with the consumer's total resources (income or total spending)
- Named after Engel (1857) — looked at food expenditures and income of Belgian families
  - See Chai and Moneta (2010) for a description of Engel's work and Lewbel (2008) for a brief summary of modern research on Engel curves

---

Example: Engel (1857) curve

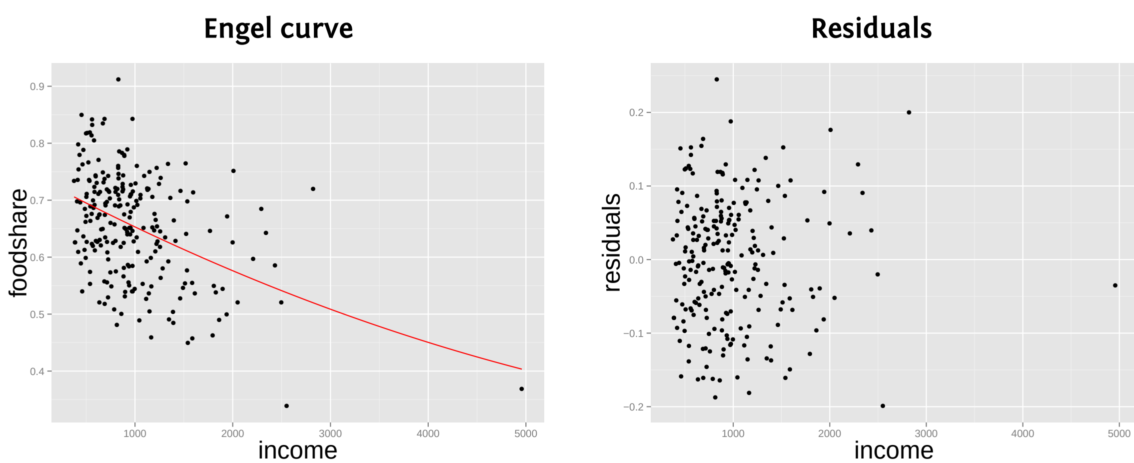
$$foodexp = \beta_0 + \beta_1 income + \beta_2 income^2 + \epsilon$$



- Income and food expenditure measured in Belgian Francs

Example: Engel (1857) curve in shares

$$\frac{foodexp}{income} = foodshare = \beta_0 + \beta_1 income + \beta_2 income^2 + \epsilon$$



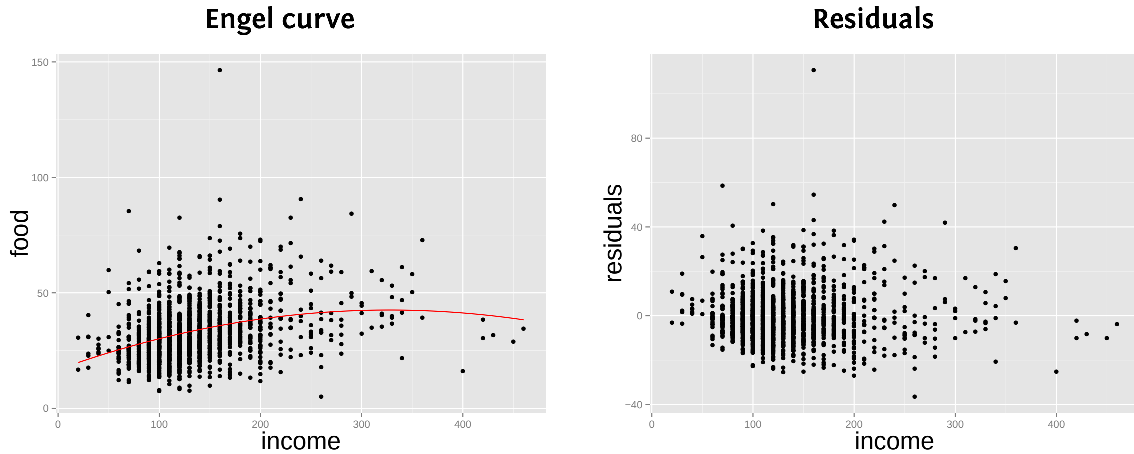
Code

Engel curves with modern data

- There are still papers about estimating Engel curves
  - Banks, Blundell, and Lewbel (1997), Blundell, Duncan, and Pendakur (1998), Blundell, Browning, and Crawford (2003)

