Bush 631-603: Quantitative Methods

Lecture 13 (04.18.2023): Uncertainty vol. III

Rotem Dvir

The Bush school of Government and Public Policy

Texas A&M University

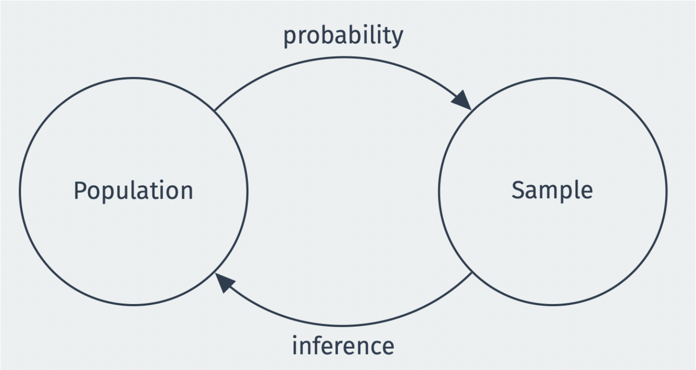
Spring 2023

What is today's plan?

- Calculating uncertainty: the full package.
- Linear regression model estimator.
- Assumptions for OLS estimators.
- Bivariate and multivariate models.
- R work: lm(), summary(lm())

Our data - our research interests

- Making inferences from data to population



Statistical hypothesis testing

- Probabilistic *proof by contradiction*
- Assume the contrast to our expectations is not possible.
- Assume $\rightarrow$ difference (sample and analyst) are zero.
- Incorrect? $\rightarrow$ differences exist.
- Senior analyst may have been wrong.
- We can never **fully** reject a hypothesis (no 100% certainty).

Procedure for hypothesis tests

- Steps for testing:
  1. Define null and alternative hyps (H_0; H_1).
  2. Select *test statistic* and level of test (α).
  3. Derive reference distribution.
  4. Calculate p-values.
  5. Make a decision: reject/retain.
- **Decision rule**:
  - **Reject null** if p-value is *below* α
  - Otherwise **retain the null** or **fail to reject**.
- Common thresholds:
  - $p \geq 0.1$: "not statistically significant".
  - $p < 0.05$: "statistically significant".
  - $p < 0.01$: "highly significant".

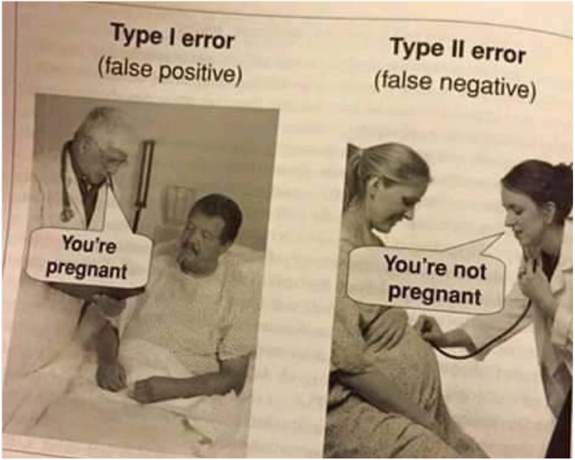
Test errors

- $p = 0.05 \rightarrow$ extreme data only happen in 5% of repeated samples (if null is true).
- $\rightsquigarrow$ 5% of time we reject null that is true!
- Types of errors:

	H_0 True	H_0 False
Retain H_0	Awesome!	Type II error
Reject H_0	Type I error	Good stuff!

Test errors

- What does these errors mean?



One sample test

- The z-statistic:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Or:

$$Z = \frac{observed - null}{SE}$$

- How many SEs away from the null guess is the sample mean?
- **Small samples problem**: uncertainty about $\bar{X}$ distribution.
- Find t-statistic instead:

$$T = \frac{\bar{X} - \mu}{\hat{SE}} \approx t_{n-1}$$

Two sample tests

- Goal: learn about population difference in means.
- Compare differences b-w multiple groups: same testing procedures.
- Define:
  - Null PATE: $H_0 : \mu_T - \mu_C = 0$
  - Alt. PATE: $H_1 : \mu_T - \mu_C \neq 0$
  - Test statistic: diff-in-means estimator.
  - z-score for *two sample z-test*.
- Are the differences in sample means just random chance?

