

Estimating Demand for (and Willingness to Review) Restaurants From Yelp.com Data

Josh Lustig and Mike Luca

Boston University

May 9, 2011

- On Yelp.com consumers review products they've consumed

- On Yelp.com consumers review products they've consumed
 - Consumers submit reviews between 1 and 5 stars, plus an optional description of their experience.

- On Yelp.com consumers review products they've consumed
 - Consumers submit reviews between 1 and 5 stars, plus an optional description of their experience.
 - Yelp users can see review information and search Yelp for products in various ways.

- On Yelp.com consumers review products they've consumed
 - Consumers submit reviews between 1 and 5 stars, plus an optional description of their experience.
 - Yelp users can see review information and search Yelp for products in various ways.
- Yelp covers a large range of markets:

- On Yelp.com consumers review products they've consumed
 - Consumers submit reviews between 1 and 5 stars, plus an optional description of their experience.
 - Yelp users can see review information and search Yelp for products in various ways.
- Yelp covers a large range of markets:
 - Usually local services: restaurants, stores, hotels, beauty/spa.

- On Yelp.com consumers review products they've consumed
 - Consumers submit reviews between 1 and 5 stars, plus an optional description of their experience.
 - Yelp users can see review information and search Yelp for products in various ways.
- Yelp covers a large range of markets:
 - Usually local services: restaurants, stores, hotels, beauty/spa.
 - Most commonly used in urban areas (from Austin, TX to Washington, DC)

- Project uses Yelp.com data at the restaurant (product) level:

	# Reviews	# 1 Star	# 2 Star	# 5 Star
Restaurant 1	20	3	3		8
Restaurant 2	30	5	5		9
.....					
.....					
.....					
Restaurant J	500	40	60		120

- Project uses Yelp.com data at the restaurant (product) level:

	# Reviews	# 1 Star	# 2 Star	# 5 Star
Restaurant 1	20	3	3		8
Restaurant 2	30	5	5		9
.....					
.....					
.....					
Restaurant J	500	40	60		120

- Data generating process?

- Project uses Yelp.com data at the restaurant (product) level:

	# Reviews	# 1 Star	# 2 Star	# 5 Star
Restaurant 1	20	3	3		8
Restaurant 2	30	5	5		9
.....					
.....					
.....					
Restaurant J	500	40	60		120

- Data generating process?
 - 1 Depends on preferences over restaurants (i.e. Restaurant Demand).

- Project uses Yelp.com data at the restaurant (product) level:

	# Reviews	# 1 Star	# 2 Star	# 5 Star
Restaurant 1	20	3	3		8
Restaurant 2	30	5	5		9
.....					
.....					
.....					
Restaurant J	500	40	60		120

- Data generating process?
 - ① Depends on preferences over restaurants (i.e. Restaurant Demand).
 - Restaurants are differentiated along many dimensions.

- Project uses Yelp.com data at the restaurant (product) level:

	# Reviews	# 1 Star	# 2 Star	# 5 Star
Restaurant 1	20	3	3		8
Restaurant 2	30	5	5		9
.....					
.....					
.....					
Restaurant J	500	40	60		120

- Data generating process?
 - 1 Depends on preferences over restaurants (i.e. Restaurant Demand).
 - Restaurants are differentiated along many dimensions.
 - Restaurants offer different levels utility and have different market shares.

- Project uses Yelp.com data at the restaurant (product) level:

	# Reviews	# 1 Star	# 2 Star	# 5 Star
Restaurant 1	20	3	3		8
Restaurant 2	30	5	5		9
.....					
.....					
.....					
Restaurant J	500	40	60		120

- Data generating process?

- 1 Depends on preferences over restaurants (i.e. Restaurant Demand).
 - Restaurants are differentiated along many dimensions.
 - Restaurants offer different levels utility and have different market shares.
 - If J is high quality, they'll attract more customers and their customers will be more satisfied.

- Project uses Yelp.com data at the restaurant (product) level:

	# Reviews	# 1 Star	# 2 Star	# 5 Star
Restaurant 1	20	3	3		8
Restaurant 2	30	5	5		9
.....					
.....					
.....					
Restaurant J	500	40	60		120

- Data generating process?

① Depends on preferences over restaurants (i.e. Restaurant Demand).

- Restaurants are differentiated along many dimensions.
- Restaurants offer different levels utility and have different market shares.
- If J is high quality, they'll attract more customers and their customers will be more satisfied.
- If this the only determinant of observed data, Yelp data implies market shares.

- Project uses Yelp.com data at the restaurant (product) level:

	# Reviews	# 1 Star	# 2 Star	# 5 Star
Restaurant 1	20	3	3		8
Restaurant 2	30	5	5		9
.....					
.....					
.....					
Restaurant J	500	40	60		120

- Data generating process?

- 1 Depends on preferences over restaurants (i.e. Restaurant Demand).
 - Restaurants are differentiated along many dimensions.
 - Restaurants offer different levels utility and have different market shares.
 - If J is high quality, they'll attract more customers and their customers will be more satisfied.
 - If this the only determinant of observed data, Yelp data implies market shares.
- 2 Depends on determinants of review behavior..

- Project uses Yelp.com data at the restaurant (product) level:

	# Reviews	# 1 Star	# 2 Star	# 5 Star
Restaurant 1	20	3	3		8
Restaurant 2	30	5	5		9
.....					
.....					
.....					
Restaurant J	500	40	60		120

- Data generating process?

- Depends on preferences over restaurants (i.e. Restaurant Demand).
 - Restaurants are differentiated along many dimensions.
 - Restaurants offer different levels utility and have different market shares.
 - If J is high quality, they'll attract more customers and their customers will be more satisfied.
 - If this the only determinant of observed data, Yelp data implies market shares.
- Depends on determinants of review behavior..
 - Willingness to submit a review may depend on restaurant characteristics.

- Project uses Yelp.com data at the restaurant (product) level:

	# Reviews	# 1 Star	# 2 Star	# 5 Star
Restaurant 1	20	3	3		8
Restaurant 2	30	5	5		9
.....					
.....					
.....					
Restaurant J	500	40	60		120

- Data generating process?

1 Depends on preferences over restaurants (i.e. Restaurant Demand).

- Restaurants are differentiated along many dimensions.
- Restaurants offer different levels utility and have different market shares.
- If J is high quality, they'll attract more customers and their customers will be more satisfied.
- If this the only determinant of observed data, Yelp data implies market shares.

2 Depends on determinants of review behavior..

- Willingness to submit a review may depend on restaurant characteristics.
- Willingness to submit a review may depend on utility restaurant provided.

- Project uses Yelp.com data at the restaurant (product) level:

	# Reviews	# 1 Star	# 2 Star	# 5 Star
Restaurant 1	20	3	3		8
Restaurant 2	30	5	5		9
.....					
.....					
.....					
Restaurant J	500	40	60		120

- Data generating process?

- Depends on preferences over restaurants (i.e. Restaurant Demand).
 - Restaurants are differentiated along many dimensions.
 - Restaurants offer different levels utility and have different market shares.
 - If J is high quality, they'll attract more customers and their customers will be more satisfied.
 - If this the only determinant of observed data, Yelp data implies market shares.
- Depends on determinants of review behavior..
 - Willingness to submit a review may depend on restaurant characteristics.
 - Willingness to submit a review may depend on utility restaurant provided.
 - If this is relevant, Yelp data does not imply market shares.

This Project

- Estimate a model of consumers' restaurant choices and review behavior.

This Project

- Estimate a model of consumers' restaurant choices and review behavior.
 - Restaurant choices maximize expected utility.

This Project

- Estimate a model of consumers' restaurant choices and review behavior.
 - Restaurant choices maximize expected utility.
 - Willingness to review depends on realized utility.

This Project

- Estimate a model of consumers' restaurant choices and review behavior.
 - Restaurant choices maximize expected utility.
 - Willingness to review depends on realized utility.
- Potentially use model to answer questions in IO and about online reviews.

This Project

- Estimate a model of consumers' restaurant choices and review behavior.
 - Restaurant choices maximize expected utility.
 - Willingness to review depends on realized utility.
- Potentially use model to answer questions in IO and about online reviews.
 - ① Better answers to IO questions with market share data obtainable from Yelp after controlling for review behavior. For example,

This Project

- Estimate a model of consumers' restaurant choices and review behavior.
 - Restaurant choices maximize expected utility.
 - Willingness to review depends on realized utility.
- Potentially use model to answer questions in IO and about online reviews.
 - ① Better answers to IO questions with market share data obtainable from Yelp after controlling for review behavior. For example,
 - Papers estimating "competitive effects" using entry behavior and market characteristics that proxy for local demand and costs.

This Project

- Estimate a model of consumers' restaurant choices and review behavior.
 - Restaurant choices maximize expected utility.
 - Willingness to review depends on realized utility.
- Potentially use model to answer questions in IO and about online reviews.
 - ① Better answers to IO questions with market share data obtainable from Yelp after controlling for review behavior. For example,
 - Papers estimating "competitive effects" using entry behavior and market characteristics that proxy for local demand and costs.
 - Ideally, would estimate local demand using information on restaurant market shares (not available)

This Project

- Estimate a model of consumers' restaurant choices and review behavior.
 - Restaurant choices maximize expected utility.
 - Willingness to review depends on realized utility.
- Potentially use model to answer questions in IO and about online reviews.
 - ① Better answers to IO questions with market share data obtainable from Yelp after controlling for review behavior. For example,
 - Papers estimating "competitive effects" using entry behavior and market characteristics that proxy for local demand and costs.
 - Ideally, would estimate local demand using information on restaurant market shares (not available)
 - Control for review behavior \Rightarrow Infer market shares from Yelp \Rightarrow Estimate richer models of local imperfect competition.