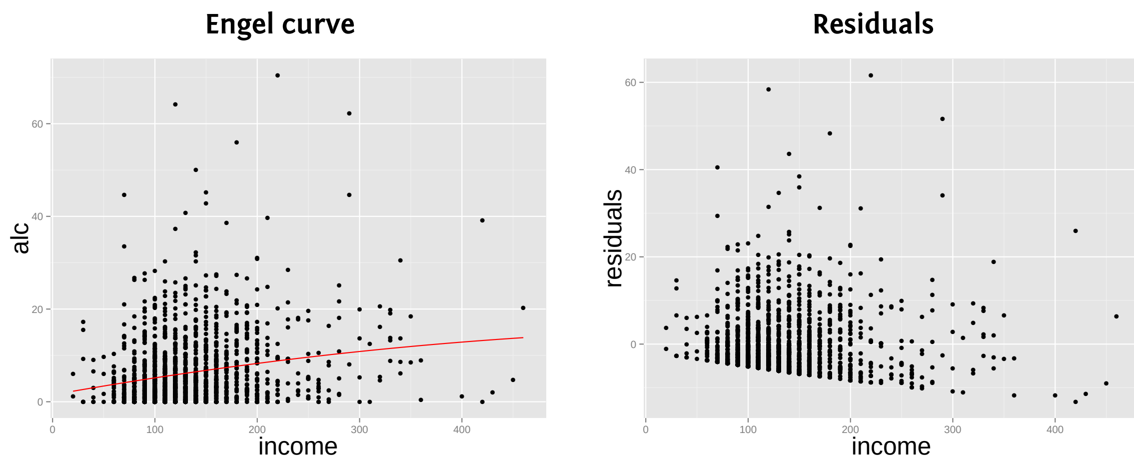
- Next slides uses data from British Family Expenditure Surveys (FES) for 1980-1982 (same data as Blundell, Duncan, and Pendakur (1998))

Engel curves with modern data: food



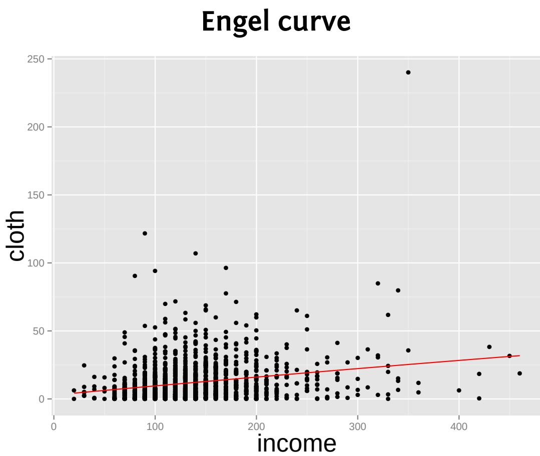
Code

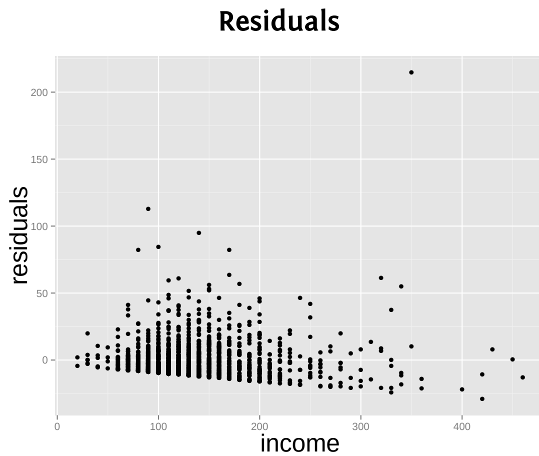
Engel curves with modern data: alcohol



Code

Engel curves with modern data: clothing





Code

2 Consequences of heteroskedasticity

Consequences of heteroskedasticity

- OLS still unbiased and consistent assuming

MLR.1 (linear model)

MLR.2 (independence) $\{(x_{1,i}, x_{2,i}, y_i)\}_{i=1}^n$ is an independent random sample

MLR.3 (rank condition) no multicollinearity: no $x_{j,i}$ is constant and there is no exact linear relationship among the $x_{j,i}$

