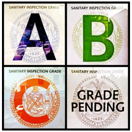
Do Monitoring and Transparency Make Food Safer?

Jason Huang

May 21, 2017

Motivation

- Sanitation at retail food establishments is a great public health concern but is also difficult for average consumers to enforce or observe
- Municipal health departments conduct regular food inspections to ensure restaurants uphold hygiene standards
  - Dearth of empirical studies that measure the efficacy of inspections on restaurant compliance
- Major cities, such as New York City, have started to make food inspection results more transparent by having restaurants post grades



  - Inspection results have become more prominent as they become integrated with Yelp

Research Questions

- Do inspections incentivize restaurants to improve their sanitation practices?
    - Do restaurants shift attention and effort away from areas that they did well toward areas that got flagged?
    - How do the magnitudes of the results vary across different types of establishments (heterogenous effects)
- How do restaurants, inspectors, and consumers respond to these letter grades?
    - Does simplifying multi-dimensional and continuous object into discrete letters cause distortions?

Main Findings

- Violation Citations
  - Marginal increase in the number of citations leads to improved conditions in subsequent inspections
    - Instead of focusing only on areas that they got cited, restaurants seem to improve in other areas as welll
  - More citations also reduces the probability that an establishment receives a complaint call
- Letter Grades
  - Probability that a restaurant receives a complaint increases sharply when it experiences a letter downgrade
    - Finding suggests a disconnect between consumer perception and the underlying sanitation quality
  - Inspection scores bunch around the threshold between letter grades

Road Map

2. Background on NYC Food Inspection Grading System

- Inspection occurs at least once a year
  - During inspection, inspector cites violations, assign scores based on severity, and sum up scores
- Total score is converted to letter grades based on the following cutoffs:
  - ≤ 13: A
  - ≥ 14 and ≤ 27: B
  - > 28: C
- Dual Inspection: anything lower than *A* during initial inspection results in a second inspection within a month
  - In the meantime, post previous grade
- After each inspection, restaurants can pursue adjudication to argue for better grades (in the meantime, post "Grade Pending")
  - Resolved within 6 weeks of initial inspection
- The health department temporally closes a restaurant if it finds critical violations

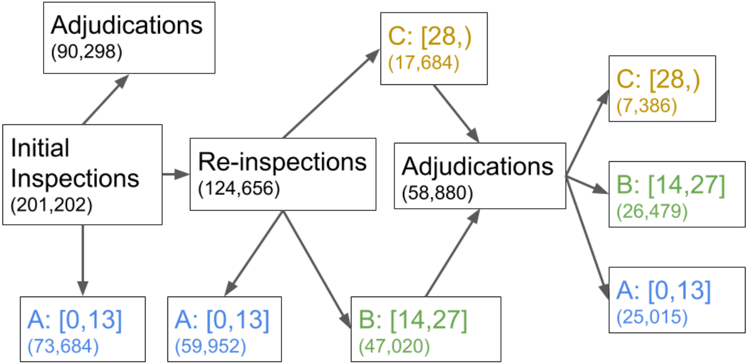
Food Inspection Data

- Universe of all food inspections conducted in NYC (2007 - 2016)
- Inspection date and inspector ID
- Individual violation codes, total score, corresponding adjudication date, and modified score
- Restaurant level info: name, address, cuisine, service type, and venue type

311 Call Data

- 311 is a phone line for non-emergency municipal services - also carries complaint calls to Department of Health and Mental Hygiene concerning restaurants
  - Examples of complaints: 'Rodents/Insects/Garbage', 'Bare Hands in Contact w/ Food', 'Food Contains Foreign Object', 'Food Spoiled',
- From 2010 to present
- Each complaint has an incident address and date
- Use fuzzy string matching on street address to the inspection data

NYC Food Inspection Pipeline



[0] The numbers in the parenthesis are the number of inspections in those steps

Impact of Inspection Results on Subsequent Inspection Scores

- Do inspection results affect sanitation?
  - Does getting a "bad" score make a restaurant work harder?
  - Does getting a "good" score cause a restaurant to slack off?
- Do the inspection results cause restaurants to reallocate efforts across multiple tasks?

