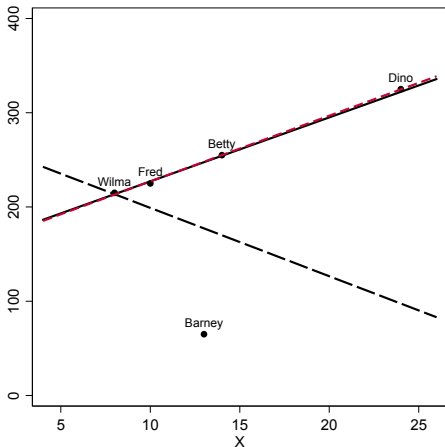


PLSC 503 – Spring 2025

Residuals & Outliers + Instrumental Variables

February 24, 2025

Discrepancy, Leverage, and Influence



Note: Solid line is the regression fit for Wilma, Fred, and Betty only.
Long-dashed line is the regression for Wilma, Fred, Betty, and Barney.
Short-dashed (red) line is the regression for Wilma, Fred, Betty and Dino.

Discrepancy, Leverage, and Influence

$$\text{Influence} = \text{Leverage} \times \text{Discrepancy}$$

Leverage

$$\begin{aligned}\hat{\mathbf{Y}} &= \mathbf{X}\hat{\boldsymbol{\beta}} \\ &= \mathbf{X}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] \\ &= \mathbf{H}\mathbf{Y}\end{aligned}$$

where

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'.$$

$$h_i = \mathbf{X}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_i'$$

Variation:

$$\widehat{\text{Var}}(\hat{u}_i) = \hat{\sigma}^2[1 - \mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i'] \quad (1)$$

$$\begin{aligned} \widehat{\text{s.e.}}(\hat{u}_i) &= \hat{\sigma}\sqrt{[1 - \mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i']} \\ &= \hat{\sigma}\sqrt{1 - h_i} \end{aligned} \quad (2)$$

“Standardized”:

$$\tilde{u}_i = \frac{\hat{u}_i}{\hat{\sigma}\sqrt{1 - h_i}} \quad (3)$$

“Studentized”: define

$$\begin{aligned}\hat{\sigma}_{-i}^2 &= \text{Variance for the } N - 1 \text{ observations } \neq i \\ &= \frac{\hat{\sigma}^2(N - K)}{N - K - 1} - \frac{\hat{u}_i^2}{(N - K - 1)(1 - h_i)}.\end{aligned}\quad (4)$$

Then:

$$\hat{u}'_i = \frac{\hat{u}_i}{\hat{\sigma}_{-i}\sqrt{1 - h_i}} \quad (5)$$

“DFBETA”:

$$D_{ki} = \hat{\beta}_k - \hat{\beta}_{k(-i)} \quad (6)$$

“DFBETAS” (the “S” is for “standardized”):

$$D_{ki}^* = \frac{D_{ki}}{\widehat{\text{s.e.}}(\hat{\beta}_{k(-i)})} \quad (7)$$

Cook's D :

$$\begin{aligned} D_i &= \frac{\tilde{u}_i^2}{K} \times \frac{h_i}{1 - h_i} \\ &= \frac{h_i \hat{u}_i^2}{K \hat{\sigma}^2 (1 - h_i)^2} \end{aligned} \quad (8)$$

```
> # No Barney OR Dino...
> summary(lm(Y~X,data=subset(flintstones,name!="Dino" & name!="Barney"))
```

Residuals:

```
      2      4      5
0.714 -2.143  1.429
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	159.286	6.776	23.5	0.027 *
X	6.786	0.619	11.0	0.058 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.67 on 1 degrees of freedom

Multiple R-squared: 0.992, Adjusted R-squared: 0.984

F-statistic: 120 on 1 and 1 DF, p-value: 0.0579

```
> # No Barney (Dino included...)
> summary(lm(Y~X,data=subset(flintstones,name!="Barney")))
```

Residuals:

	2	3	4	5
	-8.88e-16	2.63e-01	-2.11e+00	1.84e+00

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	157.368	2.465	63.8	0.00025 ***
X	6.974	0.161	43.3	0.00053 ***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 1.99 on 2 degrees of freedom

Multiple R-squared: 0.999, Adjusted R-squared: 0.998

F-statistic: 1.87e+03 on 1 and 2 DF, p-value: 0.000534

“COVRATIO”:

$$\text{COVRATIO}_i = \left[(1 - h_i) \left(\frac{N - K - 1 + \hat{u}_i'^2}{N - K} \right)^K \right]^{-1} \quad (9)$$

Example: Federal Judicial Review, 1789-2018

Dahl (1957):

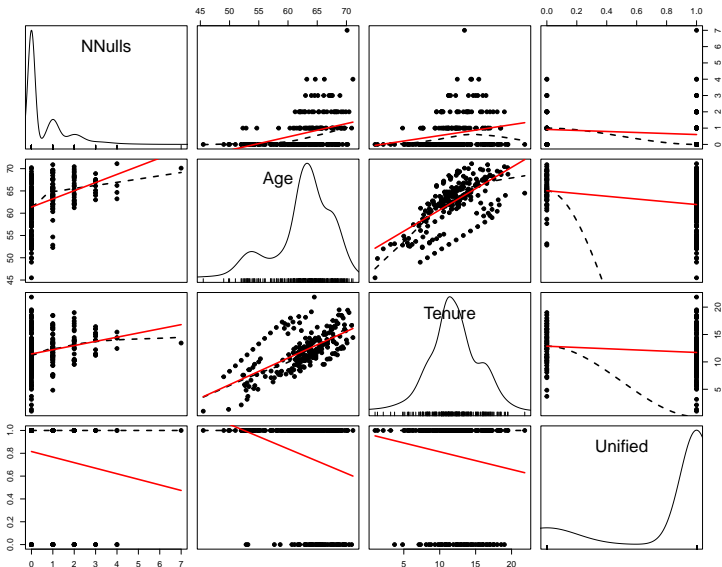
- SCOTUS gets “out of step” with the other branches → judicial review
- Older / longer-serving justices will more likely to invalidate legislation

Data:

```
> summary(NewDahl)
```

Year	NNulls	Age	Tenure	Unified
Min. :1789	Min. :0.000	Min. :45.5	Min. : 1.0	Min. :0.000
1st Qu.:1846	1st Qu.:0.000	1st Qu.:60.7	1st Qu.:10.0	1st Qu.:1.000
Median :1904	Median :0.000	Median :63.5	Median :11.8	Median :1.000
Mean :1904	Mean :0.674	Mean :62.6	Mean :12.0	Mean :0.783
3rd Qu.:1961	3rd Qu.:1.000	3rd Qu.:66.0	3rd Qu.:14.1	3rd Qu.:1.000
Max. :2018	Max. :7.000	Max. :71.1	Max. :21.8	Max. :1.000

Example: Federal Judicial Review, 1789-2018



Basic Regression...