MLR.4 (exogeneity) $\mathrm{E}[\epsilon_i | x_{1,i}, \ldots, x_{k,i}] = 0$

- Homoskedastic-only standard errors are incorrect,

$$\mathrm{Var}(\hat{\beta}_j | X) \neq \frac{\sum_{i=1}^n \sigma_\epsilon^2}{\sum_{i=1}^n \tilde{x}_{j,i}^2} \text{ and } \operatorname{plim} \frac{\sum_{i=1}^n \hat{\epsilon}_i^2}{\sum_{i=1}^n \tilde{x}_{j,i}^2} \neq \mathrm{Var}(\hat{\beta}_j)$$

- t (and F) statistics formed using homoskedasticity-only standard errors do not have t (or F) distributions (not even asymptotically)

- Hypothesis tests and confidence intervals formed using homoskedasticity-only standard errors are invalid

- OLS is not BLUE

3 $\mathrm{Var}(\hat{\beta})$ with heteroskedasticity

$\mathrm{Var}(\hat{\beta})$ **with heteroskedasticity**

- *Central limit theorem*: for i.i.d. data, if $\mathrm{E}[y_i] = \mu$ and $\mathrm{Var}(y_i) = \sigma^2$ for all i then

$$\sqrt{n}\,(\bar{y}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

- Recall how to how to show $\hat{\beta}$ is asymptotically normal in bivariate regression

$$\begin{aligned}\sqrt{n}(\hat{\beta}_1 - \beta_1) =& \sqrt{n}\left(\frac{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})y_i}{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2} - \beta_1\right) \\ =& \sqrt{n}\left(\frac{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})(\beta_0 + \beta_1 x_i + \epsilon_i)}{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2} - \beta_1\right) \\ =& \frac{\sqrt{n}\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})\epsilon_i}{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2}\end{aligned}$$

  - Already showed that $\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2 \xrightarrow{p} \mathrm{Var}(x)$
  - Need to apply CLT to $\sqrt{n}\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})\epsilon_i$. First note that

$$\sqrt{n}\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})\epsilon_i = \sqrt{n}\frac{1}{n}\sum_{i=1}^n (x_i - \mathrm{E}[x])\epsilon_i + \underbrace{(\mathrm{E}[x] - \bar{x})}_{\xrightarrow{p} 0}\underbrace{\sqrt{n}\frac{1}{n}\sum_{i=1}^n \epsilon_i}_{\xrightarrow{d} N(0, \mathrm{Var}(\epsilon))}$$

  - Let $w_i = (x_i - \mathrm{E}[x])\epsilon_i$, can apply CLT to w_i because
    * $\mathrm{E}[w_i] = \mathrm{E}[x_i \epsilon_i] = 0$
    * Observations are independent
    * Assume $\mathrm{Var}(w_i) = \mathrm{E}[(x_i - \mathrm{E}[x])^2 \epsilon_i^2]$ exists

    then $\frac{1}{\sqrt{n}}\sum_{i=1}^n (x_i - \mathrm{E}[x])\epsilon_i \xrightarrow{d} N\left(0, \mathrm{E}\left[(x_i - \mathrm{E}[x])^2 \epsilon_i^2\right]\right)$
  - Can conclude that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n (x_i - \bar{x})\epsilon_i \xrightarrow{d} N\left(0, \mathrm{E}\left[(x_i - \mathrm{E}[x])^2 \epsilon_i^2\right]\right)$$

- By Slutsky's theorem,

$$\begin{aligned}\sqrt{n}(\hat{\beta}_1 - \beta_1) =& \frac{\sqrt{n}\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})\epsilon_i}{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2} \\ &\xrightarrow{d} N\left(0, \frac{\mathrm{E}\left[(x_i - \mathrm{E}[x])^2 \epsilon_i^2\right]}{\mathrm{Var}(x)^2}\right)\end{aligned}$$

  or equivalently,

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\mathrm{E}\left[(x_i - \mathrm{E}[x])^2 \epsilon_i^2\right]}{n\mathrm{Var}(x)^2}}} \xrightarrow{d} N(0, 1)$$

- By Slutsky's theorem can replace $\sqrt{\frac{\mathrm{E}\left[(x_i - \mathrm{E}[x])^2 \epsilon_i^2\right]}{\mathrm{Var}(x)^2}}$ by consistent estimators, and

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2 \hat{\epsilon}_i^2}{n\left(\frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2\right)^2}}} \xrightarrow{d} N(0,1)$$