Two sample test

- Run a **two sample t-test** $\rightarrow$ t.test()

```
t.test(exp.dat$cont_cor1[exp.dat$trt1 == 0],
       exp.dat$cont_cor1[exp.dat$trt1 == 1])
```

```
##
##  Welch Two Sample t-test
##
## data:  exp.dat$cont_cor1[exp.dat$trt1 == 0] and exp.dat$cont_cor1[exp.dat$tr
## t = -13.697, df = 993.53, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -23.59653 -17.68267
## sample estimates:
## mean of x mean of y
##  1489.333  1509.973
```

What we did? and next...

- So far, we covered uncertainty in:
  - Sample proportions (Trump vs. the polls).
  - Sample means (Israel thermometer scores).
  - Differences in sample means (experimental data, leaders' type).
- What about our regression estimates?
- Much uncertainty about them too!

Least squared

- Assumption: model $\rightsquigarrow$ Data generation process (DGS)
- **Parameters/coefficients** (α, β): true values unknown.
- Use data to estimate $\alpha, \beta \implies \hat{\alpha}, \hat{\beta}$
- Predictions:
  - Use the *regression line*.
  - Calculate *fitted value* ($\neq$ observed value)

$$\hat{Y} = \hat{\alpha} + \hat{\beta} * x$$

Linear model elements

- *Residual/prediction error*: the difference b-w fitted and observed values.
- Real error is unknown $\Rightarrow \hat{\epsilon}$

$$\hat{\epsilon} = Y - \hat{Y}$$

Linear model estimation

Least squared:

- A method to estimate the regression line.
- Use data (values of Y & X_i).
- 'Select' $\hat{\alpha}, \hat{\beta}$ to minimize SSR.

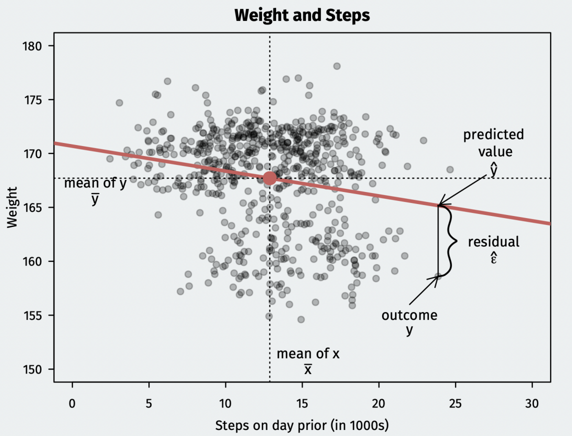
$$SSR = \sum_{i=1}^{n} \hat{\epsilon}^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta} * X_i)^2$$

Linear regression in R

Fit the model

- Syntax: lm(Y ∼ x, data = mydata)
- Y = dependent variable; x = independent variable(s).

How does it look like?



Linear models in RCT

Binary independent variable:

- Slope coefficient (β) = diff-in-means estimator.
- $\hat{\beta}$: estimated average treatment effect.

- Why works?
  - Randomization $\rightarrow$ causal interpretation
  - *Slope* (β): the average change in Y when X increases by 1 unit.

When X is binary:

- Treatment: yes or no.
- X change by 1 unit $\rightarrow$ no to yes.
- Y changes as well.

Building linear models

- Leader background and nuclear technology pursuit (2015)
- Rebel or not?
- Our model $\rightarrow$ rebel exp. & nukes technology.