```
> Fit<-with(NewDahl, lm(NNulls~Age+Tenure+Unified))
> summary(Fit)
```

Call:

```
lm(formula = NNulls ~ Age + Tenure + Unified)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.308	-0.700	-0.135	0.308	5.693

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-4.6833	1.0026	-4.67	0.0000051 ***
Age	0.0901	0.0181	4.97	0.0000013 ***
Tenure	-0.0201	0.0248	-0.81	0.42
Unified	-0.0573	0.1613	-0.36	0.72

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

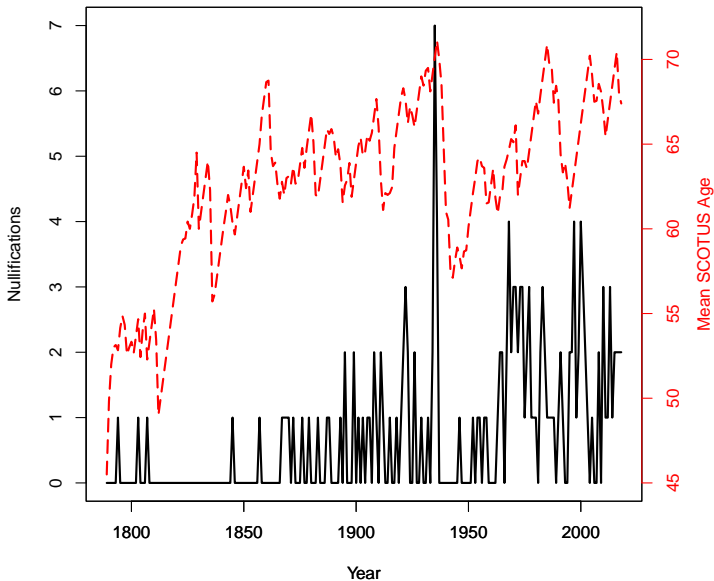
Residual standard error: 0.973 on 226 degrees of freedom

(4 observations deleted due to missingness)

Multiple R-squared: 0.152, Adjusted R-squared: 0.141

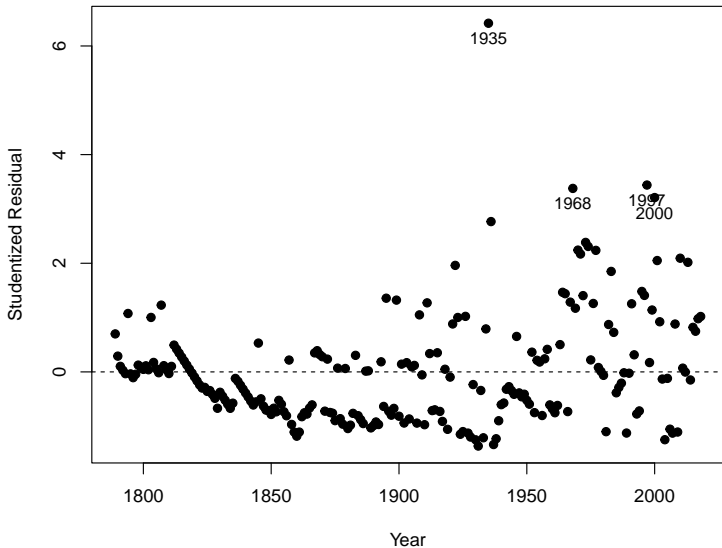
F-statistic: 13.6 on 3 and 226 DF, p-value: 0.0000000365

Federal Judicial Review and Mean SCOTUS Age



```
> FitResid<-with(NewDahl, (Fit$model$NNULLs - predict(Fit)))  
> FitStandard<-rstandard(Fit) # standardized residuals  
> FitStudent<-rstudent(Fit) # studentized residuals  
> FitCooksD<-cooks.distance(Fit) # Cook's D  
> FitDFBeta<-dfbeta(Fit) # DFBeta  
> FitDFBetaS<-dfbetas(Fit) # DFBetaS  
> FitCOVRATIO<-covratio(Fit) # COVRATIOs
```

Studentized Residuals



More About Studentized Residuals

```
> max(FitStudent)
[1] 6.418
```

```
> NewDahl$Year1935<-ifelse(NewDahl$Year==1935,1,0)
> summary(with(NewDahl, lm(NNulls~Age+Tenure+Unified+Year1935)))
```

Residuals:

Min	1Q	Median	3Q	Max
-1.250	-0.652	-0.122	0.302	3.247

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-3.9298	0.9313	-4.22	0.00003546916 ***
Age	0.0768	0.0168	4.56	0.00000846697 ***
Tenure	-0.0113	0.0229	-0.50	0.62
Unified	-0.1210	0.1490	-0.81	0.42
Year1935	5.8186	0.9066	6.42	0.00000000081 ***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

