Lecture 12: Model Specification Strategies

ECON 480 - Econometrics - Fall 2018

Ryan Safner

October 31, 2018

HOOD COLLEGE

Model Specification

Applications of Simple Multivariate OLS Models

HOOD COLLEGE

Model Specification

Model Specification

- The big challenge in applied econometrics is choosing how to specify a model to regress

HOOD COLLEGE

Model Specification

- The big challenge in applied econometrics is choosing how to specify a model to regress
- Every dataset is different, every study has a different goal

HOOD COLLEGE

Model Specification

- The big challenge in applied econometrics is choosing how to specify a model to regress
- Every dataset is different, every study has a different goal
    - there is no bright line rule, only a set of guidelines and skills that you can only learn by doing!

HOOD COLLEGE

Model Specification

- The big challenge in applied econometrics is choosing how to specify a model to regress
- Every dataset is different, every study has a different goal
    - there is no bright line rule, only a set of guidelines and skills that you can only learn by doing!
- But here are some helpful tips and frequent problems (and solutions)

HOOD COLLEGE

Model Specification: Process

1. **Identify your question of interest:** what do you want to know? What marginal effect(s) do you want to estimate?

HOOD COLLEGE

Model Specification: Process

1. **Identify your question of interest:** what do you want to know? What marginal effect(s) do you want to estimate?
2. **Think about possible sources of endogeneity:** what *other* variables would cause omitted variable bias if we left them out? Can we get data on them too?

HOOD COLLEGE

Model Specification: Process

1. **Identify your question of interest:** what do you want to know? What marginal effect(s) do you want to estimate?
2. **Think about possible sources of endogeneity:** what *other* variables would cause omitted variable bias if we left them out? Can we get data on them too?
    - Again: must BOTH (1) affect *Y* AND (2) be correlated with *X*

HOOD COLLEGE

Model Specification: Process

1. **Identify your question of interest:** what do you want to know? What marginal effect(s) do you want to estimate?
2. **Think about possible sources of endogeneity:** what *other* variables would cause omitted variable bias if we left them out? Can we get data on them too?
    - Again: must BOTH (1) affect Y AND (2) be correlated with X
    - This requires much of your economic intuitions: R^2 and statistical measures cannot tell you everything!

HOOD COLLEGE

Model Specification: Process

1. **Identify your question of interest:** what do you want to know? What marginal effect(s) do you want to estimate?
2. **Think about possible sources of endogeneity:** what *other* variables would cause omitted variable bias if we left them out? Can we get data on them too?
    - Again: must BOTH (1) affect Y AND (2) be correlated with X
    - This requires much of your economic intuitions: R^2 and statistical measures cannot tell you everything!
3. **Run multiple models and check the robustness of your results:** does the size (or direction) of your marginal effect(s) of interest change as you change your model (i.e. add more variables)?

HOOD COLLEGE

Model Specification: Process

1. **Identify your question of interest:** what do you want to know? What marginal effect(s) do you want to estimate?
2. **Think about possible sources of endogeneity:** what *other* variables would cause omitted variable bias if we left them out? Can we get data on them too?
    - Again: must BOTH (1) affect Y AND (2) be correlated with X
    - This requires much of your economic intuitions: R^2 and statistical measures cannot tell you everything!
3. **Run multiple models and check the robustness of your results:** does the size (or direction) of your marginal effect(s) of interest change as you change your model (i.e. add more variables)?
4. **Interpret your results**

HOOD COLLEGE

Model Specification: Process

1. **Identify your question of interest:** what do you want to know? What marginal effect(s) do you want to estimate?
2. **Think about possible sources of endogeneity:** what *other* variables would cause omitted variable bias if we left them out? Can we get data on them too?
   - Again: must BOTH (1) affect Y AND (2) be correlated with X
   - This requires much of your economic intuitions: R^2 and statistical measures cannot tell you everything!
3. **Run multiple models and check the robustness of your results:** does the size (or direction) of your marginal effect(s) of interest change as you change your model (i.e. add more variables)?
4. **Interpret your results**
   - Are they statistically significant?

HOOD COLLEGE

Model Specification: Process

1. **Identify your question of interest:** what do you want to know? What marginal effect(s) do you want to estimate?
2. **Think about possible sources of endogeneity:** what *other* variables would cause omitted variable bias if we left them out? Can we get data on them too?
   - Again: must BOTH (1) affect Y AND (2) be correlated with X
   - This requires much of your economic intuitions: R^2 and statistical measures cannot tell you everything!
3. **Run multiple models and check the robustness of your results:** does the size (or direction) of your marginal effect(s) of interest change as you change your model (i.e. add more variables)?
4. **Interpret your results**
   - Are they statistically significant?
   - Regardless of *statistical* significance, are they **economically meaningful**?

HOOD COLLEGE

Model Specification: Process

1. **Identify your question of interest:** what do you want to know? What marginal effect(s) do you want to estimate?
2. **Think about possible sources of endogeneity:** what *other* variables would cause omitted variable bias if we left them out? Can we get data on them too?
    - Again: must BOTH (1) affect Y AND (2) be correlated with X
    - This requires much of your economic intuitions: R^2 and statistical measures cannot tell you everything!
3. **Run multiple models and check the robustness of your results:** does the size (or direction) of your marginal effect(s) of interest change as you change your model (i.e. add more variables)?
4. **Interpret your results**
    - Are they statistically significant?
    - Regardless of *statistical* significance, are they **economically meaningful**?
        - Why should we care?

HOOD COLLEGE

Model Specification: Process

1. **Identify your question of interest:** what do you want to know? What marginal effect(s) do you want to estimate?
2. **Think about possible sources of endogeneity:** what *other* variables would cause omitted variable bias if we left them out? Can we get data on them too?
    - Again: must BOTH (1) affect Y AND (2) be correlated with X
    - This requires much of your economic intuitions: R^2 and statistical measures cannot tell you everything!
3. **Run multiple models and check the robustness of your results:** does the size (or direction) of your marginal effect(s) of interest change as you change your model (i.e. add more variables)?
4. **Interpret your results**
    - Are they statistically significant?
    - Regardless of *statistical* significance, are they **economically meaningful**?
        - Why should we care?
        - How big is "big"?

HOOD COLLEGE

Proxy Variables

- Ideally, we would want a randomized control experiment to assign individuals to treatment

HOOD COLLEGE

Proxy Variables

- Ideally, we would want a randomized control experiment to assign individuals to treatment
- But with observational data, ϵ_i depends on other factors

HOOD COLLEGE

Proxy Variables

- Ideally, we would want a randomized control experiment to assign individuals to treatment
- But with observational data, ϵ_i depends on other factors
    - *If* we can observe and measure these factors, then include them in the regression

HOOD COLLEGE

Proxy Variables

- Ideally, we would want a randomized control experiment to assign individuals to treatment
- But with observational data, ϵ_i depends on other factors
    - *If* we can observe and measure these factors, then include them in the regression
- If we can't directly measure them, often we can include variables *correlated* with these variables to proxy for the effects of them!

HOOD COLLEGE

Example
Consider test scores and class sizes again. What about learning opportunities outside of school?

- Probably a bias-causing omitted variable (affects test score and correlated with class size) but we can't measure it!

Example
Consider test scores and class sizes again. What about learning opportunities outside of school?

- Probably a bias-causing omitted variable (affects test score and correlated with class size) but we can't measure it!
- But suppose we *can* measure a variable V, and significantly, $corr(V, Z) \neq 0$

Example
Consider test scores and class sizes again. What about learning opportunities outside of school?

- Probably a bias-causing omitted variable (affects test score and correlated with class size) but we can't measure it!
- But suppose we *can* measure a variable V, and significantly, $corr(V, Z) \neq 0$
- e.g. we have data on the percent of students who get a free or subsidized lunch ('meal_pct')

HOOD COLLEGE

Example
Consider test scores and class sizes again. What about learning opportunities outside of school?

- Probably a bias-causing omitted variable (affects test score and correlated with class size) but we can't measure it!
- But suppose we *can* measure a variable V, and significantly, $corr(V, Z) \neq 0$
- e.g. we have data on the percent of students who get a free or subsidized lunch ('meal_pct')
- This is a good **proxy** for income-determined learning opportunities outside of school

HOOD COLLEGE

Example
Consider test scores and class sizes again. What about learning opportunities outside of school?

- Probably a bias-causing omitted variable (affects test score and correlated with class size) but we can't measure it!
- But suppose we *can* measure a variable V, and significantly, $corr(V, Z) \neq 0$
- e.g. we have data on the percent of students who get a free or subsidized lunch ('meal_pct')
- This is a good **proxy** for income-determined learning opportunities outside of school
  - %meal is correlated with Income, which is correlated with both class size and test score

HOOD COLLEGE

Example
Consider test scores and class sizes again. What about learning opportunities outside of school?

