Regression III: Advanced Methods

Bill Jacoby
Michigan State University

http://polisci.msu.edu/icpsr/regress3

Goals of the lecture

- Introduce the idea of "**robustness**", in particular, distributional robustness
- Discuss ***robust*** and ***resistant regression*** methods that can be used when there are unusual observations or skewed distributions
- Particular emphasis will be placed on:
  - M-Estimation
  - Bounded-Influence Regression

Defining "Robust" (1)

- Following Huber (1981) we will interpret ***robustness*** as insensitivity to small deviations from the assumptions the model imposes on the data.
- In particular, we are interested in ***distributional robustness,*** and the impact of skewed distributions and/or outliers on regression estimates.
  - In this context, "robust" refers to the shape of a distribution (specifically, when it differs from the theoretically assumed distribution
  - Although conceptually distinct, distributional robustness and outlier resistance are, for practical purposes, synonymous
  - Robust can also be used to describe standard errors that are adjusted for non-constant error variance (This is a different topic!!! We will discuss later …)

Defining "Robust" (2)

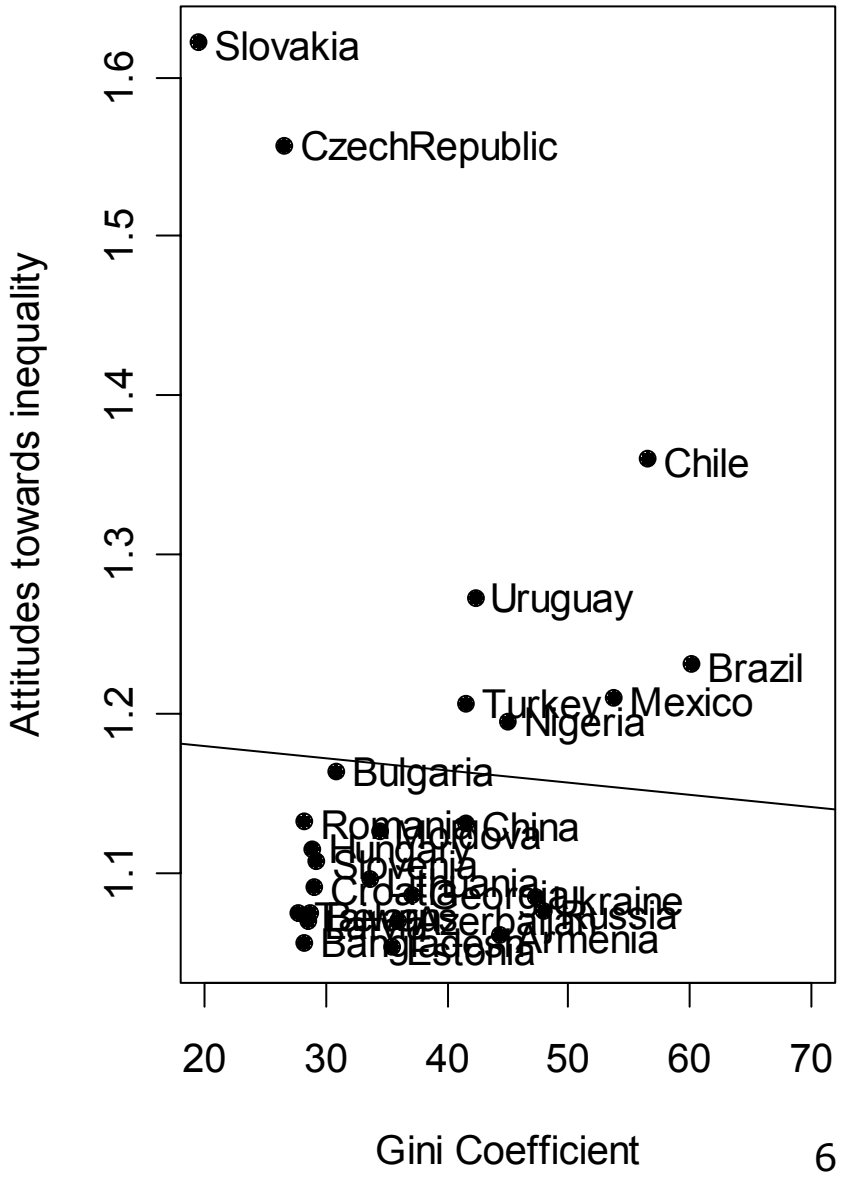
- Statistical inferences are based both on observations and on prior assumptions about underlying distributions and relationships between variables
- Although these assumptions are never exactly true, some statistical models are more sensitive to small deviations from these assumptions than others
- A model is robust if it has the following features:
  - Reasonably efficient and unbiased
  - Small deviations from the model assumptions will not substantially impair the performance of the model
  - Somewhat larger deviations will not invalidate the model completely

Influential Cases and OLS

- ***OLS is not robust to outliers***. It can produce misleading results if unusual cases go undetected—even a single case can have a significant impact on the fit of the regression surface
- Moreover, the efficiency of the OLS regression can be hindered by heavy-tailed distributions and outliers
- Diagnostics can be used to detect non-normal error distributions and influential cases. But, once they are found, what should we do?
- Investigate whether the deviations are a symptom of model failure that can be repaired by deleting cases, transformations, or adding more terms to the model
  - In cases when the unusual cases cannot be remedied this way,, ***robust regression*** can provide an alternative to OLS
  - Robust regression can also be used as a diagnostic tool

Inequality model revisited (1)

- The scatterplot clearly shows Slovakia and the Czech Rebublic as unusual cases—they are outliers in Y and have high leverage according to X, a combination that results in high influence
- The OLS fit to the data indicates that these observations significantly pull the line upward



R script for plot with case names

```
> plot(gini, secpay, pch=16,
+    xlim=c(20,70),
+    xlab="Gini Coefficient",
+    ylab="Attitudes towards inequality")
> chw <- par()$cxy[1]
> #adds space from data point
> text(gini+chw, secpay,
+    labels=row.names(Nondemo), adj=0)
> #adj=0 adjusts the text on the x axis
> abline(lm(secpay~gini))
```