Identification strategy: exploit the randomness of inspector assignment and use inspector stringencies as an instrument

Outcomes of Interests from Subsequent Inspections

- Overall inspection scores
- Whether a restaurants receives an A and does not need an re-inspection
- Whether a restaurant commits critical violations that results in temporarily closure

Empirical Strategy

$$Y_{i,t^{next}} = \beta Score_{i,t} + \delta_i + \tau_t + \tau_{t^{next}} + \varepsilon_{it}$$

- $Y_{i,t^{next}}$: outcome from next inspection: overall score, temporary closure, getting an A
- t^{next}: time period of next inspection
- $Score_{i,t^{next}}$: inspection score from the next inspection
- τ_t: time fixed effect
- δ_i: restaurant fixed effect
- ε_{it}: two-way clustered at zipcode and inspector levels[1]

[1] (Cameron at el 2009)

Empirical Strategy

A challenge of OLS is that $SCORE_{it}$ is endogenous

- If β positive, places that do poorly in the past tend to do poorly in the future
  - Restaurant FE doesn't fix problem if we have persistence
- If β is negative, cannot rule out mechanical mean reversion.

Use Inspector Assignment as Instrument

Instrument for when inspector j assigned to restaurant i

$$Z_{ij} = \frac{1}{n_j - I_{ij}} \left(\sum_{kjt \neq ijt} Y_{kjt} \right)$$

- n_j total number of inspections done by inspector j
- I_{ij} is the number of times inspector j has inspected restaurant i
- Y_{kjt} is the outcome used to calculate inspector tendencies

Inspector Stringency is Highly Predictive of Inspection Score

$$score_{it} = \gamma Z_{it} + \delta_i + \tau_t + \varepsilon_{it}$$

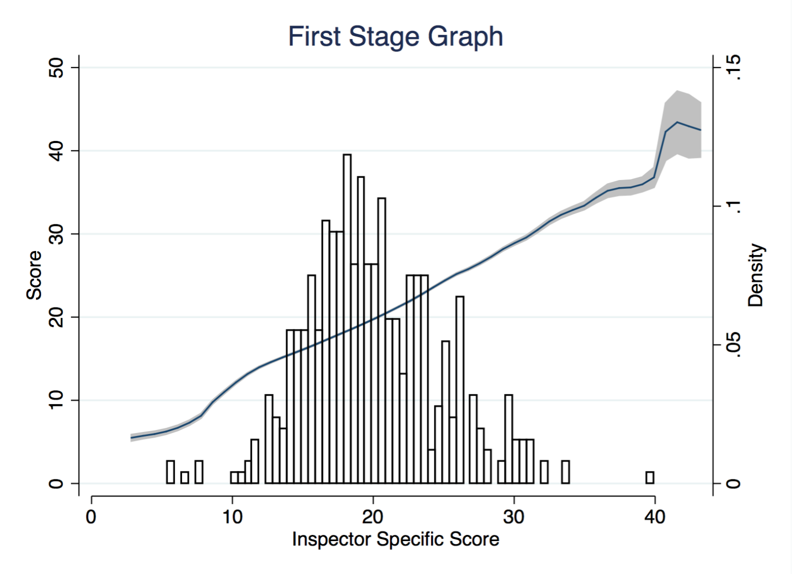
where Z_{it} is the leave-out propensity of inspector assigned to restaurant i in period t.

VARIABLES	(1) Score	(2) Score	(3) Score
Z	0.961***	0.997***	1.135***
	(0.00948)	(0.00983)	(0.0162)
Observations	330,469	330,466	325,681
R-squared	0.149	0.201	0.414
Restaurant Controls	NO	YES	NO
Restaurant FE	NO	NO	YES
F Statistics	10271	10270	4908

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

[1] First column consists of all inspections after 10/1/2010. The sample for the second column is reduced to inspections with non-empty zipcode, chain indicator, cuisine type, venue type, and service type. Standard errors are two-way clustered at the inspector and zipcode level.

Graphical Representation of First Stage



5. Assignment Process of Inspectors to Restaurants

- DOHMH claims that each inspector is randomly assigned to each inspection
- Implies:

$$Z_{ijt} = \beta X_i + \delta L_{i,t-1} + \varepsilon,$$

with X_i as restaurant characteristics (cuisine, service type, venue type, chain, etc) and $L_{i,t-1}$ as restaurant specific lag terms (previous scores and previous grades), $\beta = \delta = 0$.