- Probably a bias-causing omitted variable (affects test score and correlated with class size) but we can't measure it!
- But suppose we *can* measure a variable V, and significantly, $corr(V, Z) \neq 0$
- e.g. we have data on the percent of students who get a free or subsidized lunch ('meal_pct')
- This is a good **proxy** for income-determined learning opportunities outside of school
  - %meal is correlated with Income, which is correlated with both class size and test score
  - So this is a good *indirect* measure of Income

HOOD COLLEGE

Proxy Variables: Example II

- We've been assuming we don't have data on average district income, we would expect `meal_pct` to be strongly negatively correlated with income

	testscr	str	el_pct	avginc	meal_pct
testscr	1	-0.226	-0.644	0.712	-0.869
str	-0.226	1	0.188	-0.232	0.135
el_pct	-0.644	0.188	1	-0.307	0.653
avginc	0.712	-0.232	-0.307	1	-0.684
meal_pct	-0.869	0.135	0.653	-0.684	1

HOOD COLLEGE

Proxy Variables: Example II

- We've been assuming we don't have data on average district income, we would expect `meal_pct` to be strongly negatively correlated with income
- Just kidding, we do have data on `avginc`, but we'll only use it to confirm our suspicion:

	testscr	str	el_pct	avginc	meal_pct
testscr	1	-0.226	-0.644	0.712	-0.869
str	-0.226	1	0.188	-0.232	0.135
el_pct	-0.644	0.188	1	-0.307	0.653
avginc	0.712	-0.232	-0.307	1	-0.684
meal_pct	-0.869	0.135	0.653	-0.684	1

HOOD COLLEGE

Proxy Variables: Example II

- We've been assuming we don't have data on average district income, we would expect `meal_pct` to be strongly negatively correlated with income
- Just kidding, we do have data on `avginc`, but we'll only use it to confirm our suspicion:

	testscr	str	el_pct	avginc	meal_pct
testscr	1	-0.226	-0.644	0.712	-0.869
str	-0.226	1	0.188	-0.232	0.135
el_pct	-0.644	0.188	1	-0.307	0.653
avginc	0.712	-0.232	-0.307	1	-0.684
meal_pct	-0.869	0.135	0.653	-0.684	1

- We can see `meal_pct` is strongly (negatively) correlated with income, as expected

HOOD COLLEGE

Proxy Variables: Example III

```
proxyreg<-lm(testscr~str+el_pct+meal_pct, data=CAproxy)
summary(proxyreg)
```

```
##
## Call:
## lm(formula = testscr ~ str + el_pct + meal_pct, data = CAproxy)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.849  -5.151  -0.308   5.243  31.501
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 700.14996    4.68569 149.423  < 2e-16 ***
## str          -0.99831    0.23875  -4.181 3.54e-05 ***
## el_pct       -0.12157    0.03232  -3.762 0.000193 ***
## meal_pct     -0.54735    0.02160 -25.341  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.08 on 416 degrees of freedom
## Multiple R-squared:  0.7745, Adjusted R-squared:  0.7729
## F-statistic: 476.3 on 3 and 416 DF,  p-value: < 2.2e-16
```

HOOD COLLEGE

Proxy Variables: Example III

```
proxyreg<-lm(testscr~str+el_pct+meal_pct, data=CAproxy)
summary(proxyreg)
```

```
##
## Call:
## lm(formula = testscr ~ str + el_pct + meal_pct, data = CAproxy)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.849  -5.151  -0.308   5.243  31.501
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 700.14996    4.68569 149.423  < 2e-16 ***
## str          -0.99831    0.23875  -4.181 3.54e-05 ***
## el_pct       -0.12157    0.03232  -3.762 0.000193 ***
## meal_pct     -0.54735    0.02160 -25.341  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.08 on 416 degrees of freedom
## Multiple R-squared:  0.7745, Adjusted R-squared:  0.7729
## F-statistic: 476.3 on 3 and 416 DF,  p-value: < 2.2e-16
```

$$\widehat{\text{Test Score}} = \underset{(4.68)}{700.15} - \underset{(0.24)}{1.00}STR - \underset{(0.03)}{0.122}el\_pct - \underset{(0.02)}{0.547}meal\_pct$$

Interpretting Control Variables

$$\widehat{\text{Test Score}} = \underset{(4.68)}{700.15} - \underset{(0.24)}{1.00}STR - \underset{(0.03)}{0.122}el\_pct - \underset{(0.02)}{0.547}meal\_pct$$

- Is `meal_pct` causal?
- Getting rid of programs in districts where a large percentage of students need them would boost test scores A LOT! (So probably not causal...)

HOOD COLLEGE

$$\widehat{\text{Test Score}} = \underset{(4.68)}{700.15} - \underset{(0.24)}{1.00}STR - \underset{(0.03)}{0.122}el\_pct - \underset{(0.02)}{0.547}meal\_pct$$

- Is `meal_pct` causal?
- Getting rid of programs in districts where a large percentage of students need them would boost test scores A LOT! (So probably not causal...)
- `meal_pct` likely correlated with other things in ϵ (like outside learning opportunities!).

HOOD COLLEGE

$$\widehat{\text{Test Score}} = \underset{(4.68)}{700.15} - \underset{(0.24)}{1.00}STR - \underset{(0.03)}{0.122}el\_pct - \underset{(0.02)}{0.547}meal\_pct$$

- Is `meal_pct` causal?
- Getting rid of programs in districts where a large percentage of students need them would boost test scores A LOT! (So probably not causal...)
- `meal_pct` likely correlated with other things in ϵ (like outside learning opportunities!).
    - In fact, that's *exactly why* we included it as a variable!

HOOD COLLEGE

$$\widehat{\text{Test Score}} = \underset{(4.68)}{700.15} - \underset{(0.24)}{1.00}STR - \underset{(0.03)}{0.122}el\_pct - \underset{(0.02)}{0.547}meal\_pct$$

- Is `meal_pct` causal?
- Getting rid of programs in districts where a large percentage of students need them would boost test scores A LOT! (So probably not causal...)
- `meal_pct` likely correlated with other things in ϵ (like outside learning opportunities!).
  - In fact, that's *exactly why* we included it as a variable!
- We don't need the OLS estimate on `meal_pct` to be unbiased!

HOOD COLLEGE

$$\widehat{\text{Test Score}} = \underset{(4.68)}{700.15} - \underset{(0.24)}{1.00}STR - \underset{(0.03)}{0.122}el\_pct - \underset{(0.02)}{0.547}meal\_pct$$

- Is `meal_pct` causal?
- Getting rid of programs in districts where a large percentage of students need them would boost test scores A LOT! (So probably not causal...)
- `meal_pct` likely correlated with other things in ϵ (like outside learning opportunities!).
  - In fact, that's *exactly why* we included it as a variable!
- We don't need the OLS estimate on `meal_pct` to be unbiased!
  - We only care about getting the estimate on `str` to be unbiased!

HOOD COLLEGE

Control Variables

- A control variable is a regressor variable **note** of interest, but included to hold factors constant that, if omitted, would bias the causal effect of interest

HOOD COLLEGE

Control Variables

- A control variable is a regressor variable **note** of interest, but included to hold factors constant that, if omitted, would bias the causal effect of interest
- The control variable may still be correlated with omitted causal factors in ϵ

HOOD COLLEGE

Control Variables

- A control variable is a regressor variable **note** of interest, but included to hold factors constant that, if omitted, would bias the causal effect of interest
- The control variable may still be correlated with omitted causal factors in ϵ
    - Estimators ($\hat{\beta}$'s) on control variables can be biased and that is OK!

HOOD COLLEGE

Control Variables

- A control variable is a regressor variable **note** of interest, but included to hold factors constant that, if omitted, would bias the causal effect of interest
- The control variable may still be correlated with omitted causal factors in ϵ
  - Estimators ($\hat{\beta}$'s) on control variables can be biased and that is OK!
  - So long as we have unbiased estimators ($\hat{\beta}$'s) on the regressors we do care about!

HOOD COLLEGE

Control Variables

- A control variable is a regressor variable **note** of interest, but included to hold factors constant that, if omitted, would bias the causal effect of interest
- The control variable may still be correlated with omitted causal factors in ϵ
  - Estimators ($\hat{\beta}$'s) on control variables can be biased and that is OK!
  - So long as we have unbiased estimators ($\hat{\beta}$'s) on the regressors we do care about!
- Control variables allow us to proceed *as if* X were randomly assigned

HOOD COLLEGE

Do NOT just try to maximize R^2 or $\bar{R}^2$

- A high R^2 or $\bar{R}^2$ means that the regressors explain variation in Y

Do NOT just try to maximize R^2 or $\bar{R}^2$

- A high R^2 or $\bar{R}^2$ means that the regressors explain variation in Y
- A high R^2 or $\bar{R}^2$ does *NOT* mean you have eliminated omitted variable bias

HOOD COLLEGE

Do NOT just try to maximize R^2 or $\bar{R}^2$

- A high R^2 or $\bar{R}^2$ means that the regressors explain variation in Y
- A high R^2 or $\bar{R}^2$ does *NOT* mean you have eliminated omitted variable bias
- A high R^2 or $\bar{R}^2$ does *NOT* mean you have an unbiased estimate of a causal effect

HOOD COLLEGE

Do NOT just try to maximize R^2 or $\bar{R}^2$

- A high R^2 or $\bar{R}^2$ means that the regressors explain variation in Y
- A high R^2 or $\bar{R}^2$ does *NOT* mean you have eliminated omitted variable bias
- A high R^2 or $\bar{R}^2$ does *NOT* mean you have an unbiased estimate of a causal effect
- A high R^2 or $\bar{R}^2$ does *NOT* mean included variables are statistically significant

HOOD COLLEGE

Applications of Simple Multivariate OLS Models

Sacerdote (2004) on Peer Effects

Sacerdote, Bruce, (2004), "Peer Effects with Random Assignment: Results from Dartmouth Roommates" *Quarterly Journal of Economics* 116(2):681-704

Paper Motivation

- What determines student outcomes in college? (GPAs, fraternity enrollment, alcohol/drug use, etc)

HOOD COLLEGE

Paper Motivation

- What determines student outcomes in college? (GPAs, fraternity enrollment, alcohol/drug use, etc)
- Effects of peer groups

HOOD COLLEGE

- "Standard" way to estimate peer effects: regress student *i*'s outcomes/behavior on *other students*' outcomes/behavior

$$GPA_i = \beta_0 + \beta_1 OwnBehavior_i + \beta_2 RoommateBehavior_i + \epsilon_i$$

- "Standard" way to estimate peer effects: regress student *i*'s outcomes/behavior on *other students'* outcomes/behavior

$$GPA_i = \beta_0 + \beta_1 OwnBehavior_i + \beta_2 RoommateBehavior_i + \epsilon_i$$

- Problems with this approach:

- "Standard" way to estimate peer effects: regress student *i*'s outcomes/behavior on *other students*' outcomes/behavior

$$GPA_i = \beta_0 + \beta_1 OwnBehavior_i + \beta_2 RoommateBehavior_i + \epsilon_i$$

- Problems with this approach:
  1. Individuals **self-select** into peer groups

- "Standard" way to estimate peer effects: regress student i's outcomes/behavior on *other students'* outcomes/behavior

$$GPA_i = \beta_0 + \beta_1 OwnBehavior_i + \beta_2 RoommateBehavior_i + \epsilon_i$$

- Problems with this approach:
  1. Individuals **self-select** into peer groups
  2. If two roommates A and B influence each other, how do we disentangle causal effect of $B \longrightarrow A$ vs. $A \longrightarrow B$?