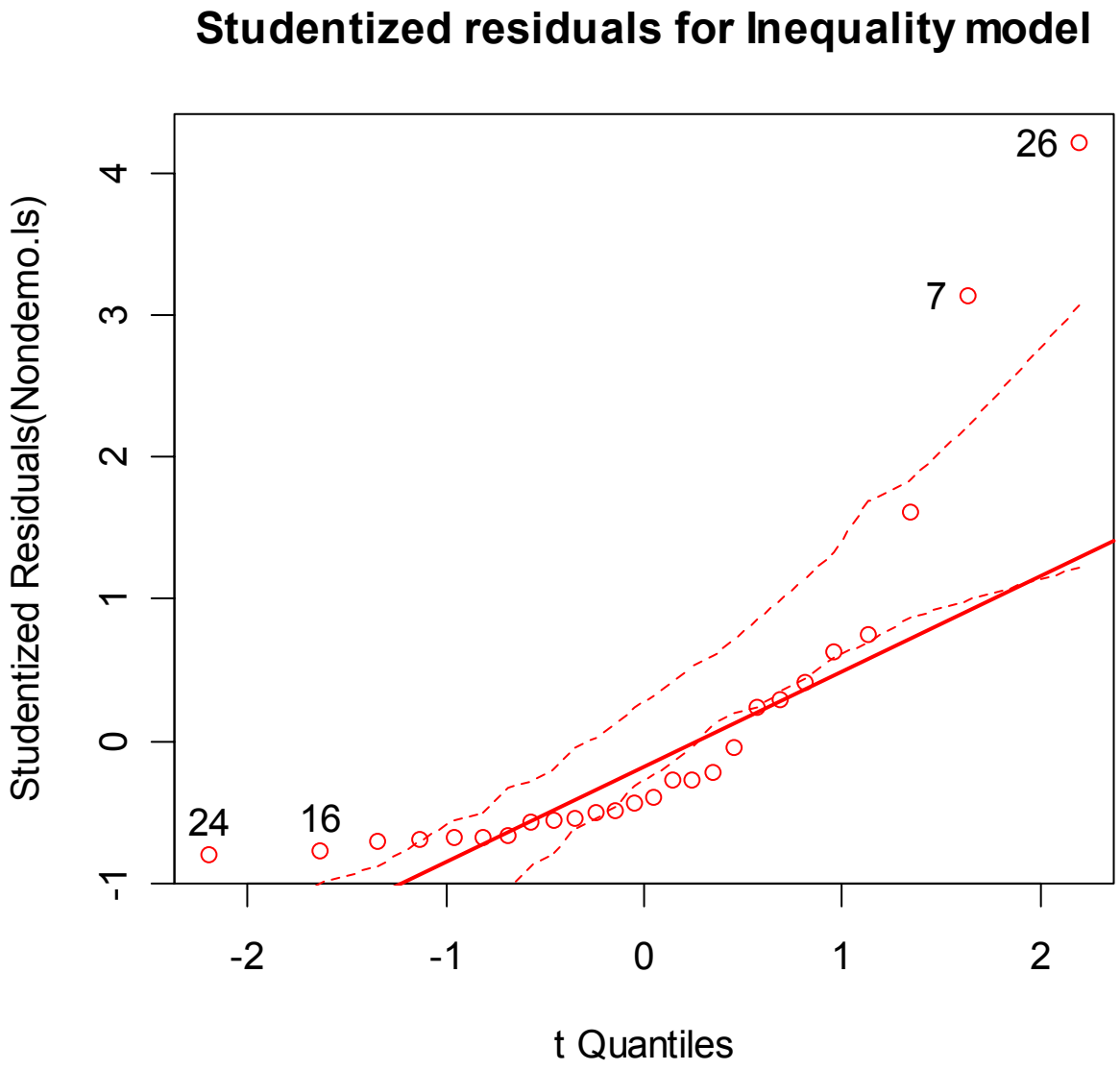
Inequality model revisited (2)
OLS model including unusual cases

```
> Nondemo.ls<-lm(secpay~gini)
   #Including all cases
> summary(Nondemo.ls)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.1947705  0.1106550  10.797 1.08e-10 ***
gini        -0.0007586  0.0028843  -0.263    0.795
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1485 on 24 degrees of freedom
Multiple R-Squared: 0.002874,   Adjusted R-squared: -0.03867
F-statistic: 0.06918 on 1 and 24 DF,  p-value: 0.7948
```

Inequality model revisited (3) Diagnostics

- Given the poor fit of OLS model, we now proceed to do some further diagnostics
- In particular, we examine the quantile comparison plot, checking for a skewed distribution
- We also look at the Cook's D to explore the impact of individual cases on the regression line (using an influence plot)
- Recall that, if this were a *multiple* regression model, we would also explore the partial-regression plots—added-variable plots—for influence)

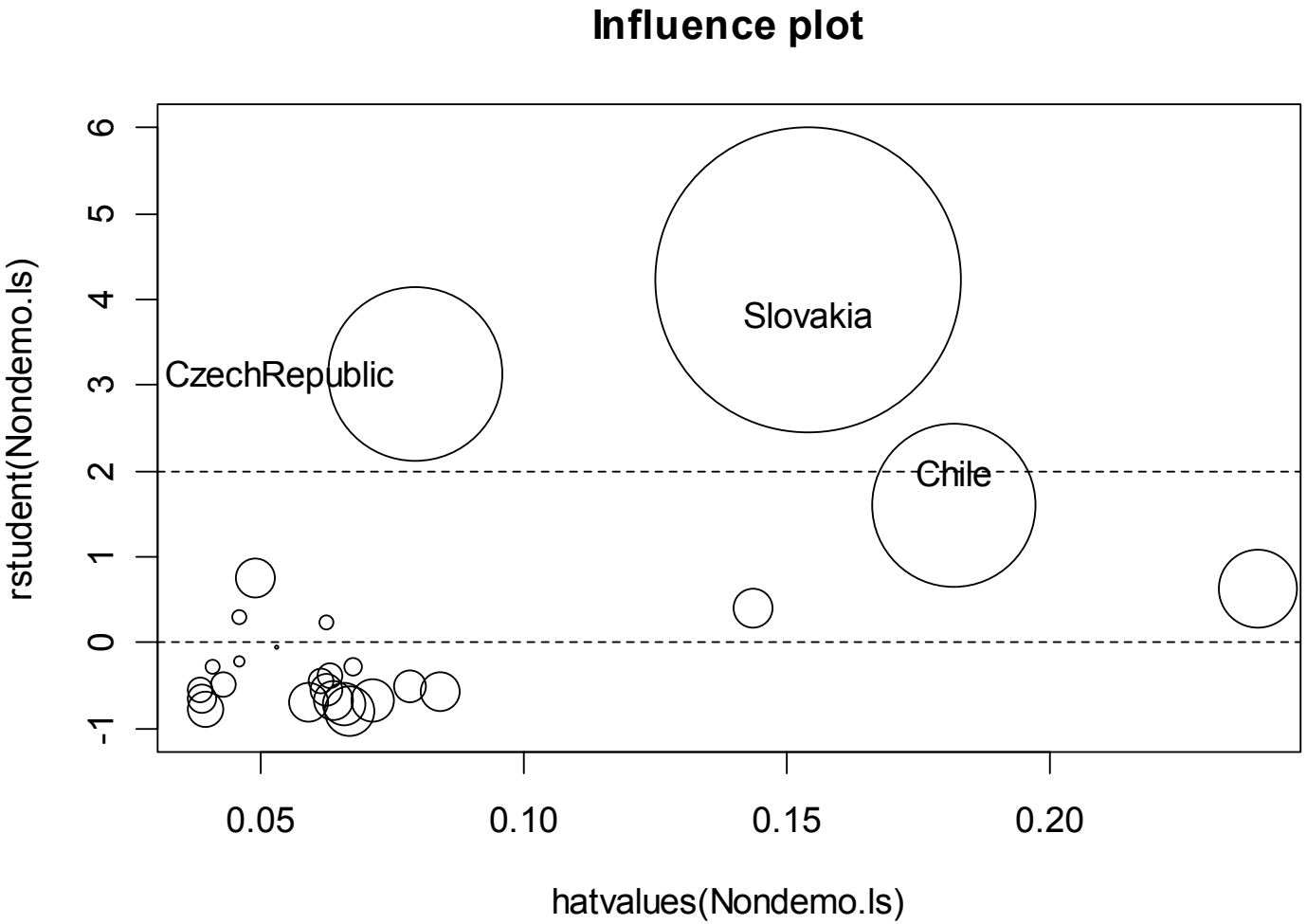
Inequality model revisited (4)
Outlier Tests



Inequality model revisited (5)
Outlier Tests

```
> library (car)
> qq.plot(Nondemo.ls, simulate=T,
        lables=row.names(Nondemo))
[1]  7 16 24 26
> row.names(Nondemo)[c(7, 16, 24, 26)]
[1] "CzechRepublic" "Estonia"
      "Bangladesh"    "Slovakia"

> outlier.test(Nondemo.ls)
 max|rstudent| df unadjusted p Bonferroni p
      4.219304 23 0.0003259723  0.008475279

Observation: 26
```

Inequality model revisited (6)
Influence Plot



- We see the Czech Republic and Slovakia standing out, but Chile also has relatively high influence

