Agent-based simulation for predicting internal displacement and return

Which models and parameters best predict the geographic distribution of internal displacement and return in an agent-based simulation?

Tyler Amos
MACS 30200
April 4 2018

Models of (Forced) Migration

- Gravity

- Radiation (Simini et al 2012), with flux between locations as:

$$T_{ij} = T_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})}$$

Case Study

Iraq, 2017- Present

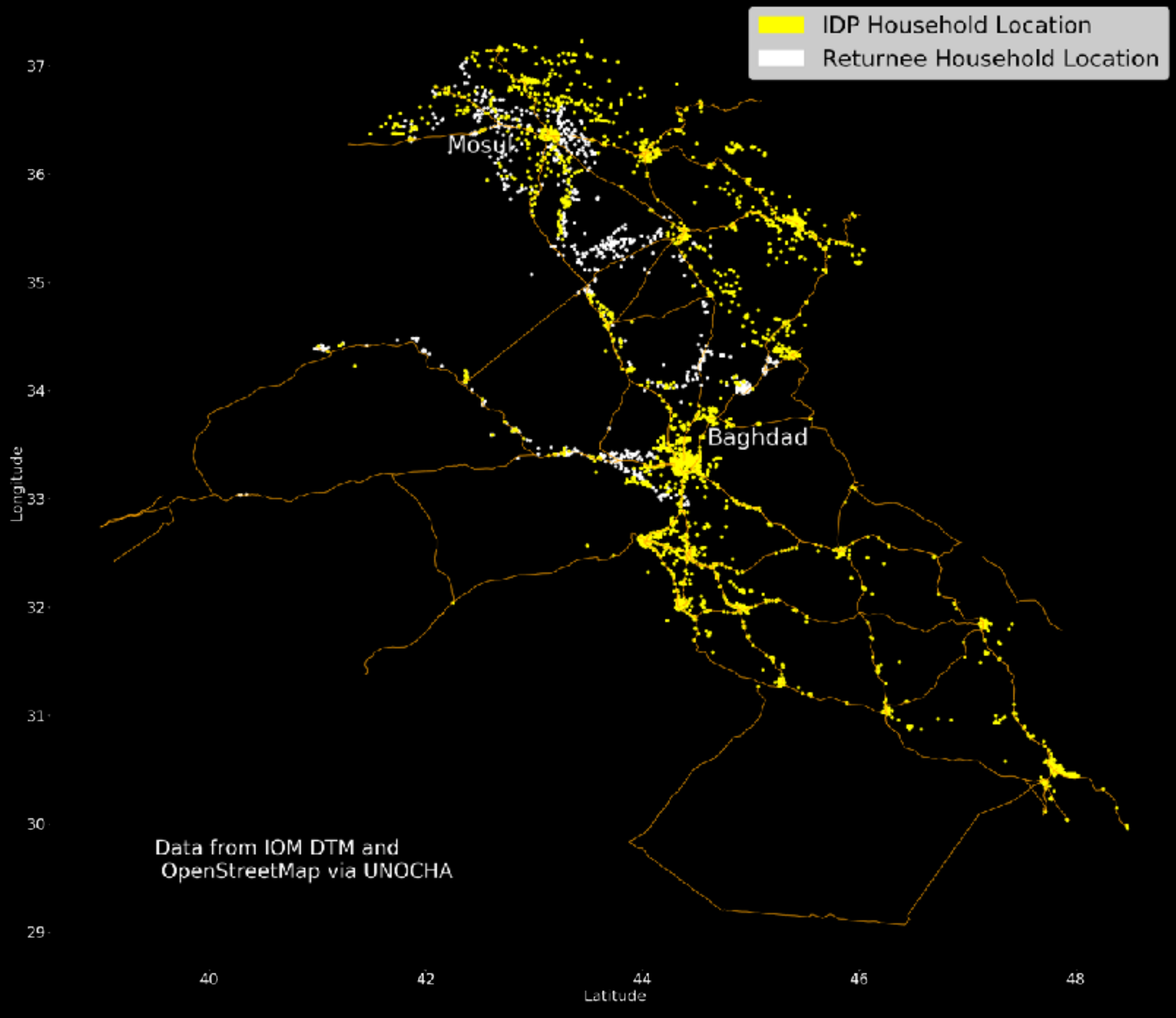
Data

- International Organization for Migration (IOM) Displacement Tracking Matrix (DTM)
- Armed Conflict Location and Event Database
- Empirical Studies of Conflict ethnic/religious mapping
- UNOCHA (e.g., Open Street Map, 4W's NGO operation mapping)