HOOD COLLEGE

- "Standard" way to estimate peer effects: regress student *i*'s outcomes/behavior on *other students'* outcomes/behavior

$$GPA_i = \beta_0 + \beta_1 OwnBehavior_i + \beta_2 RoommateBehavior_i + \epsilon_i$$

- Problems with this approach:
  1. Individuals **self-select** into peer groups
  2. If two roommates *A* and *B* influence each other, how do we disentangle causal effect of $B \longrightarrow A$ vs. $A \longrightarrow B$?
  3. Are peer effects actually driven by students' own backgrounds, or by their actual choices?

HOOD COLLEGE

$$corr(OwnBehavior, \epsilon) \neq 0$$

$$corr(RoomateBehavior, \epsilon) \neq 0$$

$$E[\epsilon | OwnBehavior, RoommateBehavior] \neq 0$$

Sacerdote's Identification Strategy

- Freshmen entering Dartmouth College are **randomly** assigned to dorms & roommates

Sacerdote's Identification Strategy

- Freshmen entering Dartmouth College are **randomly** assigned to dorms & roommates
- Removes self-selection of peer groups by shared characteristics

HOOD COLLEGE

Sacerdote's Identification Strategy

- Freshmen entering Dartmouth College are **randomly** assigned to dorms & roommates
- Removes self-selection of peer groups by shared characteristics
- Random assignment: roommate A's background characteristics are uncorrelated with roommate B's background characteristics

HOOD COLLEGE

The Relevant Institutions

- Freshmen entering Dartmouth College are randomly assigned to dorms & roommates

HOOD COLLEGE

The Relevant Institutions

- Freshmen entering Dartmouth College are randomly assigned to dorms & roommates
- Each incoming freshman fills out a questionnaire:

HOOD COLLEGE

The Relevant Institutions

- Freshmen entering Dartmouth College are randomly assigned to dorms & roommates
- Each incoming freshman fills out a questionnaire:

    1. I smoke. (Y/N)

HOOD COLLEGE

The Relevant Institutions

- Freshmen entering Dartmouth College are randomly assigned to dorms & roommates
- Each incoming freshman fills out a questionnaire:

    1. I smoke. (Y/N)
    2. I like to listen to music while studying. (Y/N)

HOOD COLLEGE

The Relevant Institutions

- Freshmen entering Dartmouth College are randomly assigned to dorms & roommates
- Each incoming freshman fills out a questionnaire:

1. I smoke. (Y/N)
2. I like to listen to music while studying. (Y/N)
3. I keep late hours. (Y/N)

HOOD COLLEGE

The Relevant Institutions

- Freshmen entering Dartmouth College are randomly assigned to dorms & roommates
- Each incoming freshman fills out a questionnaire:

1. I smoke. (Y/N)
2. I like to listen to music while studying. (Y/N)
3. I keep late hours. (Y/N)
4. I am more neat than messy. (Y/N)

HOOD COLLEGE

The Relevant Institutions

- Freshmen entering Dartmouth College are randomly assigned to dorms & roommates
- Each incoming freshman fills out a questionnaire:

1. I smoke. (Y/N)
2. I like to listen to music while studying. (Y/N)
3. I keep late hours. (Y/N)
4. I am more neat than messy. (Y/N)
5. I am (Male/Female).

HOOD COLLEGE

The Relevant Institutions

- Freshmen entering Dartmouth College are randomly assigned to dorms & roommates
- Each incoming freshman fills out a questionnaire:
    1. I smoke. (Y/N)
    2. I like to listen to music while studying. (Y/N)
    3. I keep late hours. (Y/N)
    4. I am more neat than messy. (Y/N)
    5. I am (Male/Female).
- There are $2^5 = 32$ combinatorial possibilities of answers to the questions

HOOD COLLEGE

The Relevant Institutions

- Freshmen entering Dartmouth College are randomly assigned to dorms & roommates
- Each incoming freshman fills out a questionnaire:

    1. I smoke. (Y/N)
    2. I like to listen to music while studying. (Y/N)
    3. I keep late hours. (Y/N)
    4. I am more neat than messy. (Y/N)
    5. I am (Male/Female).

- There are $2^5 = 32$ combinatorial possibilities of answers to the questions
- Students are assigned to roommates/dorms **at random**, conditional on their combination of answers to the 5 survey answers

HOOD COLLEGE

The Data

- Data from Dartmouth's database of students: history of dorm assignments & term-by-term academic performance

HOOD COLLEGE

The Data

- Data from Dartmouth's database of students: history of dorm assignments & term-by-term academic performance
- Data on pre-treatment characteristics (SAT scores, high school class rank, private/public HS, home state, academic index)

HOOD COLLEGE

The Data

- Data from Dartmouth's database of students: history of dorm assignments & term-by-term academic performance
- Data on pre-treatment characteristics (SAT scores, high school class rank, private/public HS, home state, academic index)
- Outcome variables: GPA, time to graduation, frat membership, major choice, participation in athletics

HOOD COLLEGE

The Data

- Data from Dartmouth's database of students: history of dorm assignments & term-by-term academic performance
- Data on pre-treatment characteristics (SAT scores, high school class rank, private/public HS, home state, academic index)
- Outcome variables: GPA, time to graduation, frat membership, major choice, participation in athletics
- Survey of Incoming Freshman: if student drank beer in last year and expectation of graduating with honors

HOOD COLLEGE

The Data

- Data from Dartmouth's database of students: history of dorm assignments & term-by-term academic performance
- Data on pre-treatment characteristics (SAT scores, high school class rank, private/public HS, home state, academic index)
- Outcome variables: GPA, time to graduation, frat membership, major choice, participation in athletics
- Survey of Incoming Freshman: if student drank beer in last year and expectation of graduating with honors
- Sample of 1589 students

HOOD COLLEGE

The Data

- Data from Dartmouth's database of students: history of dorm assignments & term-by-term academic performance
- Data on pre-treatment characteristics (SAT scores, high school class rank, private/public HS, home state, academic index)
- Outcome variables: GPA, time to graduation, frat membership, major choice, participation in athletics
- Survey of Incoming Freshman: if student drank beer in last year and expectation of graduating with honors
- Sample of 1589 students
- Create dummy variable for each block to control for covariates (we'll talk later about dummy variables and fixed effects like this)

HOOD COLLEGE

The Data: Summary Statistics

TABLE I

SUMMARY STATISTICS FOR SAMPLE OF DARTMOUTH ROOMMATES GRADUATING CLASSES OF 1997 AND 1998

Variable	Obs.	Mean	Std. dev.	Min	Max
freshman year GPA	1589	3.20	0.43	0.67	4.00
sophomore year GPA	1552	3.28	0.44	0.30	4.00
junior year GPA	1529	3.35	0.45	0.60	4.00
senior year GPA	1508	3.41	0.45	0.50	4.00
roommate freshman year GPA	1589	3.19	0.39	1.15	4.00
fraternity/sorority/coed house	1589	0.49	0.50	0.00	1.00
graduate late	1589	0.03	0.18	0.00	1.00
economics major	1589	0.10	0.31	0.00	1.00
social science major	1589	0.33	0.47	0.00	1.00
science major	1589	0.29	0.45	0.00	1.00
humanities major	1589	0.35	0.48	0.00	1.00
black	1589	0.05	0.22	0.00	1.00
SAT Math	1589	691.26	67.08	420.00	800.00
SAT Verbal	1589	632.86	70.07	360.00	800.00
academic score (incoming)	1589	204.20	12.88	151.00	231.00
high school class rank (incoming)	993	9.14	12.27	1.00	75.00
high school class rank missing	1589	0.38	0.48	0.00	1.00
private high school	1589	0.11	0.32	0.00	1.00
smokes (housing form)	1589	0.01	0.12	0.00	1.00
more neat than messy (housing form)	1589	0.69	0.46	0.00	1.00
stays up late (housing form)	1589	0.60	0.49	0.00	1.00
listens to music (housing form)	1589	0.47	0.50	0.00	1.00
same roommate sophomore year	1589	0.14	0.35	0.00	1.00
HS GPA	1328	3.56	0.51	2.00	4.00
Pre-Dart: drank beer in past year	1337	0.59	0.49	0.00	1.00

HOOD COLLEGE

The Data: Summary Statistics II

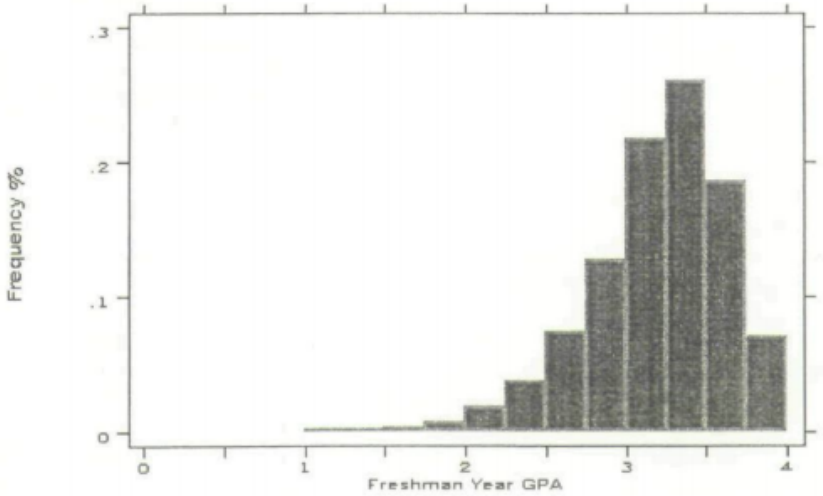


FIGURE I
Distribution of Freshman Year GPA

HOOD COLLEGE

The Data: Demonstrating Random Assignment

TABLE II

OWN PRETREATMENT CHARACTERISTICS REGRESSED ON ROOMMATE PRETREATMENT CHARACTERISTICS

EVIDENCE OF THE RANDOM ASSIGNMENT OF ROOMMATES

	(1) SAT Math (self)	(2) SAT Verbal (self)	(3) HS Academic class index	(4) HS Rank	(5) HS Academic index
roommates' math SAT scores	−0.025 (0.028)				−0.005 (0.008)
roommates' verbal SAT scores		−0.009 (0.029)			−0.005 (0.007)
roommates' HS academic scores			0.010 (0.028)		0.055 (0.056)
roommates' HS class ranks				−0.032 (0.028)	0.031 (0.042)
roommates' HS class rank missing					−0.512 (0.838)
Dummies for housing questions	yes	yes	yes	yes	yes
F-test: All roommate background coeff = 0					$F(5, 1543) = 0.50$ $P > F = .78$
R^2	.09	.03	.04	.03	.04
N	1589	1589	1589	993	1589

Standard errors are in parentheses. In cases with more than one roommate, roommate variables are averaged.