This Project

- Estimate a model of consumers' restaurant choices and review behavior.
 - Restaurant choices maximize expected utility.
 - Willingness to review depends on realized utility.
- Potentially use model to answer questions in IO and about online reviews.
 - 1 Better answers to IO questions with market share data obtainable from Yelp after controlling for review behavior. For example,
 - Papers estimating "competitive effects" using entry behavior and market characteristics that proxy for local demand and costs.
 - Ideally, would estimate local demand using information on restaurant market shares (not available)
 - Control for review behavior \Rightarrow Infer market shares from Yelp \Rightarrow Estimate richer models of local imperfect competition.
 - 2 Questions about online review require controlling for demand. For example,

This Project

- Estimate a model of consumers' restaurant choices and review behavior.
 - Restaurant choices maximize expected utility.
 - Willingness to review depends on realized utility.
- Potentially use model to answer questions in IO and about online reviews.
 - 1 Better answers to IO questions with market share data obtainable from Yelp after controlling for review behavior. For example,
 - Papers estimating "competitive effects" using entry behavior and market characteristics that proxy for local demand and costs.
 - Ideally, would estimate local demand using information on restaurant market shares (not available)
 - Control for review behavior \Rightarrow Infer market shares from Yelp \Rightarrow Estimate richer models of local imperfect competition.
 - 2 Questions about online review require controlling for demand. For example,
 - Review clustering has been observed. If clustering due to review behavior (for example, only review products you dislike), consumers' ability to infer product quality from reviews is potentially limited.

This Project

- Estimate a model of consumers' restaurant choices and review behavior.
 - Restaurant choices maximize expected utility.
 - Willingness to review depends on realized utility.
- Potentially use model to answer questions in IO and about online reviews.
 - 1 Better answers to IO questions with market share data obtainable from Yelp after controlling for review behavior. For example,
 - Papers estimating "competitive effects" using entry behavior and market characteristics that proxy for local demand and costs.
 - Ideally, would estimate local demand using information on restaurant market shares (not available)
 - Control for review behavior \Rightarrow Infer market shares from Yelp \Rightarrow Estimate richer models of local imperfect competition.
 - 2 Questions about online review require controlling for demand. For example,
 - Review clustering has been observed. If clustering due to review behavior (for example, only review products you dislike), consumers' ability to infer product quality from reviews is potentially limited.
 - But review clustering can also be explained by homogenous products.

Intuition for Identification

- Yelp data contains two types of information:

Intuition for Identification

- Yelp data contains two types of information:
 - 1 # of Reviews: Helps identify preferences for restaurants.

Intuition for Identification

- Yelp data contains two types of information:
 - 1 # of Reviews: Helps identify preferences for restaurants.
 - 2 Distribution of reviews (% 1 star, 2 star, etc.): Helps identify determinants of willingness to review..

Intuition for Identification

- Yelp data contains two types of information:
 - 1 # of Reviews: Helps identify preferences for restaurants.
 - 2 Distribution of reviews (% 1 star, 2 star, etc.): Helps identify determinants of willingness to review..
- Combined model predicts joint distribution of $(n_j, n_{j1}, \dots, n_{j5})$ across restaurants (j).

Intuition for Identification

- Yelp data contains two types of information:
 - 1 # of Reviews: Helps identify preferences for restaurants.
 - 2 Distribution of reviews (% 1 star, 2 star, etc.): Helps identify determinants of willingness to review..
- Combined model predicts joint distribution of $(n_j, n_{j1}, \dots, n_{j5})$ across restaurants (j).
- Moments from this distribution allow for ID of restaurant demand and willingness to review. Intuition:

Intuition for Identification

- Yelp data contains two types of information:
 - 1 # of Reviews: Helps identify preferences for restaurants.
 - 2 Distribution of reviews (% 1 star, 2 star, etc.): Helps identify determinants of willingness to review..
- Combined model predicts joint distribution of $(n_j, n_{j1}, \dots, n_{j5})$ across restaurants (j).
- Moments from this distribution allow for ID of restaurant demand and willingness to review. Intuition:
 - 1 If we know demand:

Intuition for Identification

- Yelp data contains two types of information:
 - 1 # of Reviews: Helps identify preferences for restaurants.
 - 2 Distribution of reviews (% 1 star, 2 star, etc.): Helps identify determinants of willingness to review..
- Combined model predicts joint distribution of $(n_j, n_{j1}, \dots, n_{j5})$ across restaurants (j).
- Moments from this distribution allow for ID of restaurant demand and willingness to review. Intuition:
 - 1 If we know demand:
 - Can infer distributions of utility different restaurants provide in equilibrium.

Intuition for Identification

- Yelp data contains two types of information:
 - 1 # of Reviews: Helps identify preferences for restaurants.
 - 2 Distribution of reviews (% 1 star, 2 star, etc.): Helps identify determinants of willingness to review..
- Combined model predicts joint distribution of $(n_j, n_{j1}, \dots, n_{j5})$ across restaurants (j) .
- Moments from this distribution allow for ID of restaurant demand and willingness to review. Intuition:
 - 1 If we know demand:
 - Can infer distributions of utility different restaurants provide in equilibrium.
 - If willingness to review unrelated to utility, these distributions should predict the reviews restaurants receive. If they don't...

Intuition for Identification

- Yelp data contains two types of information:
 - 1 # of Reviews: Helps identify preferences for restaurants.
 - 2 Distribution of reviews (% 1 star, 2 star, etc.): Helps identify determinants of willingness to review..
- Combined model predicts joint distribution of $(n_j, n_{j1}, \dots, n_{j5})$ across restaurants (j).
- Moments from this distribution allow for ID of restaurant demand and willingness to review. Intuition:
 - 1 If we know demand:
 - Can infer distributions of utility different restaurants provide in equilibrium.
 - If willingness to review unrelated to utility, these distributions should predict the reviews restaurants receive. If they don't...
 - Choose parameters (that affect review behavior) so model accurately predicts "within-restaurant" distributions of reviews.

Intuition for Identification

- Yelp data contains two types of information:
 - 1 # of Reviews: Helps identify preferences for restaurants.
 - 2 Distribution of reviews (% 1 star, 2 star, etc.): Helps identify determinants of willingness to review..
- Combined model predicts joint distribution of $(n_j, n_{j1}, \dots, n_{j5})$ across restaurants (j).
- Moments from this distribution allow for ID of restaurant demand and willingness to review. Intuition:
 - 1 If we know demand:
 - Can infer distributions of utility different restaurants provide in equilibrium.
 - If willingness to review unrelated to utility, these distributions should predict the reviews restaurants receive. If they don't...
 - Choose parameters (that affect review behavior) so model accurately predicts "within-restaurant" distributions of reviews.
 - 2 Typically, choose demand parameters so predicted market shares match observed market shares. If we know willingness to review:

Intuition for Identification

- Yelp data contains two types of information:
 - 1 # of Reviews: Helps identify preferences for restaurants.
 - 2 Distribution of reviews (% 1 star, 2 star, etc.): Helps identify determinants of willingness to review..
- Combined model predicts joint distribution of $(n_j, n_{j1}, \dots, n_{j5})$ across restaurants (j).
- Moments from this distribution allow for ID of restaurant demand and willingness to review. Intuition:
 - 1 If we know demand:
 - Can infer distributions of utility different restaurants provide in equilibrium.
 - If willingness to review unrelated to utility, these distributions should predict the reviews restaurants receive. If they don't...
 - Choose parameters (that affect review behavior) so model accurately predicts "within-restaurant" distributions of reviews.
 - 2 Typically, choose demand parameters so predicted market shares match observed market shares. If we know willingness to review:
 - Demand model will predict variation in market shares AND distributions of reviews across restaurants.

Intuition for Identification

- Yelp data contains two types of information:
 - 1 # of Reviews: Helps identify preferences for restaurants.
 - 2 Distribution of reviews (% 1 star, 2 star, etc.): Helps identify determinants of willingness to review..
- Combined model predicts joint distribution of $(n_j, n_{j1}, \dots, n_{j5})$ across restaurants (j) .
- Moments from this distribution allow for ID of restaurant demand and willingness to review. Intuition:
 - 1 If we know demand:
 - Can infer distributions of utility different restaurants provide in equilibrium.
 - If willingness to review unrelated to utility, these distributions should predict the reviews restaurants receive. If they don't...
 - Choose parameters (that affect review behavior) so model accurately predicts "within-restaurant" distributions of reviews.
 - 2 Typically, choose demand parameters so predicted market shares match observed market shares. If we know willingness to review:
 - Demand model will predict variation in market shares AND distributions of reviews across restaurants.
 - Choose demand paramters to match "across restaurant" variation in reviews.

Raw Data

- Calculated all restaurants' average review and split sample into thirds. Three "thirds" received 3.64, 3.74, and 3.80 stars in their average reviews.

Raw Data

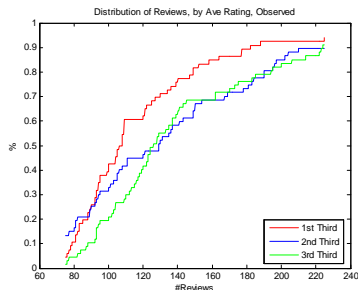
- Calculated all restaurants' average review and split sample into thirds. Three "thirds" received 3.64, 3.74, and 3.80 stars in their average reviews.
- Distributions of *#Reviews* informative about selection..

Raw Data

- Calculated all restaurants' average review and split sample into thirds. Three "thirds" received 3.64, 3.74, and 3.80 stars in their average reviews.
- Distributions of *#Reviews* informative about selection..
 - T2 restaurants higher quality than T1... meaning more customers. T2 restaurants receive many more reviews.