Validation Set

- DTM 03-2018 - 05-2018

Iraq Displacement & Return, March 2018

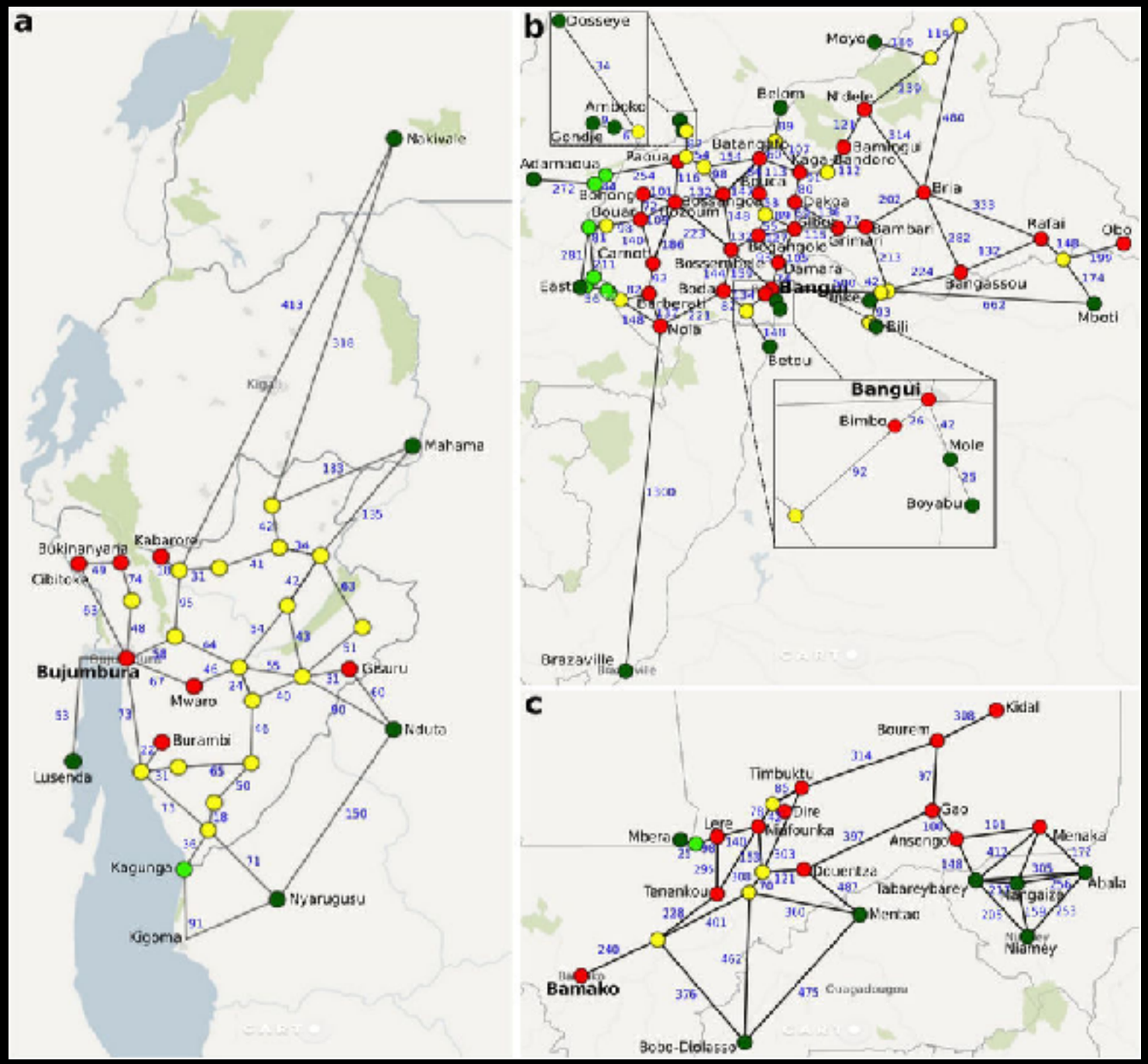


Interesting Progress in Simulation

Suleimanova, Bell, Groen (2017), Groen (2016)

- Cases: CAR, Burundi, Mali

- Claim accurate prediction of > 75% of refugee pop. after 12 “days”