R-script for the Influence Plot

```
> plot(hatvalues(Nondemo.ls),
+     rstudent(Nondemo.ls), ylim=range(-1,6),type='n',
+     main="Influence plot")
> cook<-sqrt(cookd(Nondemo.ls))
>      points(hatvalues(Nondemo.ls),
+      rstudent(Nondemo.ls), cex=20*cook/max(cook))
> abline(v=3/length(Nondemo), lty=2)
> #line for hatvalues
> abline(h=c(-2,0,2), lty=2)
> #lines for studentized residuals
> identify(hatvalues(Nondemo.ls),  rstudent(Nondemo.ls),
+    row.names(Nondemo))
```

Alternatively,

```
>library(Rcmdr)
>influence.plot(Nondemo.ls)
```

Inequality model revisited (7)
OLS model without influential cases

```
> Nondemo.ls2<-update(Nondemo.ls, subset=-c(7,26))
  #Ignoring Czech Republic and Slovakia
> summary(Nondemo.ls2)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.940766   0.052569   17.90 1.34e-14 ***
gini        0.004995   0.001336    3.74  0.00113 **

Residual standard error: 0.0627 on 22 degrees of freedom
Multiple R-Squared: 0.3887,     Adjusted R-squared: 0.3609
F-statistic: 13.99 on 1 and 22 DF,  p-value: 0.001135
```

- We now proceed to explore various robust regression methods to see how their results compare

Types of Robust and Resistant Regression

- There are many types of robust regression models. Although they work in different ways, they all give less weight to observations that would otherwise influence the regression line
- Early methods:
  - ***Least Absolute Values*** (LAV) regression or least absolute deviation regression (Minimizes |e| instead of e^2)
- Methods that we will emphasize are:
  - ***M-Estimation***
    - Huber estimates
    - Bisquare estimators
  - ***Bounded Influence Regression***
    - Least Median of Squares
    - Least-Trimmed Squares

Breakdown Point (1)

- Assume a sample, $\mathbf{Z}$, with n observations, and let $\mathbf{T}$ be a regression estimator
- In other words, applying $\mathbf{T}$ to $\mathbf{Z}$ gives us the vector of regression coefficients:

$$\mathbf{T}(\mathbf{Z}) = \hat{\beta}$$

- Imagine all possible "corrupted" samples $\mathbf{Z^0}$ that replace any subset of observations, m, in the dataset with arbitrary values (*i.e.*, influential cases)
- The maximum bias that could arise from these substitutions is:

$$\text{bias}(m; \mathbf{T}, \mathbf{Z}) = \sup_{Z'} \| \mathbf{T}(\mathbf{Z}') - \mathbf{T}(\mathbf{Z}) \|$$

where the supremum is over all possible $\mathbf{Z}'$

Breakdown Point (2)

- If the the bias $(m; \mathbf{T}, \mathbf{Z})$ is infinite, the m outliers have an arbitrarily large effect on $\mathbf{T}$. In other words, the estimator "breaks down"
- As a result, the breakdown point for an estimator $\mathbf{T}$ for a finite sample $\mathbf{Z}$ is:

$$\varepsilon_n^*(\mathbf{T}, \mathbf{Z}) = \min \left\{ \frac{m}{n}; \text{bias } (m; \mathbf{T}, \mathbf{Z}) \text{ is infinite} \right\}$$

- In other words, the breakdown point of an estimator is the smallest fraction of "bad" data (outliers or data grouped at the extreme of a tail) the estimator can tolerate without taking on values arbitrarily far from $\mathbf{T(Z)}$
- For OLS regression one unusual case is enough to influence the coefficient estimates. Its breakdown point is then

$$\varepsilon_n^*(\mathbf{T}, \mathbf{Z}) = 1/n$$

Breakdown Point (3)

- As n gets larger, $1/n$ tends towards 0, meaning that the breakdown point for OLS is 0%
- Robust and resistant regression methods attempt to limit the impact of unusual cases on the regression estimates
  - ***Least Absolute Values (LAV) regression*** is robust to outliers (unusual Y values given X), but typically fares even worse than OLS for cases with high leverage
    - If a leverage point is very far away, the LAV line will pass through it. In other words, its breakdown point is also $1/n$
  - More efficient than LAV estimators, ***M-Estimators*** are also robust to outliers but can have trouble handling cases with high leverage, meaning that the breakdown point is also $1/n$
  - ***Bounded influence methods*** have a much higher breakdown point (as high as 50%) because they effectively remove a large proportion of the cases. These methods can have trouble with small samples, however

Estimating the Centre of a Distribution

- In order to explain how robust regression works it is helpful to start with the simple case of robust estimation of the centre of a distribution
- Consider independent observations and the simple model:

$$Y_i = \mu + \varepsilon_i$$

- If the underlying distribution is normal, the sample mean is the maximally efficient estimator of μ, producing the fitted model:

$$Y_i = \bar{Y} + E_i$$

- The mean then minimizes the ***least-squares objective function***:

$$\sum_{i=1}^{n} \rho_{LS}(E_i) = \sum_{i=1}^{n} \rho_{LS}(Y_i - \hat{\mu}) \equiv \sum_{i=1}^{n} (Y_i - \hat{\mu})^2$$

- The derivative of the objective function with respect to E gives the influence function which determines the influence of observations: $\Psi_{LS}(E)' \rho^0_{LS}(E)=2E$. In other words, influence is proportional to the residual E.
- We know that compared to the median, however, the mean is sensitive to extreme cases
- As an alternative, then, we now consider the median as an estimator of μ. The median minimizes the ***least-absolute-values (LAV) objective function***:

$$\sum_{i=1}^{n} \rho_{LAV}(E_i) = \sum_{i=1}^{n} \rho_{LAV}(Y_i - \widehat{\mu}) \equiv \sum_{i=1}^{n} |Y_i - \widehat{\mu}|$$

- This method is more resistant to outliers, because in contrast to the mean, the influence of an unusual observation on the median is *bounded*

- Again taking the derivative of the objective function gives the shape of the influence function:

$$\psi_{LAV}(E) \equiv \rho'_{LAV}(E) = \begin{cases} 1 & \text{for } E > 0 \\ 0 & \text{for } E = 0 \\ -1 & \text{for } E < 0 \end{cases}$$

- The fact that the median is more resistant than the mean to outliers is a favourable characteristic.
- It is far less efficient, however. If $Y \sim N(\mu,\sigma^2)$, the sampling variance is σ^2/n; the variance for the median is $\pi\sigma^2/2n$
  - In other words the sampling variance for the median is $\pi/2$ ' 1.57 times as large as for the mean
- LAV regression, then, simply minimizes $\sum |E_i|$ where the absolute residuals are based on the median

Influence Functions for the Mean (a) and Median (b)