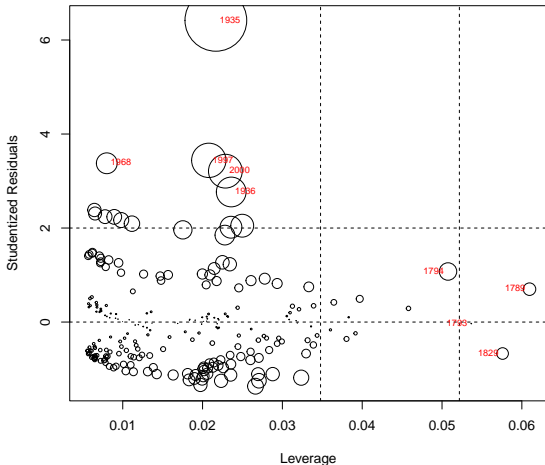
Residual standard error: 0.897 on 225 degrees of freedom

Multiple R-squared: 0.284, Adjusted R-squared: 0.271

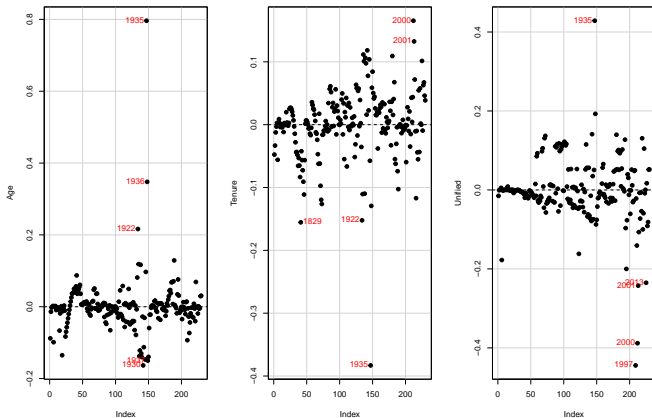
F-statistic: 22.3 on 4 and 225 DF, p-value: 1.65e-15

"Bubble Plot"

```
> influencePlot(Fit,id=list(method="noteworthy",n=4,cex=0.7,  
                           labels=NewDahl$Year,col="red"),  
               xlab="Leverage")
```

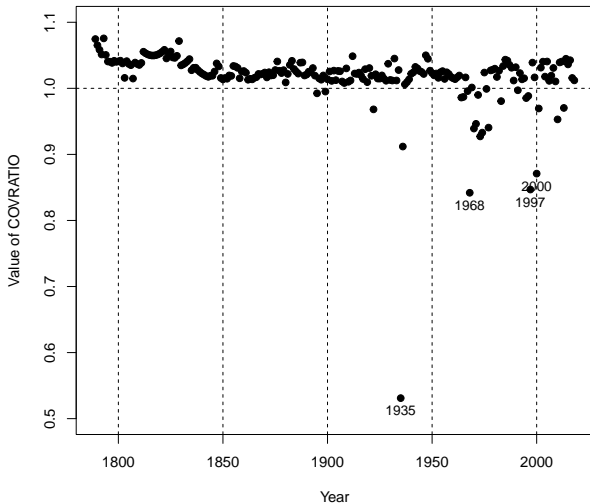


```
> dfbetasPlots(Fit,id.n=5,id.col="red",main="",pch=19,  
  layout=c(1,3),labels=NewDahl$Year)
```



COVRATIO Plot

```
> plot(FitCOVRATIO~NewDahl$Year,pch=19,ylim=c(0.5,1.1),  
      xlab="Year",ylab="Value of COVRATIO")
```



Sensitivity Analyses: Omitting Outliers

```
> out1<-c(1935) # one outlier
> LD2<-NewDahl[!(NewDahl$Year %in% out1),]
> out2<-c(1935,1968,1997,2000) # four outliers
> LD3<-NewDahl[!(NewDahl$Year %in% out2),]
> Fit2<-lm(NNulls~Age+Tenure+Unified,data=LD2)
> Fit3<-lm(NNulls~Age+Tenure+Unified,data=LD3)
```

	<i>Dependent variable:</i>		
		NNulls	
	(1)	(2)	(3)
Age	0.090*** (0.018)	0.077*** (0.017)	0.079*** (0.015)
Tenure	-0.020 (0.025)	-0.011 (0.023)	-0.019 (0.021)
Unified	-0.057 (0.161)	-0.121 (0.149)	-0.010 (0.139)
Constant	-4.683*** (1.003)	-3.930*** (0.931)	-4.130*** (0.855)
Observations	230	229	226
R ²	0.152	0.148	0.158
Adjusted R ²	0.141	0.137	0.147
Residual Std. Error	0.973 (df = 226)	0.897 (df = 225)	0.822 (df = 222)
F Statistic	13.550*** (df = 3; 226)	13.030*** (df = 3; 225)	13.930*** (df = 3; 222)

Note:

* p<0.1; ** p<0.05; *** p<0.01

Thinking About Diagnostics

"Looking"
(Art)



"Testing"
(Science)

Observational Data
Complex Data
Structure
Informative Missingness
Complex / Uncertain
Causality

Experimental Data
Simple Data Structure
No / Uninformative
Missingness
Simple / Clear Causality

Pena, E.A. and E.H. Slate. 2006. "Global Validation of Linear Model Assumptions." J. American Statistical Association 101(473):341-354.

Tests for:

- Normality in $\hat{u}s$ (via skewness & kurtosis tests)
- "Link function" (linearity / additivity)
- Constant variance and uncorrelatedness in $\hat{u}s$ ("heteroskedasticity" test)

```
> Fit<-with(NewDahl, lm(NNulls~Age+Tenure+Unified))

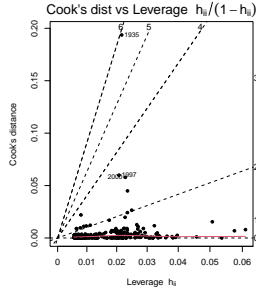
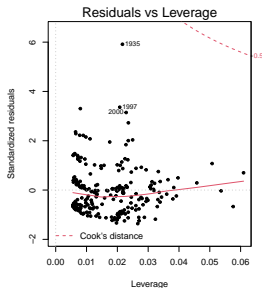
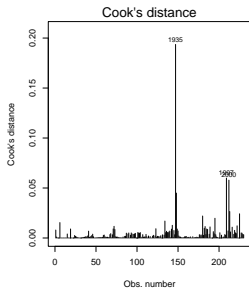
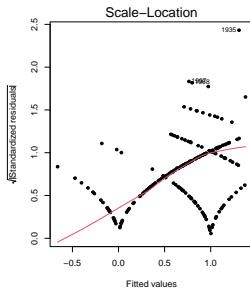
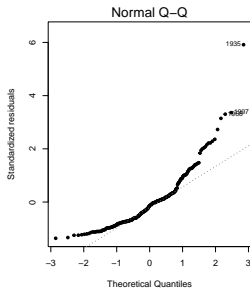
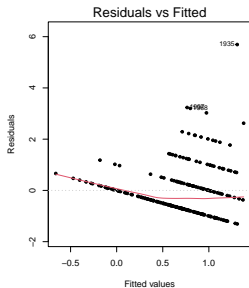
> library(gvlma)
> Nope <- gvlma(Fit)
> display.gvlmatests(Nope)
```