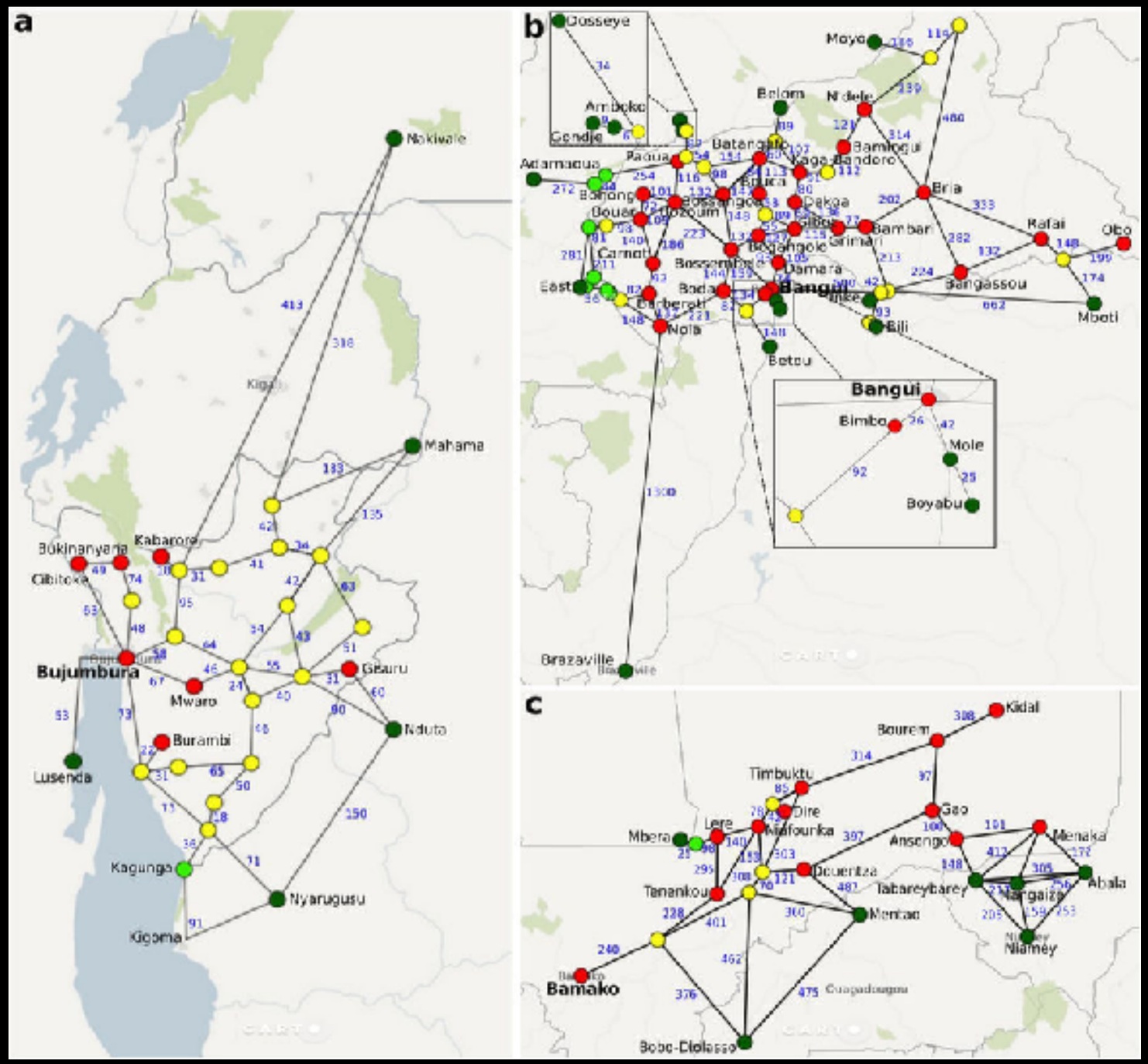
Graphic from Suleimanova et al. 2017

Interesting Progress in Simulation

Suleimanova, Bell, Groen (2017), Groen (2016)

- Cases: CAR, Burundi, Mali

- Claim accurate prediction of > 75% of refugee pop. after 12 “days”



Graphic from Suleimanova et al. 2017

Value of Contribution

A. Displacement & return —> integrated response

B. More sophisticated parameter estimation (e.g., Bayesian Network, reinforcement learning) —> more robust justification for parameter choices

C. Ties to current forced migration studies —> social scientists' added value

Value of Contribution

A. Displacement & return —> integrated response

B. More sophisticated parameter estimation (e.g., Bayesian Network, reinforcement learning) —> more robust justification for parameter choices

C. Ties to current forced migration studies —> social scientists' added value

Preliminary Hypotheses

1. A radiation model-based edge-weighting will better predict displacement and return than gravity model-based edge-weighting.

2. Additional factors such as the location of humanitarian agencies and co-ethnic population distributions will improve predictive accuracy.

Preliminary Hypotheses

1. A radiation model-based edge-weighting will better predict displacement and return than gravity model-based edge-weighting.

2. Additional factors such as the location of humanitarian agencies and co-ethnic population distributions will improve predictive accuracy.

OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

U.S. PUBLIC PENSION ASSET ALLOCATIONS

Research Question: How do fund characteristics (performance and composition) effect the asset allocations of U.S. public pension funds from 2001-2016?

By: Bethany Bailey

IMPORTANCE OF TOPIC

What factors influence asset allocation of U.S. public pensions?