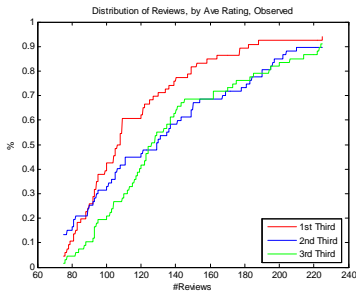
Raw Data

- Calculated all restaurants' average review and split sample into thirds. Three "thirds" received 3.64, 3.74, and 3.80 stars in their average reviews.
- Distributions of *#Reviews* informative about selection.
 - T2 restaurants higher quality than T1... meaning more customers. T2 restaurants receive many more reviews.
 - T3 restaurants higher quality than T2... meaning more customers. But T3 restaurants receive similar # of reviews.



Raw Data

- Calculated all restaurants' average review and split sample into thirds. Three "thirds" received 3.64, 3.74, and 3.80 stars in their average reviews.
- Distributions of *#Reviews* informative about selection.
 - T2 restaurants higher quality than T1... meaning more customers. T2 restaurants receive many more reviews.
 - T3 restaurants higher quality than T2... meaning more customers. But T3 restaurants receive similar # of reviews.



- Results robust to conditioning on price level and sample choice.

Preliminary Results

- Evidence suggests willingness to review depends on utility...

Preliminary Results

- Evidence suggests willingness to review depends on utility...
 - ① If your realized utility warrants a 1 star review, you submit with probability 12%.

Preliminary Results

- Evidence suggests willingness to review depends on utility...
 - 1 If your realized utility warrants a 1 star review, you submit with probability 12%.
 - 2 If your realized utility warrants a 5 star review, you submit with probability 2%.

Preliminary Results

- Evidence suggests willingness to review depends on utility...
 - ① If your realized utility warrants a 1 star review, you submit with probability 12%.
 - ② If your realized utility warrants a 5 star review, you submit with probability 2%.
- Implications?

Preliminary Results

- Evidence suggests willingness to review depends on utility...
 - 1 If your realized utility warrants a 1 star review, you submit with probability 12%.
 - 2 If your realized utility warrants a 5 star review, you submit with probability 2%.
- Implications?
 - 1 Restaurants with high market shares (i.e. high quality) are underrepresented in Yelp. But, can use model to infer market shares from observed reviews.

Preliminary Results

- Evidence suggests willingness to review depends on utility...
 - ① If your realized utility warrants a 1 star review, you submit with probability 12%.
 - ② If your realized utility warrants a 5 star review, you submit with probability 2%.
- Implications?
 - ① Restaurants with high market shares (i.e. high quality) are underrepresented in Yelp. But, can use model to infer market shares from observed reviews.
 - ② Selection leads to clustering. Review behavior causes variance of restaurants' "Ave. # of Stars" to fall by $\approx 66\%$

- Model requires very little data (so strategy is widely applicable)

Data

- Model requires very little data (so strategy is widely applicable)
- For restaurants in Boston: # Total Reviews, # 1 Star Reviews, , # 5 Star Reviews.

- Model requires very little data (so strategy is widely applicable)
- For restaurants in Boston: # Total Reviews, # 1 Star Reviews, , # 5 Star Reviews.
 - Started with restaurants whose first review observed before Sept 18, 2006 and last review observed after Aug 6, 2009

- Model requires very little data (so strategy is widely applicable)
- For restaurants in Boston: # Total Reviews, # 1 Star Reviews, , # 5 Star Reviews.
 - Started with restaurants whose first review observed before Sept 18, 2006 and last review observed after Aug 6, 2009
 - Dropped restaurants that don't serve dinner. Dropped restaurants in neighborhoods receiving very few reviews

- Model requires very little data (so strategy is widely applicable)
- For restaurants in Boston: # Total Reviews, # 1 Star Reviews, , # 5 Star Reviews.
 - Started with restaurants whose first review observed before Sept 18, 2006 and last review observed after Aug 6, 2009
 - Dropped restaurants that don't serve dinner. Dropped restaurants in neighborhoods receiving very few reviews
 - End up with 482 Boston restaurants. Use the largest 200 (account for 72% of reviews).

- Model requires very little data (so strategy is widely applicable)
- For restaurants in Boston: # Total Reviews, # 1 Star Reviews, , # 5 Star Reviews.
 - Started with restaurants whose first review observed before Sept 18, 2006 and last review observed after Aug 6, 2009
 - Dropped restaurants that don't serve dinner. Dropped restaurants in neighborhoods receiving very few reviews
 - End up with 482 Boston restaurants. Use the largest 200 (account for 72% of reviews).
- Two supplemental pieces of information:

- Model requires very little data (so strategy is widely applicable)
- For restaurants in Boston: # Total Reviews, # 1 Star Reviews, , # 5 Star Reviews.
 - Started with restaurants whose first review observed before Sept 18, 2006 and last review observed after Aug 6, 2009
 - Dropped restaurants that don't serve dinner. Dropped restaurants in neighborhoods receiving very few reviews
 - End up with 482 Boston restaurants. Use the largest 200 (account for 72% of reviews).
- Two supplemental pieces of information:
 - Total market share of inside good: about 14% of meals eaten out (American Dietetic Association, USDA)

- Model requires very little data (so strategy is widely applicable)
- For restaurants in Boston: # Total Reviews, # 1 Star Reviews, , # 5 Star Reviews.
 - Started with restaurants whose first review observed before Sept 18, 2006 and last review observed after Aug 6, 2009
 - Dropped restaurants that don't serve dinner. Dropped restaurants in neighborhoods receiving very few reviews
 - End up with 482 Boston restaurants. Use the largest 200 (account for 72% of reviews).
- Two supplemental pieces of information:
 - Total market share of inside good: about 14% of meals eaten out (American Dietetic Association, USDA)
 - # Households = 1,000,000

- Model requires very little data (so strategy is widely applicable)
- For restaurants in Boston: # Total Reviews, # 1 Star Reviews, , # 5 Star Reviews.
 - Started with restaurants whose first review observed before Sept 18, 2006 and last review observed after Aug 6, 2009
 - Dropped restaurants that don't serve dinner. Dropped restaurants in neighborhoods receiving very few reviews
 - End up with 482 Boston restaurants. Use the largest 200 (account for 72% of reviews).
- Two supplemental pieces of information:
 - Total market share of inside good: about 14% of meals eaten out (American Dietetic Association, USDA)
 - # Households = 1,000,000
- More general versions of model will incorporate identities of reviewers and restaurant characteristics (all downloadable from Yelp).

Summary Statistics

Full Sample: 200 Restaurants				
	<u>Mean</u>	<u>Std Dev</u>	<u>Minimum</u>	<u>Maximum</u>
# Reviews	137.8	65.7	71	511
% 1 Star	4.6%	0.039	0.0%	25.0%
% 2 Star	9.1%	0.056	0.8%	39.5%
% 3 Star	20.5%	0.078	5.2%	38.1%
% 4 Star	40.4%	0.085	8.6%	61.8%
% 5 Star	25.3%	0.135	1.2%	65.4%

Summary Statistics

Differences across Price Levels			
	\$	\$\$	\$\$\$-\$\$\$\$
# Restaurants	53	100	47
# Reviews	126.6	138.8	151.3
% 1 Star	5.2%	4.6%	4.4%
% 2 Star	9.0%	9.0%	10.3%
% 3 Star	20.6%	21.4%	19.9%
% 4 Star	39.6%	42.8%	37.4%
% 5 Star	25.6%	22.2%	27.9%

Summary Statistics

Restaurants with Different total # Reviews				
	<u>Quartile 1</u>	<u>Quartile 2</u>	<u>Quartile 3</u>	<u>Quartile 4</u>
# Restaurants	50	50	50	50
# Reviews	80.8	103.6	134.7	225.5
% 1 Star	5.4%	5.0%	3.8%	4.1%
% 2 Star	10.7%	8.9%	8.1%	9.0%
% 3 Star	22.2%	20.5%	20.6%	18.9%
% 4 Star	40.2%	40.9%	41.5%	38.9%
% 5 Star	21.5%	24.5%	26.0%	29.0%

Decisions Modeled

- Observed data depends on three decisions, all modeled:

Decisions Modeled

- Observed data depends on three decisions, all modeled:
 - 1 Choose a restaurant or stay home.

Decisions Modeled

- Observed data depends on three decisions, all modeled:
 - 1 Choose a restaurant or stay home.
 - 2 If restaurant chosen, whether to submit a review.

Decisions Modeled

- Observed data depends on three decisions, all modeled:
 - 1 Choose a restaurant or stay home.
 - 2 If restaurant chosen, whether to submit a review.
 - 3 If a review being submitted, how many stars to give.

Model of Restaurant Demand

- Consumers choose the restaurant (or outside good) that maximizes expected utility.

Model of Restaurant Demand

- Consumers choose the restaurant (or outside good) that maximizes expected utility.
 - Risk averse consumer i knows restaurant j will provide level of consumption (net of prices) drawn from $N(\mu_j, \sigma_j^2)$.

Model of Restaurant Demand

- Consumers choose the restaurant (or outside good) that maximizes expected utility.
 - Risk averse consumer i knows restaurant j will provide level of consumption (net of prices) drawn from $N(\mu_j, \sigma_j^2)$.
 - $EU_{ij} = \int_c -e^{-\gamma c} f(c|\mu_j, \sigma_j^2) + \varepsilon_{ij} = -e^{-\gamma(\mu_j - \frac{\gamma}{2}\sigma_j^2)} + \varepsilon_{ij}$

Model of Restaurant Demand

- Consumers choose the restaurant (or outside good) that maximizes expected utility.
 - Risk averse consumer i knows restaurant j will provide level of consumption (net of prices) drawn from $N(\mu_j, \sigma_j^2)$.
 - $EU_{ij} = \int_c -e^{-\gamma c} f(c|\mu_j, \sigma_j^2) + \varepsilon_{ij} = -e^{-\gamma(\mu_j - \frac{\gamma}{2}\sigma_j^2)} + \varepsilon_{ij}$
 - $EU_{i0} = \beta + \varepsilon_{i0}$ for the outside good.