- Similar reasoning applies to multivariate regression

$$\frac{\hat{\beta}_j - \beta_j}{\sqrt{\frac{\frac{1}{n}\sum_{i=1}^n \tilde{x}_{j,i}^2 \hat{\epsilon}_i^2}{n\left(\frac{1}{n}\sum_{i=1}^n \tilde{x}_{j,i}^2\right)^2}}} \xrightarrow{d} N(0,1)$$

- $\sqrt{\frac{\frac{1}{n}\sum_{i=1}^n \tilde{x}_{j,i}^2 \hat{\epsilon}_i^2}{n\left(\frac{1}{n}\sum_{i=1}^n \tilde{x}_{j,i}^2\right)^2}}$ is called the heteroskedasticity robust standard error or the Eicker-Huber-White standard error
  - Statistics: Eicker (1967), Huber (1967)
  - Econometrics: White (1980)

3.1 Calculating in R

R: Heteroskedasticity robust standard errors

```r
## calculating Heteroskedasticity robust standard errors
engelCurve <- lm(foodexp ~ income, data=engel)
## calculate by hand
hetSE <- sqrt(mean( (engel$income - mean(engel$income))^2 *
                    residuals(engelCurve)^2) /
              (nrow(engel)*mean( (engel$income - mean(engel$income))^2 )^2 ))

## use "sandwich" package
library(sandwich)
sqrt(vcovHC(engelCurve, type="HC0")[2,2])

## test H_0: each coefficient = 0 separately
library(lmtest)
coeftest(engelCurve, vcov=vcovHC(engelCurve,type="HC0"))
## compare with homoskedastic standard errors
coeftest(engelCurve)

## we would do F-tests with
## waldtest(engelCurve, vcov=vcovHC(engelCurve,type="HC0")) or
## lrtest(engelCurve, vcov=vcovHC(engelCurve,type="HC0"))

## Or even easier, just use lfe packaage
library(lfe)
engelCurve2 <- felm(foodexp ~ income, data=engel)
summary(engelCurve2, robust=TRUE)
```

Code

4 Examples

Growth, GDP, and education

	GDP 1960	Education 1960	Both
(Intercept)	1.796***	0.958**	0.895**
	(0.378)	(0.418)	(0.389)
	[0.456]	[0.454]	[0.428]
rgdp60	0.047		−0.485***
	(0.095)		(0.146)
	[0.083]		[0.149]
yearsschool		0.247***	0.640***
		(0.089)	(0.144)
		[0.084]	[0.154]
Num. obs.	65	65	65

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$
Homoskedastic standard errors in parentheses
Heteroskedastic standard errors in brackets

Test scores and student teacher ratios

	1	2	3
β_0	698.93***	638.73***	640.32***
	(9.47)	(7.45)	(5.77)
	[10.46]	[7.37]	[6.06]
$\frac{student}{teacher}$	−2.28***	−0.65*	−0.07
	(0.48)	(0.35)	(0.28)
	[0.52]	[0.36]	[0.29]
Income		1.84***	1.49***
		(0.09)	(0.07)
		[0.12]	[0.10]
English learners			−0.49***
			(0.03)
			[0.03]
Num. obs.	420	420	420

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$
Homoskedastic standard errors in parentheses
Heteroskedastic standard errors in brackets