- WHY PUBLIC PENSIONS?
  - Many U.S. public pensions are underfunded
    - In 2016, the median funding ratio (assets available for payments to retirees) was *71.1%* (Bloomberg)
    - In 2017, “US public pension funds lack $3.85tn that they need to pay the retirement benefits of current and retired workers” (Financial Times)
- WHY ASSET ALLOCATION?
  - Modern portfolio theory - reducing portfolio risk through diversified, uncorrelated assets

DATA

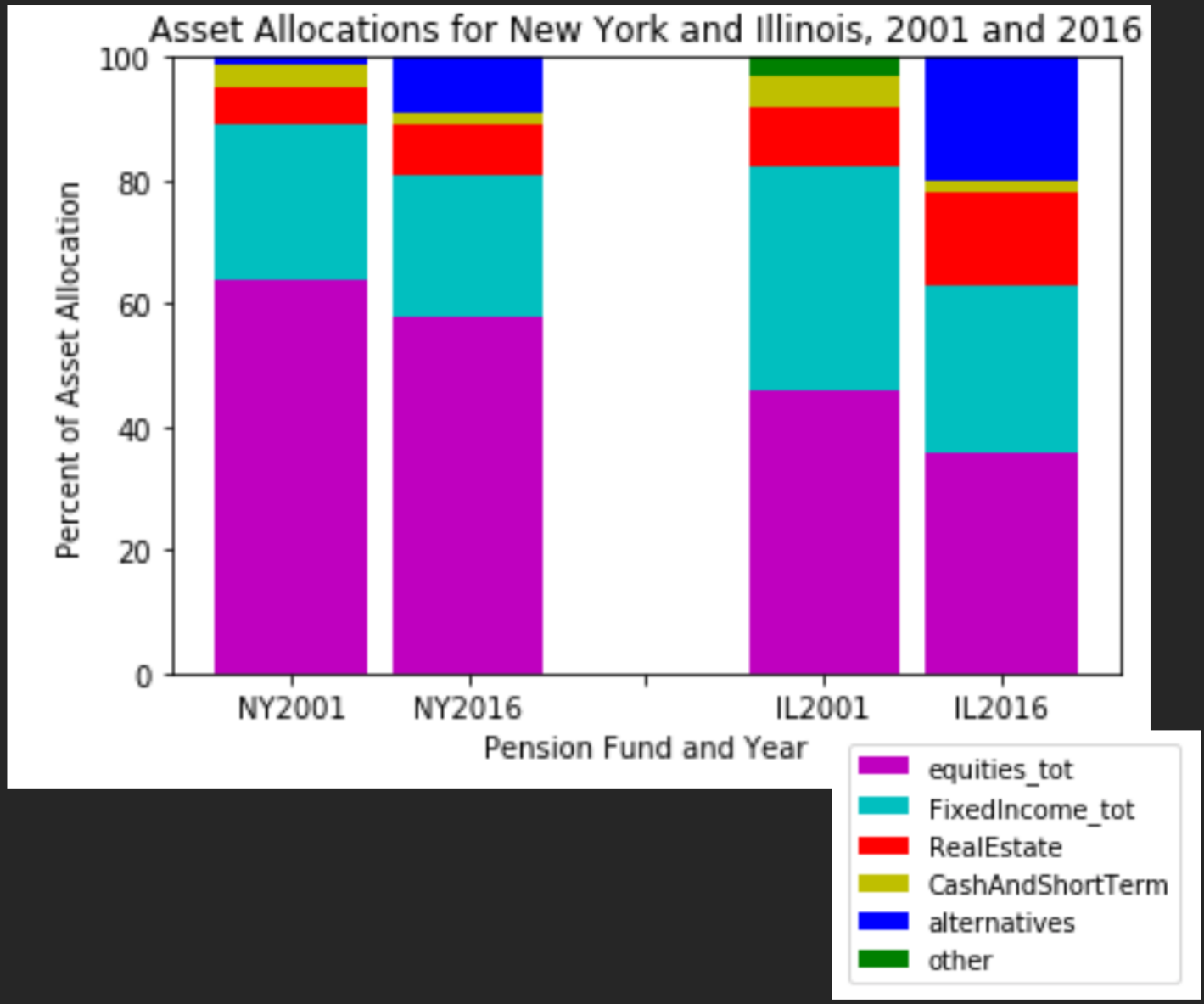
- Public Plans Database
  - Plan-level data on 170 public pension plans: 114 administered at a state level and 56 administered locally
  - Covers 2001-2016 (16 years)
  - Covers 95 percent of public pension membership and assets nationwide
  - Includes information on
    - Asset Allocation
    - Membership composition (type and quantity)
    - Funding
    - Returns
    - Etc.

VARIABLES

- Endogenous:
  - Asset Allocation
- Potential exogenous variables to consider:
  - Funded Ratio
  - Number of pensioners in plan
  - Assets under management
  - Pension industry
  - Previous year’s investment return