Coefficients Close to 0 and Insignificant

VARIABLES	Score		Inspector Stringency		Inspector Stringency	
VARIABLES	(> 50 Inspections)	se	(> 50 Inspections)	se	(> 650 Inspections)	se
last score	0.218***	(0.00746)	-0.000440	(0.00199)	-0.00193	(0.00205)
last grade = B	1.430***	(0.131)	0.0178	(0.0628)	0.0837	(0.0683)
last grade = C	1.495***	(0.206)	-0.0494	(0.0764)	0.0240	(0.0807)
last inspector propensity	-0.298***	(0.0104)	-0.00363	(0.00491)	-0.00529	(0.00551)
chain	-3.644***	(0.179)	-0.0274	(0.0644)	-0.0300	(0.0734)
Sea Food	0.301	(0.412)	-0.0687	(0.129)	-0.0648	(0.136)
Chinese	1.588***	(0.307)	-0.0952*	(0.0510)	-0.0511	(0.0546)
Pizza/Italian	0.305**	(0.118)	-0.0893***	(0.0342)	-0.0614	(0.0384)
Coffee/Tea	-2.219***	(0.162)	-0.0421	(0.0889)	-0.0514	(0.0966)
Latin	1.534***	(0.303)	-0.0714	(0.0556)	-0.0390	(0.0583)
Spanish	1.560***	(0.261)	0.0125	(0.0615)	-0.0435	(0.0691)
Caribbean	1.542***	(0.288)	-0.0637	(0.0608)	-0.0619	(0.0544)
Sandwich	0.661**	(0.297)	-0.00916	(0.0472)	0.0180	(0.0488)
Concession Stands	-4.379***	(0.722)	0.295	(0.197)	0.209	(0.211)
Fast Food Restaurant-Food Court	0.570***	(0.187)	0.0394	(0.0881)	-0.0211	(0.0868)
Restaurant	1.453***	(0.158)	0.0449	(0.0662)	0.00343	(0.0641)
Buffet Service	2.564***	(0.365)	-0.182*	(0.107)	-0.196	(0.118)
Cater Service	-1.897***	(0.609)	-0.200	(0.190)	-0.193	(0.219)
Counter Service	-0.635***	(0.185)	0.0273	(0.0613)	0.0123	(0.0684)
Take-out Service	-1.410***	(0.185)	-0.153	(0.110)	-0.173	(0.123)
Wait Service	1.205***	(0.166)	0.0418	(0.0481)	0.0404	(0.0525)
Cafeteria Service	-3.541***	(0.500)	-0.280*	(0.156)	-0.200	(0.168)
Observations	299,174		299,174		244,099	
F Statistics	101.6		2.423		2.249	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

[1] Standard errors are two-way clustered at the inspector and restaurant level.

TSLS Regression Results

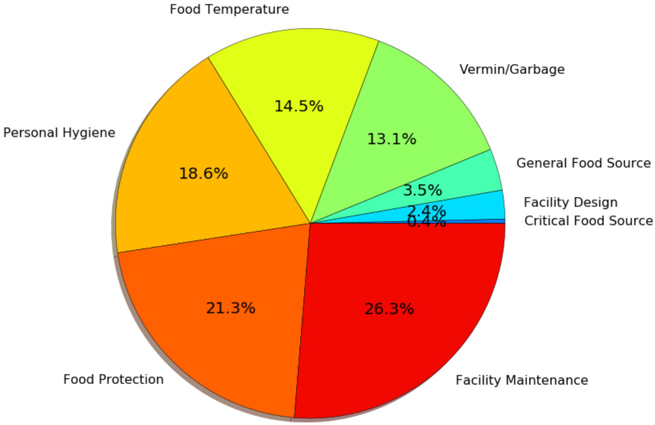
VARIABLES	(1) Score (OLS)	(2) Score (IV)	(3) Closure (OLS)	(4) Closure (IV)	(5) Grade A (OLS)	(6) Grade A (IV)
Score	-0.139***	-0.242***	4.67e-05	-0.00234***	-8.49e-05	0.00222***
	(0.00492)	(0.0162)	(0.000101)	(0.000393)	(0.000127)	(0.000556)
Observations	149,831	149,831	108,849	108,849	149,831	149,831
Inspection Date FE	YES	YES	YES	YES	YES	YES
Restaurant FE	YES	YES	YES	YES	YES	YES
Dependent mean	21.08	21.08	0.0123	0.0123	0.372	0.372

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

[1] Standard errors are two-way clustered at zipcode and inspector levels. Closure samples exclude inspections resulting in scores over 28 points.

