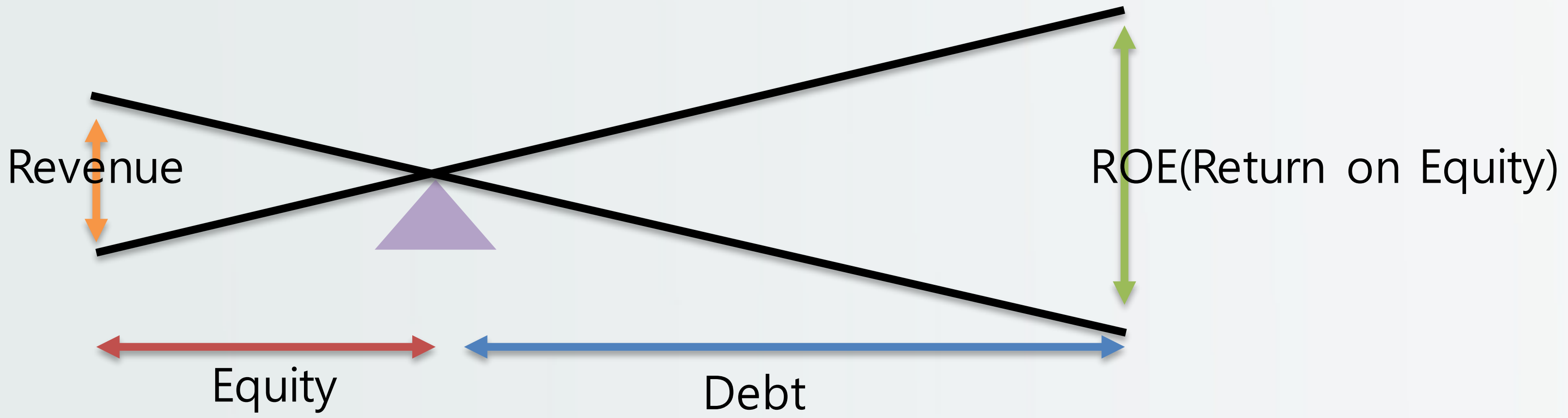
Research Proposal

Alice Mee Seon Chung

Background

➢ Company wants to pursue maximize their firm-value



➢ Company can maximize their firm-value with high-debt

➢ What factors affect on Capital Structure?

➢ The cost of Capital, Credit ratings, Growth rate, Tax exposure and so on

➢ Collateral value of asset

Research Question

What is the relationship between Asset liquidity and Capital structure in each business life cycle(i.e., star t-up, growth, mature, shake-out, decline)?

Literature Review

- Debates on impact of liquidity of a firm's assets on optimal leverage
- Asset liquidity increases optimal leverage(Shleifer and Vishny, 1992)
- The value of real assets influence on capital restructuring (Flor, 2007)

- Positive Relationship between Asset liquidity and leverage
- Asset liquidity increases optimal leverage (Sibilkov, 2009)
- Asset redeployability is important driver of leverage of firms (Campello and Giambona ,2013)

- Contribution
- Many researchers has studied this relationship by industries, by countries and by firm-sizes,
- But has not studied in each business life cycle stages!

Hypothesis

- At aggregate level, there is positive relationship between asset liquidity and leverage. Asset liquidity increases the amount of capital that firm can borrow and optimal leverage.

- Firms in early business life cycle stage have stronger positive relationship than firms in decline stage.

Data

- SEC – EDGAR (https://www.sec.gov/edgar)
- Company financial statements, annual reports data from 1984
- Offer transaction data and merger and acquisition data

- Others
- Morningstar, Bloomberg (Financial Statistics)

Computation & Methodology

- Computation

Liquidity Index

Using the two methods motivated

By Hernan O. and Gordon M.(2014)

1. The number of potential buyers for a fir m's asset minus the number of rival firms in the industry that have debt ratings.

2. The average book leverage net of cas h of rival firms in the industry

Business Life Cycle

Using Cash Flow patterns

(Dickinson, 2011)

The combination of a firm's net operating, investing, and financing cash flow and us ing the sign(positive or negative)

- Methodology

Multiple Linear Regression

Testing the relationship between liquidity index and capital structure

Testing with sub-categories of capital stru cture(i.e. leverage, secured debt, unsecu red debt)

References

Arikan, Asli M., and R. Stulz, "Corporate acquisitions, diversification, and the firm's lifecycle", Journal of Finance 71(2016), 139-194.

Campello, M., and E. Giambona. "Real Assets and Capital Structure." Journal of Financial and Quantitative Analysis, 48 (2013), 1333–1370.

Dickinson, V. "Cash Flow Patterns as a Proxy for Firm Life Cycle." The Accounting Review, 86(6)(2011), 1969-1994.

Flor, C. R. "Capital Structure and Real Assets: Effects of an Implicit Collateral to Debt Holders," Working paper(2002), Department of Accounting, Finance, and Law, University of Southern Denmark, Campusvej 55

Hernan,O., and Gordon M. "Real Asset Illiquidity and the Cost of Capital", Journal of Financial and Quantitative Analysis, 49 (2014), 1-32.

Myers, S., and R. Rajan. "The Paradox of Liquidity." Quarterly Journal of Economics, 113 (1998), 733-771.

Shleifer, A., and R. Vishny. "Liquidation Values and Debt Capacity: A Market Equilibrium Approach." Journal of Finance, 47 (1992), 1343

Sibilkov, V. "Asset Liquidity and Capital Structure." Journal of Financial and Quantitative Analysis, 44 (2009), 1173–1196.

Differential Human Capital Accumulation in Urban and Rural China

Luxi Han

Computational Social Science
University of Chicago

2014.04.05

Outline

1. Research Question
   - Research Question
   - Hypothesis
2. Literature and Background
3. Model Proposed
4. data

Outline

1. Research Question
   - Research Question
   - Hypothesis
2. Literature and Background
3. Model Proposed
4. data

Research Question

Question Defined

Research Question

What factors determine the human capital accumulation process in Urban and Rural China?

Outline

1. Research Question
   - Research Question
   - Hypothesis
2. Literature and Background
3. Model Proposed
4. data

Hypothesis

1. Credit constraint is the predoinant factor affecting human capital accumulation in rural China.
2. Intergenerational human capital transfer is the predominant factor affecting human capital accumulation in urban China.

Literature and Background

- China GINI coefficient is around 0.465; Urban rural income annual income ratio is around 2.72 : 1 (NBS, 2014)
- Evidence shows human capital is a significant contributor to this difference (Zhang et al. , 2007; Gao et al. , 2008; Guo, 2017).
- Research has been focusing on the effect of income on human capital accumulation(Zhang, 2006; Zhao, et al., 2010). But the effect of intergenerational human capital transfer is under researched. .
- Qin, et al.(2014) used overlapping generations model to study the interatction between human capial and integenerational mobility.

Models

- Overlapping Generations Model
- Simultaneous Equation and 2SLS
- Principal Component Analysis?

Data

- China Health and Nutrition Survey (CHNS)
- China Family Panel Studies (CFPS)
- Chinese Household Income Project Series (CHIPS)

Challenges and Contribution

- Heterogeneity in Parents
- Urban Rural Difference
- Quality of Education
- Endogeneity Problem
- Identification of Constratins

Identifying Over-prescription in Healthcare Claims Using Bayesian Network

HUANYE LIU

Research Question

How to use Bayesian network model to detect a doctor's over-prescription of a target drug?

References:

Sakshi Babber and Sanjay Chawla's *On Bayesian Network and Outlier Detection*, 2010

Jing Li etc.'s *A Survey on statistical methods for health care fraud detection*, 2007

Motivation

Why meaningful?

1) Over-prescription is rampant and incurs a large amount of cost/waste in the healthcare system.

2) The illegal deals between the drug company and the healthcare provider still exist in many countries.