Columns (1)–(5) are OLS. All regressions include 41 dummies representing nonempty blocks based upon responses to the housing questions.

The lack of statistical significance on the coefficients is intended to demonstrate that the assignment process resembles a randomized experiment. In earlier nonrandomly assigned classes (such as the classes of 1995–1996), own and roommate background are highly correlated.

HOOD COLLEGE

The Data: Demonstrating Random Assignment II

TABLE II

OWN PRETREATMENT CHARACTERISTICS REGRESSED ON ROOMMATE PRETREATMENT CHARACTERISTICS
EVIDENCE OF THE RANDOM ASSIGNMENT OF ROOMMATES

	(1) SAT Math (self)	(2) SAT Verbal (self)	(3) HS Academic class index	(4) HS Rank	(5) HS Academic index
roommates' math SAT scores	−0.025 (0.028)				−0.005 (0.008)
roommates' verbal SAT scores		−0.009 (0.029)			−0.005 (0.007)
roommates' HS academic scores			0.010 (0.028)		0.055 (0.056)
roommates' HS class ranks				−0.032 (0.028)	0.031 (0.042)
roommates' HS class rank missing					−0.512 (0.838)
Dummies for housing questions	yes	yes	yes	yes	yes
F-test: All roommate background coeff = 0					$F(5, 1543) = 0.50$ $P > F = .78$
R^2	.09	.03	.04	.03	.04
N	1589	1589	1589	993	1589

Standard errors are in parentheses. In cases with more than one roommate, roommate variables are averaged.

Columns (1)–(5) are OLS. All regressions include 41 dummies representing nonempty blocks based upon responses to the housing questions.

The lack of statistical significance on the coefficients is intended to demonstrate that the assignment process resembles a randomized experiment. In earlier nonrandomly assigned classes (such as the classes of 1995–1996), own and roommate background are highly correlated.

Sacerdote's Empirical Model (Basically)

$$GPA_i = \beta_0 + \beta_1 ACA_i + \beta_2 ACA_j + \epsilon_i$$

- Student i and roommate j

Sacerdote's Empirical Model (Basically)

$$GPA_i = \beta_0 + \beta_1 ACA_i + \beta_2 ACA_j + \epsilon_i$$

- Student i and roommate j
- ACA: Index of academic performance (broken down into different metrics)

HOOD COLLEGE

Sacerdote's Empirical Model (Basically)

$$GPA_i = \beta_0 + \beta_1 ACA_i + \beta_2 ACA_j + \epsilon_i$$

- Student i and roommate j
- *ACA*: Index of academic performance (broken down into different metrics)
- Other outcomes of interest (besides *GPA*: graduation, major, fraternity, athlete

HOOD COLLEGE

Regression Results

TABLE III
PEER EFFECTS IN ACADEMIC OUTCOMES

	(1) Fresh year GPA	(2) Fresh year GPA w/ dorm f.e.	(3) Senior year GPA	(4) Fresh year GPA	(5) Fresh year GPA	(6) Fresh year GPA	(7) Graduate late	(8) Econ major
Roommates' GPA	0.120**	0.068**	0.008					
	(0.039)	(0.029)	(0.026)					
HS academic score	0.014**	0.015**	0.013**				−0.0001	0.003**
(self)	(0.0008)	(0.0007)	(0.0009)				(0.0003)	(0.0006)
HS academic score	−0.001	−0.0003	0.0009				0.0003	−0.0001
(roommates')	(0.001)	(0.0009)	(0.001)				(0.0003)	(0.0006)
roommates' academic score bottom 25 percent				0.016	0.014	0.017		
				(0.028)	(0.025)	(0.025)		
roommates' academic score top 25 percent				0.060**	0.047*	0.043*		
				(0.028)	(0.026)	(0.026)		
roommates' intention to graduate w/honors (1–4)						0.082**		
						(0.037)		
own academic score bottom 25 percent					−0.284**	−0.282**		
					(0.025)	(0.025)		

- For every 1 point increase (decrease) in the roommate's GPA, a student's GPA increased (decreased) about .12 points

HOOD COLLEGE

Regression Results

TABLE III
PEER EFFECTS IN ACADEMIC OUTCOMES

	(1) Fresh year GPA	(2) Fresh year GPA w/ dorm f.e.	(3) Senior year GPA	(4) Fresh year GPA	(5) Fresh year GPA	(6) Fresh year GPA	(7) Graduate late	(8) Econ major
Roommates' GPA	0.120**	0.068**	0.008					
	(0.039)	(0.029)	(0.026)					
HS academic score	0.014**	0.015**	0.013**				−0.0001	0.003**
(self)	(0.0008)	(0.0007)	(0.0009)				(0.0003)	(0.0006)
HS academic score	−0.001	−0.0003	0.0009				0.0003	−0.0001
(roommates')	(0.001)	(0.0009)	(0.001)				(0.0003)	(0.0006)
roommates' academic score bottom 25 percent				0.016	0.014	0.017		
				(0.028)	(0.025)	(0.025)		
roommates' academic score top 25 percent				0.060**	0.047*	0.043*		
				(0.028)	(0.026)	(0.026)		
roommates' intention to graduate w/honors (1–4)						0.082**		
						(0.037)		
own academic score bottom 25 percent					−0.284**	−0.282**		
					(0.025)	(0.025)		

- For every 1 point increase (decrease) in the roommate's GPA, a student's GPA increased (decreased) about .12 points
- If you would have been a 3.0 student with a 3.0 roommate, but you were assigned to a 2.0 roommate, your GPA would be 2.88

HOOD COLLEGE

Regression Results: Academic Performance

own academic score top 25 percent					0.174** (0.025)	0.175** (0.025)		
Roommate graduate late							0.008 (0.029)	
Roommate econ major								−0.018 (0.026)
Dummies for housing questions	yes	yes	yes	yes	yes	yes	yes	yes
F test of roommate background coefficient = 0				$F = 2.31$ $P = 0.10$	$F = 1.63$ $P = 0.20$	$F = 2.74$ P = 0.04		
R^2	.24	.38	.18	.05	.19	.19	.06	.07
N	1589	1589	1441	1589	1589	1589	1589	1589

Standard errors are in parentheses and are corrected for clustering at the room level. In cases with more than one roommate, roommate variables are averaged. ** = p-value < .05. * = p-value < .10.

Regression (1) is OLS of own GPA on roommate GPA and controls. If own and roommate academic indices are excluded, the coefficient on roommate GPA falls to .111, and the standard error falls to 0.037.

Regression (2) adds dorm fixed effects. The coefficient on roommate GPA falls, but remains significant. Regression (3) is OLS of own senior year GPA on freshman year roommates' senior year GPA. Senior year GPA includes all grades in final year and excludes grades from earlier years.

Regressions (4)–(6) are OLS of own GPA on own and roommate background. These regressions use dummies for own and roommate academic index are in the bottom 25 percent, middle 50 percent (excluded category), or top 25 percent of their respective distributions. Regression (4) shows that "roommate top 25 percent" is significant in predicting own GPA. The level of significance on "roommate top 25 percent" falls to .10 when two dummies for own academic index are added. (This is regression (5).) Regression (6) shows that roommate intention to graduate with honors also predicts own GPA. This variable is a self-assessed probability of graduating with honors and is coded as a 1, 2, 3, or 4 for the responses of no chance, very little chance, some chance, or a very good chance. Regression (6) also includes a dummy for "roommate intend to graduate with honors" missing. See text for more discussion of this variable.

Regressions (7) and (8) are probits of own "graduate late" and own "major choice = econ" on roommate graduate late and roommate major choice = econ. $\partial y/\partial x$ is shown.

HOOD COLLEGE

Regression Results: Social Outcomes

TABLE V
PEER EFFECTS IN SOCIAL OUTCOMES

	(1) Member frat/ soror	(2) Member frat/ soror	(3) Member frat/ soror	(4) Varsity athlete
roommate member of fraternity/sorority/coed	0.078**	0.056		
	(0.038)	(0.037)		
dorm average of fraternity/sorority/coed		0.321**		
		(0.135)		
roommate varsity athlete				0.045
				(0.033)
HS academic score (self)	0.0098	0.0011	0.0010	−0.004**
	(0.0010)	(0.0011)	(0.0011)	(0.001)
HS academic score (roommates')	−0.0017	−0.0016	−0.0016	−0.0002
	(0.0011)	(0.0011)	(0.0011)	(0.0007)
Own use of beer in high school (0–1)			0.135**	
			(0.038)	
Roommates' use of beer in high school (0–1)			−0.025	
			(0.026)	
			(0.026)	
Dormmates' use of beer in high school (0–1)			0.287**	
			(0.146)	
Dummies for housing questions	yes	yes	yes	yes
R^2	.02	.02	.03	.05
N	1589	1589	1589	1589

Standard errors are in parentheses and are corrected for clustering at the room level. In cases with more than one roommate, roommate variables are averaged. ** = p-value < .05.

Columns (1)–(4) are Probits. $\partial y/\partial x$ is shown.

In regression (2), dorm average of frat membership excludes own observation, and standard errors are corrected for clustering at dorm level.

In regression (3), use of beer in past year is coded 0–1 as follows: 0 = not at all, occasionally or frequently = 1. Dorm use of beer excludes own room and standard errors are corrected for clustering at dorm level.

HOOD COLLEGE

Sacerdote's Conclusions

- Peer effects are very strong!