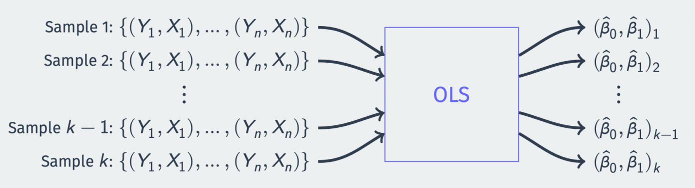
- $Y_i = \beta_0 + \beta_1 * RebelExp_i + \epsilon_i$
- P(Nukes) = rebel experience and ϵ (error).

Uncertainty in regression

- Quantify uncertainty in linear models
- Model parameters - estimators
- What estimator? **least squared**.

Least squared estimator

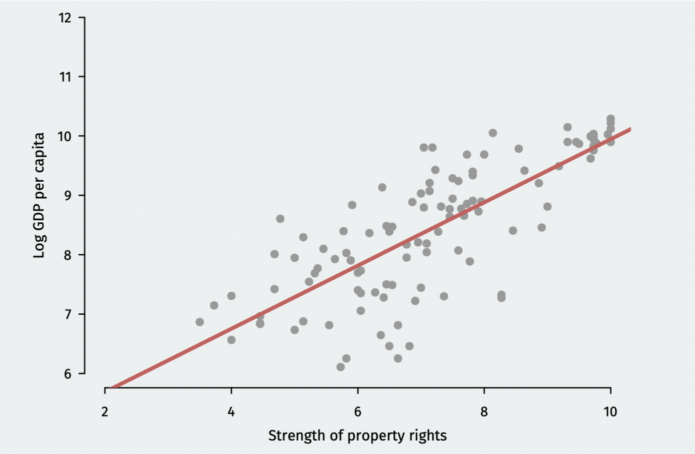
- We 'plug-in' data and get estimates.



- Estimators values are uncertain.

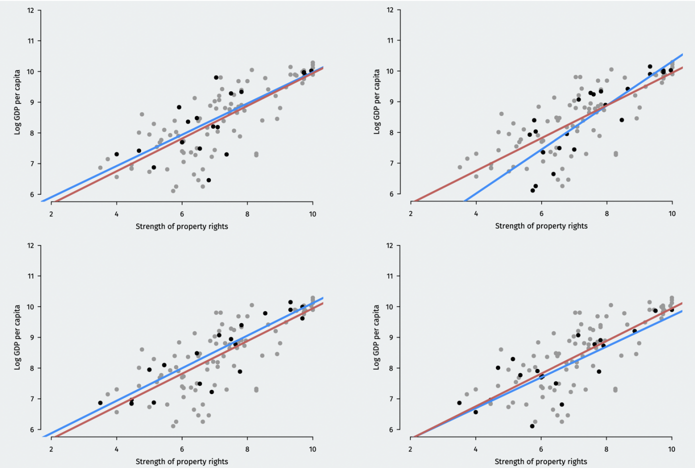
Uncertainty of least squared estimators

- Data: Relationship between strength of property rights and GDP.



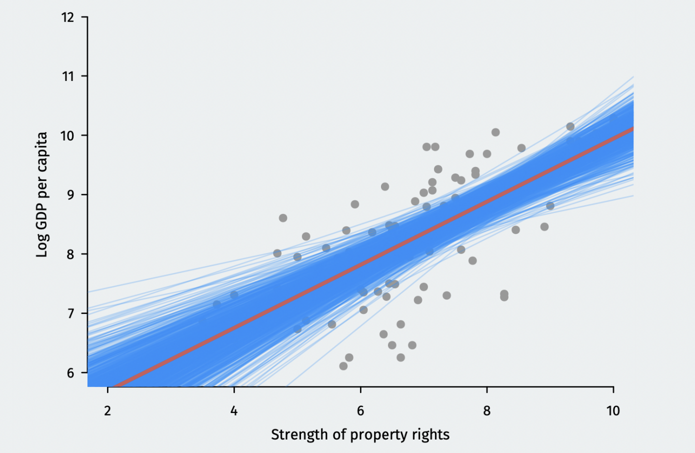
Simulation Again?