Model of Restaurant Demand

- Consumers choose the restaurant (or outside good) that maximizes expected utility.
 - Risk averse consumer i knows restaurant j will provide level of consumption (net of prices) drawn from $N(\mu_j, \sigma_j^2)$.
 - $EU_{ij} = \int_c -e^{-\gamma c} f(c|\mu_j, \sigma_j^2) + \varepsilon_{ij} = -e^{-\gamma(\mu_j - \frac{\gamma}{2}\sigma_j^2)} + \varepsilon_{ij}$
 - $EU_{i0} = \beta + \varepsilon_{i0}$ for the outside good.
- Distribution assumptions?

Model of Restaurant Demand

- Consumers choose the restaurant (or outside good) that maximizes expected utility.
 - Risk averse consumer i knows restaurant j will provide level of consumption (net of prices) drawn from $N(\mu_j, \sigma_j^2)$.
 - $EU_{ij} = \int_c -e^{-\gamma c} f(c|\mu_j, \sigma_j^2) + \varepsilon_{ij} = -e^{-\gamma(\mu_j - \frac{\gamma}{2}\sigma_j^2)} + \varepsilon_{ij}$
 - $EU_{i0} = \beta + \varepsilon_{i0}$ for the outside good.
- Distribution assumptions?
 - $\gamma\mu \sim N(0, \alpha_1)$

Model of Restaurant Demand

- Consumers choose the restaurant (or outside good) that maximizes expected utility.
 - Risk averse consumer i knows restaurant j will provide level of consumption (net of prices) drawn from $N(\mu_j, \sigma_j^2)$.
 - $EU_{ij} = \int_c -e^{-\gamma c} f(c|\mu_j, \sigma_j^2) + \varepsilon_{ij} = -e^{-\gamma(\mu_j - \frac{\gamma}{2}\sigma_j^2)} + \varepsilon_{ij}$
 - $EU_{i0} = \beta + \varepsilon_{i0}$ for the outside good.
- Distribution assumptions?
 - $\gamma\mu \sim N(0, \alpha_1)$
 - $\gamma^2\sigma^2 \sim U[0, \alpha_2]$

Model of Restaurant Demand

- Consumers choose the restaurant (or outside good) that maximizes expected utility.
 - Risk averse consumer i knows restaurant j will provide level of consumption (net of prices) drawn from $N(\mu_j, \sigma_j^2)$.
 - $EU_{ij} = \int_c -e^{-\gamma c} f(c|\mu_j, \sigma_j^2) + \varepsilon_{ij} = -e^{-\gamma(\mu_j - \frac{\gamma}{2}\sigma_j^2)} + \varepsilon_{ij}$
 - $EU_{i0} = \beta + \varepsilon_{i0}$ for the outside good.
- Distribution assumptions?
 - $\gamma\mu \sim N(0, \alpha_1)$
 - $\gamma^2\sigma^2 \sim U[0, \alpha_2]$
 - $\varepsilon \sim \text{Type1EV}$.

Model of Restaurant Demand

- Consumers choose the restaurant (or outside good) that maximizes expected utility.
 - Risk averse consumer i knows restaurant j will provide level of consumption (net of prices) drawn from $N(\mu_j, \sigma_j^2)$.
 - $EU_{ij} = \int_c -e^{-\gamma c} f(c|\mu_j, \sigma_j^2) + \varepsilon_{ij} = -e^{-\gamma(\mu_j - \frac{\gamma}{2}\sigma_j^2)} + \varepsilon_{ij}$
 - $EU_{i0} = \beta + \varepsilon_{i0}$ for the outside good.
- Distribution assumptions?
 - $\gamma\mu \sim N(0, \alpha_1)$
 - $\gamma^2\sigma^2 \sim U[0, \alpha_2]$
 - $\varepsilon \sim \text{Type1EV}$.
- Parameters to estimate: $(\beta, \alpha_1, \alpha_2)$

Model of Restaurant Demand

- Consumers choose the restaurant (or outside good) that maximizes expected utility.
 - Risk averse consumer i knows restaurant j will provide level of consumption (net of prices) drawn from $N(\mu_j, \sigma_j^2)$.
 - $EU_{ij} = \int_c -e^{-\gamma c} f(c|\mu_j, \sigma_j^2) + \varepsilon_{ij} = -e^{-\gamma(\mu_j - \frac{\gamma}{2}\sigma_j^2)} + \varepsilon_{ij}$
 - $EU_{i0} = \beta + \varepsilon_{i0}$ for the outside good.
- Distribution assumptions?
 - $\gamma\mu \sim N(0, \alpha_1)$
 - $\gamma^2\sigma^2 \sim U[0, \alpha_2]$
 - $\varepsilon \sim \text{Type1EV}$.
- Parameters to estimate: $(\beta, \alpha_1, \alpha_2)$
 - Note: Not separately IDing γ from α .

Model of Restaurant Demand

- Consumers choose the restaurant (or outside good) that maximizes expected utility.
 - Risk averse consumer i knows restaurant j will provide level of consumption (net of prices) drawn from $N(\mu_j, \sigma_j^2)$.
 - $EU_{ij} = \int_c -e^{-\gamma c} f(c|\mu_j, \sigma_j^2) + \varepsilon_{ij} = -e^{-\gamma(\mu_j - \frac{\gamma}{2}\sigma_j^2)} + \varepsilon_{ij}$
 - $EU_{i0} = \beta + \varepsilon_{i0}$ for the outside good.
- Distribution assumptions?
 - $\gamma\mu \sim N(0, \alpha_1)$
 - $\gamma^2\sigma^2 \sim U[0, \alpha_2]$
 - $\varepsilon \sim \text{Type1EV}$.
- Parameters to estimate: $(\beta, \alpha_1, \alpha_2)$
 - Note: Not separately IDing γ from α .
- Feasible Model/Extensions:

Model of Restaurant Demand

- Consumers choose the restaurant (or outside good) that maximizes expected utility.
 - Risk averse consumer i knows restaurant j will provide level of consumption (net of prices) drawn from $N(\mu_j, \sigma_j^2)$.
 - $EU_{ij} = \int_c -e^{-\gamma c} f(c|\mu_j, \sigma_j^2) + \varepsilon_{ij} = -e^{-\gamma(\mu_j - \frac{\gamma}{2}\sigma_j^2)} + \varepsilon_{ij}$
 - $EU_{i0} = \beta + \varepsilon_{i0}$ for the outside good.
- Distribution assumptions?
 - $\gamma\mu \sim N(0, \alpha_1)$
 - $\gamma^2\sigma^2 \sim U[0, \alpha_2]$
 - $\varepsilon \sim \text{Type1EV}$.
- Parameters to estimate: $(\beta, \alpha_1, \alpha_2)$
 - Note: Not separately IDing γ from α .
- Feasible Model/Extensions:
 - Allow μ_j and σ_j^2 to be i specific and depend on j 's observable and unobservable characteristics.

Model of Restaurant Demand

- Consumers choose the restaurant (or outside good) that maximizes expected utility.
 - Risk averse consumer i knows restaurant j will provide level of consumption (net of prices) drawn from $N(\mu_j, \sigma_j^2)$.
 - $EU_{ij} = \int_c -e^{-\gamma c} f(c|\mu_j, \sigma_j^2) + \varepsilon_{ij} = -e^{-\gamma(\mu_j - \frac{\gamma}{2}\sigma_j^2)} + \varepsilon_{ij}$
 - $EU_{i0} = \beta + \varepsilon_{i0}$ for the outside good.
- Distribution assumptions?
 - $\gamma\mu \sim N(0, \alpha_1)$
 - $\gamma^2\sigma^2 \sim U[0, \alpha_2]$
 - $\varepsilon \sim \text{Type1EV}$.
- Parameters to estimate: $(\beta, \alpha_1, \alpha_2)$
 - Note: Not separately IDing γ from α .
- Feasible Model/Extensions:
 - Allow μ_j and σ_j^2 to be i specific and depend on j 's observable and unobservable characteristics.
 - Luca's JMP strategy. Identify causal effect of Yelp on μ_j and σ_j^2

Model of Review Decision

- After visiting restaurant j , i realizes a level of consumption and utility (RU):

Model of Review Decision

- After visiting restaurant j , i realizes a level of consumption and utility (RU):
 - $RU_{ij} = -e^{-\gamma c} + \varepsilon_{ij}$ where $c \sim N(\mu_j, \sigma_j^2)$

Model of Review Decision

- After visiting restaurant j , i realizes a level of consumption and utility (RU):
 - $RU_{ij} = -e^{-\gamma c} + \varepsilon_{ij}$ where $c \sim N(\mu_j, \sigma_j^2)$
 - Or, $RU_{ij} = EU_{ij} + \omega_{ij}$ where ω drawn from a distribution that depends on (μ_j, σ_j^2)

Model of Review Decision

- After visiting restaurant j , i realizes a level of consumption and utility (RU):
 - $RU_{ij} = -e^{-\gamma c} + \varepsilon_{ij}$ where $c \sim N(\mu_j, \sigma_j^2)$
 - Or, $RU_{ij} = EU_{ij} + \omega_{ij}$ where ω drawn from a distribution that depends on (μ_j, σ_j^2)
- Realized utility (RU_{ij}) flexibly enters review decision:

$$\begin{aligned}u_{i,Review} &= f(RU_{ij}) + v_i \\ &= \phi_0 + \phi_1 RU_{ij} + \phi_2 RU_{ij}^2 + v_i\end{aligned}$$

Model of Review Decision

- After visiting restaurant j , i realizes a level of consumption and utility (RU):
 - $RU_{ij} = -e^{-\gamma c} + \varepsilon_{ij}$ where $c \sim N(\mu_j, \sigma_j^2)$
 - Or, $RU_{ij} = EU_{ij} + \omega_{ij}$ where ω drawn from a distribution that depends on (μ_j, σ_j^2)
- Realized utility (RU_{ij}) flexibly enters review decision:

$$\begin{aligned}u_{i,Review} &= f(RU_{ij}) + v_i \\ &= \phi_0 + \phi_1 RU_{ij} + \phi_2 RU_{ij}^2 + v_i\end{aligned}$$