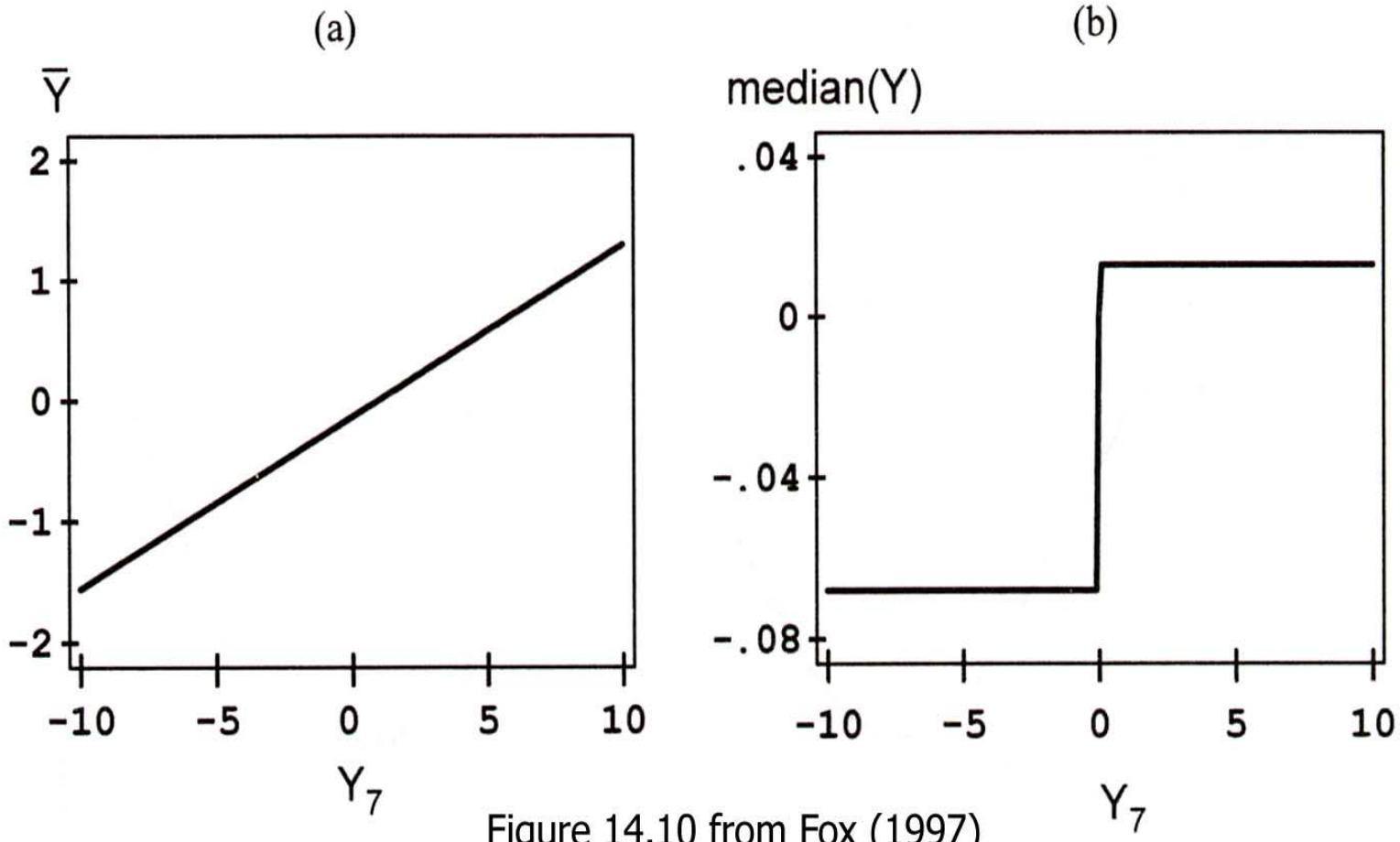


Figure 14.10 from Fox (1997)

LAV Regression in R (1)

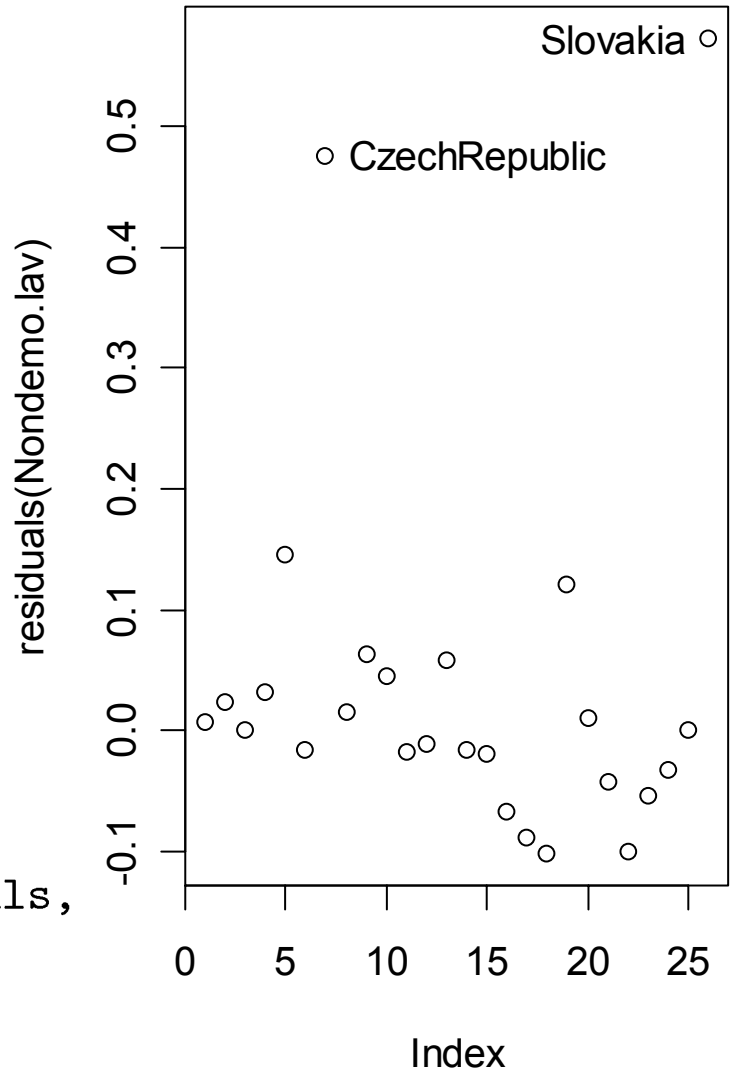
```
> library(quantreg)
> Nondemo.lav<-rq(secpay~gini,.5)
> summary(Nondemo.lav)

Call: rq(formula = secpay ~ gini, tau = 0.5)

Coefficients:
             Value    Std. Error t value  Pr(>|t|)
(Intercept)  0.96145  0.07298    13.17505  0.00000
gini         0.00450  0.00199     2.26642  0.03272
```

- The LAV model is very resistant here because the influence is largely due to high Y-values—The slope for Gini is nearly the same as the OLS model excluding the Czech Republic and Slovakia

LAV Regression in R (2)

- A plot of the residuals from the LAV regression shows that Slovakia and the Czech Republic are far from the fitted line, indicating that they would be influential in a regular OLS regression

```
> library(quantreg)
> Nondemo.lav<-rq(secpay~gini,.5)
> plot(residuals(Nondemo.lav))
> identify(1:26, Nondemo.lav$residuals,
          rownames(Nondemo))
```



M-Estimation: Huber Estimates (1)

- A good compromise between the efficiency of the least-squares and the robustness of the least-absolute values estimators is the ***Huber objective function***
- At the centre of the distribution the Huber function behaves like the OLS function, but at the extremes it behaves like the LAV function:

$$\rho_H(E) = \begin{cases} \frac{1}{2}E^2 & \text{for } E \leq k \\ k|E| - \frac{1}{2}k^2 & \text{for } E > k \end{cases}$$

- The influence function is determined by taking the derivative:

$$\psi_H(E) = \begin{cases} k & \text{for } E > k \\ E & \text{for } E \leq k \\ -k & \text{for } E < -k \end{cases}$$

- The tuning constant, k, defines the centre and tails

M-Estimation: Huber weights
k, the tuning constant

- The tuning constant is expressed as a multiple of the *scale* (the spread) of Y, k=cS, where S is the measure of the scale of Y (*i.e.*, the spread)
  - We could use the standard deviation as a measure of scale, but it is more influenced by extreme observations than is the mean
  - Instead, we use the *median absolute deviation:*

$$\text{MAD} \equiv \text{median}|Y_i - \hat{\mu}|$$

- The median of Y serves as an initial estimate of μ-hat, thus allowing us to define S´ MAD/.6745, which ensures that S estimates σ when the population is normal
- Using k=1.345 (1.345/.6745 is about 2) produces 95% efficiency relative to the sample mean when the population is normal and gives substantial resistance to outliers when it is not
- A smaller k gives more resistance