Sacerdote's Conclusions

- Peer effects are very strong!
- Important influences in Freshman year performance (GPA) and activities (joining a social organization)

Sacerdote's Conclusions

- Peer effects are very strong!
- Important influences in Freshman year performance (GPA) and activities (joining a social organization)
- Not important for choosing a major

Duggan and Levitt (2002) on Corruption in Sumo

Duggan, Mark and Steven D. Levitt (2002), "Winning Isn't Everything: Corruption in Sumo Wrestling," *American Economic Review* 92(5): 1594-1605

HOOD COLLEGE

Paper Motivation

- How can we *understand* and *detect* corruption?

Paper Motivation

- How can we *understand* and *detect* corruption?
- A very important and consequential economic problem, nearly impossible to measure

- Transparency International: Corruption Perceptions Index (CPI)

- Transparency International: Corruption Perceptions Index (CPI)
  - Index of how people *perceive* corruption in their country -168 countries

HOOD COLLEGE

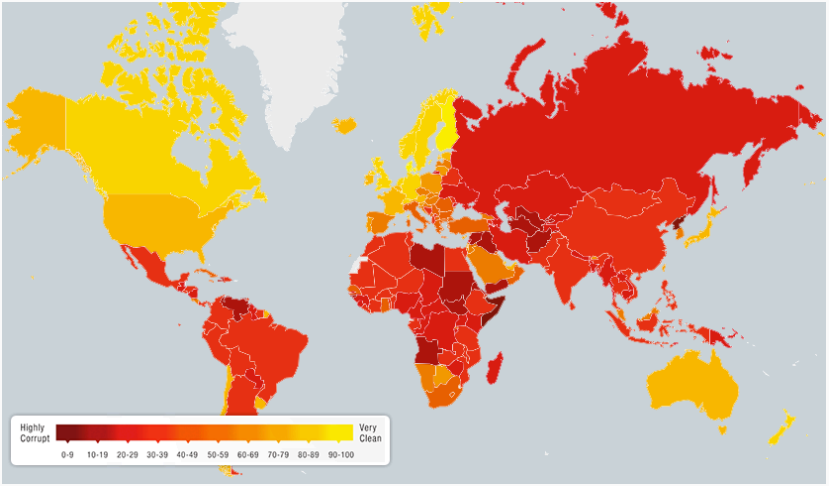
- Transparency International: Corruption Perceptions Index (CPI)
  - Index of how people *perceive* corruption in their country -168 countries
  - Draws on a composite of 4 expert opinion surveys

HOOD COLLEGE

- Transparency International: Corruption Perceptions Index (CPI)
  - Index of how people *perceive* corruption in their country -168 countries
  - Draws on a composite of 4 expert opinion surveys
  - Scale from 0-100, 0: high levels of perceived corruption; 100: low levels of perceived corruption

HOOD COLLEGE

Some Background: CPI II



Transparency International CPI 2015

Some Background: CPI III

Rank	CPI2015	Country
1	91	Denmark
2	90	Finland
3	89	Sweden
4	88	New Zealand
5	87	Netherlands
5	87	Norway
7	86	Switzerland
8	85	Singapore
9	83	Canada
10	81	Germany
10	81	Luxembourg
10	81	United Kingdom

Transparency International CPI 2015

HOOD COLLEGE

158	17	Haiti
158	17	Guinea-Bissau
158	17	Venezuela
161	16	Iraq
161	16	Libya
163	15	Angola
163	15	South Sudan
165	12	Sudan
166	11	Afghanistan
167	8	Korea (North)
167	8	Somalia

Transparency International CPI 2015

HOOD COLLEGE

- 2000 year history, national sport of Japan, extremely ritualistic

Duggan and Levitt (2002): Relevant Institutions

- 2000 year history, national sport of Japan, extremely ritualistic
- Japan is a country with low corruption (CPI: 75, Rank 18th best)

Duggan and Levitt (2002): Relevant Institutions

- 2000 year history, national sport of Japan, extremely ritualistic
- Japan is a country with low corruption (CPI: 75, Rank 18th best)
- Good data available

Duggan and Levitt (2002): Relevant Institutions

- 2000 year history, national sport of Japan, extremely ritualistic
- Japan is a country with low corruption (CPI: 75, Rank 18th best)
- Good data available
- Situation is ripe for cheating! So when/why does it happen?

- Tournaments (*basho*), 66 wrestlers (*rikishi*), 15 bouts each

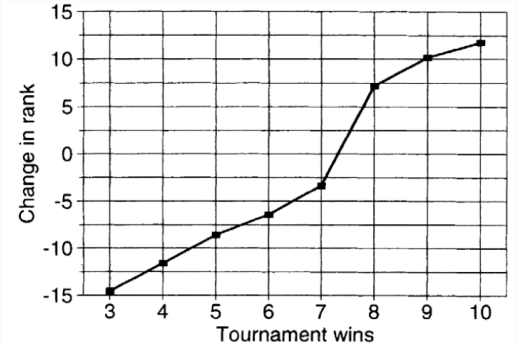


FIGURE 1. PAYOFF TO TOURNAMENT WINS

Duggan and Levitt (2002): Relevant Institutions II

- Tournaments (*basho*), 66 wrestlers (*rikishi*), 15 bouts each
- Wrestlers with 8+ wins (*kachi-koshi*) *move up* in rankings (*banzuke*)

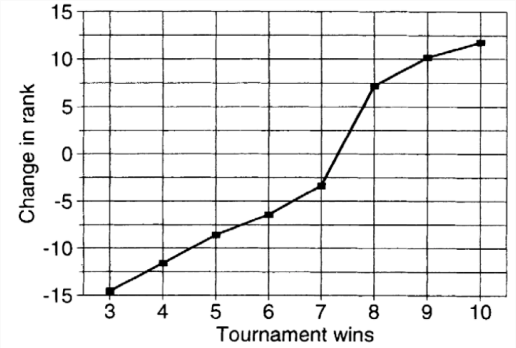


FIGURE 1. PAYOFF TO TOURNAMENT WINS

Duggan and Levitt (2002): Relevant Institutions II

- Tournaments (*basho*), 66 wrestlers (*rikishi*), 15 bouts each
- Wrestlers with 8+ wins (*kachi-koshi*) *move up* in rankings (*banzuke*)
- Those with a losing record ($<$8 wins) (*maki-koshi*) *fall* in rankings

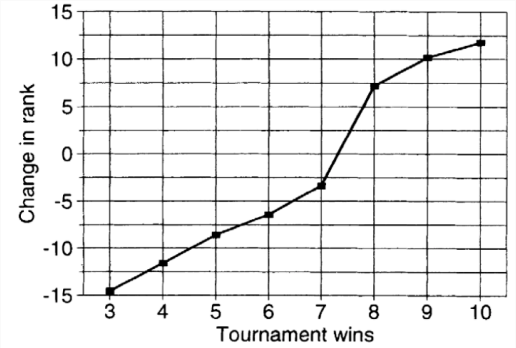


FIGURE 1. PAYOFF TO TOURNAMENT WINS

Duggan and Levitt (2002): Relevant Institutions III

- A marginal win generates a 2.5 rank increase

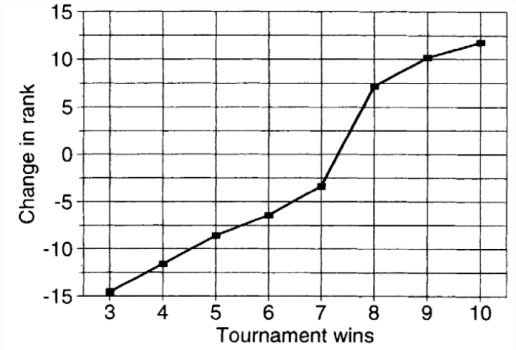


FIGURE 1. PAYOFF TO TOURNAMENT WINS

Duggan and Levitt (2002): Relevant Institutions III

- A marginal win generates a 2.5 rank increase
- But movement from 7 to 8 wins produces almost an 11 rank increase!

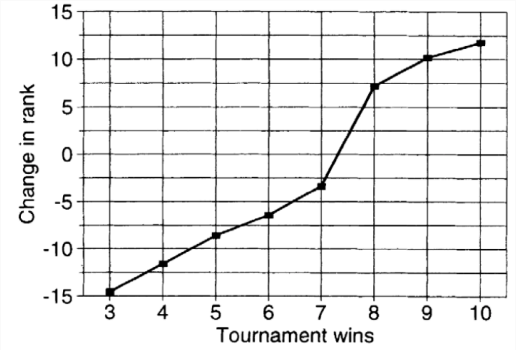


FIGURE 1. PAYOFF TO TOURNAMENT WINS

Duggan and Levitt (2002): Relevant Institutions III

- A marginal win generates a 2.5 rank increase
- But movement from 7 to 8 wins produces almost an 11 rank increase!
- Rank signals prestige, moving up a single rank is worth about $3,000/year

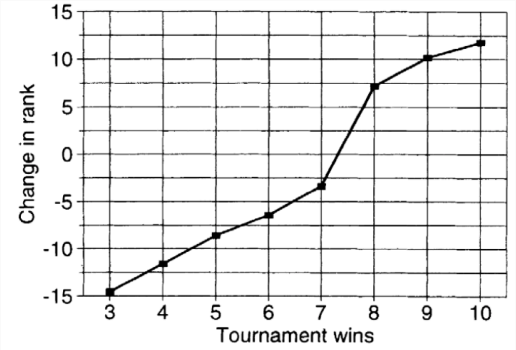


FIGURE 1. PAYOFF TO TOURNAMENT WINS

Duggan and Levitt (2002): Relevant Institutions III

- A marginal win generates a 2.5 rank increase
- But movement from 7 to 8 wins produces almost an 11 rank increase!
- Rank signals prestige, moving up a single rank is worth about $3,000/year
- Top 5th-10th ranked wrestlers make $250,000/year