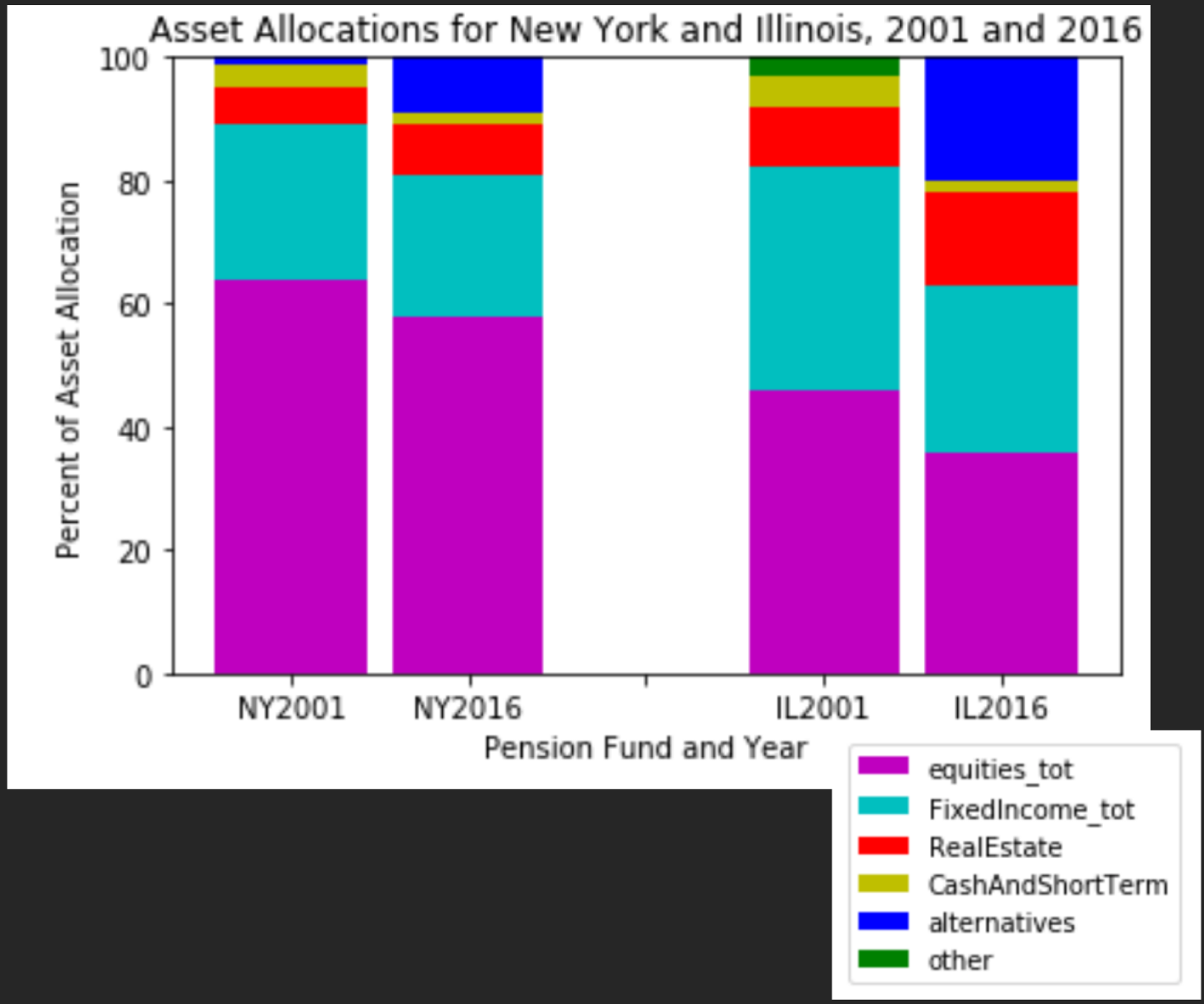
DEFINING ASSET ALLOCATION: BREAKDOWN

- % total assets invested in:
  - Equities
  - Fixed Income
  - Alternatives
  - Real Estate
  - Cash/Short-Term
  - Other
- Vary by year and plan
  - NY has higher funded ratio than IL



DEFINING ASSET ALLOCATION CONT.

- Operationalizing asset allocation
  - Classifier
    - High (1) and low (0) equity to fixed income ratio
  - Continuous
    - Finer-grained ratio
    - Scale of allocation profiles
  - Different Models?
    - Predict 6 categories



MODEL/THEORY

- Let’s do a horse race!
  - Logistic Regression vs. Random Forest vs. Neural Net
- Take model strengths/weaknesses into account
  - Classification vs. continuous model of asset allocation
  - Adding more features (Neural Net vs. Random Forest)

POTENTIAL ISSUES WITH MODEL

- Data is relatively small (170 obs/year)
  - Include resampling methods
- Time-Series data
  - Will need to account for changes in markets and investment strategies over time
- Omitted Variable Bias and correlations between variables
  - Need to think hard about the exogenous variables I use

PLACE IN EXISTING LITERATURE

- Pennacchi and Rastad, 2010
  - Analyzed factors that effect risk in 125 public pension funds, as measured by tracking error (higher risk assuming by managers following low performance)

- Weller and Wegner, 2008:
  - Have public sector pension plan managers acted “imprudently” to chase returns after encountering underfunding? (No)

CONTRIBUTION

- Looking at overall asset allocation (not a measure of risk or prudent behavior)
- Different exogenous variables
- Different models
- Data years are more recent (pensions have become more underfunded in recent years)

OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

Automation and Labor Demand: Estimating Task Based Models with Microeconomic Data