ASSESSMENT OF THE LINEAR MODEL ASSUMPTIONS
 USING THE GLOBAL TEST ON 4 DEGREES-OF-FREEDOM:
 Level of Significance = 0.05

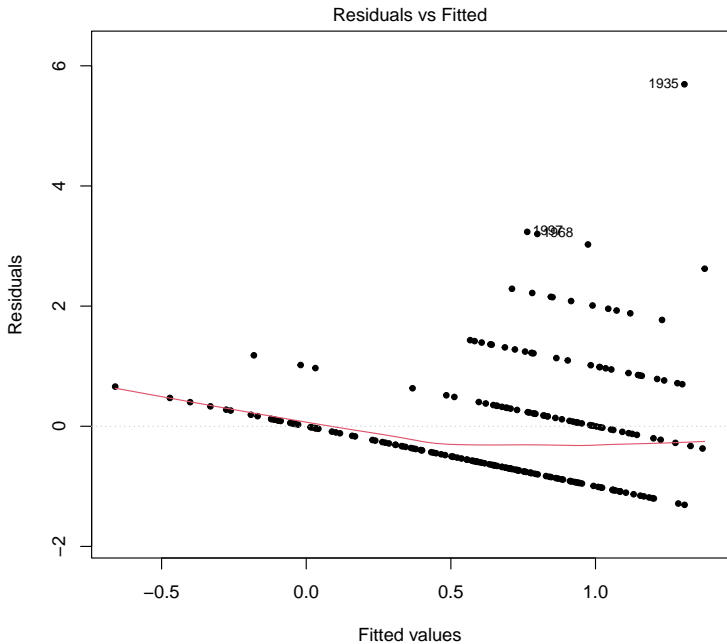
Call:
 gvlma(x = Fit)

	Value	p-value	Decision
Global Stat	454.87	0.00e+00	Assumptions NOT satisfied!
Skewness	122.09	0.00e+00	Assumptions NOT satisfied!
Kurtosis	283.21	0.00e+00	Assumptions NOT satisfied!
Link Function	5.35	2.07e-02	Assumptions NOT satisfied!
Heteroscedasticity	44.23	2.92e-11	Assumptions NOT satisfied!