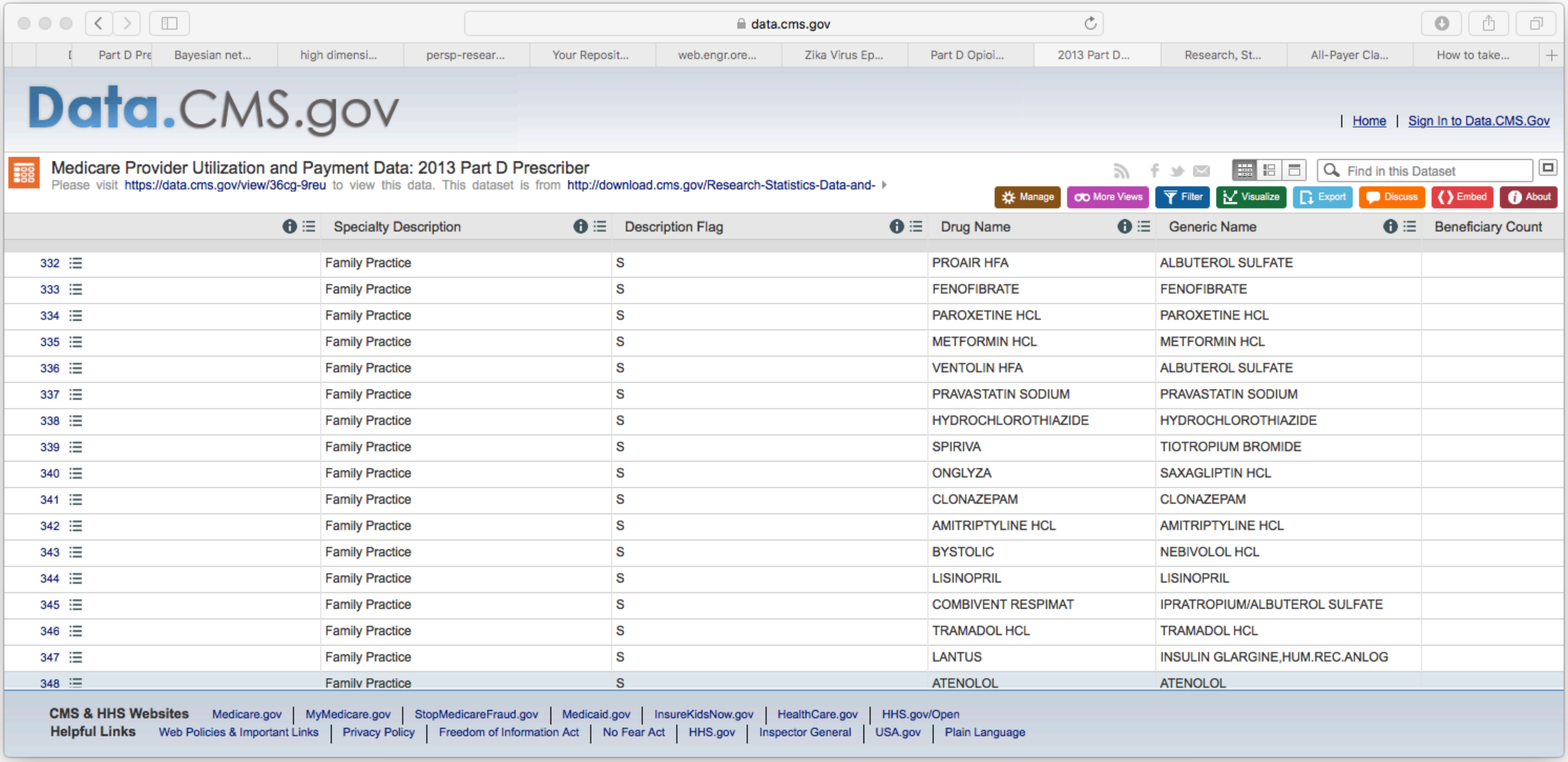
Therefore, we need effective computer aided methods to measure and detect the over-prescription behavior.

Data

Data used in this project can be downloaded from the U.S. Centers for Medicare and Medicaid Services (CMS.com) website:

https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Part-D-Prescriber.html

In particular, they published a dataset of prescriptions under Medicare Part D in 2013



	Specialty Description	Description Flag	Drug Name	Generic Name	Beneficiary Count
332	Family Practice	S	PROAIR HFA	ALBUTEROL SULFATE	
333	Family Practice	S	FENOFIBRATE	FENOFIBRATE	
334	Family Practice	S	PAROXETINE HCL	PAROXETINE HCL	
335	Family Practice	S	METFORMIN HCL	METFORMIN HCL	
336	Family Practice	S	VENTOLIN HFA	ALBUTEROL SULFATE	
337	Family Practice	S	PRAVASTATIN SODIUM	PRAVASTATIN SODIUM	
338	Family Practice	S	HYDROCHLOROTHIAZIDE	HYDROCHLOROTHIAZIDE	
339	Family Practice	S	SPIRIVA	TIOTROPIUM BROMIDE	
340	Family Practice	S	ONGLYZA	SAXAGLIPTIN HCL	
341	Family Practice	S	CLONAZEPAM	CLONAZEPAM	
342	Family Practice	S	AMITRIPTYLINE HCL	AMITRIPTYLINE HCL	
343	Family Practice	S	BYSTOLIC	NEBIVOLOL HCL	
344	Family Practice	S	LISINOPRIL	LISINOPRIL	
345	Family Practice	S	COMBIVENT RESPIMAT	IPRATROPIUM/ALBUTEROL SULFATE	
346	Family Practice	S	TRAMADOL HCL	TRAMADOL HCL	
347	Family Practice	S	LANTUS	INSULIN GLARGINE,HUM.REC.ANLOG	
348	Family Practice	S	ATENOLOL	ATENOLOL	

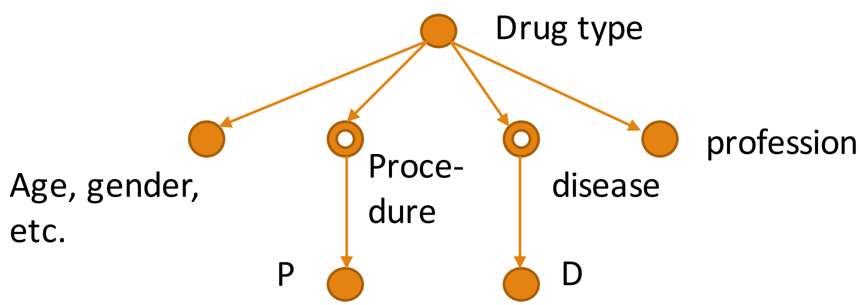
Theory to interpret the data

We use Bayesian statistics to capture the probabilistic dependency among different variables, and detect outliers from the data based on their small posterior probabilities.

Analysis and Computational tools

1)Bayesian Network

Among many outliers/anomaly detection tools, Bayesian network model can directly map the relationship between multiple variables to a graph. And there are ready-to-use algorithms for model learning and inference.



Analysis and Computational tools

2) Dynamic Programming

We use dynamic programming to accumulate each individual posterior probability of using a target drug r prescribed by one doctor to detect if the doctor's over-prescription of this particular drug.

We need to calculate the probability that the expected number of prescriptions for the target drug r based on the model inference is greater than the number of observed ones:

$\Pr(e_r \geq o_r) = \sum_{k=o_r}^{T} \Pr(e_r = k)$=$\sum_{k=o_r}^{T} Pr_{k,i=T}$

Where $Pr_{ki} = \Pr(drug_T = r|y_T)\, Pr_{k-1,i-1} + \Pr(drug_T \neq r|y_T)\, Pr_{k,i-1}$

Learning with your spouse: Does the similarity of spouse's occupation affect individual's earning?

Ningyin Xu

MACSS, University of Chicago

April 5th, 2017

Table of Contents

Research Question

To what extent the occupational similarity between one's partner and herself(himself) would influence her(his) income?

- occupational similarity
- partner/spouse
- marriage
- control variables

Hypothesis

Controlling one's education-level, age, sex, and occupation, I expect with higher similarity of spouse's occupation, one's after-marriage income would be relatively higher comparing to one who share lower similarity in occupation with her spouse.

Literature Review

- Human Capital Theory
- Marriage premiums and penalty (Korenman and Neumark 1991; Hundley 2000; Hotchkiss and Moore 1999)
- Earnings and spouse's career status (Landau and Arthur 1992)
- Occupations and marital stability (Kammen and Adams 2014)

Data

PSID: Household-level data

Table: Summary Statistics of Key Variables (PSID 2003 family-level data)

Variable	Mean	S.D.
Age of Man When Married	29.300	8.621
Age of Woman When Married	27.125	8.018
Years Married	13.840	10.922
Years Education Head	13.454	2.548
Years Education Spouse	13.533	2.420
Earnings in 1000s Head	49.072	76.637
Earnings in 1000s Spouse	27.790	23.372
Moved last year(=1)	0.280	0.449
Variance of Husband's Occupation Earnings	0.749	2.240
Variance of Wife's Occupation Earnings	0.592	1.521
Spouses Employed Same Industry(=1)	0.114	0.317
Distance between Spouses' Occupations 'Abilities' File	0.123	0.050
Distance between Spouses' Occupations 'Activities' File	0.196	0.071
Distance between Spouses' Occupations 'Knowledge' File	0.270	0.093
Distance between Spouses' Occupations 'Skills' File	0.159	0.067

Sample Size = 2032
(Kammen and Adams 2014)

Methodology

- human capital model
- Occupation similarity level: O*Net Content Model
- Occupation similarity and mobility

"Ego at a price"
Empirical study on biased self-belief with monetary stakes

Wanlin Ji

The University of Chicago, 2017

Research Question

- To what extent are people biased (over-confidence or inferiority) in the process of adjusting self-evaluation towards their Intelligence Quotient, after receiving imperfect signals about their IQ scores?
- Possible answer: Over-confident(As beliefs related their abilities, people are easier influenced by positive signal rather than negative signals), but to what degree?