Ariel Boyarsky
University of Chicago
aboyarsky@uchicago.edu

April 3, 2018

Introduction

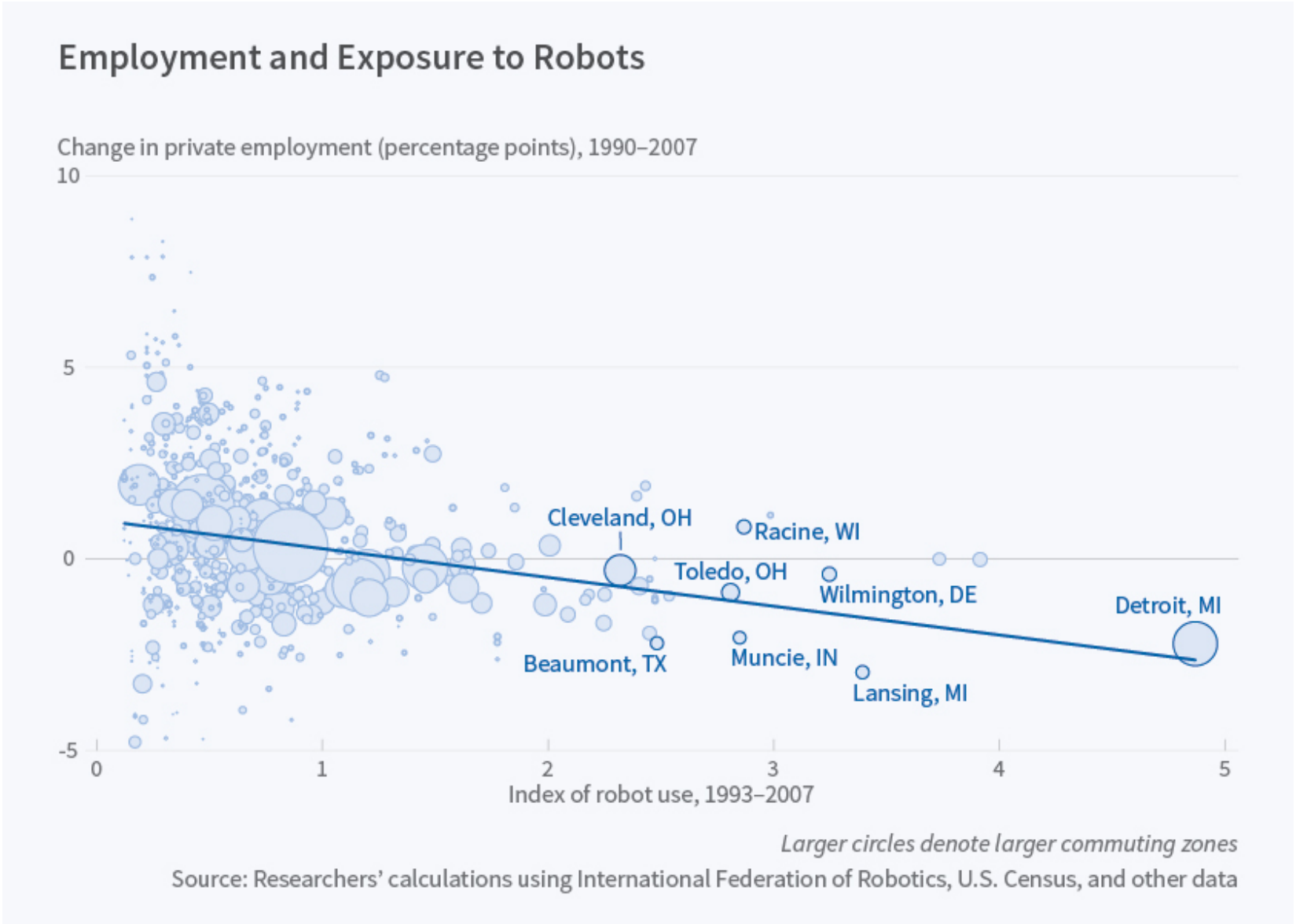
Research Question

How do current task based models of labor automation preform against empirical evidence at the firm level?

Motivation:

- Rising productivity/declining labor share
  - "Labor will become less and less important... More and more workers will be replaced by machines. I do not see that new industries can employ everybody who wants a job" - Wassily Leontief, 1952
- Trade vs. Automation and policy questions
- "Future of Work"
- Firm-level data

Declining Labor Share



Literature Review

- Zeira (1998)
- Acemoglu and Autor (2011)
- Brynjolfsson and McAfee (2012)
- Acemoglu and Restrepo (2016)
- **Acemoglu and Restrepo (2017)**
  - Henceforth, (AR17)
  - Empirical
- **Acemoglu and Restrepo (2018)**
  - Task Based vs. Factor Augmenting

Contributions

- Evaluation of task based model using firm-level data
- Revised empirical model specification to incorporate different data
- Structural estimation of parameters (GMM)
- Estimation of aggregate labor demand effect (Maybe)