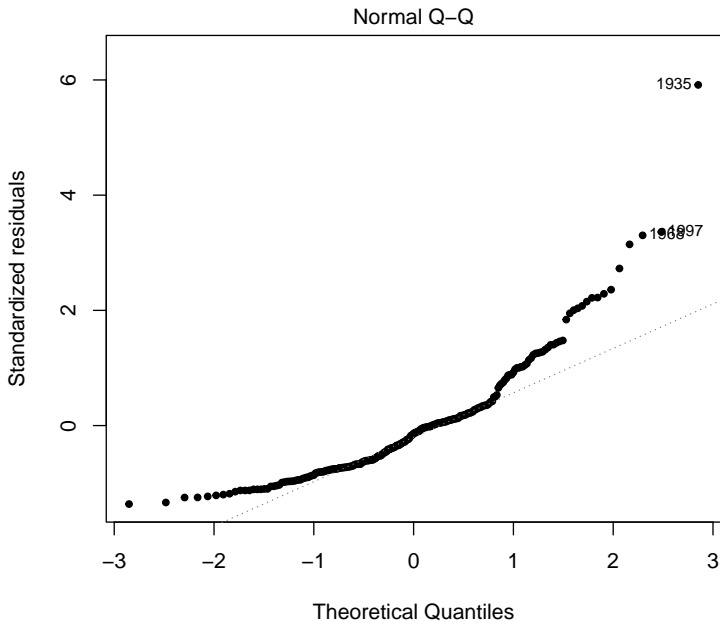
Another Approach: `plot(fit)`



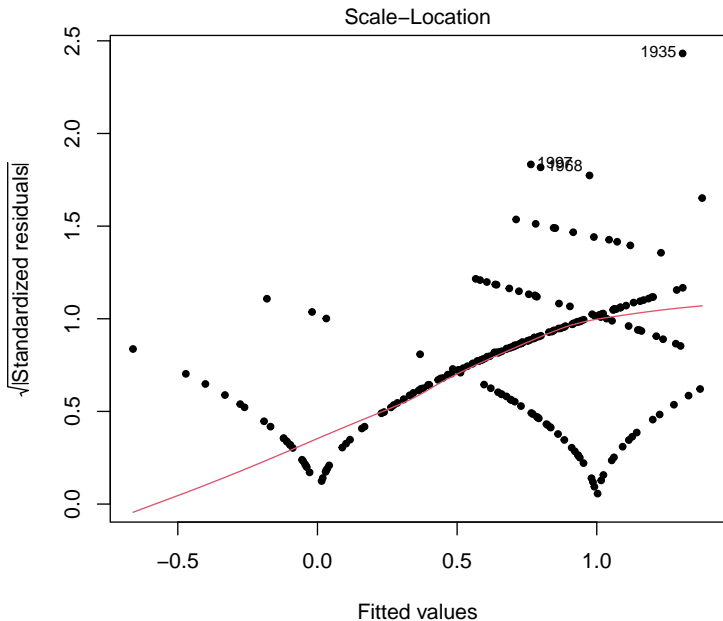
#1: Residuals vs. Fitted Values

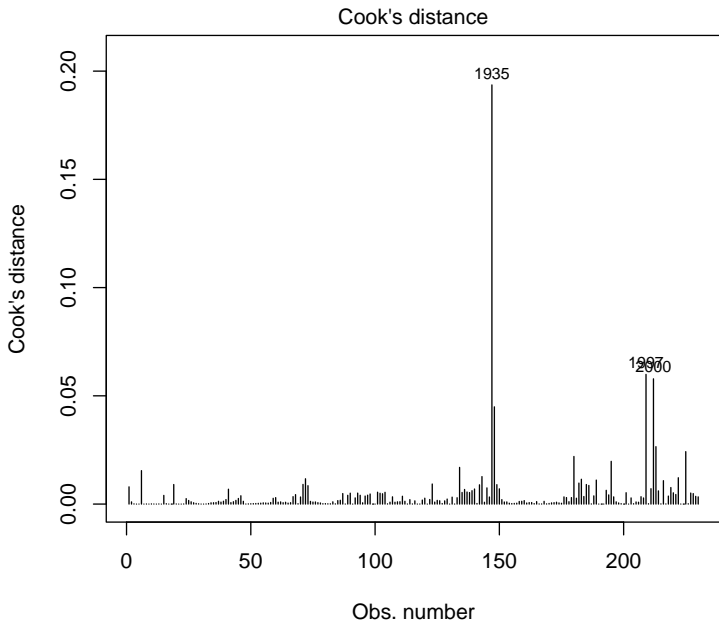


#2: Q-Q Plot of \hat{u} s

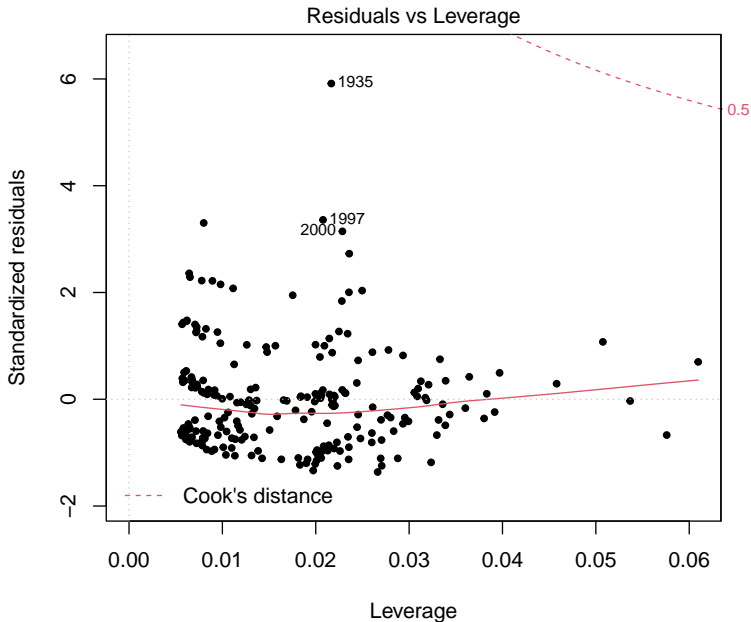


"Scale-Location" Plot

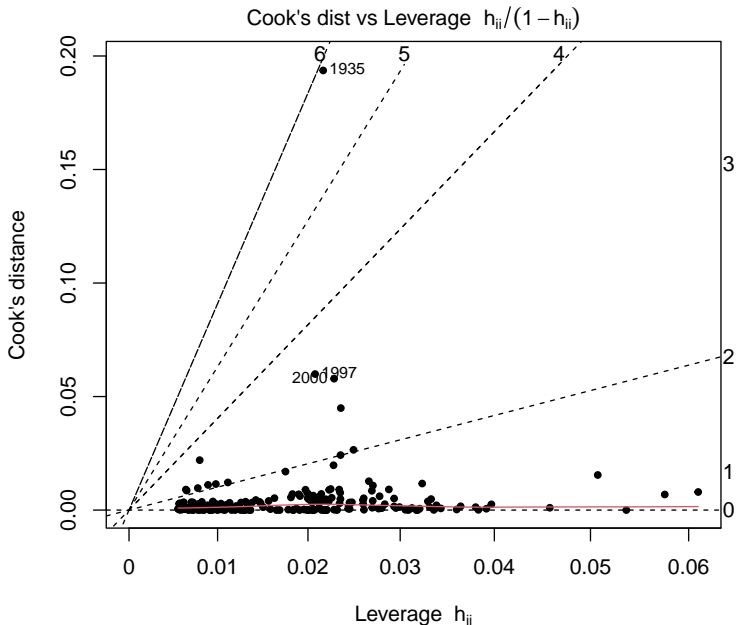




Residuals vs. Leverage



Cook's D vs. Leverage



Simultaneity, Endogeneity, and Instrumental Variables

Consider:

$$Y_1 = \mathbf{X}_1\beta_1 + \gamma_1 Y_2 + \mathbf{u}_1$$

$$Y_2 = \mathbf{X}_2\beta_2 + \gamma_2 Y_1 + \mathbf{u}_2$$

Rewrite:

$$\begin{aligned} Y_1 &= \mathbf{X}_1\beta_1 + \gamma_1[\mathbf{X}_2\beta_2 + \gamma_2 Y_1 + \mathbf{u}_2] + \mathbf{u}_1 \\ &= \mathbf{X}_1\beta_1 + \gamma_1(\mathbf{X}_2\beta_2) + \gamma_1\gamma_2 Y_1 + \gamma_1\mathbf{u}_2 + \mathbf{u}_1 \\ Y_1 - \gamma_1\gamma_2 Y_1 &= \mathbf{X}_1\beta_1 + \gamma_1(\mathbf{X}_2\beta_2) + \gamma_1\mathbf{u}_2 + \mathbf{u}_1 \\ (1 - \gamma_1\gamma_2)Y_1 &= \mathbf{X}_1\beta_1 + \gamma_1(\mathbf{X}_2\beta_2) + \gamma_1\mathbf{u}_2 + \mathbf{u}_1 \\ Y_1 &= \mathbf{X}_1 \left(\frac{1}{1 - \gamma_1\gamma_2} \beta_1 \right) + \mathbf{X}_2 \left(\frac{\gamma_1}{1 - \gamma_1\gamma_2} \beta_2 \right) + \left(\frac{\gamma_1\mathbf{u}_2 + \mathbf{u}_1}{1 - \gamma_1\gamma_2} \right) \\ &= \Delta_1 \mathbf{X}_1 + \Delta_2 \mathbf{X}_2 + \mathbf{e} \end{aligned}$$