Multi-Dimensionality of Food Inspections

- Individual Violation Codes Groped into Eight Groups:



Multi-dimensional 1st Stage

Instrument Construction:

$$Z_{ijg} = \frac{1}{n_j - I_{ij}} \left(\sum_{kjgt \neq ijgt} \#Cited_{ijgt} \right)$$

1st Stage:

$$\#Cited_{ijgt} = \sum_{g' \in \mathcal{G}} \theta_{gg'} Z_{ijg'} + \varepsilon_{ijgt}$$

- $\#Cited_{ijgt}$ is the number of group g violation that inspector j finds in restaurant i at time t and $\mathcal{G}$ is the set of violation groups
- $\theta_{gg'}$: measures how inspector's propensity to find violation in group g' relates to one's probability of finding violation in group g

Multi-task First Stage Equations

VARIABLES	(1) Facility Maintenance	(2) Food Protection	(3) Personal Hygiene	(4) Food Temperature	(5) Vermin/Garbage	(6) Gen. Food Source	(7) Facility Design	(8) Crit. Food Source
Facility Maintenance	1.051***	-0.0550**	0.0115	-0.0256*	-0.0104	0.00857	0.00276	0.000821
	(0.0177)	(0.0227)	(0.0147)	(0.0132)	(0.0139)	(0.00553)	(0.00395)	(0.00139)
Food Protection	-0.00719	0.726***	-0.00429	0.0709*	-0.0911***	0.00237	-0.0105	0.00328
	(0.0372)	(0.0419)	(0.0287)	(0.0388)	(0.0284)	(0.0126)	(0.00883)	(0.00391)
Personal Hygiene	0.00395	-0.127***	0.952***	-0.00446	-0.0686***	0.00179	-0.00719*	-0.00161
	(0.0197)	(0.0199)	(0.0192)	(0.0193)	(0.0128)	(0.00524)	(0.00415)	(0.00185)
Food Temperature	0.000328	-0.0810***	-0.0890***	0.808***	-0.0638***	0.00215	-0.0150**	-0.00488***
	(0.0236)	(0.0297)	(0.0179)	(0.0226)	(0.0182)	(0.00836)	(0.00609)	(0.00181)
Vermin/Garbage	0.0116	0.121**	-0.118***	-0.186***	1.011***	-0.0250	-0.0274**	-0.00804*
	(0.0486)	(0.0562)	(0.0419)	(0.0601)	(0.0383)	(0.0178)	(0.0136)	(0.00466)
Gen. Food Source	-0.00784	-0.0505	-0.0105	-0.0591	-0.0410	0.959***	-0.0233*	0.00318
	(0.0456)	(0.0451)	(0.0393)	(0.0407)	(0.0281)	(0.0220)	(0.0129)	(0.00365)
Facility Design	-0.0288	-0.112	-0.0860	0.0153	-0.116*	-0.0878**	0.845***	-0.0102
	(0.0915)	(0.130)	(0.0950)	(0.0930)	(0.0680)	(0.0347)	(0.0256)	(0.0105)
Crit. Food Source	-0.0660	0.0496	-0.487*	-0.115	0.145	-0.0241	-0.0630	0.921***
	(0.333)	(0.440)	(0.260)	(0.329)	(0.244)	(0.0868)	(0.0686)	(0.0454)
Observations	149,831	149,831	149,831	149,831	149,831	149,831	149,831	149,831
dependent mean	1.013	0.840	0.863	0.671	0.518	0.135	0.0827	0.0149

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Multi-dimensional 2nd Stage

$$\#Cited_{igt^{next}} = \sum_{g' \in \mathcal{G}} \beta_{gg'} \#Cited_{igt} + \delta_i + \tau_t + \varepsilon_{igt},$$

where

- $\#Cited_{igt}$: number of group g violations that restaurant i receives on date t
- τ_t: date fixed effect
- δ_i: restaurant fixed effect