- Submit a review if $u_{i,Review} > 0$ where $v \sim Type1EV$

Model of Review Decision

- After visiting restaurant j , i realizes a level of consumption and utility (RU):
 - $RU_{ij} = -e^{-\gamma c} + \varepsilon_{ij}$ where $c \sim N(\mu_j, \sigma_j^2)$
 - Or, $RU_{ij} = EU_{ij} + \omega_{ij}$ where ω drawn from a distribution that depends on (μ_j, σ_j^2)
- Realized utility (RU_{ij}) flexibly enters review decision:

$$\begin{aligned}u_{i,Review} &= f(RU_{ij}) + v_i \\ &= \phi_0 + \phi_1 RU_{ij} + \phi_2 RU_{ij}^2 + v_i\end{aligned}$$

- Submit a review if $u_{i,Review} > 0$ where $v \sim Type1EV$
- Parameters to estimate: (ϕ_0, ϕ_1, ϕ_2)

Model of Review Decision

- After visiting restaurant j , i realizes a level of consumption and utility (RU):
 - $RU_{ij} = -e^{-\gamma c} + \varepsilon_{ij}$ where $c \sim N(\mu_j, \sigma_j^2)$
 - Or, $RU_{ij} = EU_{ij} + \omega_{ij}$ where ω drawn from a distribution that depends on (μ_j, σ_j^2)
- Realized utility (RU_{ij}) flexibly enters review decision:

$$\begin{aligned}u_{i,Review} &= f(RU_{ij}) + v_i \\ &= \phi_0 + \phi_1 RU_{ij} + \phi_2 RU_{ij}^2 + v_i\end{aligned}$$

- Submit a review if $u_{i,Review} > 0$ where $v \sim Type1EV$
- Parameters to estimate: (ϕ_0, ϕ_1, ϕ_2)
- Feasible Model: Allow (ϕ_0, ϕ_1, ϕ_2) to depend on i 's and j 's characteristics.

Model of Review Content Decision

- If consumer reviews j , decision on how many stars to give depends on realized utility:

$$1 \text{ star} \Leftrightarrow RU_{ij} < \tau_1$$

$$2 \text{ star} \Leftrightarrow \tau_1 \leq RU_{ij} < \tau_2$$

$$3 \text{ star} \Leftrightarrow \tau_2 \leq RU_{ij} < \tau_3$$

$$4 \text{ star} \Leftrightarrow \tau_3 \leq RU_{ij} < \tau_4$$

$$5 \text{ star} \Leftrightarrow \tau_4 \leq RU_{ij}$$

Model of Review Content Decision

- If consumer reviews j , decision on how many stars to give depends on realized utility:

$$1 \text{ star} \Leftrightarrow RU_{ij} < \tau_1$$

$$2 \text{ star} \Leftrightarrow \tau_1 \leq RU_{ij} < \tau_2$$

$$3 \text{ star} \Leftrightarrow \tau_2 \leq RU_{ij} < \tau_3$$

$$4 \text{ star} \Leftrightarrow \tau_3 \leq RU_{ij} < \tau_4$$

$$5 \text{ star} \Leftrightarrow \tau_4 \leq RU_{ij}$$

- Parameters to estimate: $(\tau_0, \tau_1, \tau_2, \tau_3)$

Model of Review Content Decision

- If consumer reviews j , decision on how many stars to give depends on realized utility:

$$1 \text{ star} \Leftrightarrow RU_{ij} < \tau_1$$

$$2 \text{ star} \Leftrightarrow \tau_1 \leq RU_{ij} < \tau_2$$

$$3 \text{ star} \Leftrightarrow \tau_2 \leq RU_{ij} < \tau_3$$

$$4 \text{ star} \Leftrightarrow \tau_3 \leq RU_{ij} < \tau_4$$

$$5 \text{ star} \Leftrightarrow \tau_4 \leq RU_{ij}$$

- Parameters to estimate: $(\tau_0, \tau_1, \tau_2, \tau_3)$
- Feasible Model: Allow $(\tau_0, \tau_1, \tau_2, \tau_3)$ to depend on i 's and j 's characteristics.

Estimation

- Estimate the model via non-linear least squares with nested fixed point algorithms.

Estimation

- Estimate the model via non-linear least squares with nested fixed point algorithms.
- n_{js} equal # of "s" star reviews received by j . $r(j)$ is j 's "rank" in n (if $r(j) = 3$, only two restaurants received more reviews than j).

Estimation

- Estimate the model via non-linear least squares with nested fixed point algorithms.
- n_{js} equal # of "s" star reviews received by j . $r(j)$ is j 's "rank" in n (if $r(j) = 3$, only two restaurants received more reviews than j).
- $E(n_{r(j)s} | \Theta)$ equals expected number of "s" star reviews j receives given model and $\Theta \equiv (\alpha, \beta, \phi, \tau)$

Estimation

- Estimate the model via non-linear least squares with nested fixed point algorithms.
- n_{js} equal # of "s" star reviews received by j . $r(j)$ is j 's "rank" in n (if $r(j) = 3$, only two restaurants received more reviews than j).
- $E(n_{r(j)s} | \Theta)$ equals expected number of "s" star reviews j receives given model and $\Theta \equiv (\alpha, \beta, \phi, \tau)$
- Estimating equation is:

$$n_{js} = E(n_{r(j)s} | \Theta) + v_{js}(n_{js}, \Theta)$$

Estimation

- Estimate the model via non-linear least squares with nested fixed point algorithms.
- n_{js} equal # of "s" star reviews received by j . $r(j)$ is j 's "rank" in n (if $r(j) = 3$, only two restaurants received more reviews than j).
- $E(n_{r(j)s} | \Theta)$ equals expected number of "s" star reviews j receives given model and $\Theta \equiv (\alpha, \beta, \phi, \tau)$
- Estimating equation is:

$$n_{js} = E(n_{r(j)s} | \Theta) + v_{js}(n_{js}, \Theta)$$

- $v_{js}(n_{js}, \Theta)$ includes:

Estimation

- Estimate the model via non-linear least squares with nested fixed point algorithms.
- n_{js} equal # of "s" star reviews received by j . $r(j)$ is j 's "rank" in n (if $r(j) = 3$, only two restaurants received more reviews than j).
- $E(n_{r(j)s}|\Theta)$ equals expected number of "s" star reviews j receives given model and $\Theta \equiv (\alpha, \beta, \phi, \tau)$
- Estimating equation is:

$$n_{js} = E(n_{r(j)s}|\Theta) + v_{js}(n_{js}, \Theta)$$

- $v_{js}(n_{js}, \Theta)$ includes:
 - Prediction error. $E(n_{r(j)s}|\Theta)$ integrates over distributions of restaurant/consumer characteristics. n_{js} determined by their true realizations.

Estimation

- Estimate the model via non-linear least squares with nested fixed point algorithms.
- n_{js} equal # of "s" star reviews received by j . $r(j)$ is j 's "rank" in n (if $r(j) = 3$, only two restaurants received more reviews than j).
- $E(n_{r(j)s}|\Theta)$ equals expected number of "s" star reviews j receives given model and $\Theta \equiv (\alpha, \beta, \phi, \tau)$
- Estimating equation is:

$$n_{js} = E(n_{r(j)s}|\Theta) + v_{js}(n_{js}, \Theta)$$

- $v_{js}(n_{js}, \Theta)$ includes:
 - Prediction error. $E(n_{r(j)s}|\Theta)$ integrates over distributions of restaurant/consumer characteristics. n_{js} determined by their true realizations.
 - Simulation error. Integrals in $E(n_{r(j)s}|\Theta)$ are simulated.

Estimation

- Estimate the model via non-linear least squares with nested fixed point algorithms.
- n_{js} equal # of "s" star reviews received by j . $r(j)$ is j 's "rank" in n (if $r(j) = 3$, only two restaurants received more reviews than j).
- $E(n_{r(j)s} | \Theta)$ equals expected number of "s" star reviews j receives given model and $\Theta \equiv (\alpha, \beta, \phi, \tau)$
- Estimating equation is:

$$n_{js} = E(n_{r(j)s} | \Theta) + v_{js}(n_{js}, \Theta)$$

- $v_{js}(n_{js}, \Theta)$ includes:
 - Prediction error. $E(n_{r(j)s} | \Theta)$ integrates over distributions of restaurant/consumer characteristics. n_{js} determined by their true realizations.
 - Simulation error. Integrals in $E(n_{r(j)s} | \Theta)$ are simulated.
 - By construction, at Θ^* (the truth), $E(v_{js}(n_{js}, \Theta) | \text{Data}, \Theta^*) = 0$ (Hansen 1982, Pakes & Pollard 1989, Berry 1992).