M-Estimation: Biweight Estimates

- ***Biweight estimates*** behave somewhat differently than Huber weights, but are calculated in a similar manner
- The **biweight objective function** is especially resistant to observations on the extreme tails:

$$\rho_{BW}(E) = \begin{cases} \frac{k^2}{6}\left\{1 - \left[1 - \left(\frac{E}{k}\right)^2\right]^3\right\} & \text{for } |E| \leq k \\ \frac{k^2}{6} & \text{for } |E| > k \end{cases}$$

- The influence function, then, tends toward zero:

$$\psi_{BW}(E) = \begin{cases} E\left[1 - \left(\frac{E}{k}\right)^2\right]^2 & \text{for } |E| \leq k \\ 0 & \text{for } |E|\ k \end{cases}$$

- For this function a tuning constant of k=4.685/.6745, or about 7 MADS, produces 95% efficiency when sampling from a normal population

Weight Functions for Various Estimators

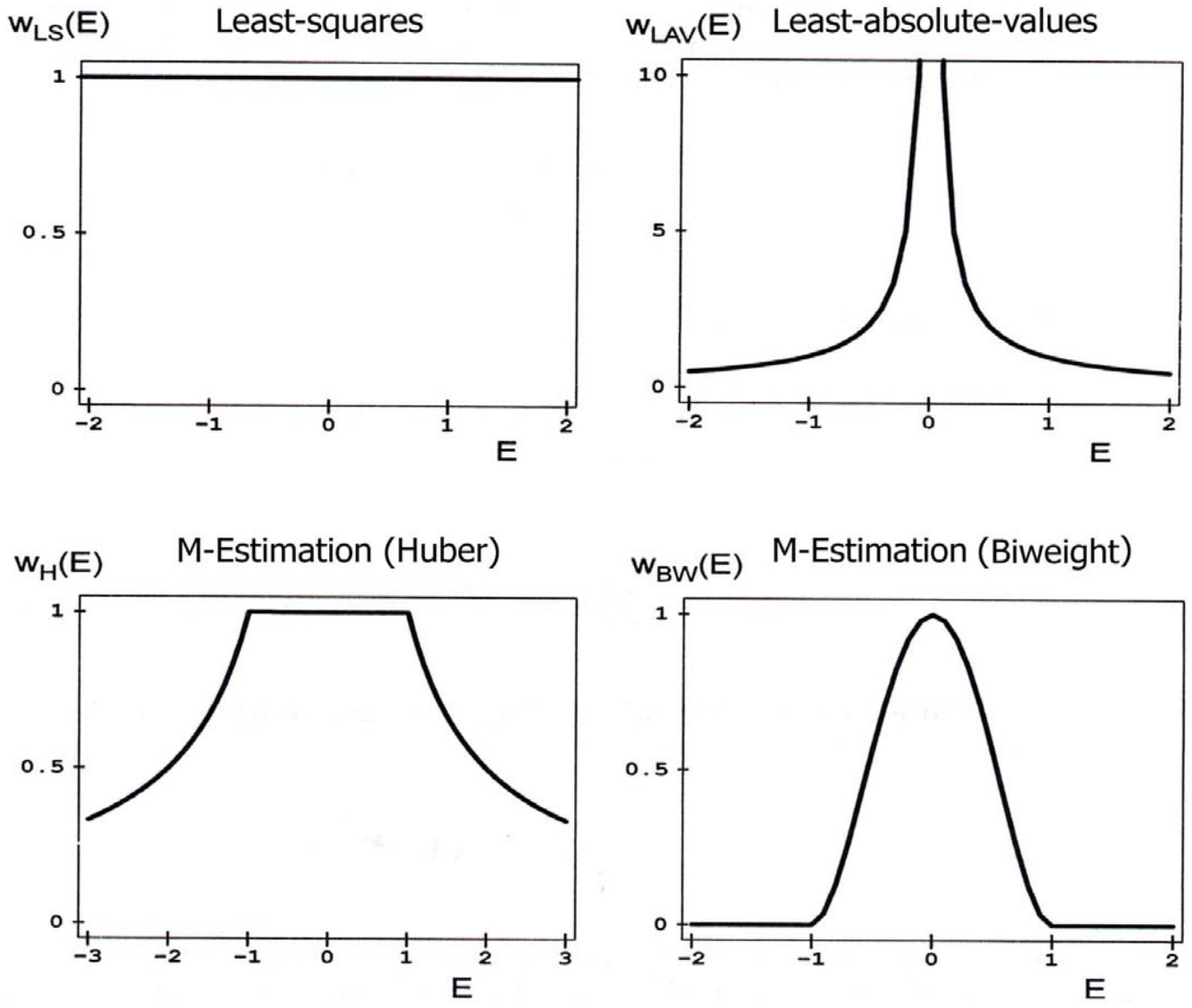


Figure 14.13 from Fox (1997)

M-Estimation and Regression (1)

- Since regression is based on the mean, it is easy to extend the idea of M-estimation to regression
- The estimated linear model is:

$$Y_i = \boldsymbol{x}_i\mathbf{b} + \boldsymbol{\varepsilon}_i$$

- The M-estimator then minimizes the objective function:

$$\sum_{i=1}^{n} \rho(E_i) = \sum_{i=1}^{n} \rho(Y_i - \boldsymbol{x}_i'\mathbf{b})$$

- Differentiating and setting the derivative to zero produces the shape of the weight function:

$$\sum_{i=1}^{n} \psi(Y_i - \boldsymbol{x}_i'\mathbf{b})\boldsymbol{x}_i = 0$$

M-Estimation and Regression (2)

- We now have a system of k+1 equations, for which we simply replace ψ with the weight function, $w(E_i)$:

$$\sum_{i=1}^{n} w_i(Y_i - \boldsymbol{x}_i'\mathbf{b})\boldsymbol{x}_i = 0$$

- In other words, the ***solution assigns a different weight to each case depending on the size of their residual***, and thus minimizes the the weighted sum of squares

$$\sum w_i^2 E_i^2$$

- The w_i weights depend on the residuals in the model—*i.e.*, we don't know them until after fitting an initial regression—so an iterative solution (using **Iterative Reweighted Least Squares, IRLS**) is required.

M-Estimation and Regression (3) Weighted Least Squares

- The likelihood for a WLS model is:

$$L(\beta, \sigma_\varepsilon^2) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \boldsymbol{X}\beta)'\Sigma(\mathbf{y} - \boldsymbol{X}\beta)\right]$$

  where Σ is the covariance matrix of the errors

- The maximum likelihood estimators are then defined as:

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

$$\widehat{\boldsymbol{\sigma}_\epsilon^2} = \frac{\sum(E_i/w_i)^2}{n}$$

- This is equivalent to minimizing the weighted sum of squares $\sum w_i^2 E_i^2$, giving greater weight to observations with smaller residuals from a preliminary OLS

M-Estimation and Regression (4) Iterative Reweighted Least Squares

- Initial estimates of **b** are selected using **weighted least squares**
- The residuals from this model are used to calculate an estimate of the scale of the residuals $S^{(0)}$ and the weights $w_i^{(0)}$
- The model is then refit with several iterations minimizing the weighted sum of squares to obtain new estimates of **b**:

$$\mathbf{b}^{(l)} = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}\mathbf{y}$$

where l is for the iteration counter; in ith row of the model matrix are $\boldsymbol{x}_i'$; and $\boldsymbol{W} \equiv \text{diag}\{w_i^{(l-1)}\}$