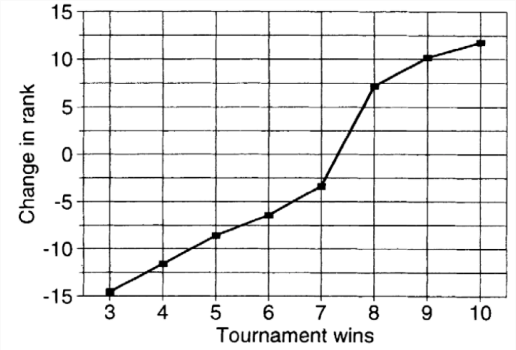


FIGURE 1. PAYOFF TO TOURNAMENT WINS

Duggan and Levitt (2002): Relevant Institutions IV

- Consider 2 wrestlers: A (8-6) vs B (7-7) going into final (15th) match

Duggan and Levitt (2002): Relevant Institutions IV

- Consider 2 wrestlers: A (8-6) vs B (7-7) going into final (15th) match
- Return to winning for B (7-7) is much higher than for A (8-6)

Duggan and Levitt (2002): Relevant Institutions IV

- Consider 2 wrestlers: A (8-6) vs B (7-7) going into final (15th) match
- Return to winning for B (7-7) is much higher than for A (8-6)
- A (8-6) throws the match to B (7-7), who must return the favor in later tournaments if A finds himself in the same 7-7 position

HOOD COLLEGE

Duggan and Levitt (2002): Data

- All official top-rank sumo matches from January 1989-January 2000

HOOD COLLEGE

Duggan and Levitt (2002): Data

- All official top-rank sumo matches from January 1989-January 2000
- Six tournaments per year, nearly 70 wrestlers per tournament

Duggan and Levitt (2002): Data

- All official top-rank sumo matches from January 1989-January 2000
- Six tournaments per year, nearly 70 wrestlers per tournament
- Tournaments last 15 days with one match per wrestler

HOOD COLLEGE

Duggan and Levitt (2002): Data

- All official top-rank sumo matches from January 1989-January 2000
- Six tournaments per year, nearly 70 wrestlers per tournament
- Tournaments last 15 days with one match per wrestler
- 64,000 wrestler-matches

HOOD COLLEGE

The Theoretical vs. Actual Probability of Winning

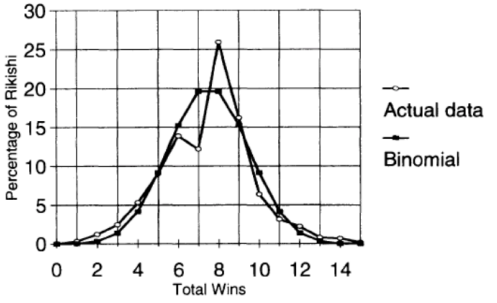


FIGURE 2. WINS IN A SUMO TOURNAMENT (ACTUAL VS. BINOMIAL)

- Theoretical (binomial) probability of winning 8 times: 19.6%

The Theoretical vs. Actual Probability of Winning

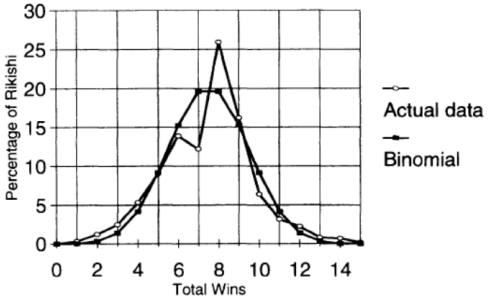


FIGURE 2. WINS IN A SUMO TOURNAMENT (ACTUAL VS. BINOMIAL)

- Theoretical (binomial) probability of winning 8 times: 19.6%
- Actual probability (from data): 26%

The Theoretical vs. Actual Probability of Winning

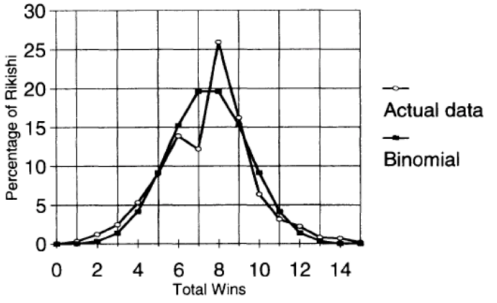


FIGURE 2. WINS IN A SUMO TOURNAMENT (ACTUAL VS. BINOMIAL)

- Theoretical (binomial) probability of winning 8 times: 19.6%
- Actual probability (from data): 26%
- Much higher probability for 8 wins than it should be! (& lower for 7)

The Theoretical vs. Actual Probability of Winning

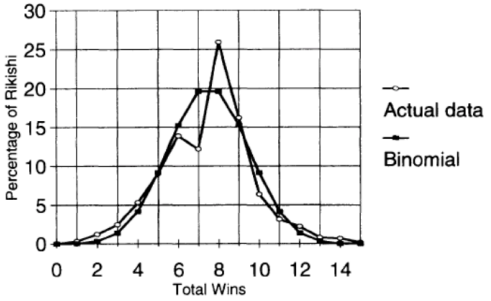


FIGURE 2. WINS IN A SUMO TOURNAMENT (ACTUAL VS. BINOMIAL)

- Theoretical (binomial) probability of winning 8 times: 19.6%
- Actual probability (from data): 26%
- Much higher probability for 8 wins than it should be! (& lower for 7)
  - Could this be from rampant cheating!?

$$(1) \quad Win_{ijtd} = \beta \mathbf{Bubble}_{ijtd} + \gamma Rankdiff_{ijt} + \lambda_{ij} + \delta_{it} + \epsilon_{ijtd}$$

- *Win* $=$ 1 if wrestler *i* beats wrestler *j* in tournament *t* on day *d*

$$(1) \quad Win_{ijtd} = \beta \mathbf{Bubble}_{ijtd} + \gamma Rankdiff_{ijt} + \lambda_{ij} + \delta_{it} + \epsilon_{ijtd}$$

- *Win* $=$ 1 if wrestler *i* beats wrestler *j* in tournament *t* on day *d*
- Bubble =1 if wrestler (*i*) is on margin (7-7), -1 if opponent (*j*) is on margin, =0 if neither are on margin

$$(1) \quad Win_{ijtd} = \beta \mathbf{Bubble}_{ijtd} + \gamma Rankdiff_{ijt} + \lambda_{ij} + \delta_{it} + \epsilon_{ijtd}$$

- *Win* $=$ 1 if wrestler *i* beats wrestler *j* in tournament *t* on day *d*
- Bubble =1 if wrestler (*i*) is on margin (7-7), -1 if opponent (*j*) is on margin, =0 if neither are on margin
- *Rankdiff*: difference in rank between wrestlers

HOOD COLLEGE

$$(1) \quad Win_{ijtd} = \beta \mathbf{Bubble}_{ijtd} + \gamma Rankdiff_{ijt} + \lambda_{ij} + \delta_{it} + \epsilon_{ijtd}$$

- *Win* $=$ 1 if wrestler *i* beats wrestler *j* in tournament *t* on day *d*
- Bubble =1 if wrestler (*i*) is on margin (7-7), -1 if opponent (*j*) is on margin, =0 if neither are on margin
- *Rankdiff*: difference in rank between wrestlers
- Wrestler λ and tournament δ fixed effects

HOOD COLLEGE

Initial Results

Table 1—Excess Win Percentages for Wrestlers on the Margin for Achieving an Eighth Win, by Day of the Match

On the Margin on:	(1)	(2)	(3)	(4)	(5)	(6)
Day 15	0.244	0.249	0.249	0.255	0.260	0.264
	(0.019)	(0.019)	(0.018)	(0.019)	(0.022)	(0.022)
Day 14	0.150	0.155	0.152	0.157	0.168	0.171
	(0.016)	(0.016)	(0.016)	(0.016)	(0.019)	(0.019)
Day 13	0.096	0.107	0.110	0.118	0.116	0.125
	(0.016)	(0.016)	(0.016)	(0.016)	(0.019)	(0.019)
Day 12	0.038	0.061	0.064	0.082	0.073	0.076
	(0.017)	(0.018)	(0.017)	(0.018)	(0.020)	(0.021)
Day 11	0.000	0.018	0.015	0.025	0.010	0.012
	(0.018)	(0.018)	(0.018)	(0.018)	(0.021)	(0.021)
Rank difference	—	0.0053	—	0.0020	—	−0.0020
		(0.0003)		(0.0003)		(0.0004)
Constant	0.500	0.500	—	—	—	—
	(0.000)	(0.000)				
R^2	0.008	0.018	0.030	0.031	0.0634	0.0653
Number of observations	64,272	62,708	64,272	62,708	64,272	62,708
Wrestler and opponent fixed effects	No	No	Yes	Yes	Yes	Yes
Wrestler-opponent interactions	No	No	No	No	Yes	Yes

HOOD COLLEGE

- Frequency of rigging increases as tournament nears end (day 15)

Initial Results

Table 1—Excess Win Percentages for Wrestlers on the Margin for Achieving an Eighth Win, by Day of the Match

On the Margin on:	(1)	(2)	(3)	(4)	(5)	(6)
Day 15	0.244	0.249	0.249	0.255	0.260	0.264
	(0.019)	(0.019)	(0.018)	(0.019)	(0.022)	(0.022)
Day 14	0.150	0.155	0.152	0.157	0.168	0.171
	(0.016)	(0.016)	(0.016)	(0.016)	(0.019)	(0.019)
Day 13	0.096	0.107	0.110	0.118	0.116	0.125
	(0.016)	(0.016)	(0.016)	(0.016)	(0.019)	(0.019)
Day 12	0.038	0.061	0.064	0.082	0.073	0.076
	(0.017)	(0.018)	(0.017)	(0.018)	(0.020)	(0.021)
Day 11	0.000	0.018	0.015	0.025	0.010	0.012
	(0.018)	(0.018)	(0.018)	(0.018)	(0.021)	(0.021)
Rank difference	—	0.0053	—	0.0020	—	−0.0020
		(0.0003)		(0.0003)		(0.0004)
Constant	0.500	0.500	—	—	—	—
	(0.000)	(0.000)				
R^2	0.008	0.018	0.030	0.031	0.0634	0.0653
Number of observations	64,272	62,708	64,272	62,708	64,272	62,708
Wrestler and opponent fixed effects	No	No	Yes	Yes	Yes	Yes
Wrestler-opponent interactions	No	No	No	No	Yes	Yes

HOOD COLLEGE

- Frequency of rigging increases as tournament nears end (day 15)

- Two alternative hypotheses to explain results:

- Two alternative hypotheses to explain results:
    1. Match rigging (corruption)