$$Y_1 = \mathbf{x}_1 \left(\frac{1}{1 - \gamma_1 \gamma_2} \beta_1 \right) + \mathbf{x}_2 \left(\frac{\gamma_1}{1 - \gamma_1 \gamma_2} \beta_2 \right) + \left(\frac{\gamma_1 \mathbf{u}_2 + \mathbf{u}_1}{1 - \gamma_1 \gamma_2} \right)$$

means

$$\frac{\partial Y_1}{\partial X_\ell} = \frac{\beta_\ell}{1 - \gamma_1 \gamma_2}.$$

But

$$\hat{\Delta}_1 \neq \hat{\beta}_1.$$

For (e.g.)

$$Y_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \gamma_1 Y_2 + \mathbf{u}_1$$

we have:

$$E(Y_2, \mathbf{u}_1) = \frac{\gamma_2}{1 - \gamma_1\gamma_2} \sigma_{\mathbf{u}}^2$$

Result:

- Bias (unless $\gamma_2 = 0$)
- Inconsistency

Consider a variable W that affects both X and Y , so that the true data-generating process is:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + e_i$$

Omitting W from the equation yields:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

where

$$u_i = \beta_2 W_i + e_i.$$

This also results in $\text{Cov}(X_i, u_i) \neq 0 \dots$

Options:

- OLS
- Lagged Variables
- Instrumental Variables (sometimes, ‘Two-Stage Least Squares’ / “2SLS”)
- Systems of Equations / 3SLS / etc.

Instrumental Variables (IV)

Recall that a simple linear model:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{u}$$

gives us:

$$\hat{\beta}_{OLS} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}.$$

Suppose $\text{Cov}(\mathbf{X}, \mathbf{u}) \neq \mathbf{0}$, but we have \mathbf{Z} with

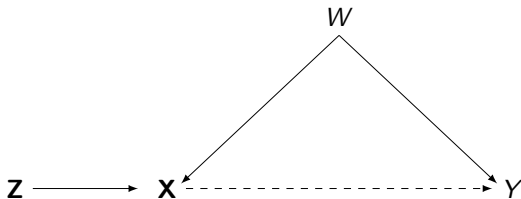
- $\text{Cov}(\mathbf{Z}, \mathbf{X}) \neq \mathbf{0}$ and
- $\text{Cov}(\mathbf{Z}, \mathbf{u}) = \mathbf{0}$.

Then:

$$\begin{aligned}\hat{\beta}_{IV} &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Y} \\ &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'(\mathbf{X}\beta + \mathbf{u}) \\ &= \beta + (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{u}\end{aligned}$$

is consistent.

For an unmeasured / unmodeled *confounder* W :



Consistency requires that:

1. $\text{Cov}(\mathbf{X}, \mathbf{Z}) \neq 0$ (\mathbf{Z} is a *strong instrument*)
2. \mathbf{Z} is *exogenous* / *unconfounded* by W [i.e., $\text{Cov}(\mathbf{Z}, W) = 0$]
3. \mathbf{Z} has no independent effect on Y , except through W (the *exclusion restriction*)

“Two-stage least squares”:

- Regress endogenous \mathbf{X} s variables on $\{\mathbf{Z}, \mathbf{X}\}$
- Generate $\hat{\mathbf{X}}$ s
- Regress Y on $\hat{\mathbf{X}}$ to get β_{2SLS} .
- Adjust standard error estimates

IV Estimation: A Simulation

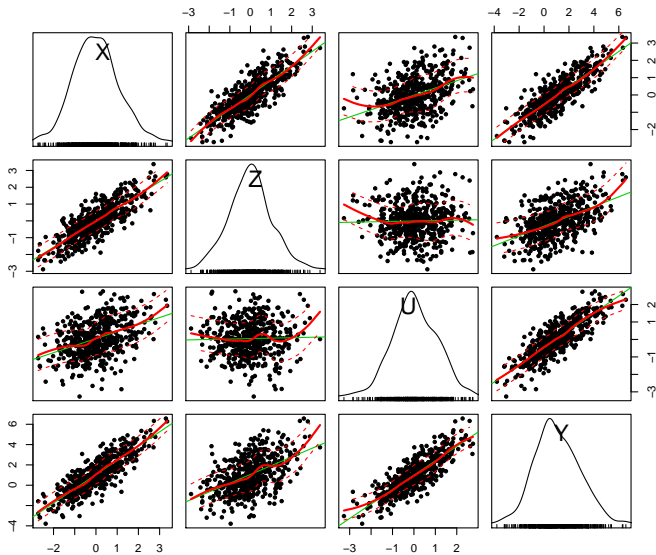
```
library(MASS)
library(sem)
library(car)

seed<-1337
set.seed(seed)

mu<-c(0,0,0) # <== X, Z, U
Sigma<-matrix(c(1,0.8,0.4,0.8,1,0,0.4,0,1),
              nrow=3,byrow=TRUE)          # Cor(X,Y)=0.8, etc.
Vars<- mvrnorm(500,mu,Sigma)
colnames(Vars)<-c("X","Z","U")
Vars<-data.frame(Vars)

Vars$Y<- 1 + Vars$X + Vars$U
```


Plots...