Intuition behind biased belief

- Innate: Direct utility today from having a positive belief about yourself;
  Instrumental: Helps you work harder or perform better or convince others more

Theoretical and Empirical Literature

- Theoretical Hypothesis: Rational judgement is common pattern. Do people really hold biased belief towards themselves?
  Seda Erta, Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback, Journal of Economic Behavior Organization, 2011, vol. 80, issue 3, pages 532-545
- Burks, Carpenter et al(2013): Overconfidence and Social Signalling.
  Eil and Rao (2011):Asymmetric processing of objective information about yourself Favorable news: subjects roughly Bayesian (slightly optimistic) Unfavorable news: discounted, noisy posterior beliefs

Methodology: Experiment with Mechanical Turk

- Why Mechanical Turk? (Fast response; Cheap and easy; Automated analysis)
- Experimental design:
  1. Elicitation of initial confidence
  2. IQ test
  3. Elicitation of post-test confidence(Set the baseline bias degree)
  4. Four binary signals, each correct with 75 percent probability(Randomization for imperfect information; explanatory variable)
  5. Confidence elicited after each signal(Response variables)
- Monetary stakes: elicit beliefs that are close to real decision
- Why IQ test as our measurement? Authentic; Quantitative;

Basic Model: Bayes Rule

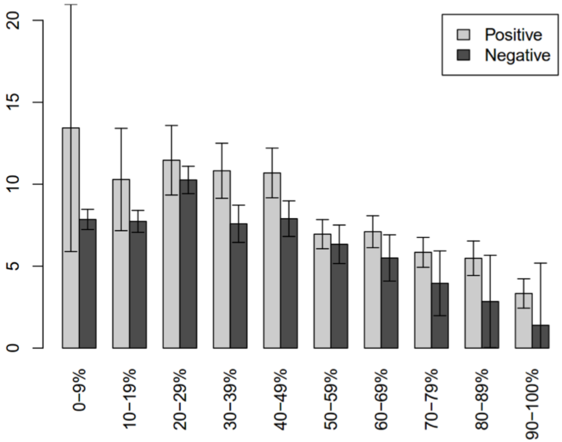
- $\text{logit}(\mu t) = logit(\mu t1) + I(t = H)H + I(st = L)L$
  Assumptions about the evolution of μ

Invariant updates; Fully responsive; Time independent response

Predicted result

- Hypothesis: This belief updating process follow Bayes Rules in the big picture, however, we expect it shows bias in different degrees than past studies, and possibly in following three dimensions.
- Dimensions: Invariance, Stability, Asymmetry

Predicted result



Open Science

- Open information on data and codes: Github
  Open source toolbox: R, Python
  Open Experimental platform: Mechanical Turk

Privacy & Innovation

Esha Banerjee

What is the influence of data privacy regulations on innovation across countries?

Privacy Innovation Conundrum

Privacy

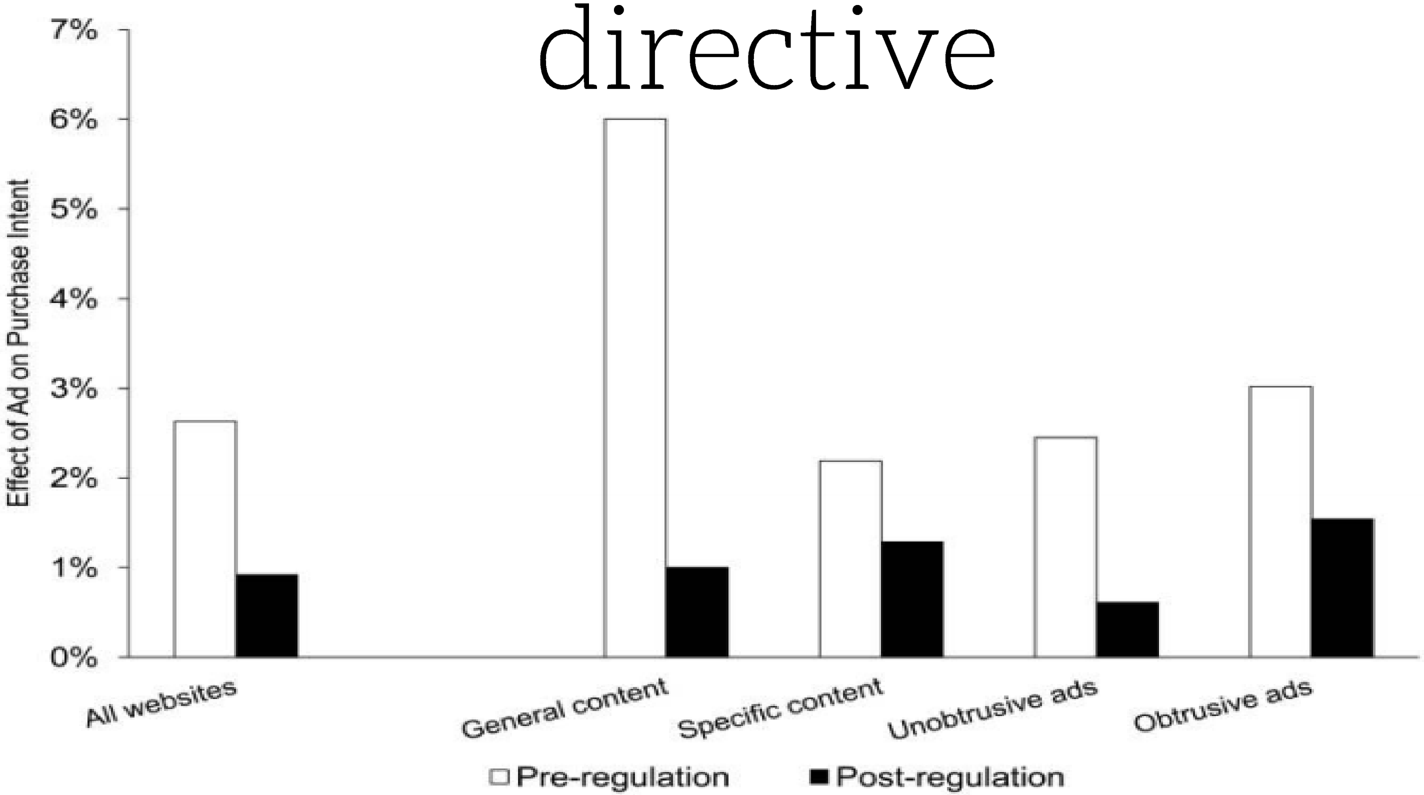
stifles

OR

drives

Innovation ?

Influence of ePrivacy directive



Ad effectiveness in the European Union before and after regulation.

Source: Privacy and Innovation (Goldfarb, Tucker 2012)

Building on

- Privacy Regulation and Market Structure: James Campbell, Avi Goldfarb, Catherine Tucker (2015)
- The influence of regulations on innovation: A quantitative assessment for OECD countries Knut Blind (2012)

Innovation Metrics

- Patent, copyrights, trademark
- Research and development expenditure as a percentage of GDP
- Number of domestically domiciled high-tech public companies
- Professionals, including Ph.D. students, engaged in R&D per 1 million population
- Postsecondary education

World Bank & Census Data

Privacy Metrics

- Privacy Laws
- Constitutional protection
- Data Sharing
- Data Retention
- Penalties for violation of laws
- Open Data Resources

Text Analysis of laws of country & assigning a categorical according to strictness of regulations

Methodology

- Text Analysis
- Classification
- Simple & Weighted Regression Analysis

Questions?

Perspectives on Computational Research: Proposal

To Move or not to Move: Disciplines and Geographical Mobility from Origin to Occupation

Xingyun Wu

Apr. 5, 2017

Research Question

- What is the effect of educational discipline on geographical mobility from original location to occupational location ?
    - Original location: place of high school
- To be more specific:
    - 1. Is occupational location related to original location?
    - 2. If yes, does the relationship remain the same across all disciplines (majors)?

Literature Review

- Theoretical Framework
  - Social Stratification
    - Education, Occupation
  - Geographical Mobility: Internal Migration
- Empirical Studies

Hypotheses

- H_1: Locational Dependence Hypothesis
  - H_1a: Occupational location is influenced by place of origin
  - H_1b: Occupational location is not significantly influenced by place of origin
- H_2: Discipline Interference Hypothesis
  - H_2a: The geographical mobility is relevant to disciplines of college education
  - H_2a: The geographical mobility is not relevant to disciplines of college education

Data & Method

- Data:
  - The 2015 National Survey of College Graduates (NSCG)
    - longitudinal biennial survey
- Method:
  - Supervised learning:
    - SVM or tree-based model

Data Glimpse

Variable:	**NBAMED**
Description:	Field of study of for first bachelor's degree - best code
Question:	What is the major field of study for your first BA degree? (best code)
Domain:	All respondents
SESTAT Variable:	J_ED_BA_MAJOR_ED_CAT_NEW
SOURCE (Survey Question Number):	CGO: SYSTEM/CGN: SYSTEM

Value/Description	Unweighted	Weighted
116710: Computer and information sciences	679	347,933
116730: Computer science	2,648	1,036,957
116740: Computer systems analysis	59	47,587
116760: Information services and systems	745	344,713
116770: OTHER computer and information sciences	200	107,179
128410: Applied mathematics	261	70,527
128420: Mathematics, general	2,567	862,592
128430: Operations research	48	18,305
128440: Statistics	178	55,738
128450: OTHER mathematics	45	16,845
216050: Animal sciences	423	180,284
216060: Food sciences and technology	178	45,369
216070: Plant sciences	369	111,708
216080: OTHER agricultural sciences	163	71,692
226310: Biochemistry and biophysics	736	236,030
226320: Biology, general	4,616	2,070,699
226330: Botany	112	29,224
226340: Cell and molecular biology	341	126,355

Variable:	**EMST**
Description:	State/country code for employer
Question:	Derived within SESTAT from reported information
Domain:	Working during the week of February 1, 2015
SESTAT Variable:	E_JOB_EMPLR_ST_CTRY_CD
SOURCE (Survey Question Number):	CGO: SYSTEM/CGN: SYSTEM

Value/Description	Unweighted	Weighted
085: New England region, state not specified	4,684	2,733,292
086: Middle Atlantic region, state not specified	11,144	6,694,018
087: East North Central region, state not specified	11,518	6,789,947
088: West North Central region, state not specified	5,588	3,305,209
089: South Atlantic region, state not specified	14,014	8,754,737
090: East South Central region, state not specified	2,717	2,104,318
091: West South Central region, state not specified	6,873	4,322,341
092: Mountain region, state not specified	4,899	3,033,113
093: Pacific region, state not specified	14,726	7,754,033
096: U.S. Territory, not specified	580	414,047
148: Europe, not specified	19	10,579
245: Asia, not specified	20	6,395
304: North America, not specified	10	1,655
315: Mexico	5	10,995
318: Central America, not specified	3	1,345
353: Caribbean, not specified	1	1,735
389: South America, not specified	2	62
462: Africa, not specified	6	921
527: Oceania, not specified	3	1,574

Difficulties

- Data: NSCG 👉 NLS? CPS?
  - the accuracy of location information?
  - biased to science and engineering workforce?
- Method:
  - The ability of SVM in dealing with a relatively large sample size and small number of features
    - Possible solutions:
      - adjust the model?
      - use tree-based model instead?