Estimation

- Estimate the model via non-linear least squares with nested fixed point algorithms.
- n_{js} equal # of "s" star reviews received by j . $r(j)$ is j 's "rank" in n (if $r(j) = 3$, only two restaurants received more reviews than j).
- $E(n_{r(j)s}|\Theta)$ equals expected number of "s" star reviews j receives given model and $\Theta \equiv (\alpha, \beta, \phi, \tau)$
- Estimating equation is:

$$n_{js} = E(n_{r(j)s}|\Theta) + v_{js}(n_{js}, \Theta)$$

- $v_{js}(n_{js}, \Theta)$ includes:
 - Prediction error. $E(n_{r(j)s}|\Theta)$ integrates over distributions of restaurant/consumer characteristics. n_{js} determined by their true realizations.
 - Simulation error. Integrals in $E(n_{r(j)s}|\Theta)$ are simulated.
 - By construction, at Θ^* (the truth), $E(v_{js}(n_{js}, \Theta)|Data, \Theta^*) = 0$ (Hansen 1982, Pakes & Pollard 1989, Berry 1992).
- Choose Θ to minimize $\sum_j \sum_s (n_{js} - E(n_{r(j)s}|\Theta))^2$

Estimation Intuition

- Given Θ , $E(n_{r(j)_s}|\Theta)$ equals the probability that one consumer chooses j , realizes utility $RU \in [\tau_{s-1}, \tau_s]$, and choose to submit a review, multiplied by #Households:

$$E(n_{r(j)_s}|\Theta) = P(\text{InsideGood}|\Theta) * P(\text{choose } r(j)|IG, \Theta) * P(\text{Review \& } RU \in [\tau_{s-1}, \tau_s] | r(j), \Theta) * \#HH$$

Estimation Intuition

- Given Θ , $E(n_{r(j)_s}|\Theta)$ equals the probability that one consumer chooses j , realizes utility $RU \in [\tau_{s-1}, \tau_s]$, and choose to submit a review, multiplied by #Households:

$$E(n_{r(j)_s}|\Theta) = P(\text{InsideGood}|\Theta) * P(\text{choose } r(j)|IG, \Theta) * P(\text{Review \& } RU \in [\tau_{s-1}, \tau_s] | r(j), \Theta) * \#HH$$

- Model and assumptions imply:

Estimation Intuition

- Given Θ , $E(n_{r(j)_s}|\Theta)$ equals the probability that one consumer chooses j , realizes utility $RU \in [\tau_{s-1}, \tau_s]$, and choose to submit a review, multiplied by #Households:

$$E(n_{r(j)_s}|\Theta) = P(\text{InsideGood}|\Theta) * P(\text{choose } r(j)|IG, \Theta) * P(\text{Review \& } RU \in [\tau_{s-1}, \tau_s] | r(j), \Theta) * \#HH$$

- Model and assumptions imply:
 - $P(\text{InsideGood}|\Theta)$ depends on (β, μ, σ^2)

Estimation Intuition

- Given Θ , $E(n_{r(j)_s}|\Theta)$ equals the probability that one consumer chooses j , realizes utility $RU \in [\tau_{s-1}, \tau_s]$, and choose to submit a review, multiplied by #Households:

$$E(n_{r(j)_s}|\Theta) = P(\text{InsideGood}|\Theta) * P(\text{choose } r(j)|IG, \Theta) * P(\text{Review \& } RU \in [\tau_{s-1}, \tau_s] | r(j), \Theta) * \#HH$$

- Model and assumptions imply:
 - $P(\text{InsideGood}|\Theta)$ depends on (β, μ, σ^2)
 - $P(\text{choose } r(j)|IG, \Theta)$ depends on (μ, σ^2)

Estimation Intuition

- Given Θ , $E(n_{r(j)_s}|\Theta)$ equals the probability that one consumer chooses j , realizes utility $RU \in [\tau_{s-1}, \tau_s]$, and choose to submit a review, multiplied by #Households:

$$E(n_{r(j)_s}|\Theta) = P(\text{InsideGood}|\Theta) * P(\text{choose } r(j)|IG, \Theta) * P(\text{Review \& } RU \in [\tau_{s-1}, \tau_s] | r(j), \Theta) * \#HH$$

- Model and assumptions imply:

- 1 $P(\text{InsideGood}|\Theta)$ depends on (β, μ, σ^2)
- 2 $P(\text{choose } r(j)|IG, \Theta)$ depends on (μ, σ^2)
- 3 $P(\text{Review \& } RU \in [\tau_{s-1}, \tau_s] | r(j), \Theta)$ only depends on $(\phi, \mu_{r(j)}, \sigma_{r(j)}^2, \tau)$

Estimation Intuition

- Identify β using information on market share of inside good:

Estimation Intuition

- Identify β using information on market share of inside good:
 - Given $(\mu, \sigma^2, \phi, \tau)$, a unique β satisfies:

$$\frac{1}{7} = \sum_j P(\text{InsideGood} | \beta, \mu, \sigma^2)$$

Estimation Intuition

- Identify β using information on market share of inside good:
 - Given $(\mu, \sigma^2, \phi, \tau)$, a unique β satisfies:

$$\frac{1}{7} = \sum_j P(\text{InsideGood} | \beta, \mu, \sigma^2)$$

- Identify τ using variation in total # of reviews "across s":

Estimation Intuition

- Identify β using information on market share of inside good:
 - Given $(\mu, \sigma^2, \phi, \tau)$, a unique β satisfies:

$$\frac{1}{7} = \sum_j P(\text{InsideGood} | \beta, \mu, \sigma^2)$$

- Identify τ using variation in total # of reviews "across s ":
 - Given $\mu, \sigma^2, \beta, \phi$, there exists a unique τ such that:

$$\sum_j n_{js} = \sum_j E(n_{r(j)s} | \Theta) \text{ for all } s$$

Estimation Intuition

- Identify β using information on market share of inside good:
 - Given $(\mu, \sigma^2, \phi, \tau)$, a unique β satisfies:

$$\frac{1}{7} = \sum_j P(\text{InsideGood} | \beta, \mu, \sigma^2)$$

- Identify τ using variation in total # of reviews "across s ":
 - Given $\mu, \sigma^2, \beta, \phi$, there exists a unique τ such that:

$$\sum_j n_{js} = \sum_j E(n_{r(j)s} | \Theta) \text{ for all } s$$

- Identify ϕ_1 using total number of reviews. Given $\mu, \sigma^2, \beta, \phi_2, \phi_3$ a unique ϕ_1 satisfies:

$$\sum_{js} n_{js} = \sum_{js} E(n_{r(j)s} | \Theta)$$

Estimation Intuition

- Remaining variation is "across $r(j)$ " and "within $r(j)$ "

Estimation Intuition

- Remaining variation is "across $r(j)$ " and "within $r(j)$ "
- Identify distribution of μ, σ^2 (i.e. α) using "across $r(j)$ " variation in reviews.

Estimation Intuition

- Remaining variation is "across $r(j)$ " and "within $r(j)$ "
- Identify distribution of μ, σ^2 (i.e. α) using "across $r(j)$ " variation in reviews.
 - Choice of α implies different distributions of (μ, σ^2) . Affects variation in $P(\text{choose } r(j) | IG, \Theta)$ across $r(j)$.

Estimation Intuition

- Remaining variation is "across $r(j)$ " and "within $r(j)$ "
- Identify distribution of μ, σ^2 (i.e. α) using "across $r(j)$ " variation in reviews.
 - Choice of α implies different distributions of (μ, σ^2) . Affects variation in $P(\text{choose } r(j) | IG, \Theta)$ across $r(j)$.
 - More heterogeneity in $P(\text{choose } r(j) | IG, \Theta)$ across $r(j) \Rightarrow$ more heterogeneity in predicted n across $r(j)$.

Estimation Intuition

- Remaining variation is "across $r(j)$ " and "within $r(j)$ "
- Identify distribution of μ, σ^2 (i.e. α) using "across $r(j)$ " variation in reviews.
 - Choice of α implies different distributions of (μ, σ^2) . Affects variation in $P(\text{choose } r(j) | IG, \Theta)$ across $r(j)$.
 - More heterogeneity in $P(\text{choose } r(j) | IG, \Theta)$ across $r(j) \Rightarrow$ more heterogeneity in predicted n across $r(j)$.
 - Choose α so predicted distribution of n matches observed distribution.

Estimation Intuition

- Remaining variation is "across $r(j)$ " and "within $r(j)$ "
- Identify distribution of μ, σ^2 (i.e. α) using "across $r(j)$ " variation in reviews.
 - Choice of α implies different distributions of (μ, σ^2) . Affects variation in $P(\text{choose } r(j) | IG, \Theta)$ across $r(j)$.
 - More heterogeneity in $P(\text{choose } r(j) | IG, \Theta)$ across $r(j) \Rightarrow$ more heterogeneity in predicted n across $r(j)$.
 - Choose α so predicted distribution of n matches observed distribution.
- Identify (ϕ_2, ϕ_3) by predicting "within $r(j)$ " review distributions.

Estimation Intuition

- Remaining variation is "across $r(j)$ " and "within $r(j)$ "
- Identify distribution of μ, σ^2 (i.e. α) using "across $r(j)$ " variation in reviews.
 - Choice of α implies different distributions of (μ, σ^2) . Affects variation in $P(\text{choose } r(j) | IG, \Theta)$ across $r(j)$.
 - More heterogeneity in $P(\text{choose } r(j) | IG, \Theta)$ across $r(j) \Rightarrow$ more heterogeneity in predicted n across $r(j)$.
 - Choose α so predicted distribution of n matches observed distribution.
- Identify (ϕ_2, ϕ_3) by predicting "within $r(j)$ " review distributions.
 - Given $(\beta, \mu, \sigma^2, \tau)$, distribution of reviews "within $r(j)$ " determined by:

$$P(\text{Review} \& RU \in [\tau_{s-1}, \tau_s] | \phi, \mu_{r(j)}, \sigma_{r(j)}^2, \tau) = \int_{\tau_{s-1}}^{\tau_s} P(\text{review} | u, \phi) P(u | \mu_{r(j)}, \sigma_{r(j)}^2) du$$

Estimation Intuition

- Remaining variation is "across $r(j)$ " and "within $r(j)$ "
- Identify distribution of μ, σ^2 (i.e. α) using "across $r(j)$ " variation in reviews.
 - Choice of α implies different distributions of (μ, σ^2) . Affects variation in $P(\text{choose } r(j) | IG, \Theta)$ across $r(j)$.
 - More heterogeneity in $P(\text{choose } r(j) | IG, \Theta)$ across $r(j) \Rightarrow$ more heterogeneity in predicted n across $r(j)$.
 - Choose α so predicted distribution of n matches observed distribution.
- Identify (ϕ_2, ϕ_3) by predicting "within $r(j)$ " review distributions.
 - Given $(\beta, \mu, \sigma^2, \tau)$, distribution of reviews "within $r(j)$ " determined by:

$$P(\text{Review} \& RU \in [\tau_{s-1}, \tau_s] | \phi, \mu_{r(j)}, \sigma_{r(j)}^2, \tau) = \int_{\tau_{s-1}}^{\tau_s} P(\text{review} | u, \phi) P(u | \mu_{r(j)}, \sigma_{r(j)}^2) du$$