Plain Old OLS...

```
> OLS<- lm(Y~X,data=Vars)
> summary(OLS)
```

Call:

```
lm(formula = Y ~ X, data = Vars)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.3809	-0.6058	-0.0102	0.6320	2.9470

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.04770	0.04209	24.89	<2e-16 ***
X	1.40254	0.04005	35.02	<2e-16 ***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.9413 on 498 degrees of freedom

Multiple R-squared: 0.7112, Adjusted R-squared: 0.7106

F-statistic: 1226 on 1 and 498 DF, p-value: < 2.2e-16

Two-Stage Least Squares

```
> TSLS<-tsls(Y~I(X),data=Vars,instruments=~Z)
> summary(TSLS)
```

2SLS Estimates

Model Formula: $Y \sim I(X)$

Instruments: $\sim Z$

Residuals:

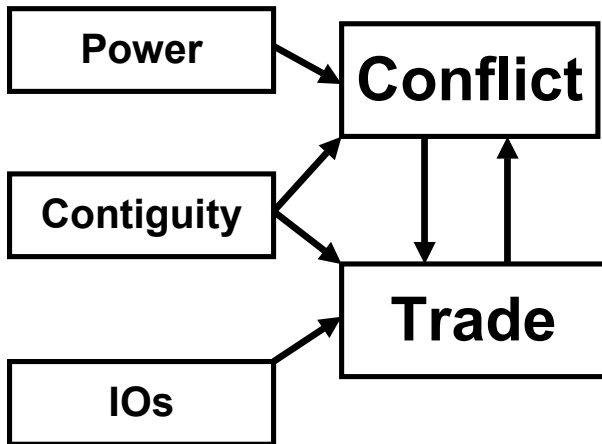
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-3.29300	-0.68210	-0.06139	0.00000	0.76270	2.70300

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.0491828	0.0456017	23.00754	< 2.22e-16 ***
I(X)	1.0302012	0.0536909	19.18763	< 2.22e-16 ***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 1.0196738 on 498 degrees of freedom

IV: A (Toy) Example



```
> summary(IRData)
```

dyadid	logdisputes	logtrade	I0s
Min. : 2020	Min. :-0.6931	Min. :-0.6931	Min. : 4.579
1st Qu.:135155	1st Qu.: -0.6931	1st Qu.: 2.4079	1st Qu.:19.500
Median :220484	Median : -0.6931	Median : 5.5786	Median :27.704
Mean :275526	Mean : -0.2627	Mean : 4.6518	Mean :30.891
3rd Qu.:385710	3rd Qu.: 0.0000	3rd Qu.: 7.1248	3rd Qu.:39.289
Max. :900920	Max. : 3.4965	Max. :11.5037	Max. :93.700

contiguity	capratio	GDPgrowth
Min. :0.0000	Min. : 1.081	Min. : -9.0800
1st Qu.:0.0000	1st Qu.: 4.849	1st Qu.: -0.2923
Median :0.0000	Median : 26.577	Median : 0.8363
Mean :0.3207	Mean : 196.310	Mean : 0.5097
3rd Qu.:1.0000	3rd Qu.: 144.035	3rd Qu.: 1.7106
Max. :1.0000	Max. :7451.982	Max. : 7.0460

Ordinary Regression

```
> OLSWar<-lm(logdisputes~logtrade+contiguity+capratio,data=IRData)
> summary(OLSWar)
```

Call:

```
lm(formula = logdisputes ~ logtrade + contiguity + capratio,
    data = IRData)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.828	-0.326	-0.269	-0.090	3.455

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.4253192	0.0602014	-7.06	3.5e-12 ***
logtrade	0.0085581	0.0105739	0.81	0.419
contiguity	0.4622674	0.0712406	6.49	1.5e-10 ***
capratio	-0.0001296	0.0000647	-2.00	0.045 *

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.853 on 813 degrees of freedom

Multiple R-squared: 0.083, Adjusted R-squared: 0.0796

F-statistic: 24.5 on 3 and 813 DF, p-value: 3.35e-15

2SLS “By-Hand” (stage one)

```
> ITrade<-lm(logtrade~contiguity+IOs+capratio)
> summary(ITrade)
```

Residuals:

Min	1Q	Median	3Q	Max
-6.0385	-1.7666	0.4139	1.6154	7.6029

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.7319793	0.1912570	3.827	0.000140 ***
contiguity	1.3386037	0.1816041	7.371	4.17e-13 ***
IOs	0.1218373	0.0055313	22.027	< 2e-16 ***
capratio	-0.0013913	0.0001626	-8.555	< 2e-16 ***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 2.239 on 813 degrees of freedom

Multiple R-squared: 0.5535, Adjusted R-squared: 0.5519

F-statistic: 335.9 on 3 and 813 DF, p-value: < 2.2e-16

2SLS “By-Hand” (stage two)

```
> IVWarByHand<-with(IRData, lm(logdisputes~capratio+contiguity+
+                               (ITrade$fitted.values)))
> summary(IVWarByHand)
```

Call:

```
lm(formula = logdisputes ~ capratio + contiguity + (ITrade$fitted.values))
```

Residuals:

Min	1Q	Median	3Q	Max
-1.006	-0.362	-0.278	-0.049	3.530

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.1515180	0.0832287	-1.82	0.06905 .
capratio	-0.0002664	0.0000705	-3.78	0.00017 ***
contiguity	0.6263774	0.0788444	7.94	6.5e-15 ***
ITrade\$fitted.values	-0.0558374	0.0171921	-3.25	0.00121 **

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 0.848 on 813 degrees of freedom

Multiple R-squared: 0.094, Adjusted R-squared: 0.0907

F-statistic: 28.1 on 3 and 813 DF, p-value: <2e-16

2SLS, Automagically

```
> library(AER)
> TwoSLSWar<-ivreg(logdisputes~contiguity+capratio+I(logtrade),
  instruments=~contiguity+capratio+IOs)
> summary(TwoSLSWar)
```