How Current Citations Affect Subsequent Citations

VARIABLES	(1) Facility Maintenance	(2) Food Protection	(3) Personal Hygiene	(4) Food Temperature	(5) Vermin/Garbage	(6) Gen. Food Source	(7) Facility Design	(8) Crit. Food Source
Facility Maintenance	-0.236***	0.00733	0.000184	-0.00506	-0.000922	-0.00159	0.000981	-0.000509
	(0.00961)	(0.0119)	(0.0107)	(0.00944)	(0.00766)	(0.00380)	(0.00295)	(0.00140)
Food Protection	-0.0776**	-0.177***	-0.00731	-0.0142	-0.0397*	-0.0100	0.00372	0.00780*
	(0.0378)	(0.0306)	(0.0410)	(0.0253)	(0.0227)	(0.0113)	(0.0108)	(0.00400)
Personal Hygiene	0.0222	-0.00484	-0.189***	-0.00771	-0.000551	0.00458	0.00129	0.00217
	(0.0156)	(0.0103)	(0.0148)	(0.00957)	(0.00892)	(0.00489)	(0.00320)	(0.00173)
Food Temperature	0.0148	-0.0261	0.0133	-0.214***	-0.00923	0.00613	-0.00622	-0.00153
	(0.0216)	(0.0163)	(0.0162)	(0.0134)	(0.0128)	(0.00722)	(0.00573)	(0.00243)
Vermin/Garbage	0.132***	-0.0784*	0.00134	0.0446	-0.174***	0.0128	-0.0105	-0.00126
	(0.0439)	(0.0404)	(0.0544)	(0.0318)	(0.0283)	(0.0159)	(0.0138)	(0.00404)
Gen. Food Source	-0.0114	-0.00139	-0.0456	-0.0103	-0.0395*	-0.191***	-0.0147	-0.000895
	(0.0291)	(0.0348)	(0.0374)	(0.0334)	(0.0202)	(0.0199)	(0.00933)	(0.00395)
Facility Design	0.0265	-0.195**	-0.00742	0.0168	-0.0235	0.00421	-0.199***	0.0169*
	(0.0787)	(0.0925)	(0.0776)	(0.0551)	(0.0612)	(0.0296)	(0.0239)	(0.00977)
Crit. Food Source	0.274	-0.284	0.279	0.125	-0.0346	-0.0233	-0.0643	-0.185***
	(0.248)	(0.207)	(0.244)	(0.195)	(0.142)	(0.0848)	(0.0682)	(0.0298)
Observations	149,831	149,829	149,831	149,829	149,829	149,831	149,829	149,831
Dependent mean	0.989	0.802	0.842	0.645	0.503	0.125	0.0694	0.0126

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

[1] Standard errors are two-way clustered at zipcode and inspector levels

Concerns of Empirical Strategy

Exclusion Restriction

- Inspectors adjust their grading, given the identities of the previous inspectors
  - Unlikely, given geographically dispersed and random assignment, inspectors need to mentally track the tendencies of hundreds of inspectors
- Inspectors affect restaurant outcomes through channels other than the inspections

Monotonicity condition

- The inspection score is strictly increasing in inspector stringency

Monotonicity Tests

- Because inspections are multi-dimensional, monotonicity might not hold
  - Ex. A frozen yogurt establishment may get a better score from a more stringent inspector if that inspector cares only about hot food being kept above a certain temperature
- Two empirical implications:
  - Test 1: Stringent inspectors should be strict for different types of restaurants
    - Run first stage for various sub-samples (baseline sample)
  - Test 2: Inspectors who are strict for one type of restaurants should be strict for other types
    - Recalculate inspector stringency for each sub-sample with inspection results outside of that sub-sample (inverse-sample)

Monotonicity Test Results

Baseline-Sample

	(1)	(2)	(3)	(4)
VARIABLES	1st Quartile	2nd Quartile	3nd Quartile	4th Quartile
Estimate	0.407***	0.681***	0.852***	1.231***
	(0.0270)	(0.0240)	(0.0213)	(0.0256)
Observations	80,172	81,817	81,901	86,433