References

Bures R M. Residential mobility, migration, and life course change: a study of family, work, and mobility in later mid-life[D]. Brown University, 1998.

Groen J A. The effect of college location on migration of college-educated labor[J]. Journal of Econometrics, 2004, 121(1): 125-142.

Wolniak G C, Seifert T A, Reed E J, et al. College majors and social mobility[J]. Research in Social Stratification and Mobility, 2008, 26(2): 123-139.

Geist C, McManus P A. Geographical mobility over the life course: Motivations and implications[J]. Population, Space and Place, 2008, 14(4): 283-303.

Porter J R, Howell F M. Geographical sociology: Theoretical foundations and methodological applications in the sociology of location[M]. Springer Science & Business Media, 2012.

Research Topic

How retail location can affect business?

Zhuo Leng

Apr 5, 2017

Why I want to study this?

Research Background

background

- Facing challenges from E-Commerce giants, such as Amazon
- However, the vast majority of retail sales in U.S still happen offline
- Understanding retail business and help them fight against E-Commerce became an interesting topic.

Sales revenue graph

Research Question

Research Question

How retail location factors (demographic factors, spatial pattern, etc.) affect retail business outcome (e.g. sales) of different types shopping center in U.S?

Literature Review

Duggal, Niti . "Retail Location Analysis: A Case Study of Burger King & McDonald's in Portage & Summit Counties, Ohio ." Ohiolink. December 2007.

Kean, Rita, LuAnn Gaskill, and Larry Leistritz. "Effects of community characteristics, business environment, and competitive strategies on rural retail business performance." Proquest. 1998.

Research Contribution

- Use demographic variables to predict retail sales
- Help business owners to conduct strategic planning based on their geometric location and objectives
- Identify retail locations with high revenue protential

How to answer my question?

Research Data Source

Data Source

- U.S Census Data

- Major Shopping Center Data in U.S.

http://doc.arcgis.com/en/esri-demographics/data/shopping-centers.htm

the Major Shopping Centers variables include center name, type of center, total retail sales, distance to the nearest competing center, distance to the nearest major city, and total number of stores, longitude, latitude, etc.

Research Data Source

Variable list

- ~ 100 variables

Esri Demographic and Business Data List

Major Shopping Centers

Item	Description
Major Shopping Center Mall Database	
MALLCODE	Unique Identifier code - used to cross reference with store data
MALLNAME	Mall /Shopping Center Name
MALLCOUNTY	County where project is located
MALLLOCA	Intersecting streets where project is located
MALLCITY	City where project is located
MALLSTATE	State where project is located
MALLZIP	ZIP Code where project is located
GLA	Gross Leasable area (Sq.Ft.)
SITESIZE	# of acres
TOTSALES	total retail sales (including anchor stores)
DISTONMALL	Distance to nearest competing center
TYPEMALL	Type of center ("O"=Open, "E"=Enclosed)
LEVELS	# of Levels
SHAPE	Shape code for design (see Shape Tab)
DATEOPENED	Year Opened/To Open
SPACEAVAIL	Is space available - Yes/No?
TENANTNEED	Types of tenants needed
EXPANSION	Expansion planned - Yes/No?
WHENEXPAND	When will expansion be completed
DISTONCITY	Distance to nearest major city

Research Data Source

Variable list

- Store Type Identify

STORETYPE	Description
3A	Anchor
4G	Barbers, Beauty
3M	Children's Apparel
5T	Entertainment
3U	Food & Restaurants
4A	Gifts, Cards, Books

Esri Demographic and Business Data List

Major Shopping Centers

Item	Description
5H	Hi-Tech
3W	Jewelry
3J	Men's Wear
5Z	Miscellaneous
5S	Services
3D	Shoes
4M	Specialty Store
4B	Temporary Tenant
3S	Unisex/Family Clothing
3G	Women's Wear

CENTER_CLASS_ID	CENTER_CLASS_NAME
CC	Community Center

CENTER_CLASS_ID	CENTER_CLASS_NAME
CC	Community Center
EC	Entertainment Center
LC	Lifestyle/Specialty Center
OF	Other
PC	Power Center
RC	Regional Center
SR	Super Regional Center
UU	Unknown
VR	Value Retail Center

DESIGN_TYPE_ID	DESIGN_DESCRIPTION
E	Enclosed
O	Open
U	Unspecified

SHAPE_ID	SHAPE_DESCRIPTION
A	A-Shaped
B	B-Shaped
C	Cross-Shaped
D	Dumbbell-Shaped
E	E-Shaped
F	F-Shaped
G	G-Shaped
H	H-Shaped
I	Irregular-Shaped
L	L-Shaped
M	M-Shaped
N	N-Shaped
O	Round-Shaped
P	P-Shaped
Q	Square-Shaped
R	Rectangle-Shaped
S	Strip-Rectangle
T	T-Shaped
U	U-Shaped
V	V-Shaped
W	W-Shaped
X	X-Shaped
Y	Y-Shaped