Call:

```
ivreg(formula = logdisputes ~ contiguity + capratio + I(logtrade) |
  contiguity + capratio + IOs, data = IRData)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-0.1515180	0.0856218	-1.77	0.07717	.
contiguity	0.6263774	0.0811114	7.72	3.4e-14	***
capratio	-0.0002664	0.0000725	-3.67	0.00025	***
I(logtrade)	-0.0558374	0.0176864	-3.16	0.00165	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.872 on 813 degrees of freedom

Multiple R-Squared: 0.0412, Adjusted R-squared: 0.0376

Wald test: 26.6 on 3 and 813 DF, p-value: <2e-16

Weak Instruments

```
> OLSTrade<-lm(logtrade~logdisputes+contiguity+IOs)
> summary(OLSTrade)
```

Residuals:

Min	1Q	Median	3Q	Max
-6.2467	-2.2067	0.4275	1.6659	6.1264

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.191111	0.182875	1.045	0.296
logdisputes	0.408116	0.095067	4.293	1.98e-05 ***
contiguity	1.357557	0.193109	7.030	4.38e-12 ***
IOs	0.133778	0.005614	23.831	< 2e-16 ***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 2.312 on 813 degrees of freedom

Multiple R-squared: 0.5241, Adjusted R-squared: 0.5223

F-statistic: 298.4 on 3 and 813 DF, p-value: < 2.2e-16

Weak Instruments (continued)

```
> TwoSLSTrade<-ivreg(logtrade~contiguity+IOs+I(logdisputes),  
  instruments=~contiguity+capratio+IOs)  
> summary(TwoSLSTrade)
```

Call:

```
ivreg(formula = logtrade ~ contiguity + IOs + I(logdisputes) |  
  contiguity + capratio + IOs, data = IRData)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.1501	0.8512	2.53	0.0117	*
contiguity	-2.7276	1.5262	-1.79	0.0743	.
IOs	0.1720	0.0205	8.41	<2e-16	***
I(logdisputes)	7.3712	2.4520	3.01	0.0027	**

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 6.37 on 813 degrees of freedom

Multiple R-Squared: -2.62, Adjusted R-squared: -2.63

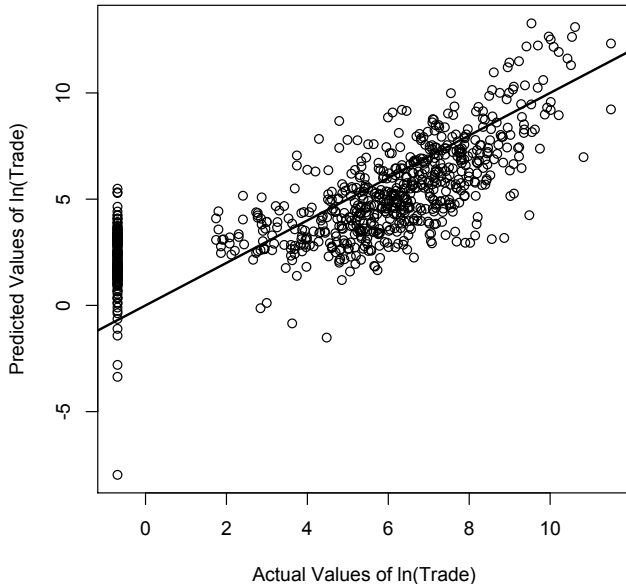
Wald test: 41.5 on 3 and 813 DF, p-value: <2e-16

	<i>Dependent variable:</i>			
	<u>ln(Disputes)</u>		<u>ln(Trade)</u>	
	<i>OLS</i>	<i>IV</i>	<i>OLS</i>	<i>IV</i>
ln(Trade)	0.009 (0.011)			
ln(Disputes)			0.408*** (0.095)	
Contiguity	0.462*** (0.071)	0.626*** (0.081)	1.358*** (0.193)	-2.728* (1.526)
Capability Ratio	-0.0001** (0.0001)	-0.0003*** (0.0001)		
l(logtrade)		-0.056*** (0.018)		
IOs			0.134*** (0.006)	0.172*** (0.020)
l(logdisputes)				7.371*** (2.452)
Constant	-0.425*** (0.060)	-0.152* (0.086)	0.191 (0.183)	2.150** (0.851)
Observations	817	817	817	817
R ²	0.083	0.041	0.524	-2.616
Adjusted R ²	0.080	0.038	0.522	-2.630
Residual Std. Error (df = 813)	0.853	0.872	2.312	6.372
F Statistic (df = 3; 813)	24.530***		298.400***	

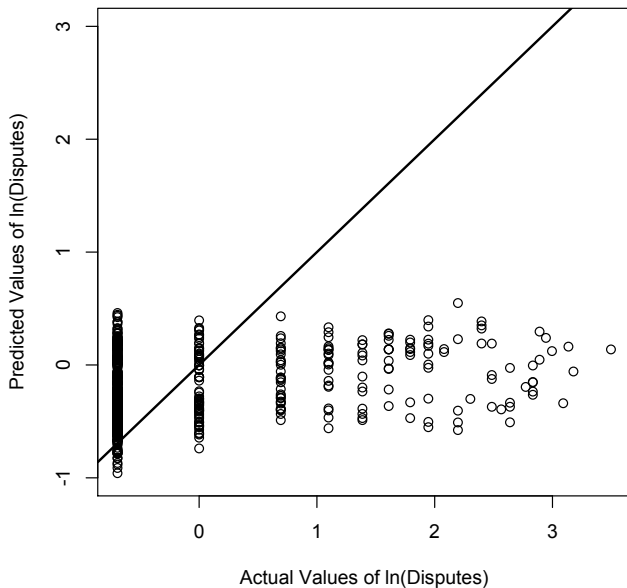
Note:

* p<0.1; ** p<0.05; *** p<0.01

Pretty Good Instrument (Trade)



Crappy Instrument (War)



Things to ask yourself when you see someone attempting an instrumental variables analysis:

1. How strong is the relationship between the instrument (Z) and the instrumented predictor (X)?
2. Is the instrument (Z) plausibly unconfounded with X ?
3. Does the instrument meet the exclusion restriction? That is, is the only way that Z influences Y via its effect on X ?