- Sample 30 countries and calculate lm(GDP ~ Property.rights)



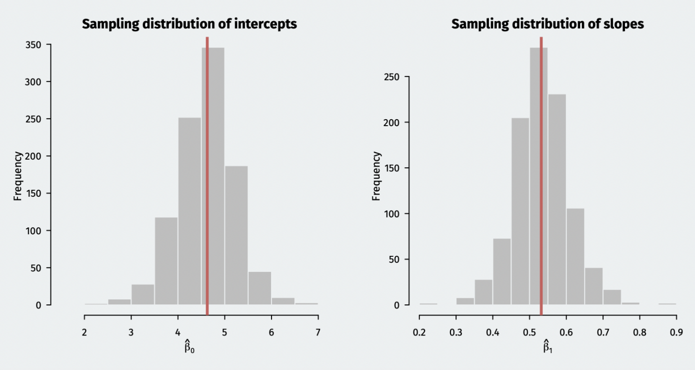
Simulation Again?

- Multiple iterations of the model within the data.



OLS sampling distributions

- Variations of intercept ($\hat{\beta}_0$) and slope ($\hat{\beta}_1$)



Least squared estimator

- Uncertainty in *least squared* estimator:
  - Generate reference distribution.
  - Calculate SEs.
  - Construct 95% CIs.
  - Run hypotheses tests.
  - Results are ‘statistically significant’, or not.
  - Is our estimator different than zero? (reject the null)

Assumptions

- Assumptions for regression estimates:

(1) Exogeneity: mean of ϵ_i does not depend on X_i

$$E(\epsilon_i|X_i) = E(\epsilon_i) = 0$$

(2) Homoskedasticity: variance of ϵ_i does not depend on X_i

$$V(\epsilon_i|X_i) = V(\epsilon_i) = \sigma^2$$

Problem of exogenous factors

- Confounders between X_i and Y_i
- Factors in ϵ_i that are related to X_i
- Why?

- Business background (X_i) $\rightarrow$ defense spending (Y_i)
- Socioeconomic background $\rightarrow \epsilon_i$
- But Socioeconomic background $\rightarrow$ Business experience, so...
- Is Y_i due to business experience?

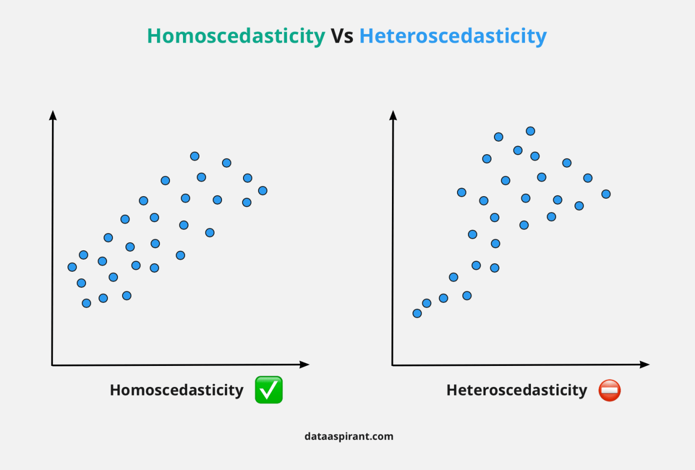
Problem of exogenous factors

- RCTs $\rightarrow$ **no exogeneity** problem.
- Randomized treatments!

- Severe issue for observational studies.
- Rebel background $\rightarrow$ nuclear weapons pursuit.
- Perhaps more conflicts $\rightarrow$ pursue advanced technology.

Homoskedas... what?

- When spread of Y_i depends on X_i



OLS properties

$$Y_i = \beta_0 + \beta_1 * X_i + \epsilon_i$$

- Our estimates: $\hat{\beta}_0, \hat{\beta}_1$ are r.v.s.
- Equal to true value? (population parameters)
- How spread are they around their center?
- Estimate the SE $\rightarrow \hat{SE}(\hat{\beta}_1)$
- Next? construct CIs...
- Run hypotheses tests.