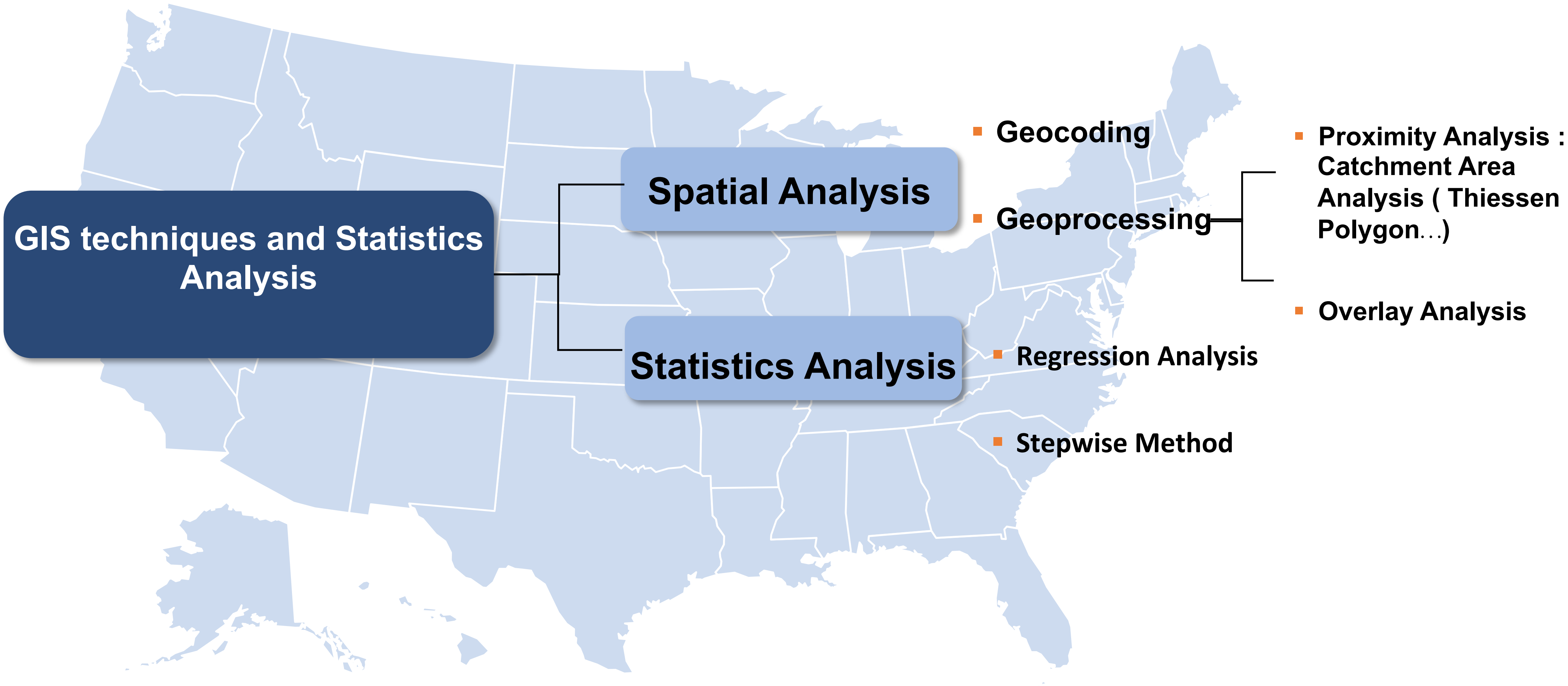
Research Data Source

Variables I will use

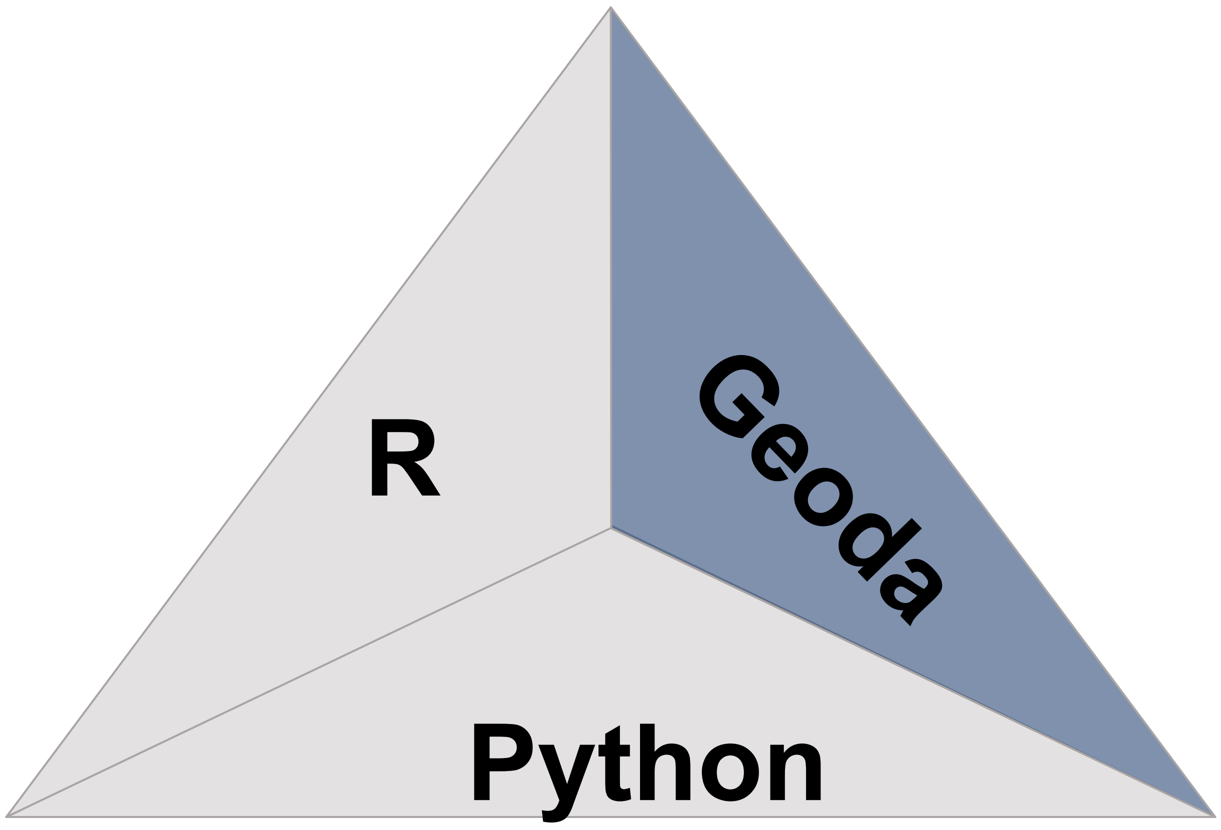
- Zip Code
- Community Area
- Total Retail sales
- Nearest Major City

- Distance to Nearest Major City
- Distance to Nearest Competing center
- Geographic Coordinate
- Other Geodemographic Data

Analyses


GIS techniques and Statistics Analysis
Spatial Analysis
Geocoding
Geoprocessing
Proximity Analysis : Catchment Area Analysis (Thiessen Polygon…)
Overlay Analysis
Statistics Analysis
Regression Analysis
Stepwise Method

Computational tool



THANKS

Zhuo Leng

Research Question Proposal

Weijiia Li

2017-04-05

What is the relationship between short interest and stock returns?

What is ‘short interest’

“A short interest is the quantity of stock shares that investors have sold short but not yet covered or closed out.”

-- Investopedia

An increase in short interest means more people believe the stock price will decrease.

Important Investment Application

“If heavily shorted stocks underperform the market, an investor should avoid stocks with a high short interest ratio when selecting stocks for a portfolio.

If an investor already owns a stock that develops sustained high short interest, the clear and strong advice is to sell the stock immediately, unless adverse tax consequences are present.”

-- D. Rapach, et al, “Short Interest and Aggregate Stock Returns”, Journal of Financial Economics, forthcoming, 2016

Will the stock price drop?

- *“We find that the higher the short interest ratio, the lower is the subsequent performance.”*

  -- P. Asquith, et al, “Short Interest, Institutional Ownership, And Stock Returns,”, Journal of Financial Economics, 2005

- *“We show that short interest is arguably the strongest known predictor of aggregate stock returns. A one-standard-deviation increase in SII corresponds to a six to seven percentage point decrease in the future annualized market excess return.”*

  -- D. Rapach, et al, “Short Interest and Aggregate Stock Returns”, Journal of Financial Economics, forthcoming, 2016

IS THIS TRUE?

DATA

★ Compustats

Company Name	Short Interest Ratio	Market Cap, $mm	Market-to-book	Prior 12-month Return	Industry
NYSE-Amex					
CVD Equipment Corp	99.3%	$653 mm	1.10	-42.37%	Semiconductor processor
UAL (United Airlines)	58.3%	$638 mm	-0.18	-85.48%	Airline
Prepaid Legal Services	54.7%	$401 mm	11.46	50.00%	Legal services
Federal Agricultural Mortgage	33.5%	$269 mm	-0.17	-30.86%	Consumer finance
Fleetwood Enterprises	33.3%	$304 mm	2.74	-34.34%	Recreational vehicle producer
Jo Ann Stores	32.7%	$293 mm	1.01	258.38%	Retail fabric stores
Metris	32.1%	$517 mm	0.82	-79.00%	Credit card issuer
BMC Industries MN	30.2%	$260 mm	0.44	-34.74%	Electronics
SWS Group	29.2%	$338 mm	1.33	-10.80%	Finance
Fleming Companies	28.9%	$976 mm	1.45	-70.85%	Supermarket supplier
Northwestern Corp.	28.7%	$464 mm	-1.02	-56.64%	Electricity and gas distribution
Nautilus Group	28.5%	$1,077 mm	5.32	-43.37%	Exercise equipment
Sunrise Assisted Living	28.1%	$602 mm	1.29	1.93%	Retirement housing
Action Performance	27.9%	$562 mm	2.45	-32.88%	Motorsports merchandise
Administaff	26.6%	$280 mm	2.41	-76.70%	Temporary staffing
Univision Communications	26.2%	$5,025 mm	3.23	-9.74%	Spanish-language TV
American Italian Pasta	25.7%	$916 mm	3.08	-2.16%	Pasta producer and marketer
Chico Fas	24.4%	$1,490 mm	6.21	112.68%	Women's clothing retailer
Salton	23.3%	$157 mm	0.64	-19.63%	Small appliance retailer
Duane Reade	23.3%	$812 mm	2.46	-42.37%	Drug store chain
Footstar	23.2%	$492 mm	1.74	-76.55%	Shoe retailer
Nasdaq					
Biosite	48.3%	$415 mm	3.85	72.74%	Diagnostic product developer
Cognizant Tech Solutions	46.1%	$452 mm	2.73	123.02%	IT outsourcing
Polymedica	45.0%	$310 mm	1.58	24.96%	Medical products retailer
Eresearch Technology	43.1%	$176 mm	4.34	120.05%	Cardiac clinical research
Cabot Microelectronics	39.6%	$1,045 mm	4.89	-12.99%	Polishing compound producer
Neoware Systems	37.7%	$146 mm	4.01	676.54%	Thin client appliances/software
FPIC Insurance Group	37.6%	$141 mm	0.85	-53.65%	Liability insurance provider
American Capital Strategies	33.0%	$1,058 mm	1.54	-17.53%	Buyout fund
Sirius Satellite Radio	31.6%	$289 mm	9.54	-87.78%	Satellite radio
Expedia	31.1%	$1,337 mm	2.75	131.13%	Internet travel agency
New Century Financial	31.0%	$864 mm	2.24	67.51%	Subprime mortgages
Silicon Laboratories	29.5%	$1,366 mm	8.77	13.70%	Integrated circuit designer
THQ	29.3%	$1,177 mm	2.88	-53.21%	Video games for PCs
J2 Global Comm	28.9%	$178 mm	3.14	460.99%	Communications services
Zix Corp.	28.7%	$97 mm	10.12	-42.52%	E-mail management/ protection
AAIPharma	28.0%	$411 mm	4.13	-16.54%	Drug marketing
XM Satellite Radio Holdings	27.4%	$661 mm	1.12	-78.26%	Satellite radio
Hot Topic Inc.	26.6%	$848 mm	5.27	27.67%	Clothing retailer for teenagers
Invision Technologies	25.8%	$403 mm	1.70	15.03%	Airline security systems
Shuffle Master	25.8%	$330 mm	6.56	7.78%	Gambling industry supplier
Cell Therapeutics	25.3%	$181 mm	4.17	-64.91%	Biotech (cancer)
Kroll Inc	25.3%	$656 mm	1.44	30.03%	Risk consulting and security