- Two alternative hypotheses to explain results:
  1. Match rigging (corruption)
  2. Effort: wrestlers on margin (7-7) just fight harder! Wrestlers with 8 wins are more complacent (already made it)

- Two alternative hypotheses to explain results:
  1. Match rigging (corruption)
  2. Effort: wrestlers on margin (7-7) just fight harder! Wrestlers with 8 wins are more complacent (already made it)
- To test, look for evidence of **reciprocity** agreements over time

- Two alternative hypotheses to explain results:
  1. Match rigging (corruption)
  2. Effort: wrestlers on margin (7-7) just fight harder! Wrestlers with 8 wins are more complacent (already made it)
- To test, look for evidence of **reciprocity** agreements over time
  - If these tacit agreements to rig matches exist, wrestlers from stable A should have very high win rates when on the margin against wrestlers from stable B, and vice versa

HOOD COLLEGE

Interaction Effects

Table 2—Determinants of Excess Win Likelihoods for Wrestlers on the Bubble

Variable	(1)	(2)	(3)
Wrestler on bubble	0.126	0.117	0.155
	(0.026)	(0.026)	(0.029)
Wrestler on bubble interacted with:			
High media scrutiny	−0.188	−0.177	−0.146
	(0.071)	(0.071)	(0.080)
Opponent in running for a prize this tournament	−0.149	−0.129	−0.156
	(0.047)	(0.046)	(0.052)
Number of meetings between two opponents in the last year	−0.0048	−0.0031	−0.0024
	(0.0082)	(0.0081)	(0.0096)
Wrestler on bubble in his last year of competing	−0.0361	−0.0195	−0.0346
	(0.0398)	(0.0395)	(0.0493)
Years in sumo for wrestler on bubble	0.0077	0.0077	0.0091
	(0.0036)	(0.0036)	(0.0043)
Winning percentage in other bubble matches between these two stables	0.272	0.293	—
	(0.059)	(0.058)	
R^2	0.016	0.074	0.246
Wrestler and opponent fixed effects?	No	Yes	Yes
Wrestler-opponent interactions?	No	No	Yes

HOOD COLLEGE

- Last row (before R^2): wrestler's success strongly increases with overall success rates of playing wrestlers on the bubble from

Interaction Effects

Table 2—Determinants of Excess Win Likelihoods for Wrestlers on the Bubble

Variable	(1)	(2)	(3)
Wrestler on bubble	0.126	0.117	0.155
	(0.026)	(0.026)	(0.029)
Wrestler on bubble interacted with:			
High media scrutiny	−0.188	−0.177	−0.146
	(0.071)	(0.071)	(0.080)
Opponent in running for a prize this tournament	−0.149	−0.129	−0.156
	(0.047)	(0.046)	(0.052)
Number of meetings between two opponents in the last year	−0.0048	−0.0031	−0.0024
	(0.0082)	(0.0081)	(0.0096)
Wrestler on bubble in his last year of competing	−0.0361	−0.0195	−0.0346
	(0.0398)	(0.0395)	(0.0493)
Years in sumo for wrestler on bubble	0.0077	0.0077	0.0091
	(0.0036)	(0.0036)	(0.0043)
Winning percentage in other bubble matches between these two stables	0.272	0.293	—
	(0.059)	(0.058)	
R^2	0.016	0.074	0.246
Wrestler and opponent fixed effects?	No	Yes	Yes
Wrestler-opponent interactions?	No	No	Yes

HOOD COLLEGE

- Last row (before R^2): wrestler's success strongly increases with overall success rates of playing wrestlers on the bubble from

TABLE 3—WIN PERCENTAGES IN PRECEDING AND SUBSEQUENT MATCHES
(For Two Wrestlers Who Meet When One is on the Margin in the Final Three Days of a Tournament)

	All Matches on the Margin		Only Matches in Which the Wrestler on the Margin Wins		Only Matches in Which the Wrestler on the Margin Loses	
Variable	(1)	(2)	(3)	(4)	(5)	(6)
One or two matches prior to the bubble match	−0.002 (0.009)	0.005 (0.012)	0.020 (0.011)	0.019 (0.017)	−0.041 (0.016)	−0.035 (0.022)
Bubble match	0.151 (0.010)	0.164 (0.014)	—	—	—	—
First meeting after bubble match	−0.073 (0.011)	−0.062 (0.015)	−0.082 (0.015)	−0.079 (0.020)	−0.056 (0.020)	−0.040 (0.027)
Second meeting after bubble match	−0.002 (0.013)	0.005 (0.016)	0.031 (0.017)	0.028 (0.022)	−0.061 (0.023)	−0.039 (0.030)
Three or more meetings after bubble match	−0.010 (0.006)	0.012 (0.011)	0.013 (0.007)	0.022 (0.014)	−0.045 (0.008)	−0.013 (0.017)
Constant	0.500 (0.000)	—	0.500 (0.000)	—	0.500 (0.000)	—
Wrestler-opponent interactions?	No	Yes	No	Yes	No	Yes
R^2	0.008	0.271	0.002	0.279	0.002	0.279

HOOD COLLEGE

-Wrestlers who won a bubble match previously tend to do worse when playing same opponent - throw the match to them in reciprocity!

Fisman and Miguel (2007) on U.N. Parking Tickets

Fisman, Raymond and Edward Miguel, (2007), "Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets," *Journal of Political Economy* 115(6): 1020-1048

- What determines the level of corruption?

- What determines the level of corruption?
    1. Legal environment of country

- What determines the level of corruption?
    1. Legal environment of country
    2. Culture or social norms

- What determines the level of corruption?
    1. Legal environment of country
    2. Culture or social norms
- How to identify the true source(s)?

- U.N. Diplomats are given immunity from prosecution or lawsuits in the U.S.

- U.N. Diplomats are given immunity from prosecution or lawsuits in the U.S.
- Reciprocal agreements with other countries, designed to protect diplomats in unfriendly environments

- U.N. Diplomats are given immunity from prosecution or lawsuits in the U.S.
- Reciprocal agreements with other countries, designed to protect diplomats in unfriendly environments
- Diplomatic license plates in NYC are identified, get ticket, but no way to enforce

- U.N. Diplomats are given immunity from prosecution or lawsuits in the U.S.
- Reciprocal agreements with other countries, designed to protect diplomats in unfriendly environments
- Diplomatic license plates in NYC are identified, get ticket, but no way to enforce
- "The best free parking pass in town"

Fisman and Miguel (2007): The Data

- Between 11/1987 and 12/2002, 150,000 unpaid parking tickets, fines totaling $18,000,000

HOOD COLLEGE

Fisman and Miguel (2007): The Data

- Between 11/1987 and 12/2002, 150,000 unpaid parking tickets, fines totaling $18,000,000
- 30 Days to pay a fine, afterwards a 110% penalty. After 70 days, recorded as unpaid violation

HOOD COLLEGE

Fisman and Miguel (2007): The Data

- Between 11/1987 and 12/2002, 150,000 unpaid parking tickets, fines totaling $18,000,000
- 30 Days to pay a fine, afterwards a 110% penalty. After 70 days, recorded as unpaid violation
- Individual violation-level data: license plate, name, country of origin, date & time of violation, fine, amount paid (if any)

HOOD COLLEGE

Fisman and Miguel (2007): The Data

- Between 11/1987 and 12/2002, 150,000 unpaid parking tickets, fines totaling $18,000,000
- 30 Days to pay a fine, afterwards a 110% penalty. After 70 days, recorded as unpaid violation
- Individual violation-level data: license plate, name, country of origin, date & time of violation, fine, amount paid (if any)
- 43% were violations of “no standing/loading zone”

HOOD COLLEGE

Fisman and Miguel (2007): The Data

- Between 11/1987 and 12/2002, 150,000 unpaid parking tickets, fines totaling $18,000,000
- 30 Days to pay a fine, afterwards a 110% penalty. After 70 days, recorded as unpaid violation
- Individual violation-level data: license plate, name, country of origin, date & time of violation, fine, amount paid (if any)
- 43% were violations of “no standing/loading zone”
- 20% of cases, the car was registered to the diplomatic mission (not personal)

HOOD COLLEGE

Fisman and Miguel (2007): The Data

- Between 11/1987 and 12/2002, 150,000 unpaid parking tickets, fines totaling $18,000,000
- 30 Days to pay a fine, afterwards a 110% penalty. After 70 days, recorded as unpaid violation
- Individual violation-level data: license plate, name, country of origin, date & time of violation, fine, amount paid (if any)
- 43% were violations of “no standing/loading zone”
- 20% of cases, the car was registered to the diplomatic mission (not personal)
- Scale fines by the size of the country’s mission

HOOD COLLEGE

- Becker's (1968) rational crime model says with no punishment $\Longrightarrow$ rational for all diplomats to never pay parking fines

- Becker's (1968) rational crime model says with no punishment $\implies$ rational for all diplomats to never pay parking fines
- But large variation in data! Unpaid fines are strongly correlated with country's score on corruption index!

Fisman and Miguel (2007): Identification Strategy

- Becker's (1968) rational crime model says with no punishment $\implies$ rational for all diplomats to never pay parking fines
- But large variation in data! Unpaid fines are strongly correlated with country's score on corruption index!
- Home country corruption norms are an important predictor of diplomats breaking the law

HOOD COLLEGE

- Becker's (1968) rational crime model says with no punishment $\implies$ rational for all diplomats to never pay parking fines
- But large variation in data! Unpaid fines are strongly correlated with country's score on corruption index!
- Home country corruption norms are an important predictor of diplomats breaking the law
    - Low corruption countries' diplomats tend to pay the fine even where they are not legally compelled to

- Becker's (1968) rational crime model says with no punishment $\Longrightarrow$ rational for all diplomats to never pay parking fines
- But large variation in data! Unpaid fines are strongly correlated with country's score on corruption index!
- Home country corruption norms are an important predictor of diplomats breaking the law
    - Low corruption countries' diplomats tend to pay the fine even where they are not legally compelled to
    - High corruption countries' diplomats rack up massive fines

- **Natural experiment**: post-9/11, NYC began cracking down on enforcement

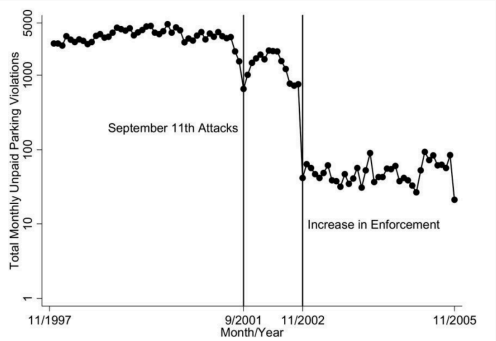


Fig. 1.—Total monthly New York City parking violations by diplomats, 1997–2005 (vertical axis on log scale).