Putting everything together

- Hypotheses:
  - $H_0 : \beta_1 = 0$
  - $H_a : \beta_1 \neq 0$
- Our estimators: $\hat{\beta}_0, \hat{\beta}_1$
- SE and CIs:
  - $\hat{\beta}_0 \pm 1.96 * \hat{SE}(\hat{\beta}_0)$
  - $\hat{\beta}_1 \pm 1.96 * \hat{SE}(\hat{\beta}_1)$
- Hypotheses test:
  - Test statistic: $\frac{\hat{\beta}_1 - \hat{\beta}_1^*}{\hat{SE}(\hat{\beta}_1)} \sim$ N(0,1)
  - $\hat{\beta}_1$ is **statistically significant** if $p < 0.05$ (reject null H_0).

Now with data

- Rebel experience and pursuit of nuclear tech (2015)

```
head(nukes, n=9)
```

```
## # A tibble: 9 x 76
##   ccode idacr  year leadid30 leader~1 startdate  inday inmonth inyear starty
##   <dbl> <chr> <dbl> <chr>    <chr>    <date>     <dbl>   <dbl>  <dbl> <date>
## 1     2 USA    1945 A2.9-43  Rooseve~ 1933-03-04     4       3   1933 1945-0
## 2     2 USA    1945 A2.9-46  Truman   1945-04-12    12       4   1945 1945-0
## 3     2 USA    1946 A2.9-46  Truman   1945-04-12    12       4   1945 1946-0
## 4     2 USA    1947 A2.9-46  Truman   1945-04-12    12       4   1945 1947-0
## 5     2 USA    1948 A2.9-46  Truman   1945-04-12    12       4   1945 1948-0
## 6     2 USA    1949 A2.9-46  Truman   1945-04-12    12       4   1945 1949-0
## 7     2 USA    1950 A2.9-46  Truman   1945-04-12    12       4   1945 1950-0
## 8     2 USA    1951 A2.9-46  Truman   1945-04-12    12       4   1945 1951-0
## 9     2 USA    1952 A2.9-46  Truman   1945-04-12    12       4   1945 1952-0
## # ... with 66 more variables: enddate <date>, outday <dbl>, outmonth <dbl>,
## #   outyear <dbl>, yearlyduration <dbl>, entry <dbl+lbl>, exit <dbl+lbl>,
## #   pursuit <dbl>, initiation <dbl>, explore <dbl>, bombprgm <dbl>,
## #   pursuitjg <dbl>, pursuitsw <dbl>, rebel <dbl>, milservice <dbl>,
## #   jcrevolutionary <dbl>, revolutionaryleader <dbl>, irregular <dbl>,
## #   fiveyear <dbl>, polity2 <dbl>, total <dbl>, spally <dbl>, NCA67 <dbl>,
## #   gdpcap <dbl>, lngdpcap <dbl>, npt <dbl>, openness <dbl>, rivalry <dbl>,
```

Rebels and Nukes (2015)

- OLS regression models in R

```r
lm(pursuit ~ rebel, data = nukes)
```

```
##
## Call:
## lm(formula = pursuit ~ rebel, data = nukes)
##
## Coefficients:
## (Intercept)        rebel
##     0.01051      0.03767
```

Rebels and Nukes (2015)

- Simple/bivariate regression

```
summary(lm(pursuit ~ rebel, data = nukes))
```

```
##
## Call:
## lm(formula = pursuit ~ rebel, data = nukes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.04819 -0.04819 -0.01051 -0.01051  0.98949
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.010513   0.002295   4.582 4.68e-06 ***
## rebel       0.037673   0.003513  10.725  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1598 on 8460 degrees of freedom
##   (390 observations deleted due to missingness)
## Multiple R-squared:  0.01341,    Adjusted R-squared:  0.0133
## F-statistic:   115 on 1 and 8460 DF,  p-value: < 2.2e-16
```

Rebels and Nukes (2015)