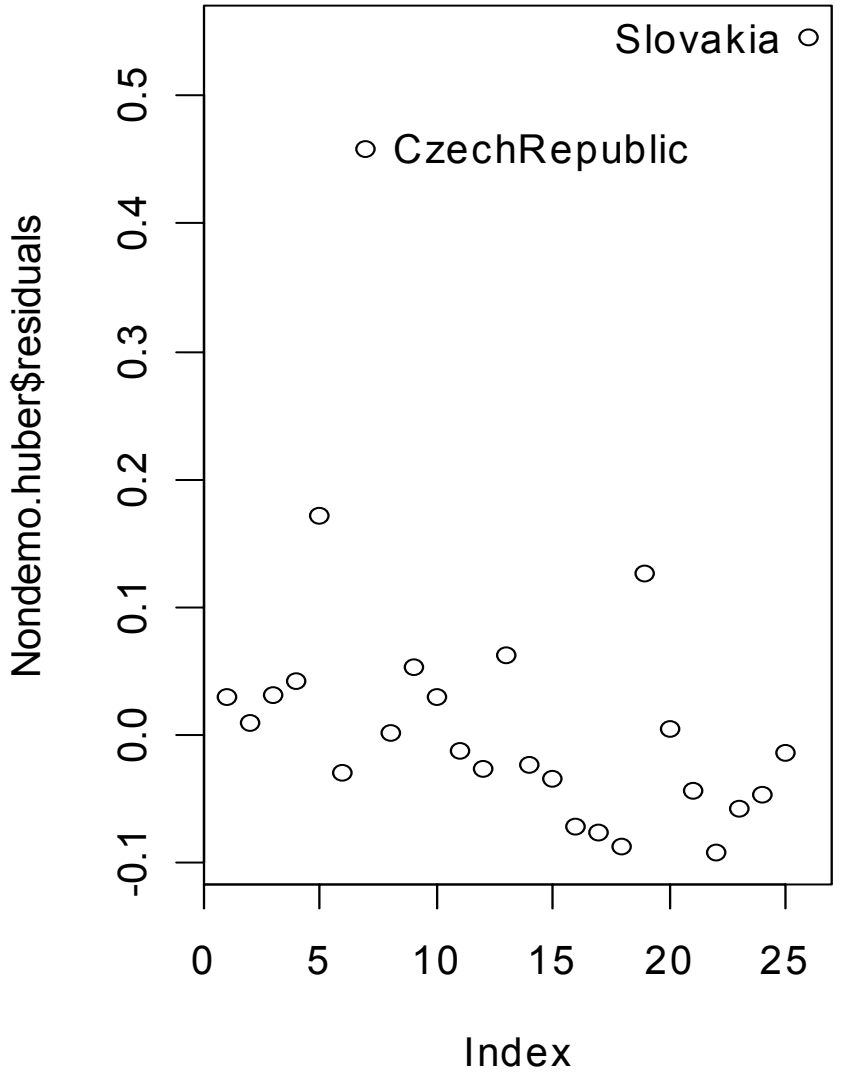
- This process is continued until the model converges (i.e., $\mathbf{b}^{(l)}$' $\mathbf{b}^{(l\text{-}1)}$)

M-Estimation in R (1)
Huber Weights

```
> library (MASS)
> Nondemo.huber<-rlm(secpay~gini)
   #M-Estimation with Huber (all cases)
> summary(Nondemo.huber)

Coefficients:
             Value   Std. Error t value
(Intercept)  1.0169  0.0563     18.0643
gini         0.0031  0.0015      2.0837

Residual standard error: 0.06363 on 24 degrees of freedom
```

M-Estimation in R (2) Plotting the Residuals

- As we would expect, the Czech Republic and Slovakia have the largest residuals, meaning that they would have the largest influence on the regular OLS regression line

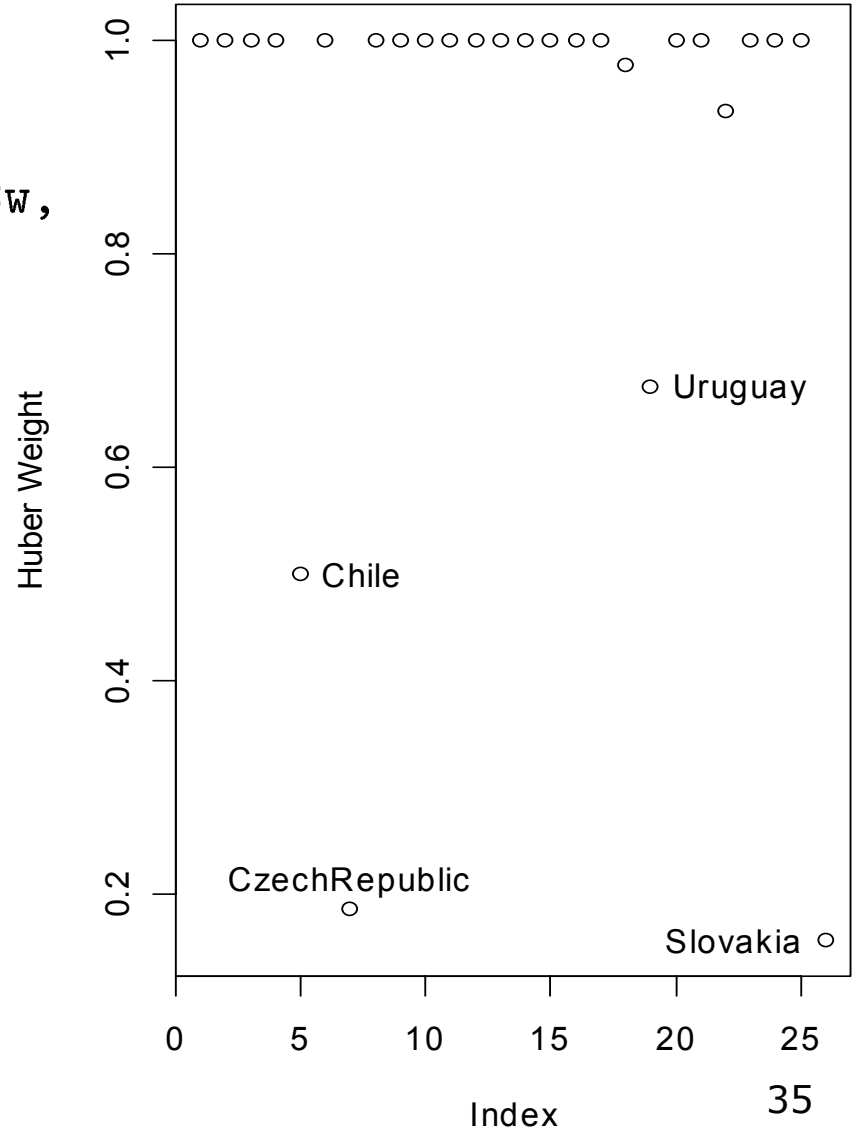



```
> plot(Nondemo.huber$residuals)
> identify(1:26,
      Nondemo.huber$residuals,
      rownames(Nondemo))
```

M-Estimation in R (3)
Examining the weights

```
> plot(Nondemo.huber$w,
   ylab="Huber Weight")
> identify(1:26, Nondemo.huber$w,
   rownames(Nondemo))
```

- The plot shows the weight given to each case
- In line with the previous graph, the Czech Republic and Slovakia received the least weight of all the observations
- ***Note***:This technique can also be used to assess influential cases