Theoretical Model: Automation in Autarky (AR17)

CES Production:

$$Y_c = (\sum_{i \in \mathcal{I}} \alpha_i Y_{ci}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma-1}{\sigma}} \text{ for } c \in \mathcal{C} \text{ and } \sum_{i \in \mathcal{I}} \alpha = 1$$

Supply of labor and robots with ϵ and η as elasticities of supply of labor and robots,

$$W_c = \mathcal{W}_c Y_c L_c^{\epsilon}, \epsilon \geq 0$$

$$Q_c = \mathcal{Q}_c (\frac{R_c}{Y_c})^{\eta}, \eta \geq 0$$

Demand for labor satisfies:

$$d \ln L_c^d = -\sum_{i \in \mathcal{I}} l_{ci} \frac{dM_i}{1 - M_i} - \sigma \sum_{i \in \mathcal{I}} l_{ci} \ln P_{Xci} + d \ln Y_c$$

l_{ci} is share of employment in industry, i, in cz, c. M_i share of tasks automated. $P_X ci$ is price of output X in industry, i, cz, c. σ denotes elasticity of substitution across goods produced in different industries.

Empirical Model (AR17)

Impact of robots on employment is given by,

$$d \ln L_c = -\frac{1+\eta}{1+\epsilon} \sum_{i \in \mathcal{I}} l_{ci} \frac{dM_i}{1-M_i} + \frac{1+\eta}{1+\epsilon} \pi_c \sum_{i \in \mathcal{I}} l_{ci} \frac{s_{icL}}{s_{cL}} \frac{dM_i}{1-M_i}$$

Where,

$$\sum_{i \in \mathcal{I}} \frac{s_{icL}}{s_{cL}} \frac{dM_i}{1-M_i} \approx \frac{1}{\gamma} \sum_{i \in \mathcal{I}} l_{ci} \frac{dR_i}{L_i} = \text{exposure to robots}$$

Then,

$$d \ln L_c = \beta_c^L \sum_{i \in \mathcal{I}} l_{ci} \frac{dR_i}{L} + \epsilon_c^L$$

Where,

$$\beta_c^L = (\frac{1+\eta}{1+\epsilon} \pi_c - \frac{1+\eta}{1+\epsilon}) \frac{1}{\gamma}$$

γ denotes productivity of labor ($\gamma > 0$). s_{icL} denotes share of labor in the output of industry i in commuting zone c. π_c is cost saving by substituting robots for labor. R_i is robot penetration in industry i.

Data

Table: US Projects Data 1985-2015 - Conway

Date Entered	Firm Name	City	County	State	NAICS	Investment ($ Million USD)	Jobs
31-12-2015	PUBLIX SUPER MARKETS, INC.	Orlando	Orange	FL	445110	16	48

Table: Concentration of industrial robots in all - Brookings

Metro Area	Total Industrial robots, 2010	Total Industrial robots, 2015
Elkhart-Goshen, IN	1778	4355

Table: County Level Returns - IRS

Year	FIPS	Number of Returns	Aggregate Wage ($ USD Thousands)
2015	12095	1501130	129840547

Structural Estimation

Our model,

$$\ln L_j = \beta_j^L \sum_{m \in \mathcal{MSA}} l_{jm} \frac{R_m}{L_m} + \epsilon_m$$

Where $m \in \mathcal{MSA}$ represents a metropolitan statistical area, and j is a firm in the US projects data.

Procedure

1. Restrict to manufacturing sectors.
2. Merge firm data with robot exposure per MSA by MSA, and total labor in MSA.
3. Estimate β_j^L using a method of moments equating first two sample moments to theoretical moment.
4. Test robustness of estimation and attempt to estimate aggregate effect from parameters (Interpret parameters).

$\Longrightarrow$ **firm level** gain from substitution of automation for labor.

Expected Problems/Future Work

Problems

- Date of Entry in Conway Data is inaccurate
- Conway data may not be complete enough
- How to interpret β and test against aggregate level results

Future *(of)* Work

- How can we revise macro models of automation to learn about micro-level phenomena?
- What can we learn/predict about the future of work from micro-level insight?
- 2SLS to test for causal effect of robot exposure on labor demand

OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

Manipulate in Nordic Electricity market?