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Inverse-Sample

	(1)	(2)	(3)	(4)
VARIABLES	1st Quartile	2nd Quartile	3nd Quartile	4th Quartile
Estimate	0.369***	0.612***	0.779***	1.609***
	(0.0340)	(0.0247)	(0.0254)	(0.0674)
Observations	75,188	81,328	81,481	85,851

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Baseline-Sample

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Manhattan	Bronx	Brooklyn	Queens	Staten-Isl
Estimate	1.005***	1.082***	0.985***	0.969***	0.944***
	(0.0156)	(0.0275)	(0.0198)	(0.0206)	(0.0286)
Observations	131,900	31,010	79,664	77,186	10,507

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Inverse-Sample

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Manhattan	Bronx	Brooklyn	Queens	Staten-Is.
Estimate	0.990***	1.082***	0.954***	0.934***	0.938***
	(0.0269)	(0.0344)	(0.0238)	(0.0248)	(0.0308)
Observations	128,977	30,334	76,973	74,933	10,468

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Baseline-Sample

	(1)	(2)	(3)	(4)	(5)
VARIABLES	American	Pizza/Italian	Chinese	Coffee	Japanese
Estimate	0.957***	0.994***	1.086***	0.796***	1.096***
	(0.0241)	(0.0180)	(0.0402)	(0.0354)	(0.0262)
Observations	75,330	38,432	38,785	13,303	10,915

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Inverse-Sample

	(1)	(2)	(3)	(4)	(5)
VARIABLES	American	Pizza/Italian	Chinese	Coffee	Japanese
Estimate	0.929***	0.990***	1.070***	0.782***	1.098***
	(0.0298)	(0.0199)	(0.0449)	(0.0363)	(0.0273)
Observations	74,954	38,312	38,424	13,292	10,910

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Baseline-Sample

	(1)	(2)	(3)
VARIABLES	Counter Service	Takeout Service	Wait Service
Estimate	1.055***	0.894***	1.127***
	(0.0177)	(0.0140)	(0.0178)
Observations	127,981	127,365	73,381

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Inverse-Sample

	(1)	(2)	(3)
VARIABLES	Counter Service	Takeout Service	Wait Service
Estimate	1.064***	0.810***	1.160***
	(0.0269)	(0.0181)	(0.0236)
Observations	126,011	124,867	72,645

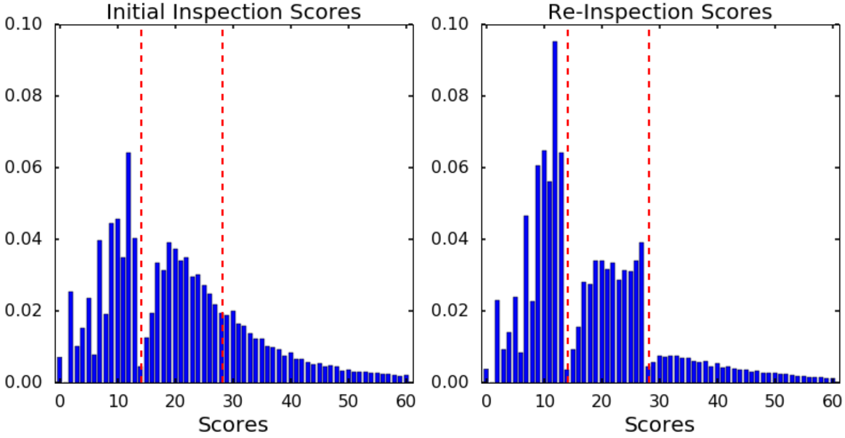
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

[1] To reduce noise, only inspections conducted by inspectors who have done at least 50 inspections remain in the regressions.

Issues of Mapping from Multi-dimensional Inspection Results into Discrete Grades

- For inspectors and restaurants, creates discontinuities in incentives
- For consumers, sanitation qualities of restaurants with the same grade are indistinguishable.