NEW METHOD

Old:

★ Four-factor linear regression
★ Time series (VAR)

Proposing:

★ Random forest

Random Forest

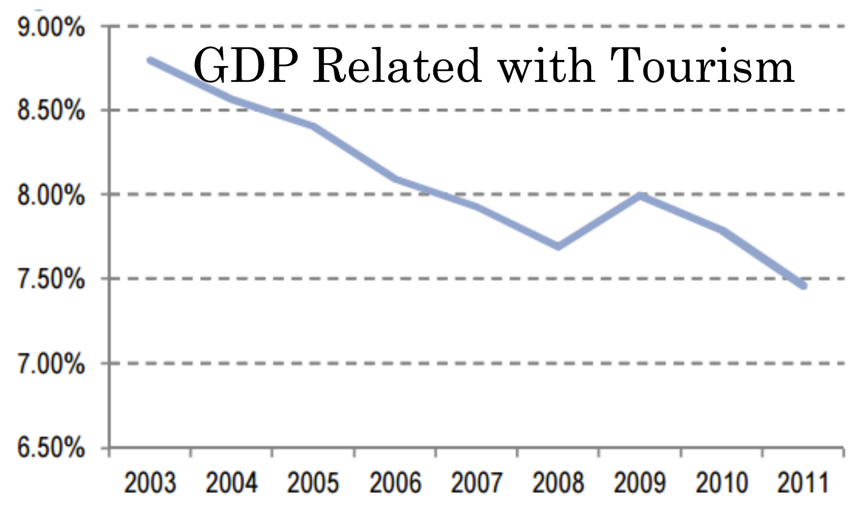
- Explanatory variable
  - Put sorted stocks into deciles to generate 10 portfolios.
  - A matrix of means of short interest of the portfolios at each time
  - Other momentums (size, growth, etc.)
- Response variable
  - Categorise stock return into large rise(2), rise(1), no change(0), drop(-1), and large drop(2)
- Two models
  - One with short interest matrix and one without
- Make variable importance plots
- Compare accuracy of predictions using 10 fold cross validation

Thank You

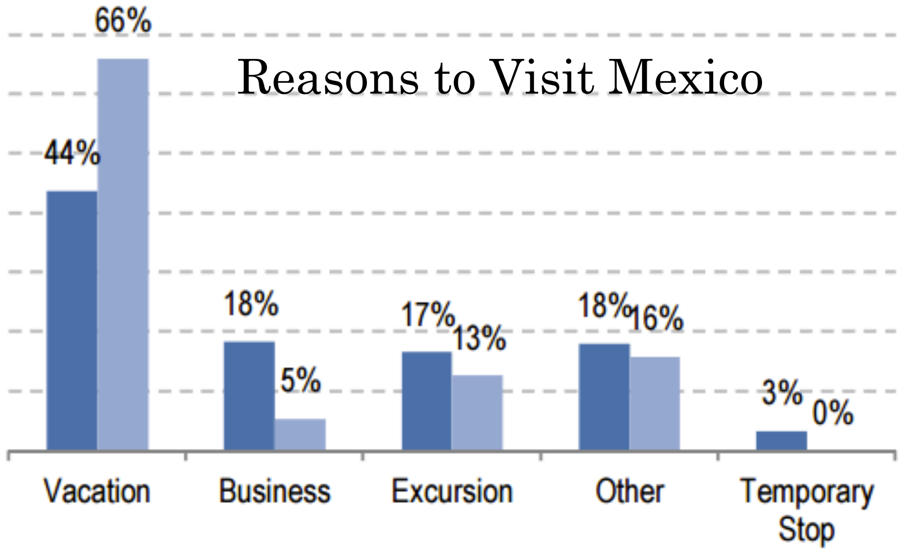
The Effect of Crime Perception in Tourism Towards Mexico: Differentiated Impacts

Rodrigo Valdés Ortiz

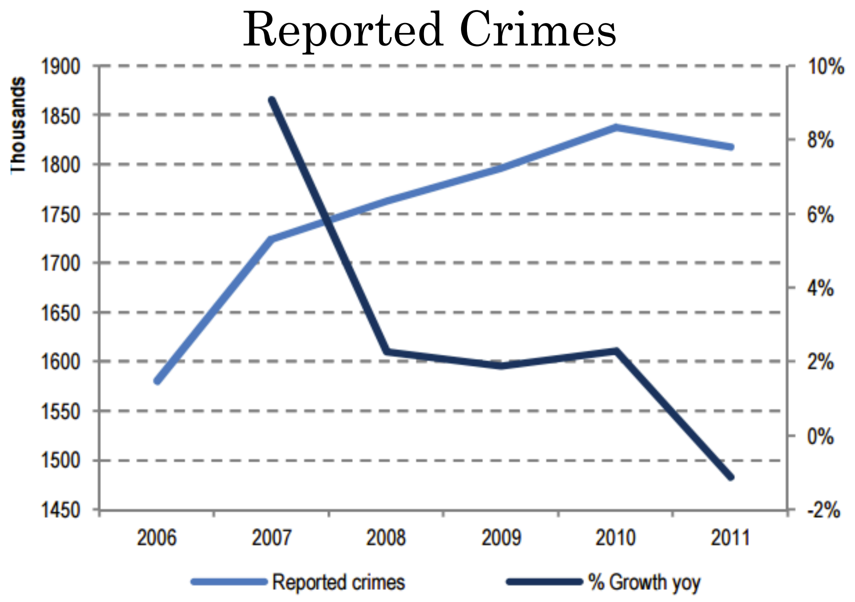
Motivation



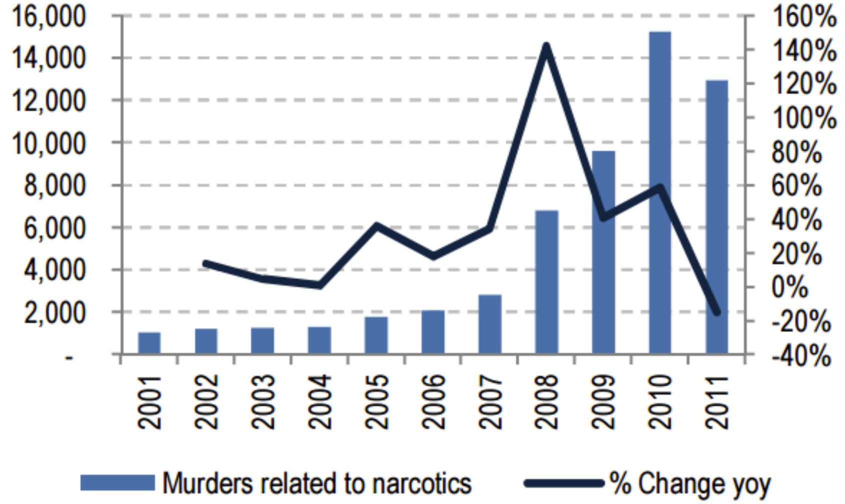
Source: INEGI.



Source: DataTur, Sectur. Data as of 2010.



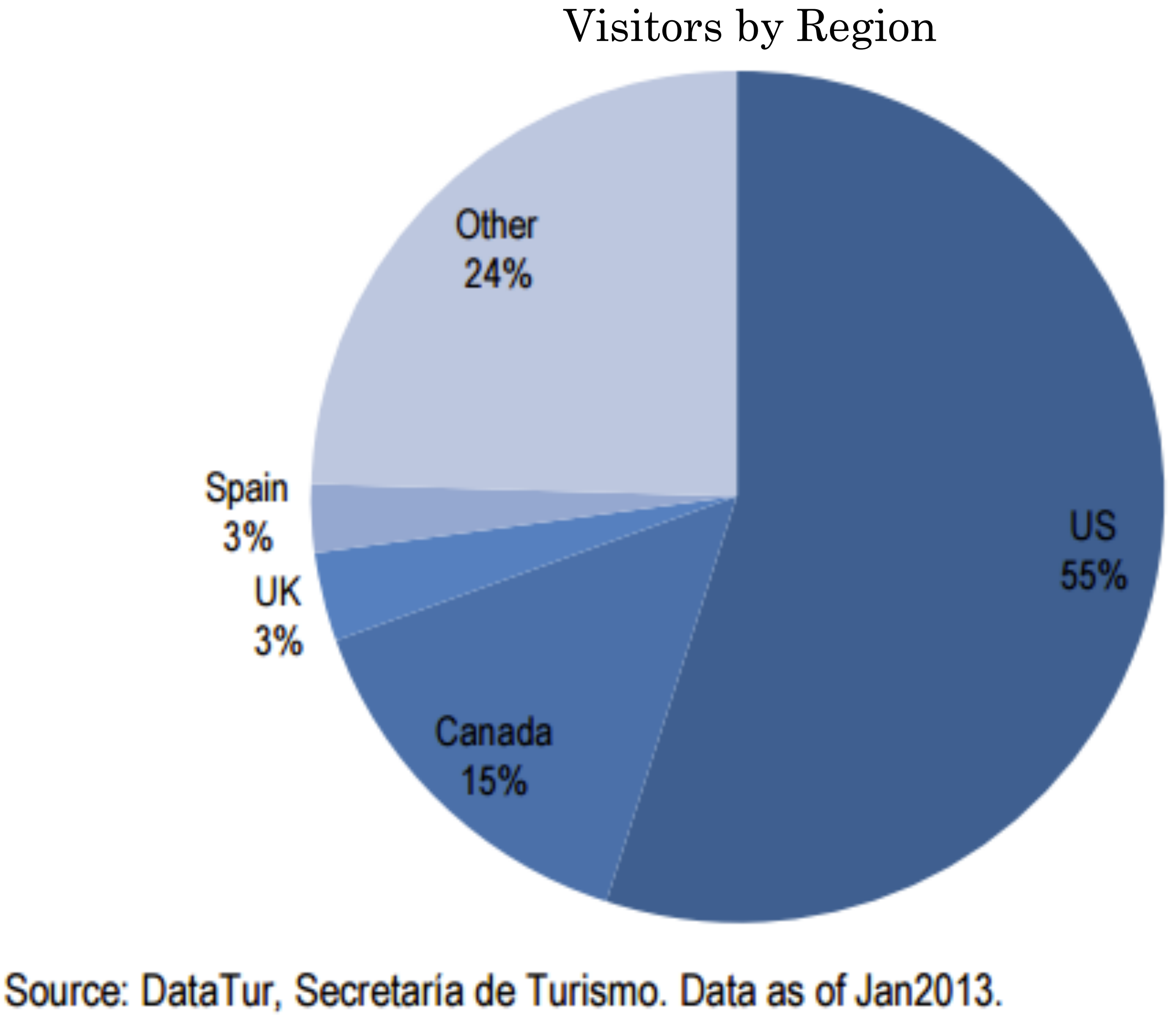
Source: ICESI.