---

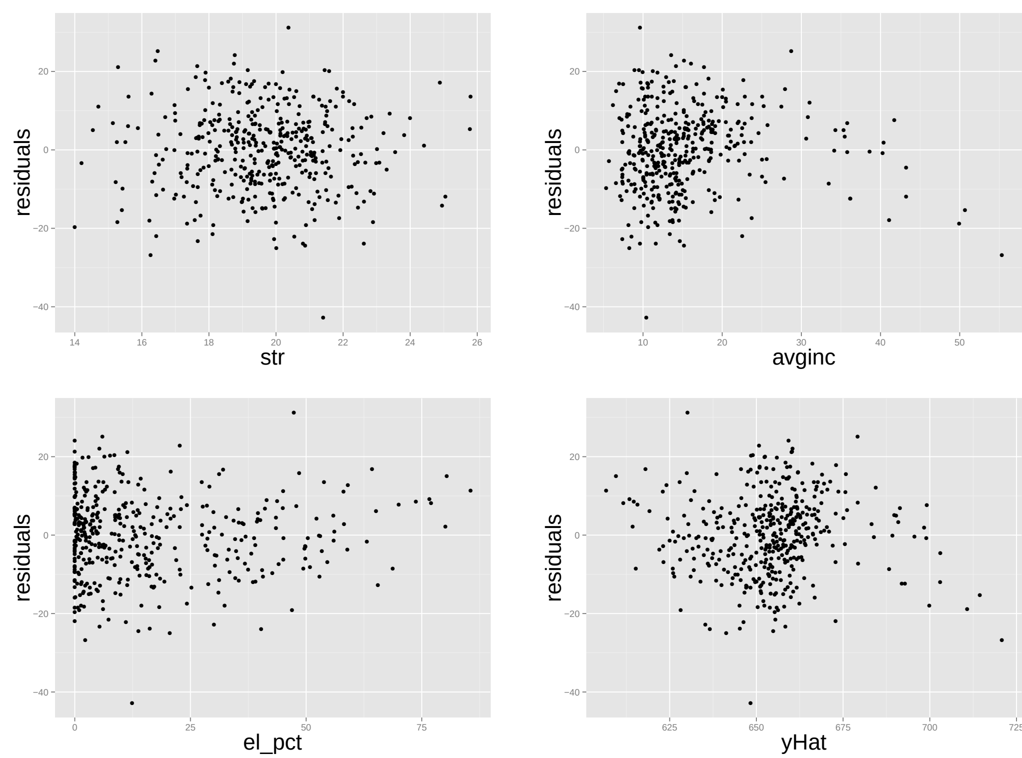
[0] Code
[0] Code

5 Detecting heteroskedasticity

Detecting heteroskedasticity

- Generally best to assume heteroskedasticity and use heteroskedasticity consistent standard errors
- If you have homoskedasticity, but use heteroskedastic standard errors, the heteroskedaticity consistent standard error converges to the homoskedicity-only standard error, so in large samples it will make no difference
- But if you assume homoskedastic when there is heteroskedasticity, your standard errors will be inconsistent, and your hypothesis tests and confidence intervals will be invalid
- Nonetheless occasionally might want to check for heteroskedasticity
  - Visually: plot $\hat{\epsilon}_i$ against $x_{j,i}$ and/or $\hat{y}_i$
  - Formally: can test H_0 : homoskedasticity, see Wooldridge (2013) chapter 8 for details

Example: Test scores and student teacher ratios



[0] Code

6 Heteroskedasticity and efficiency

Heteroskedasticity and efficiency

- The Gauss-Markov theorem requires homoskedasticity
- When there is heteroskedasticity, OLS is *not* the best linear unbiased estimator
- What is the best linear unbiased estimator with heteroskedasticity?
- Model: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$
- Assume MLR.1-4
- Suppose we know $\mathrm{Var}(\epsilon_i|x_i) = \sigma^2 h(x_i)$ for some known function h
- Consider dividing each observation by $\sqrt{h(x_i)}$

$$\begin{aligned} \frac{y_i}{\sqrt{h(x_i)}} =& \beta_0 \frac{1}{\sqrt{h(x_i)}} + \beta_1 \frac{x_i}{\sqrt{h(x_i)}} + \frac{\epsilon_i}{\sqrt{h(x_i)}} \\ y_i^* =& \beta_0 x_{0,i}^* + \beta_1 x_{1,i}^* + \epsilon_i^* \end{aligned} \tag{1}$$

- Then
  - Assumptions MLR.1 (linear model), MLR.2 (independent observations), MLR.3 (rank condition), and MLR.4 $\mathrm{E}[\epsilon_i^*|x_i^*] = 0$ hold for (1)
  - (1) is homoskedastic

$$\begin{aligned} \mathrm{Var}(\epsilon_i^*|x_i^*) =& \mathrm{Var}\left(\frac{\epsilon_i}{\sqrt{h(x_i)}} \middle| x_i \right) \\ =& \frac{1}{h(x_i)} h(x_i)\sigma^2 = \sigma^2 \end{aligned}$$

- So OLS of (1) is BLUE
- This is called *weighted least squares* (WLS)
- If do not know $h(x_i)$ but can estimate it this is called *feasible generalized least squares* (FGLS)
- In most empirical work do not know $h(x_i)$ and often cannot estimate it well, so usually just use OLS

---

7 Standard errors for dependent data

Standard errors for dependent data

- Assume:

MLR.1 (linear model)

MLR.2 **(not too dependent)** $\{(x_{j,i}, , y_i)\}_{i=1}^n$ **is not too dependent**