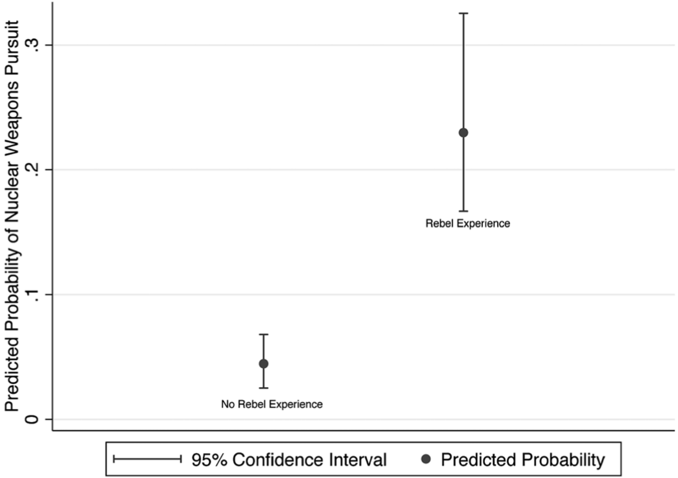
- Multivariate regression: account for confounders

```
summary(lm(pursuit ~ rebel + milservice + polity2, data = nukes))
```

```
##
## Call:
## lm(formula = pursuit ~ rebel + milservice + polity2, data = nukes)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.06587 -0.04408 -0.02544 -0.01020  0.99682
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0073899  0.0027782   2.660  0.00783 **
## rebel       0.0320096  0.0044238   7.236 5.08e-13 ***
## milservice  0.0217914  0.0045106   4.831 1.38e-06 ***
## polity2     0.0004679  0.0002801   1.670  0.09489 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1672 on 7684 degrees of freedom
##   (1164 observations deleted due to missingness)
## Multiple R-squared:  0.01596,    Adjusted R-squared:  0.01558
## F-statistic: 41.54 on 3 and 7684 DF,  p-value: < 2.2e-16
```

OLS coefficient interpretation

- Rebel experience and nuclear technology (2015)



OLS Multivariate regression

- **Remember**: correlation does not mean causation.
- Multiple confounders $\rightarrow$ same process:
  - CIs are constructed the same for all $\hat{\beta}_j$.
  - Hypothesis tests also run the same for all $\hat{\beta}_j$.
  - p-values have the same interpretation.
- Interpretation of $\hat{\beta}_j$:
  - A change in Y_i is associated with a one-unit increase in X_i when...
  - All other variables are held constant (at mean value, usually).

OLS regression models: FP research

- Joint military exercises and conflict (2021)

JME and conflict

- Under what conditions violence is more likely? who will initiate?
- Outcome conditioned by alliance partnership.
- Use two-stage model:
  1. Selection into conflict.
  2. Effects of JMEs.
- Data: directed dyad-year (1973-2003).

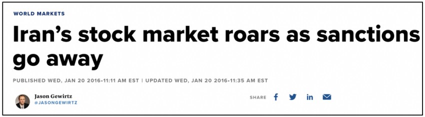
JME and military conflict

Table 2. Main Results for the Effects of JMEs and Alliances on Conflict Escalation.

	Targets		Participants	
	Model 1:	Model 2:	Model 3:	Model 4:
JME	−0.311***		−0.573***	
	(0.100)		(0.101)	
Non-Ally JME		−0.050		−0.148
		(0.146)		(0.141)
Ally JME		−0.443***		−0.823***
		(0.117)		(0.124)
Alliances	0.013*	0.016**	−0.009	−0.004
	(0.007)	(0.007)	(0.008)	(0.008)
Joint Democracy	−0.753***	−0.745***	−0.730***	−0.720***
	(0.092)	(0.092)	(0.089)	(0.089)
CINC	9.042***	8.901***	10.800***	10.597***
	(1.114)	(1.114)	(1.063)	(1.063)
UNGA	−0.055	−0.050	−0.047	−0.041
	(0.045)	(0.045)	(0.044)	(0.044)
Trade	0.00001	0.00001	0.00001	0.00001
	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Lagged DV	6.631***	6.623***	6.171***	6.159***
	(0.092)	(0.092)	(0.092)	(0.092)
Constant	−6.970***	−6.984***	−6.945***	−6.967***
	(0.272)	(0.272)	(0.271)	(0.271)
N	541,920	541,920	541,920	541,920
AIC	7,757.394	7,753.953	8,415.870	8,402.368
Log Likelihood	−3,839.697	−3,836.977	−4,168.935	−4,161.184