- Much variation in $\int_{\tau_{s-1}}^{\tau_s} P(u | \mu_{r(j)}, \sigma_{r(j)}^2) du$ across $r(j)$ s. Choose (ϕ_2, ϕ_3) to match "within $r(j)$ " review distributions

Parameter Estimates

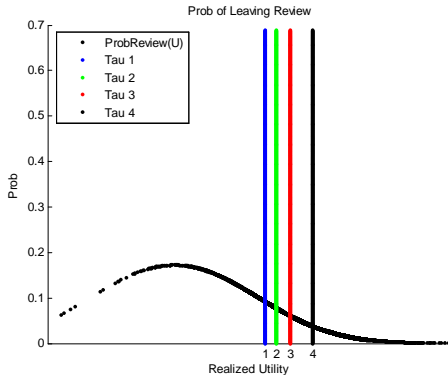
- Evidence suggests consumers less likely to submit reviews when they realize high utility

	Parameter Estimates		
	<u>N=100</u>	<u>N=150</u>	<u>N=200</u>
Var_Mu	0.01	0.019	0.015
Var_Sigma	0.103	0.071	0.0748
Beta_OG	2.637	2.3475	2.182
Tau_1	0.221	0.2989	0.3731
Tau_2	0.3257	0.396	0.4708
Tau_3	0.4458	0.515	0.5884
Tau_4	0.6481	0.7121	0.7869
Phi_1	-1.7395	-2.2548	-2.0675
Phi_2	-1.617	-1.134	-1.3344
Phi_3	-1.436	-1.334	-0.8957

Parameter Estimates

Review Probability if Utility Warrants "S" Star Review

	<u>1 Star</u>	<u>2 Star</u>	<u>3 Star</u>	<u>4 Star</u>	<u>5 Star</u>
Probability	0.118	0.082	0.067	0.047	0.021



Model Fit

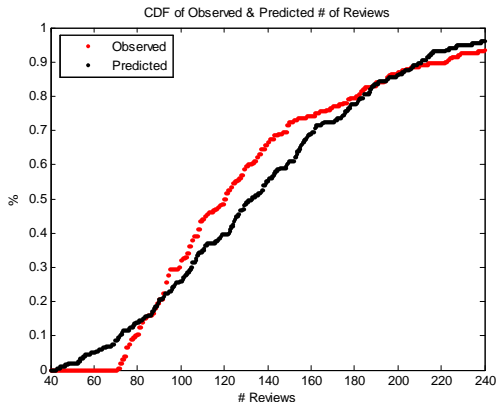
- Model should be able to explain different moments of distribution of $(n, n_1, n_2, \dots, n_5)$:

Model Fit

- Model should be able to explain different moments of distribution of $(n, n_1, n_2, \dots, n_5)$:
 - Almost perfectly predicts total numbers of reviews.

Model Fit

- Model should be able to explain different moments of distribution of $(n, n_1, n_2, \dots, n_5)$:
 - Almost perfectly predicts total numbers of reviews.
 - Overpredicts the variance of n a little.



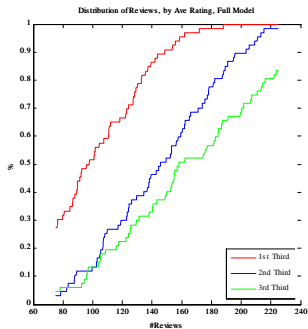
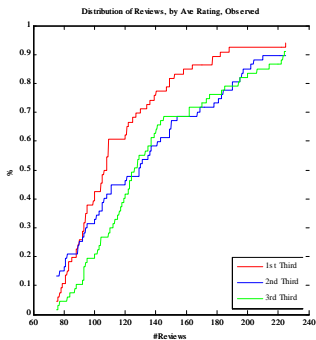
Model Fit

- Model should be able to explain distribution of (n_1, n_2, \dots, n_5) conditional on n .

Distribution of Reviews: Restaurants with different Total N					
	Observed				
	<u>% 1 Star</u>	<u>% 2 Star</u>	<u>% 3 Star</u>	<u>% 4 Star</u>	<u>% 5 Star</u>
Quartile 1	0.054	0.1065	0.2224	0.4018	0.2152
Quartile 2	0.0498	0.0888	0.2047	0.4092	0.2475
Quartile 3	0.0381	0.0808	0.2059	0.415	0.2601
Quartile 4	0.0413	0.0903	0.1893	0.389	0.2901
	Predicted - Full Model				
	<u>% 1 Star</u>	<u>% 2 Star</u>	<u>% 3 Star</u>	<u>% 4 Star</u>	<u>% 5 Star</u>
Quartile 1	0.0538	0.1087	0.2421	0.3904	0.205
Quartile 2	0.0422	0.1008	0.2123	0.407	0.2377
Quartile 3	0.051	0.0906	0.1957	0.4023	0.2604
Quartile 4	0.0393	0.0847	0.1906	0.3965	0.289

Model Fit

- Model should be able to explain distribution of n conditional on average review received.



Estimated Review Behavior and Model Fit

- Estimated model explains aggregate distribution of (n_1, n_2, \dots, n_5) and distribution of (n_1, n_2, \dots, n_5) conditional on n . Can models without $Prob(\text{Review}|u)$ do as well?

Estimated Review Behavior and Model Fit

- Estimated model explains aggregate distribution of (n_1, n_2, \dots, n_5) and distribution of (n_1, n_2, \dots, n_5) conditional on n . Can models without $Prob(\text{Review}|u)$ do as well?
- First Model:

Estimated Review Behavior and Model Fit

- Estimated model explains aggregate distribution of (n_1, n_2, \dots, n_5) and distribution of (n_1, n_2, \dots, n_5) conditional on n . Can models without $Prob(\text{Review}|u)$ do as well?
- First Model:
 - ① Take original model, set $\phi_2 = \phi_3 = 0$ (don't allow $Prob(\text{Review})$ to vary with utility).

Estimated Review Behavior and Model Fit

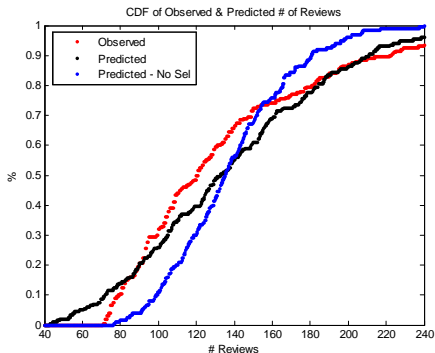
- Estimated model explains aggregate distribution of (n_1, n_2, \dots, n_5) and distribution of (n_1, n_2, \dots, n_5) conditional on n . Can models without $Prob(Review|u)$ do as well?
- First Model:
 - 1 Take original model, set $\phi_2 = \phi_3 = 0$ (don't allow $Prob(Review)$ to vary with utility).
 - 2 Look for parameters that minimize $\sum_j \sum_s (n_{js} - E(n_{r(j)s}|\Theta))^2$. (still asking model to explain joint distribution of (n_1, n_2, \dots, n_5)).

Estimated Review Behavior and Model Fit

- Estimated model explains aggregate distribution of (n_1, n_2, \dots, n_5) and distribution of (n_1, n_2, \dots, n_5) conditional on n . Can models without $Prob(\text{Review}|u)$ do as well?
- First Model:
 - 1 Take original model, set $\phi_2 = \phi_3 = 0$ (don't allow $Prob(\text{Review})$ to vary with utility).
 - 2 Look for parameters that minimize $\sum_j \sum_s (n_{js} - E(n_{r(j)s}|\Theta))^2$. (still asking model to explain joint distribution of (n_1, n_2, \dots, n_5)).
- The model can explain the variation in (n_1, n_2, \dots, n_5) across n , but not the variation in n .

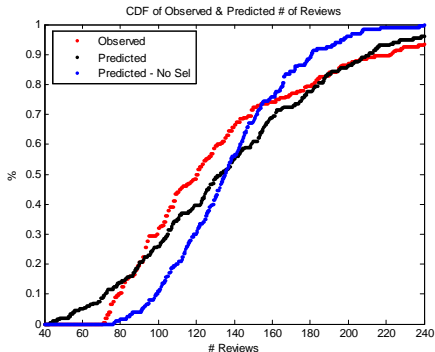
Distribution of Reviews: Restaurants with different Total N					
	Observed				
	% 1 Star	% 2 Star	% 3 Star	% 4 Star	% 5 Star
Quartile 1	0.054	0.1065	0.2224	0.4018	0.2152
Quartile 2	0.0498	0.0888	0.2047	0.4092	0.2475
Quartile 3	0.0381	0.0808	0.2059	0.415	0.2601
Quartile 4	0.0413	0.0903	0.1893	0.389	0.2901
	Predicted - No Selection, Model 1				
	% 1 Star	% 2 Star	% 3 Star	% 4 Star	% 5 Star
Quartile 1	0.0528	0.1061	0.2305	0.3981	0.2124
Quartile 2	0.0483	0.0953	0.2056	0.4052	0.2456
Quartile 3	0.0417	0.0887	0.2031	0.3953	0.2712
Quartile 4	0.0405	0.0802	0.1852	0.4005	0.2936

Estimated Review Behavior and Model Fit



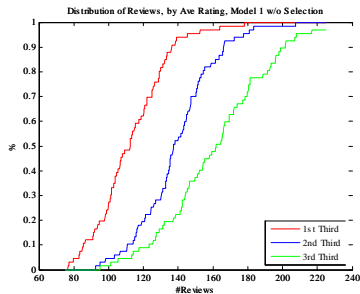
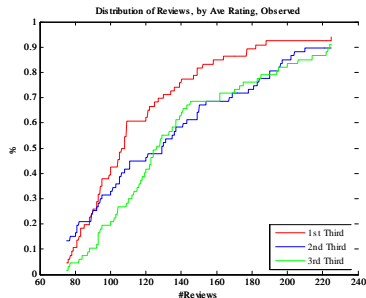
- $Var(n_{obs}) = 4.31 * 10^3$ $Var(n_{Full}) = 4.33 * 10^3$
 $Var(n_{NoSelection}) = 1.82 * 10^3$

Estimated Review Behavior and Model Fit



- $Var(n_{obs}) = 4.31 * 10^3$ $Var(n_{Full}) = 4.33 * 10^3$
 $Var(n_{NoSelection}) = 1.82 * 10^3$
- To generate variation in n , need to increase variation in EU . This leads to more heterogeneity in "within-firm" review distributions. Model can't explain this with $\phi_2 = \phi_3 = 0$.