- **Natural experiment**: post-9/11, NYC began cracking down on enforcement
- Diplomats with 3+ unpaid parking tickets had diplomat plates revoked

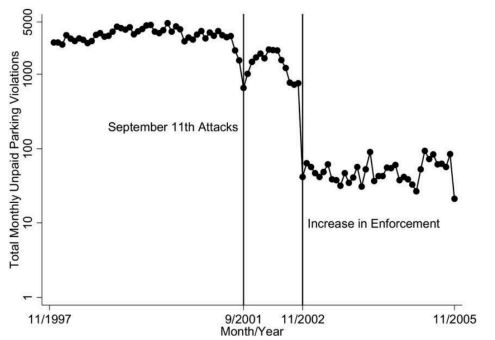


Fig. 1.—Total monthly New York City parking violations by diplomats, 1997–2005 (vertical axis on log scale).

- **Natural experiment**: post-9/11, NYC began cracking down on enforcement
- Diplomats with 3+ unpaid parking tickets had diplomat plates revoked
- Led to immediate 98% reduction in unpaid parking tickets

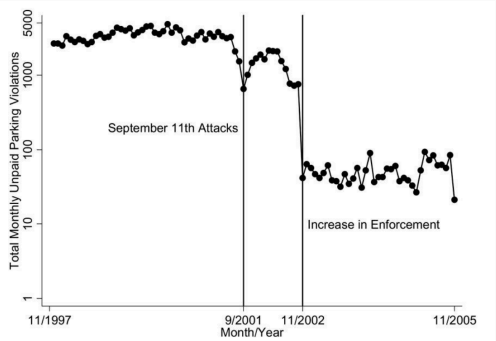


Fig. 1.—Total monthly New York City parking violations by diplomats, 1997–2005 (vertical axis on log scale).

Fisman and Miguel (2007): Identification Strategy II

- **Natural experiment**: post-9/11, NYC began cracking down on enforcement
- Diplomats with 3+ unpaid parking tickets had diplomat plates revoked
- Led to immediate 98% reduction in unpaid parking tickets
- So enforcement matters as well as corruption norms

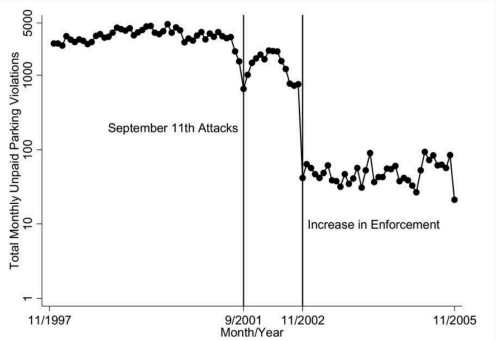


Fig. 1.—Total monthly New York City parking violations by diplomats, 1997–2005 (vertical axis on log scale).

Fisman and Miguel (2007): Data

TABLE 1

AVERAGE UNPAID ANNUAL NEW YORK CITY PARKING VIOLATIONS PER DIPLOMAT, NOVEMBER 1997 TO NOVEMBER 2005

Parking Violations Rank	Country Name	Violations per Diplomat, Pre-enforcement (11/1997–11/2002)	Violations per Diplomat, Postenforcement (11/2002–11/2005)	UN Mission Diplomats in 1998	Corruption Index, 1998	Country Code
1	Kuwait	249.4	.15	9	−1.07	KWT
2	Egypt	141.4	.33	24	.25	EGY
3	Chad	125.9	.00	2	.84	TCD
4	Sudan	120.6	.37	7	.75	SDN
5	Bulgaria	119.0	1.64	6	.50	BGR
6	Mozambique	112.1	.07	5	.77	MOZ
7	Albania	85.5	1.85	3	.92	ALB
8	Angola	82.7	1.71	9	1.05	AGO
9	Senegal	80.2	.21	11	.45	SEN
10	Pakistan	70.3	1.21	13	.76	PAK
11	Ivory Coast	68.0	.46	10	.35	CIV
12	Zambia	61.2	.15	9	.56	ZMB
13	Morocco	60.8	.40	17	.10	MAR
14	Ethiopia	60.4	.62	10	.25	ETH
15	Nigeria	59.4	.44	25	1.01	NGA
16	Syria	53.3	1.36	12	.58	SYR
17	Benin	50.4	6.50	8	.76	BEN
18	Zimbabwe	46.2	.86	14	.13	ZWE
19	Cameroon	44.1	2.86	8	1.11	CMR
20	Montenegro and Serbia	38.5	.05	6	.97	YUG
21	Bahrain	38.2	.65	7	−.41	BHR
22	Burundi	38.2	.11	3	.80	BDI

HOOD COLLEGE

TABLE 2
Descriptive Statistics

Variable	Mean	Standard Deviation	Observations
	A. Country-Level Data		
Unpaid New York City parking violations:[a]			
11/1997–11/2002	977.9	2,000.9	149
11/2002–11/2005	11.3	18.4	149
Unpaid and paid New York City parking violations, 11/1997–11/2002[a]	1,066.2	2,021.4	149
After-hours New York city parking violations, 11/1997–11/2002 [a]	40.6	72.7	149
Diplomats in the country UN mission, 1998[b]	11.8	11.1	149
Number of license plates registered to the country's UN mission, 2006[c]	10.5	14.0	139
Country corruption index, 1998[d]	.01	1.01	149
Log per capita income (1998 US$)[e]	7.35	1.59	149
Average government wage/country per capita income, early 1990s[f]	2.83	2.38	92
Log weighted distance between populations[g]	9.12	.41	149
Log total trade with the United States (1998 US$)[h]	20.3	2.7	146
Received U.S. economic aid (indicator), 1998[i]	.69	.46	147
Received U.S. military aid (indicator), 1998[i]	.63	.49	147

$$\text{Unpaid Violations} = \beta_0 + \beta_1 \text{Corruption} + \beta_2 \text{Enforcement} + \beta_3 \text{Diplomats} + \ldots + \beta_k \text{Controls}$$

Fisman and Miguel (2007): Results

TABLE 3
Country Characteristics and Unpaid New York City Parking Violations, November 1997 to November 2005

	Dependent Variable: Unpaid Parking Violations				
	(1)	(2)	(3)	(4)	(5)
Country corruption index, 1998	.48*** (.18)	.57*** (.22)	.57*** (.21)	.56** (.28)	.57* (.30)
Postenforcement period indicator (post-11/2002)	−4.41*** (.21)	−4.41*** (.21)	−4.21*** (.13)	−4.43*** (.20)	−4.41*** (.21)
Country corruption index × postenforcement period					−.01 (.28)
Diplomats	.05** (.02)	.04** (.02)	.05*** (.02)	.05** (.02)	.04** (.02)
Log per capita income (1998 US$)		.06 (.14)	.09 (.14)	64.2* (36.9)	.06 (.14)
Africa region indicator variable			2.86*** (.48)		
Asia region indicator variable			1.99*** (.50)		
Europe region indicator variable			2.24*** (.55)		
Latin America region indicator variable			1.67*** (.56)		
Middle East region indicator variable			3.23*** (.60)		
Oceania region indicator variable			1.51** (.64)		
Log per capita income (1998 US$) polynomials (quadratic, cubic, quartic)	No	No	No	Yes	No
Observations	298	298	298	298	298
Log pseudolikelihood	−1,570.21	−1,570.07	−1,547.69	−1,567.56	−1,570.07

HOOD COLLEGE

Fisman and Miguel (2007): Results II

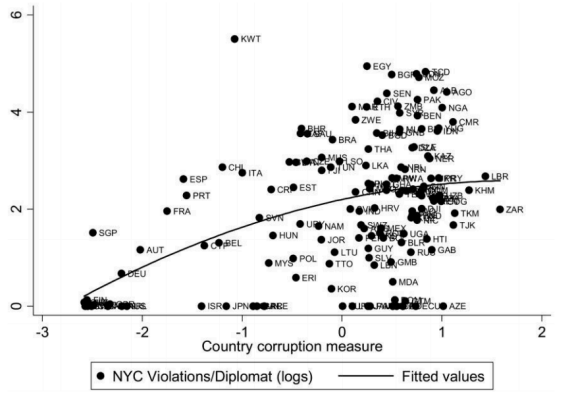


FIG. 2.—Country corruption and unpaid New York City parking violations per diplomat (in logs), pre-enforcement (November 1997 to November 2002). Country abbreviations are presented in table 1. The line is the quadratic regression fit. The *y*-axis is log(1 + Annual NYC Parking Violations/Diplomat).

Fisman and Miguel (2007): Results III

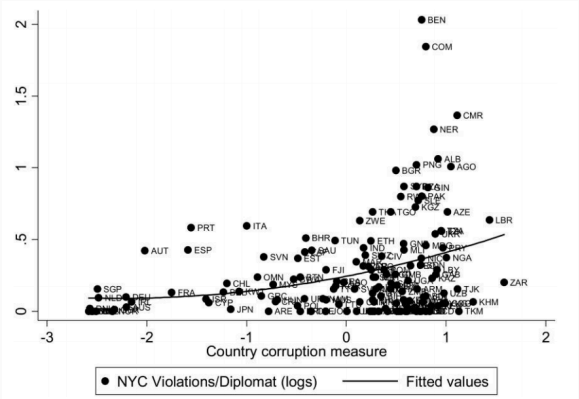


FIG. 3.—Country corruption and unpaid New York City parking violations per diplomat (in logs), postenforcement (November 2002 to November 2005). Country abbreviations are presented in table 1. The line is the quadratic regression fit. The *y*-axis is log(1 + Annual NYC Parking Violations/Diplomat).