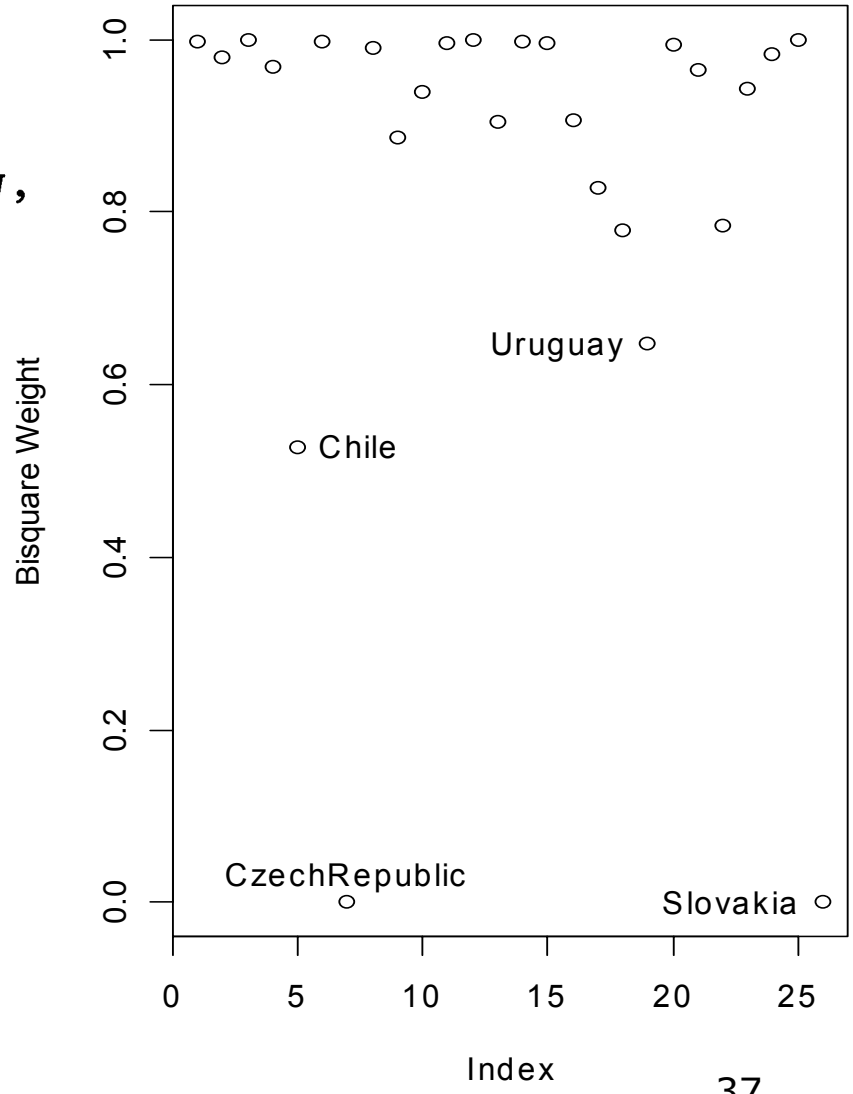
M-Estimation in R (4)
Bisquare Weights

```
> library(MASS)
> Nondemo.bisq<-rlm(secpay~gini, data=Nondemo, method='MM')
> summary(Nondemo.bisq)

Residuals:
      Min        1Q    Median        3Q       Max
-0.097533 -0.023654  0.004203  0.046373  0.578463

Coefficients:
            Value   Std. Error t value
(Intercept)  0.9546  0.0489    19.5070
gini         0.0046  0.0013     3.5759

Residual standard error: 0.06068 on 24 degrees of freedom
```

M-Estimation in R (5)
Examining the weights

```
> plot(Nondemo.bisq$w,
    ylab="Bisquare Weight")
> identify(1:26, Nondemo.bisq$w,
    rownames(Nondemo))
[1]  5  7 19 26
```

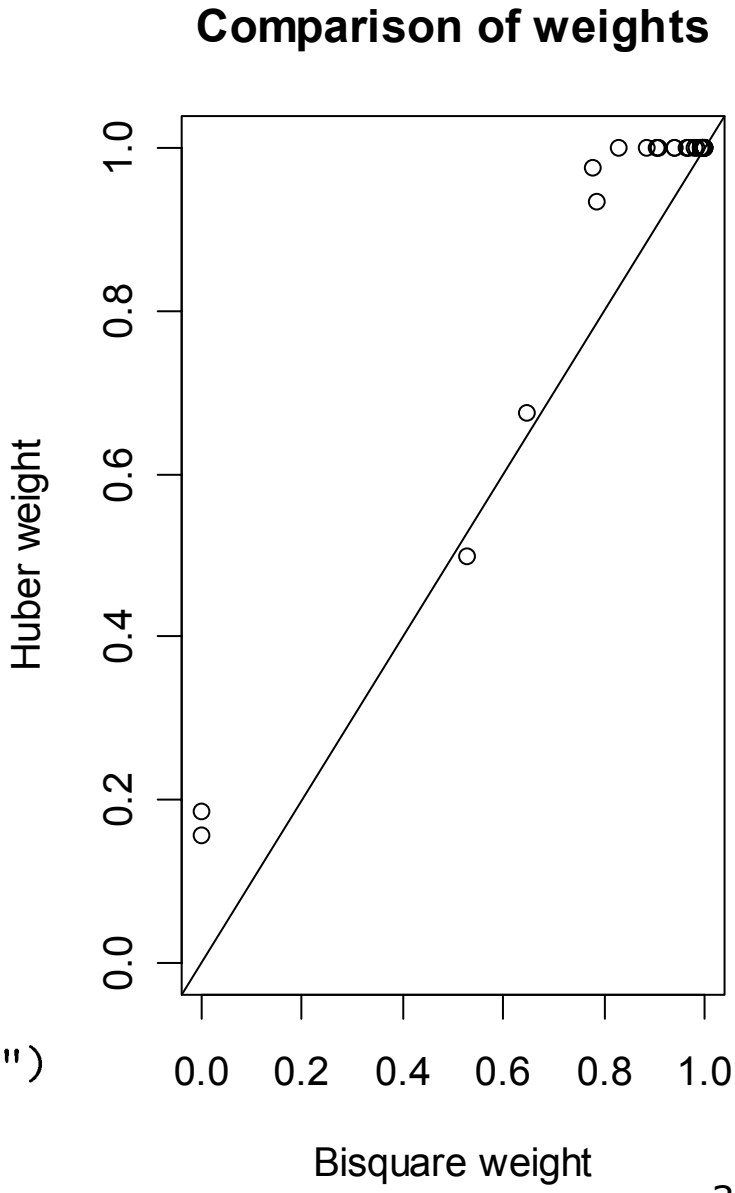
- Again the weights accorded to observations are as expected, and the ***relative size*** of the weights match those of the Huber estimates



Comparing the Huber and Bisquare weights

- Notice that although the relative weights are similar, the Huber method gives a weight of 1 to far more cases than does the bisquare weight

```
> plot(Nondemo.bisq$w,
+      Nondemo.huber$w,
+      xlim=c(0,1), ylim=c(0,1),
+      xlab="Bisquare weight",
+      ylab="Huber weight",
+      main="Comparison of weights")
> abline(0,1)
```



Bounded-Influence Regression and the Breakdown Point

- M-Estimators are statistically equally efficient as OLS if the distribution is normal, while at the same time are more robust with respect to influential cases
- Nonetheless, M-estimation can still be influenced by a single very extreme X-value—*i.e.*, like OLS, it still has a breakdown point of 0
- **Least-trimmed-squares (LTS)** estimators can have a breakdown point up to 50%—i.e., half the data can be influential in the OLS sense before the LTS estimator is seriously affected
  - Least-trimmed-squares essentially proceeds with OLS after ***eliminating the most extreme positive or negative residuals***

Bounded Influence Regression: Least-Trimmed Squares

- Least-Trimmed Squares orders the squared residuals from smallest to largest: $(E^2)_{(1)}$, $(E^2)_{(2)}$, …, $(E^2)_{(n)}$
- It then calculates **b** that **minimizes the sum of only the smaller half of the residuals**:

$$\sum_{i=1}^{m} (E^2)_{(i)}$$

where $m = [n/2] + 1$; the square brackets indicate rounding *down* to the nearest integer