Note: Coefficients Represent Logistic Regression Coefficients.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Targeting the stock market



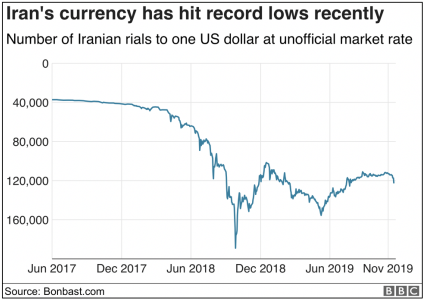
WORLD MARKETS

Iran's stock market roars as sanctions go away

PUBLISHED WED, JAN 20 2016•11:11 AM EST | UPDATED WED, JAN 20 2016•11:35 AM EST

Jason Gewirtz
@JASONGEWIRTZ

SHARE



Targeting the stock market

- How sanctions affect stock markets' in targeted countries (2021).
- Imposing costs on stock market $\rightarrow$ behavior change.
- Account for types of sanctions.
- The cumulative effects of sanctions over time.
- Data: monthly stock market values for 66 countries (1990-2005)

Targeting the stock market

- Types of sanctions matter:
  - Import: restrict access to global markets and reduce firm revenues.
  - Also harm exporters: investment shifts away from losing firms.
  - Export: limits on exports thus loss of hard currency.
  - Less efficient as import firms make-up for lost capital and goods.
- Example: Iraqi oil boycott (1990).
- Cumulative sanctions regime:
  - More is better.
  - But decreasing marginal effect.
  - Initial sanctions are more useful
  - Target adjusts to additional restrictions.

Targeting the stock market

- Empirical analysis:
  - OLS regression models.
  - ADL: account for time lags.
- Results:
  - Negative effect on stocks.
  - Type matters, as well as number of sanctions.
  - Sender state also matters.
- Models 1&2: full and reduced set of controls.
- Models 3-5: sanctions types.
- Models 6&7: Comparing G20 to non-G20 countries.

International Aid and civilian casualties

Apr 13, 2016

Balochistan: Pakistan Army Kills Over 35 Civilians and Carries Out Mass Abductions

International Aid and civilian casualties

- Are civilians facing risks due to aid distribution?
- Two mechanisms:
  1. Persuasion: reduce incentives to target civilians (military).
  2. Predation: adverse incentives for resource capturing and extended collective violence (development).
- Data: military and ODA flows in 135 countries (1989-2011).

Military and development aid flows

Variables	1(a) U.S. military aid	1(b) Development aid	1(c) Full model	1(d) Lagged DV	1(e) Excluding outliers
OSV (t-1)				0.000**	0.0148**
				(0.000)	(0.00666)
U.S. military aid (logged, lagged)	−0.338***		−0.368***	−0.348***	−0.187**
	(0.109)		(0.097)	(0.101)	(0.090)
Development aid (logged, lagged)		0.237**	0.366***	0.371***	0.269**
		(0.117)	(0.135)	(0.136)	(0.130)
State strength	−0.000	−0.002***	0.000	−0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Polity2	−0.256***	−0.117*	−0.167**	−0.151*	−0.009
	(0.075)	(0.069)	(0.079)	(0.079)	(0.045)
Rebel OSV (lag)	−0.000	0.001	−0.000	−0.001	−0.001
	(0.001)	(0.002)	(0.000)	(0.001)	(0.001)
Intrastate conflict	4.717***	4.858***	5.230***	5.463***	3.653***
	(0.646)	(0.634)	(0.709)	(0.816)	(0.621)
Trade openness	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Previous regime change	2.107***	2.071***	2.036***	2.002***	2.045***
	(0.380)	(0.520)	(0.359)	(0.368)	(0.382)
Oil production	−0.299***	−0.237***	−0.223**	−0.235***	−0.109
	(0.087)	(0.074)	(0.091)	(0.090)	(0.097)
Ethnic exclusion	0.754***	0.776***	0.765***	0.724***	0.263
	(0.210)	(0.202)	(0.215)	(0.224)	(0.172)
Ethnic fractionalization	0.624	0.362	−0.201	−0.163	−0.286
	(0.799)	(0.852)	(0.814)	(0.817)	(0.901)
Constant	4.112**	−2.553***	2.301	1.965	−0.300
	(1.759)	(0.975)	(1.791)	(1.831)	(1.523)
Observations	2,032	2,791	2,032	2,032	2,005