Source: ICESI, PGR, J.P. Morgan.

Source: all graphs from Mexico 101, JP Morgan.

Motivation



Source: all graphs from Mexico 101, JP Morgan.

Research Question

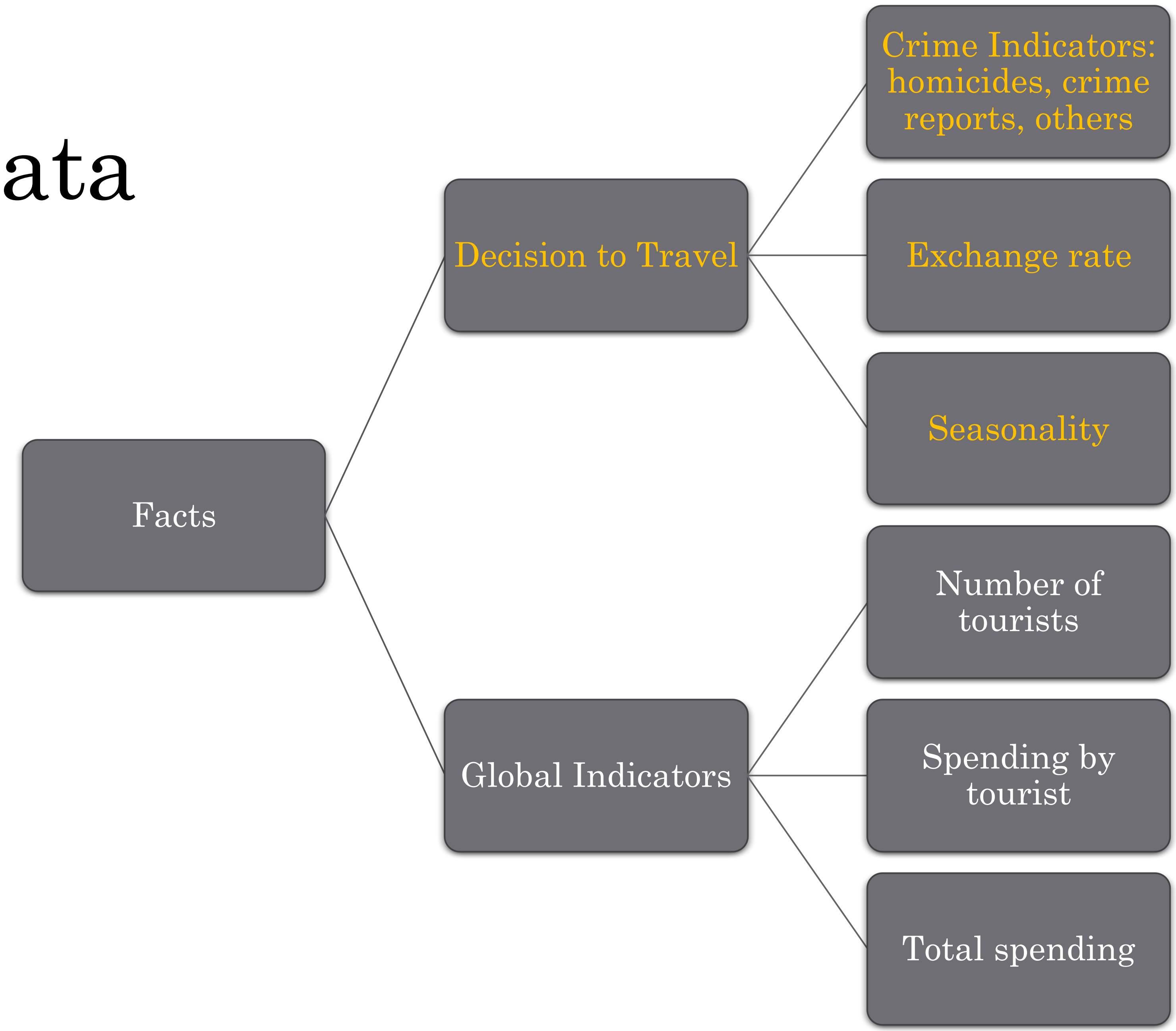
- What are the effects of perception of crime on the tourism towards Mexico?
  - How does perception varies by characteristics of the individuals, such as gender, ethnicity, age, or education?

Data

- Twitter
- Instagram
- Tumblr

- New York Times
- Boston Globe

Data


Facts
Decision to Travel
Global Indicators
Crime Indicators: homicides, crime reports, others
Exchange rate
Seasonality
Number of tourists
Spending by tourist
Total spending

Theory and Computational Tools

Content analysis

- Sentiment analysis

Statistics and Economics

- Decisions to travel (microeconomics)
- Regressions
- Time series

Web scrapping

Machine learning

- Infer characteristics

Possible Findings

- Does the perception differ by ethnicity, age, and place of origin?
- How the perception changes by social network? Which of them responds more quickly to events?
- Does the perception overreacts in the short term?
- Is perception in newspapers aligned with perception in social networks?
- Is perception more important than the facts?

Thanks!

- Criticize welcome.

The Struggle of Taxi Industry in the Age of Sharing Economy

Unveiling the optimal driving pattern and operation strategy for cabdrivers in post-2013 Chicago area

Dongping Zhang

M.A. in Computational Social Science
The University of Chicago

April 5, 2017

The Taxi-Uber War

Table of Contents

Background

Why post-2013?

- Uber launched in Chicago on September 22, 2011
- Lyft launched in Chicago on May 11, 2013

Table of Contents

Research Question

What am I studying?

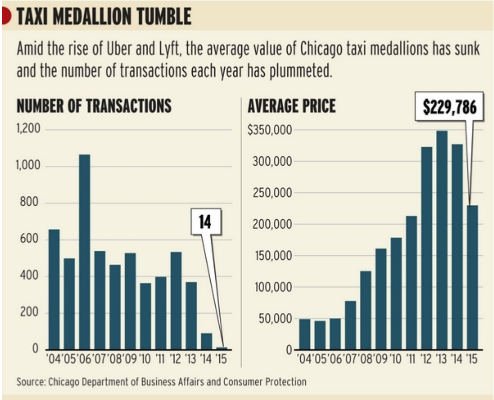
What does spatially and temporally quantified daily digital traces of the most productive cabdrivers imply the optimal operation pattern and driving strategy in post-2013 Chicago area?

Lit Review

- Liu, L., Andris, C., & Ratti, C. (2010). Uncovering cabdrivers' behavior patterns from their digital traces. *Computers, Environment and Urban Systems, 34*(6), 541-548.

Table of Contents

Taxi Medallion Sales Trend



Motivation

Taxi Medallion Price

- Taxi Medallion and related assets were worth $2.5 billion in Chicago and is regarded as one of the "cash-cows"
- Average price for a medallion in Chicago was less than $230,000 in 2015
- 30% drop of medallion sales price

Some Facts in 2015

Number of Drivers

- 156,661 active drivers in ride-hailing industry in Chicago
- 12,955 active taxi chauffeur licenses in Chicago

Number of Taxi Trips

Annual Taxi trips plummeted by 35% since 2014

- 2014: 31,013,591 trips
- 2015: 27,395,382 trips
- 2016: 19,874,714 trips

Motivation

What is the contribution?

- A systematic study of large scale cabdrivers' behaviors in a real and complex city context
- Recognizing high-level human behaviors and decision-makings from their daily digital traces
- The invasion of Uber, Lyft, and other ride-hailing service companies is gradually driving the traditional taxi industry out of business
- Policy implication to alleviate the burden of the traditional taxi industry
- Aiming to uncover some new insight of urban transportation and human mobility

Table of Contents

Data

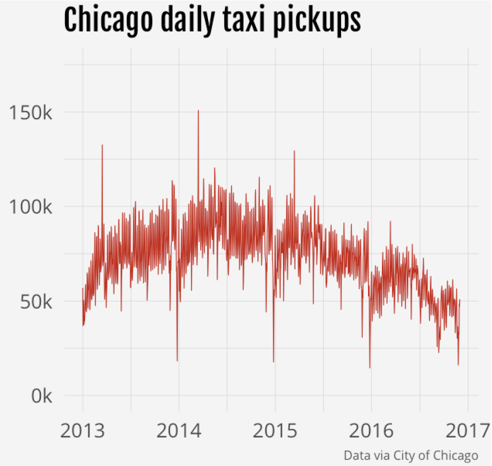
Where did I acquired the data?