Estimated Review Behavior and Model Fit



- With $\phi_2 = \phi_3 = 0$, model can't explain smaller gap between 2nd and 3rd thirds. Restaurants in 3rd third are higher quality. With $\phi_2 = \phi_3 = 0$, predicted to get many more reviews than restaurants in 2nd third.

Estimated Review Behavior and Model Fit

- Second model only attempts to explain variation in n :

Estimated Review Behavior and Model Fit

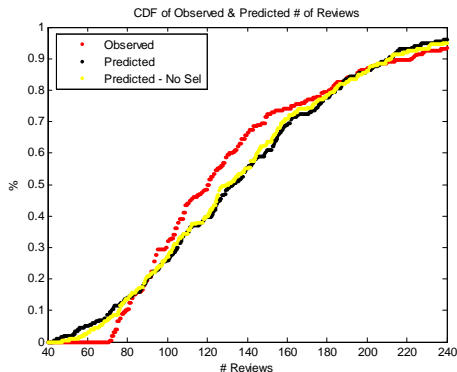
- Second model only attempts to explain variation in n :
 - 1 Take original model, set $\phi_2 = \phi_3 = 0$ (don't allow $Prob(Review)$ to vary with utility).

Estimated Review Behavior and Model Fit

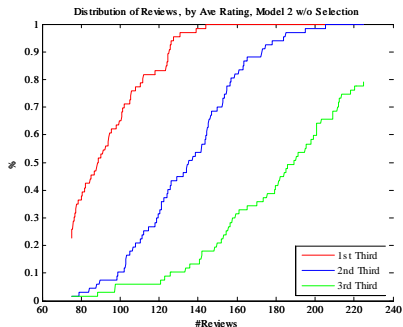
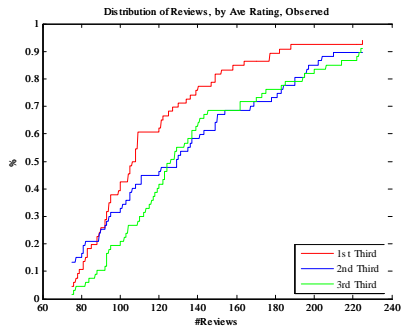
- Second model only attempts to explain variation in n :
 - 1 Take original model, set $\phi_2 = \phi_3 = 0$ (don't allow $Prob(Review)$ to vary with utility).
 - 2 Look for parameters that minimize $\sum_j \{ \sum_s n_{js} - \sum_s E(n_{r(j)s} | \Theta) \}^2$.

Estimated Review Behavior and Model Fit

- Second model only attempts to explain variation in n :
 - 1 Take original model, set $\phi_2 = \phi_3 = 0$ (don't allow $Prob(\text{Review})$ to vary with utility).
 - 2 Look for parameters that minimize $\sum_j \{ \sum_s n_{js} - \sum_s E(n_{r(j)s} | \Theta) \}^2$.
- Now, model can explain variation in n , but not the variation in (n_1, n_2, \dots, n_5) across n .



Estimated Review Behavior and Model Fit



- To generate observed variation in n , model needs much variation in EU . With $\phi_2 = \phi_3 = 0$, this extra variation implies highest quality restaurants get many more reviews than lowest quality.

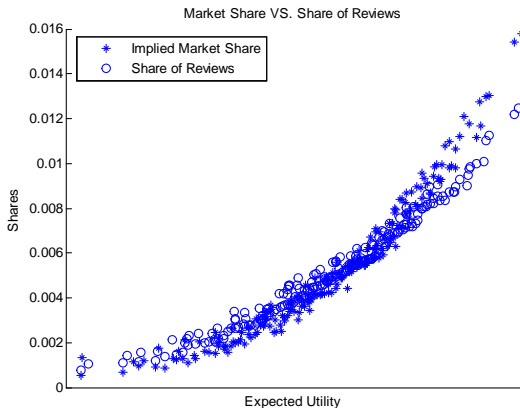
Estimated Review Behavior and Model Fit

Distribution of Reviews: Restaurants with different Total N					
	Observed				
	<u>% 1 Star</u>	<u>% 2 Star</u>	<u>% 3 Star</u>	<u>% 4 Star</u>	<u>% 5 Star</u>
Quartile 1	0.054	0.1065	0.2224	0.4018	0.2152
Quartile 2	0.0498	0.0888	0.2047	0.4092	0.2475
Quartile 3	0.0381	0.0808	0.2059	0.415	0.2601
Quartile 4	0.0413	0.0903	0.1893	0.389	0.2901
	Predicted - No Selection, Model 2				
	<u>% 1 Star</u>	<u>% 2 Star</u>	<u>% 3 Star</u>	<u>% 4 Star</u>	<u>% 5 Star</u>
Quartile 1	0.0701	0.1214	0.2471	0.3932	0.1683
Quartile 2	0.0555	0.1011	0.2166	0.4107	0.2161
Quartile 3	0.0392	0.0881	0.21	0.3958	0.2668
Quartile 4	0.0308	0.0747	0.1823	0.3994	0.3127

- Same Intuition: Higher quality restaurants tend to get more reviews. With $\phi_2 = \phi_3 = 0$, model predicts restaurants getting most n get reviews that are too good.

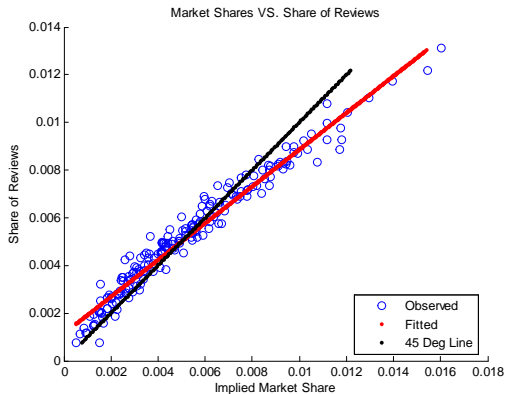
Implications

- High quality restaurants (those offering high expected utility) will be under represented in Yelp. Implies review shares can't be used as proxies for market shares.



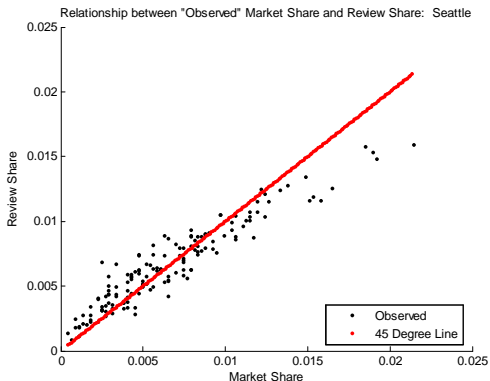
Implications

- Estimated model makes it possible to adjust for this selection effect..



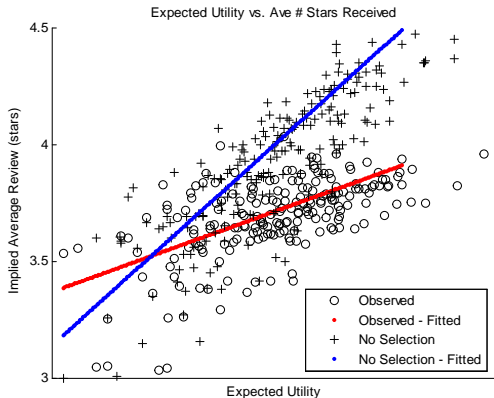
Implications

- Mike has some information on market shares for Seattle restaurants. Sees similar pattern in the data:



Implications

- If consumers less willing to review high quality restaurants, consumers have to do some extra work to infer quality from review information.



Implications

- Results suggest "clustering" in reviews partially explained by consumers' relative unwillingness to review high quality products.

Relationship b/w True Quality and Yelp Reviews

	<u>Ave Review Observed</u>	<u>Ave Review No Selection</u>
Mean (#Stars)	3.679	3.914
Var(#Stars)	0.03	0.086
Max(#Stars)-Min(#Stars)	0.963	1.437
Corr(EU,#Stars)	0.617	0.902

Note: # Stars = Average number of stars restaurant receives across reviews

Conclusion

- Project attempts to jointly estimate willingness to review products and their underlying demand.

Conclusion

- Project attempts to jointly estimate willingness to review products and their underlying demand.
 - Main idea: Can identify both by exploiting relationships between # and quality of reviews received

Conclusion

- Project attempts to jointly estimate willingness to review products and their underlying demand.
 - Main idea: Can identify both by exploiting relationships between # and quality of reviews received
- Goal is to use model to answer questions about local imperfect competition and online review markets.

Conclusion

- Project attempts to jointly estimate willingness to review products and their underlying demand.
 - Main idea: Can identify both by exploiting relationships between # and quality of reviews received
- Goal is to use model to answer questions about local imperfect competition and online review markets.
- Next step: Estimate more general model.