Note. Robust standard errors in parentheses.
***$p < .01$, **$p < .05$, *$p < .1$

What to do with reg models?

- Regression models:
  - Useful tool to assess causality.
  - Pack **a lot** of information.
  - Can be hard to interpret.
- So, what to do?
  - Substantive results.
  - Predictions!!
  - Sub-groups and effects by types.

Show meaningful results!

Reg models to presentations

- Predictions → quantity of interest

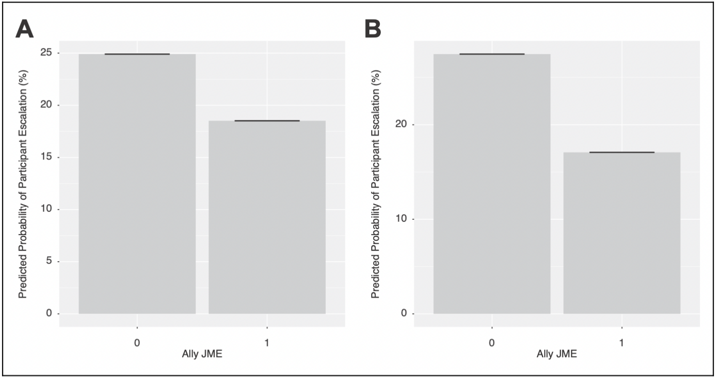
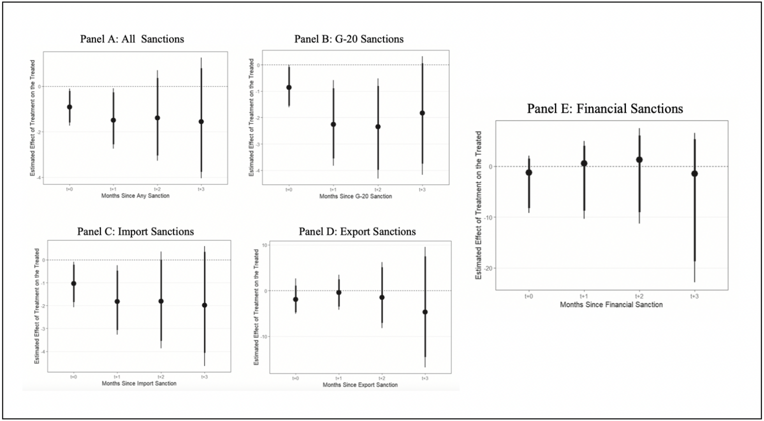


Figure 3. Predicted probability of *Escalation* as a function of *Ally JME*, with 95 percent confidence intervals. Results obtained from a Heckman selection model and are conditional upon conflict onset. (A) Targets. (B) Participants.

Reg models to presentations

- Predicting sanction types effectiveness



Wrapping up Week 13

- Summary:
  - Testing uncertainty: the full package.
  - Linear regression model estimator.
  - Assumptions for OLS estimators.
  - Bivariate and multivariate models.
  - Interpretation of β coefficient.
  - Reading a regression table.