Chicago Open Data Portal powered by Socrata

What are my variables of interest that are also included in the raw dataset?

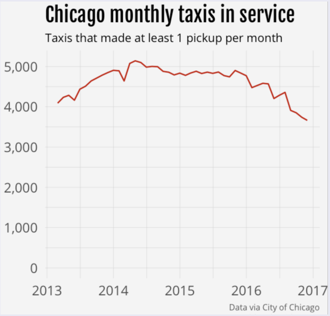
- Trip ID
- Taxi ID
- Trip Start Timestamp
- Trip End Timestamp
- Trip Miles
- Fare, Tips, Tolls, Extras
- Pickup Coordinates
- Dropoff Coordinates

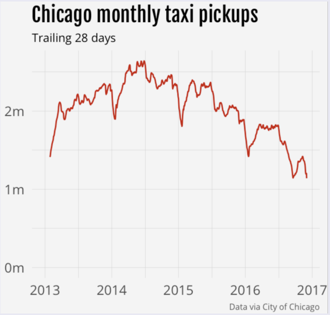
Exploratory Data Analysis 1



Exploratory Data Analysis 2

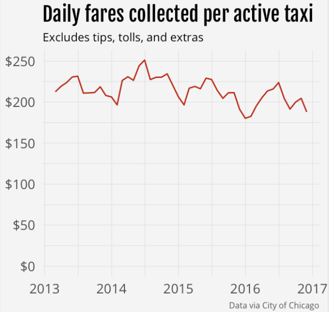
Monthly Taxis in Service & Monthly Taxi Pickups





Exploratory Data Analysis 3

Taxis Daily Fares & Trips per Day



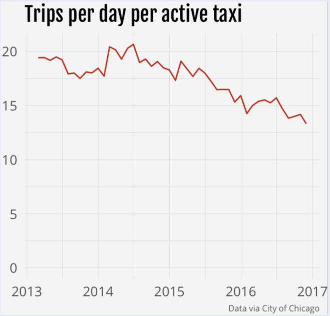


Table of Contents

Types of Analyses

What analyses would I conduct?

- ESDA (points vs. aggregating pickups and drop-offs by areal unit)
- Spatial Regression Analysis to capture spatial dependency
- 3-Dimensional Flow Analysis
  - pickup points, direction, drop-off points
- Ratio of real path length over shortest path length
- Ratio of real path travel time over shortest path travel time

Computational Tools

What computational tools will I use?

- SQL – RDBMS
- Python
- R
- GeoDa – ESDA
- GeoDaSpace – SRA

Table of Contents

Summary

- To identify a subset of most productive cabdrivers who have consistently higher daily income from 78,283,687 trip history
- To spatially and temporally quantify the daily digital traces of those most productive cabdrivers
- To uncover potential driving patterns and operation strategies of those most productive cabdrivers

Reference

- Liu, L., Andris, C., & Ratti, C. (2010). Uncovering cabdrivers' behavior patterns from their digital traces. *Computers, Environment and Urban Systems, 34*(6), 541-548.
- Huang, H., Zhang, D., Zhu, Y., Li, M., & Wu, M. Y. (2012). A Metropolitan Taxi Mobility Model from Real GPS Traces. *J. UCS, 18*(9), 1072-1092.
- Li, B., Zhang, D., Sun, L., Chen, C., Li, S., Qi, G., & Yang, Q. (2011, March). Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on* (pp. 63-68). IEEE.

Beyond the Poverty Line: Key Predictors of Maternal Mortality in India

Sushmita V Gopalan

Motivation

Allocation of scarce resources - how to allocate public health expenditure?

Currently, most maternal welfare schemes in India are targeted at Below-Poverty-Line families, i.e. income is the cut-off point for access to these schemes.

However, studies (Meesham et al. 1999, Claeson et al. 2000, Padmanaban 2009) suggest that family income might not be the best predictor of maternal mortality.

Institutional factors at the district or state level could be better predictors

-Tamil Nadu's success (Padmanaban 2009)

- Access to maternal health interventions (Claeson et al. 2000)

Research Question

1) What are the key predictors of maternal mortality in India?

2) How do we use these to identify an at-risk cluster and target maternal welfare schemes at them?
    - Challenge - how to we ascertain whether it was access to welfare schemes that lowered risk?

Data

- National Family Health Survey (NFHS):
  - ~20,000 families surveyed, once a decade
  - 4 waves of data available - newest in 2015-16 (very few research outputs!)
  - Individual-level variables such as previous pregnancies, lifestyle, usage of maternal healthcare, income, etc.

- District Level Health Survey (DLHS):
  - every 5 years
  - district-level variables such as sanitation, spread of govt. hospitals, per capita income, access to public health schemes, etc.

- Sample Registration System (annual) : district-level maternal mortality rates

Methods

Logistic Regression

Non-Linear Relationships? (Song et al. 2004, Liu et al. 2014)

Principal Component Analysis to reduce dimensionality

Decision Trees for Prediction

Single Layer Perceptron

Comparison between predictive ability of individual-level characteristics and district-level characteristics

What’s new?

Very little research on NFHS IV data (2015-16)

Very few studies applying machine learning to predicting health outcomes from socioeconomic variables

Questions?

War and Migration Patterns

Haylee Ham

How does war affect migrant flow and the selection of migrant destinations?

Data

Bilateral migration flows between 196 countries from 1990-2010[1]

World Bank Development Indicators[2]

Theory

Neo-classical: Harris-Todaro model, migrants maximize utility based on income and probability of finding work (Harris, 1970)

“Welfare magnate effect”: the generosity of the destination country’s welfare program is a main indicator (Borjas, 1987)

Network effects: relationships in the destination may be overcoming language and cultural barriers (Pedersen, 2008)

Where does this paper fit in: By including all bilateral relationships, I examine both close and far-proximity migration and also look at global effects of war

Methods

$$f_{i,\,j,\,t} = g(X_{i,\,t}, X_{j,\,t}, dist_{i,\,j}, war_{i,\,t}, war_{j,\,t})$$

Generalized Linear Regression: inference

- ➔ Comparing estimates in peacetime with estimates in wartime within the same i, j pair of countries

Decision Trees: prediction

- ➔ Identify predictive characteristics for number of migrants in peacetime and wartime

Questions?

References

[1] http://data.worldbank.org/data-catalog/world-development-indicators

[2] http://www.global-migration.info/

Harris, John R., and Michael P. Todaro. “Migration, Unemployment and Development: A Two-Sector Analysis.” The American Economic Review, vol. 60, no. 1, 1970, pp. 126–142., www.jstor.org/stable/1807860.

Borjas, George J. “Self-Selection and the Earnings of Immigrants.” The American Economic Review, vol. 77, no. 4, 1987, pp. 531–553., www.jstor.org/stable/1814529.

Pedersen, Peder J. , Mariola Pytlikova, and Nina Smith. “Selection and network effects—Migration flows into OECD countries 1990–2000.” European Economic Review, vol. 52, no. 7, 2008, pp. 1160–1186., http://dx.doi.org/10.1016/j.euroecorev.2007.12.002.

The Impact of Postponing Announced Unconventional Fiscal Policy on Consumption Expenditure

Xinzhu Sun

March 2017

Citations and Innovation

- Time Consistency and Optimal Policy Design (V.V.Chari, 1998)
- Time-consistent Optimal Fiscal Policy (Paul Klein, 2003)
- Fiscal Policy under Loose Commitment (Davide Debortoli etc., 2009)
- The effect of unconventional fiscal policy on consumption expenditure (Francesco D'Acunto etc., 2016)

A Nature Experiment from Japan

- In 2009, to win the election, Japanese government promised not to raise the consumption tax for the next four years. However, in 2012 June, they ate their own words and passed a bill to increase the tax to 8% in April 2014 and 10% in October 2015. Due to Japan's economic situation, the government has decided to delay the tax increase to 10% until April 2017. In 2016, a second postponement was announced, which pushes the increase to October 2019.

Data source

- World Bank
- Penn World Table
- Bureau of Statistics, Ministry of Internal Affairs and Dommunications
- Bank of Japan
- Ministry of Economy, Trade and Industry
- etc.

Questions

Thank you!

Research Proposal

Yang Hou

Research Question

- Does asset still determine individual trade-policy preferences towards trade barriers in this post-financial crisis, anti-globalization administration situation?

Previous Work

- Scheve, Slaughter (2000)
  - Main theoretical framework (factor type, asset)
- Irwin(1994, 1996) and Magee (1978)
  - Evidence for preferences determined by industry
- Beaulieu (1996, 1998), Balistreri (1997), Rogowski (1987, 1989)
  - Evidence for factor type dominants

Dataset

- American National Election Study
- Census Occupation Code
- Census Industry Code
- Housing value in every county

Methodology

- Logistic regression
- Possible tree-based method depending on the dataset