Scores Clearly Bunch Around Grade Thresholds



Letter Grades and Consumer Perception

- Restaurants with worse inspection scores receive more complaint calls
- The effect driven mostly by the discrete letter grades
    - Visible jumps around the letter grade thresholds
    - Effect concentrated around the month of the letter grade change (robust to placebo tests)

Sample Construction

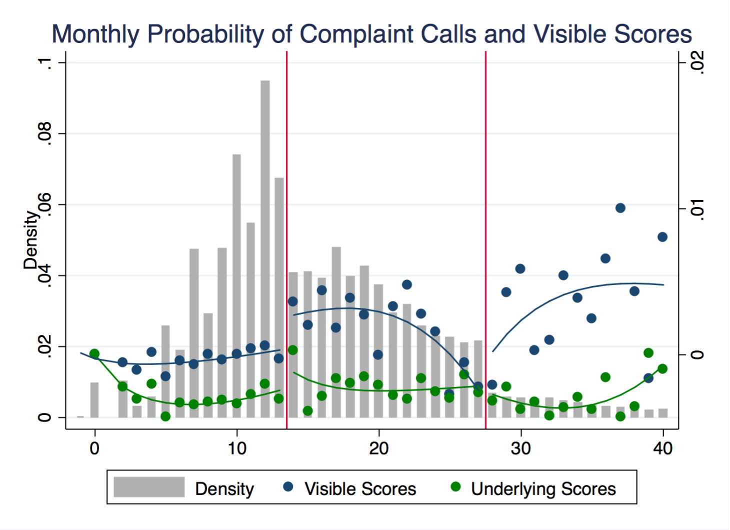
- Convert from inspection level data to a restaurant-month level panel data
- When an inspection or adjudication occurs in the middle of a month:
  - Consider only calls that were made in between the latest event and the end of the month
  - Calculate a weight variable as the fraction of days since the event to the end of the month

- Cross Sectional: use semi-parametric regression of probability of receiving at least one complaint call in a month on inspection scores

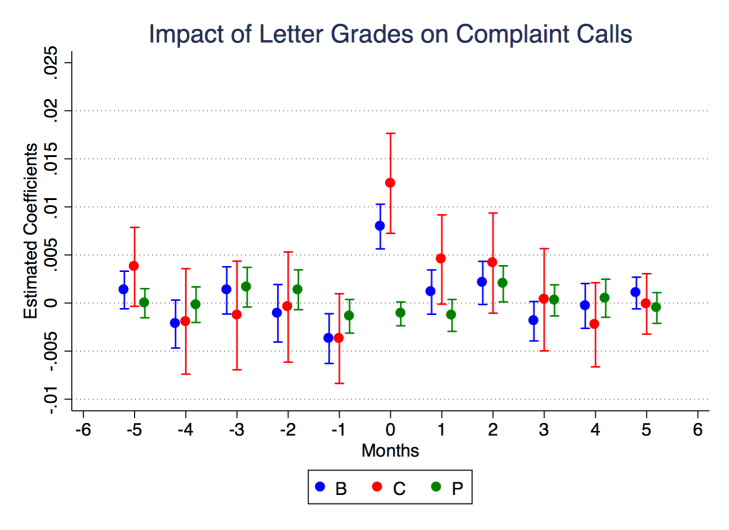
$$Pr(Called_{i,t}) = \delta_i + \tau_t + \sum_{s=0}^{50} \beta_s \mathbf{1}_{\{SCORE_{it}=s\}} + \varepsilon_{it}$$

- Longitudinal: add lead and lag months as placebo tests

$$Pr(Called_{i,t}) = \delta_i + \tau_t + \sum_{l=-5}^{5} \sum_{g \in \mathcal{G}} \beta_{gl} \mathbf{1}_{\{grade_{i,t+l}=g\}} + \varepsilon_{it}$$



[1] The grey density is the histogram of scores after adjudication. Visible scores are ones given post adjudication. Underlying scores are the ones given during cycle-inspections and may not contribute to letter grades.



[1] B, C, and P refers to letter grade B, C, and Grade Pending, respectively.

Conclusion and Discussion

Conclusion

- Marginally more citations leads to improved subsequent inspection results
  - Some complementarity in cleanliness (ex. better refrigerator leads to few food temperature and food protection violations).
- Both restaurants and inspectors respond by bunching around the threshold
- Customer perception changes significantly once a restaurant experiences a grade change.

Policy Implication

- Instead of using posting letter grades, post the numerical scores