MLR.3 (rank condition) no multicollinearity: no $x_{j,i}$ is constant and there is no exact linear relationship among the $x_{j,i}$

MLR.4 (strict exogeneity) $\mathrm{E}[\epsilon_i | \{x_{1,j}, ..., x_{k,j}\}_{j=1}^n] = 0$

- As with heteroskedasticity, when observations are not independent (and not too dependent):
  - OLS remains unbiased and consistent
  - Usual homoskedastic-only or heteroskedasticity robust standard errors are incorrect
  - If you do not correct standard errors, hypothesis tests and confidence intervals are invalid
  - OLS is not BLUE

The assumption that observations are not too dependent is deliberately vague. Exactly what "not too dependent" means varies with the form the dependence (see next slide). To see why such an assumption is needed, suppose we just observe a single variable y_i and want to estimate $\mathrm{E}[Y]$. Imagine an extreme form of dependence, y_1 is randomly drawn from some distribution, but then all other y_i are set equal to y_1,

$$y_1 = y_2 = \cdots = y_n.$$

Then $\bar{y}$ is an unbiased estimate of $\mathrm{E}[Y]$, but it is not consistent because $\bar{y} = y_1$ will not become close to $\mathrm{E}[Y]$ as the sample size increases. This situation is typical. Too much dependence is not often a problem for unbiasedness, but you lose consistency (and asymptotic normality) with too much dependence.

To see that some dependence should be okay, consider a situation where for i odd, y_i is i.i.d from some distribution, and for i even, $y_i = y_{i-1}$. In this case, the sample mean of all observations is exactly equal to the sample mean of just the odd observations,

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n/2}\sum_{j=0}^{n/2} y_{2j+1} \equiv \bar{y}^{odd}.$$

Since the odd observations are independent, $\bar{y}^{odd}$ is consistent and asymptotically normal. Then, since $\bar{y} = \bar{y}^{odd}$, it is also consistent and asymptotically normal. However, the variance of $\bar{y}$ is larger than in the independent case. In particular,

$$\mathrm{Var}(\bar{y}) = \mathrm{Var}(\bar{y}^{odd}) = \frac{\mathrm{Var}(Y)}{n/2}.$$

This result is also typical. When observations are not too dependent (which will mean observations that are far apart are approximately independent), then estimators are consistent and asymptotically normal, but with a different (and usually larger) variance.

Standard errors for dependent data

- To correct standard errors for dependent data, we need to know something about the form of dependence
- Common forms of dependence:
  - *Clustering*: observations can be organized into clusters (groups); pairs of observations in the same cluster are dependent, pairs of observations in different clusters are independent
  - *Autocorrelation* (or serial correlation): observations are taken over time; y_t is correlated with y_{t-k}
- Can correct standard errors for clustering and autocorrelation
- Clustered standard errors and autocorrelation consistent standard errors are often much larger than standard errors that assume independence

---

7.1 Clustering

Clustering

- *Clustering*: observations can be organized into clusters (groups); pairs of observations in the same cluster are dependent, pairs of observations in different clusters are independent

---

Clustering - examples

- Blimpo (2014) (from 2015 midterm)
  - Observations of students test scores, about 1500 students from 100 schools
  - Students' test scores in the same school unlikely to be independent, but reasonable to think students in different schools are independent
- Rodrik (2013) (from the 2013 midterm)

$$\Delta v_{ij} = \beta_u \log v_{ij} + \delta_0 + \delta_1 \text{Industry}_{1,ij} + \cdots + \\ + \cdots + \delta_{J-1} \text{Industry}_{J-1,ij} + \epsilon_{ij}$$

  - Data on 20 industries from 118 countries
  - $\log v_{ij}$ = log labor productivity in manufacturing industry i in country j ten years in the past
  - Δv_{ij} be the annual growth rate of labor productivity in industry i in country j over the past ten years

- Unrealistic to assume to assume that industries within the same country are independent, so $\mathrm{E}[\epsilon_{ij}\epsilon_{ik}] \neq 0$
- More reasonable to assume that observations in different countries are independent

---

Clustering

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij}$$

- J clusters, $n(j)$ observations in each cluster
- Assume observations in different clusters are independent (observations in the same cluster can be arbitrarily dependent)
- Then