- By using only the 50% of the data that fits closest to the original OLS line, LTS completely ignores extreme outliers
- On the other hand, this method can misrepresent the trend in the data if it is characterized by clusters of extreme cases or if the data set is relatively small

LTS Regression in R

```
> library(lqs)
> Nondemo.lts<-ltsreg(secpay~gini, data=Nondemo)
> Nondemo.lts
Call:
lqs.formula(formula = secpay ~ gini, data = Nondemo, method = "lts")

Coefficients:
(Intercept)         gini
   1.099496    -0.000566

Scale estimates 0.04184 0.03401
```

- We see here, that the LTS model performs poorly because of the small sample size—the slope for Gini is negative despite that the Czech Republic and Slovakia are not used in its calculation

Bounded Influence Regression: Least- Median Squares

- An alternative bounded influence method is ***Least Median Squares***
- Rather than minimize the sum of the least squares function, this model minimizes the median of the squared residuals, E_i^2
- LMS is very robust with respect to outliers both in terms of X and Y values
- It performs poorly form the point of view of asymptotic efficiency, however

LMS Regression in R

```
>library(lqs)
> Nondemo.lms<-lmsreg(secpay~gini, data=Nondemo)
> Nondemo.lms
Call:
lqs.formula(formula = secpay ~ gini, data = Nondemo, method = "lms")

Coefficients:
(Intercept)         gini
   1.136918    -0.001630

Scale estimates 0.03736 0.03417
```

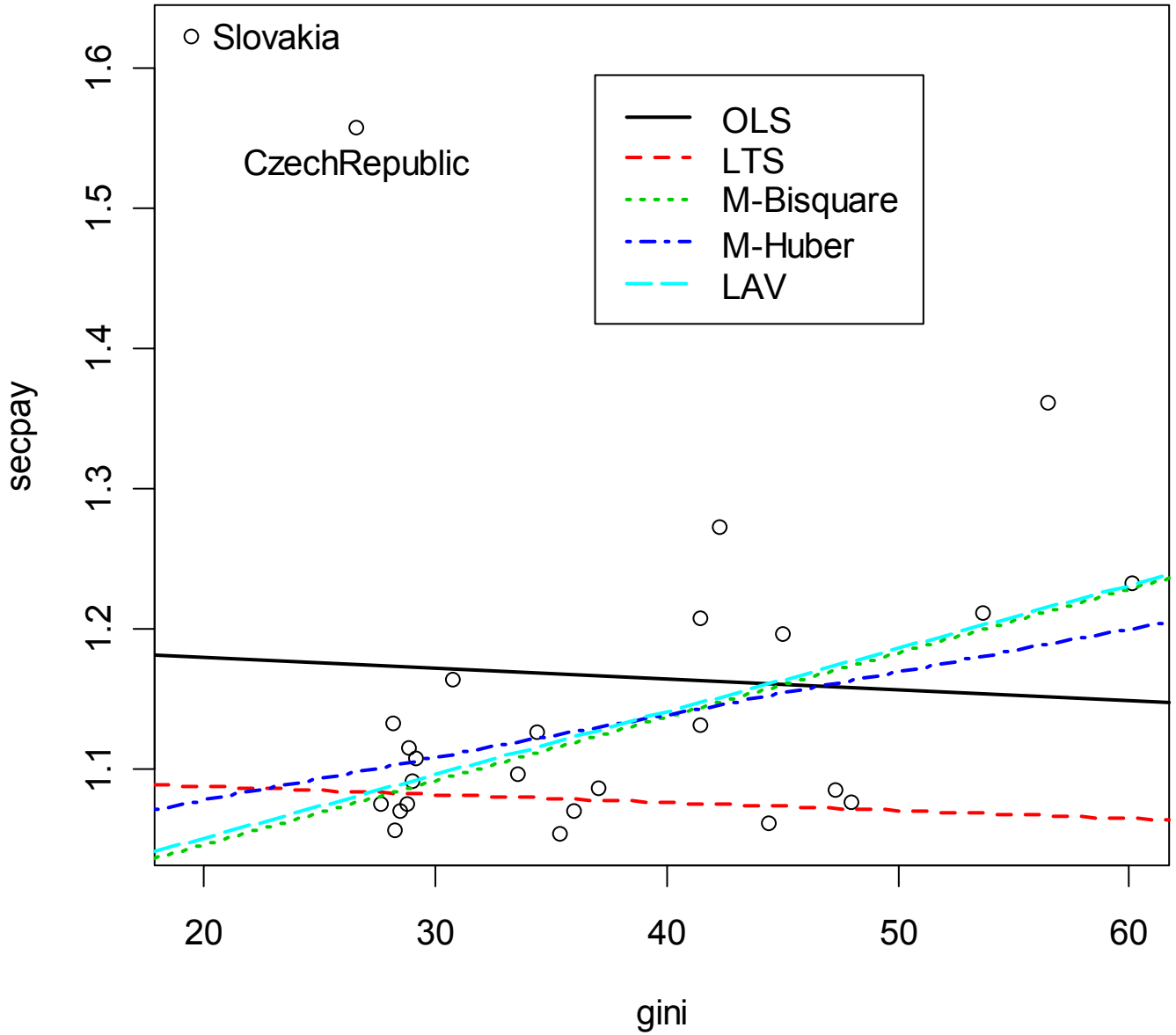
- Like the LTS model, this model performs poorly because the sample size is so small

Comparing the models (1)

	Est.	S.E.	t-value	S_E	R^2
OLS+outliers	-.00075	.002	-.26	.149	.002
OLS-outliers	.0049	.001	3.74	.062	.389
M (Huber)	.0031	.0015	2.08	.063	
M (Bisquare)	.0046	.0013	3.57	.060	
LTS	-.0005				
LMS	-.0016				

- Both M-Estimation techniques give very similar estimates to the OLS with the two influential cases deleted
  - Bisquare estimates were slightly less affected by the influential cases than the Huber estimates
- The bounded influence models (LTS and LMS) did not work well, however—taking out half the data has established a completely different pattern
- ***LESSON***: Don't just fit these models blindly—look at the patterns in the data to make sure you pick the method that works best

Comparison of Different Models



R script for comparison plot

```
>plot(gini,secpay)
>identify(gini, secpay, row.names(Nondemo))
>abline(Nondemo.ls, lwd=2, col=1, lty=1)
>abline(Nondemo.lts, lwd=2, col=2, lty=2)
>abline(Nondemo.bisq, lwd=2, col=3, lty=3)
>abline(Nondemo.huber, lwd=2, col=4, lty=4)
>abline(Nondemo.lav, lwd=2, col=5, lty=5)
>legend(locator(1), lty=1:5, lwd=2, col=1:5,
      legend=c('OLS', 'LTS', 'M-Bisquare', 'M-Huber', 'LAV'))
>title("Comparison of Different Models")
```

Summary and Conclusions (1)

- Separated points can have a strong *influence* on statistical models
  - Unusual cases can substantially influence the fit of the OLS model—Cases that are both *outliers* and *high leverage* exert *influence* on both the slopes and intercept of the model
  - Outliers may also indicate that our model fails to capture important characteristics of the data
- Efforts should be made to remedy the problem of unusual cases before proceeding to robust regression
- If robust regression is used, careful attention must be paid to the model—different procedures can give completely different answers

Summary and Conclusions (2)

- No one robust regression technique is best for all data
- There are some considerations, but even these do not hold up all the time:
  - LAV regression should generally be avoided because it is less efficient than other techniques and often not very resistant
  - Bounded influence regression models, which can have a breaking point as high as 50%, often work very well with large datasets
    - They tend to perform poorly with small datasets, however
  - M-Estimation is typically better for small datasets, but its standard errors are not reliable for small samples
    - This can be overcome by using boot-strapping to obtain new estimates of the standard errors