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\sum_{j=1}^{J}\left(\sum_{i=1}^{n(j)}(x_{ij}-\bar{x})\hat{\epsilon}_{ij}\right)^2}{\frac{1}{n}\left(\sum_{i,j}(x_{ij}-\bar{x})^2\right)^2}}} \xrightarrow{d} N(0,1)$$

  as $J \to \infty$
- Can use these standard errors and perform t and F tests as usual
- These standard errors allow heteroskedasticity and clustering
- This is an asymptotic result for a large number of clusters, many empirical applications have relatively few clusters, there is no consesus on the best thing to do when you have few clusters, but there are number of papers about it, see Angrist and Pischke (2009) chapter 8
- For clustered standard errors in R use the felm from the lfe package

---

7.2 Autocorrelation

Autocorrelation

- Observations over time

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

- Unlikely that ϵ_t independent of ϵ_{t-k}
- Need to correct standard error of $\hat{\beta}$

- Correct formula called *Newey-West* or *autocorrelation consistent* or *HAC* (heteroskedasticity and autocorrelation consistent) standard errors

$$\widehat{se}(\hat{\beta}_1) = \frac{\sum_{\ell=-(T^{1/3})}^{T^{1/3}} \left[(1 - \left| \frac{\ell}{T^{1/3}} \right|) \frac{1}{T} \sum_{t=1}^{T} (x_t - \bar{x}) \hat{\epsilon}_t (x_{t+\ell} - \bar{x}) \hat{\epsilon}_{t+\ell} \right]}{\left(\sum_{t=1}^{T} (x_t - \bar{x})^2 \right)^2}$$

- See Wooldridge (2013) section 12.5 for details, or Mikusheva and Schrimpf (2007) lectures 2 & 3, & recitation 2 for more discussion

---

References

References

Angrist, J.D. and J.S. Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Angrist, Joshua D and Jörn-Steffen Pischke. 2014. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton University Press.

Banks, James, Richard Blundell, and Arthur Lewbel. 1997. "Quadratic Engel curves and consumer demand." *Review of Economics and Statistics* 79 (4):527–539.

Blimpo, Moussa P. 2014. "Team Incentives for Education in Developing Countries: A Randomized Field Experiment in Benin." *American Economic Journal: Applied Economics* 6 (4):90–109. URL http://www.aeaweb.org/articles.php?doi=10.1257/app.6.4.90.

Blundell, Richard, Alan Duncan, and Krishna Pendakur. 1998. "Semiparametric estimation and consumer demand." *Journal of Applied Econometrics* 13 (5):435–461.

Blundell, Richard W, Martin Browning, and Ian A Crawford. 2003. "Nonparametric Engel curves and revealed preference." *Econometrica* 71 (1):205–240.

Chai, Andreas and Alessio Moneta. 2010. "Retrospectives engel curves." *The Journal of Economic Perspectives* 24 (1):225–240.

Eicker, Friedhelm. 1967. "Limit theorems for regressions with unequal and dependent errors." In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1. 59–82.

Engel, Ernst. 1857. "Die Productions- und Consumptionsverhaeltnisse des Koenigsreichs Sachsen." *Zeitschrift des Statistischen Bureaus des Koniglich Sachsischen Ministeriums des Inneren* (8-9).

Huber, Peter J. 1967. "The behavior of maximum likelihood estimates under nonstandard conditions." In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1. 221–33.

Lewbel, Arthur. 2008. "Engel curve." In *The New Palgrave Dictionary of Economics*, edited by Steven N. Durlauf and Lawrence E. Blume. Basingstoke: Palgrave Macmillan. URL http://www.dictionaryofeconomics.com/article?id=pde2008_E000085.

Mikusheva, Anna and Paul Schrimpf. 2007. "14.384 Time Series Analysis, Fall 2007 (revised 2009)." URL http://ocw.mit.edu/courses/economics/14-384-time-series-analysis-fall-2013/lecture-notes/.

Rodrik, Dani. 2013. "Unconditional Convergence in Manufacturing." *The Quarterly Journal of Economics* 128 (1):165–204. URL http://qje.oxfordjournals.org/content/128/1/165.abstract.

Stock, J.H. and M.W. Watson. 2009. *Introduction to Econometrics, 2/E*. Addison-Wesley.

White, Halbert. 1980. "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica: Journal of the Econometric Society* :817–838URL http://www.jstor.org/stable/1912934.

Wooldridge, J.M. 2013. *Introductory econometrics: A modern approach*. South-Western.