Repeated Sequential Game, Market Power and Collusion

Philip Xinyu Cao[1]

[1]Computational Social Science
Chicago

Computational Research, 2018

Motivation

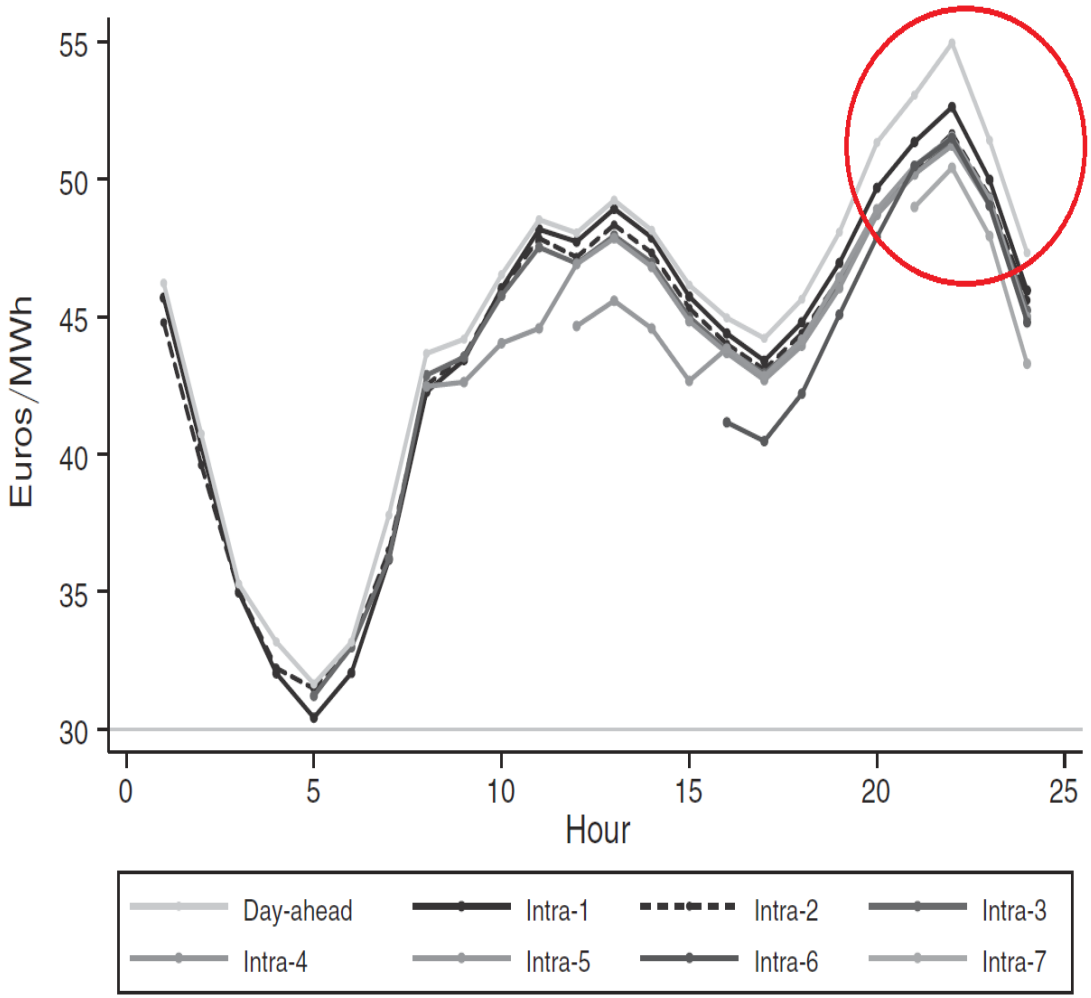


Figure: Day-ahead Market Premium In Iberian Market: Ito and Reguant (2016)

Motivation

- Electricity Market are Particular Subject to Market Power
  - Electricity are hard to storage and the capacity of the power plant is limited during certain period
  - Demand is hard to predict and inelastic to short term price change
- Electricity Market are central to economics activities
- Market Power analysis are important for policy implications

Theoretical Set Up

Sequential Market(Ito and Reguant 2016)

Residual Demand

We set up the demand as residual demand, since there are some supply are fixed, for example the supply from renewable energy market that are guaranteed by the government, nuclear energy:

$$D_1(p_1) = A - b_1 p_1$$
$$D_2(p_2) = (p_1 - p_2) b_2$$

One of the micro foundation of the residual demand is that fringe firm agree to supply electricity as long as the price is higher than marginal cost.

Monopolist Problem

Stage 2

In stage 2: The monopolist try to maximize its profit, which are determined by quantity q_1, q_2, residual demand, D_1, D_2 and costs $C(q_1, q_2)$:

$$\max_{p_2} p_2 q_2 - C(q_1 + q_2)$$

$$s.t.\ q_2 = D_2(p_1, p_2)$$

Stage 1

In stage 1: we use the idea of backward induction to calculate the monopolist maximization problem:

$$\max_{p_1} p_1 q_1 + p_2(p_1) q_2(p_1) - C(q_1 + q_2(p_1))$$

$$s.t.\ q_1 = D_1(p_1)$$

Model Extension

- There are some extensions to the model. First of all, Electricity market is a auction market repeated infinite many times, a typical set up of the repeated game. Using this model, we are able to identify potential collusion and other market manipulate behavior.

Empirical Test

Ito and Reguant (2016) Predicts Froward Market Premium

Under the assumption of monopolist exercise market power. The model predicts that there are premium between the real time market and day-head market.

- We shall use the estimated parameter from our repeated game model, we could simulate the counterfactual, when there is no collusion, and thus no market inefficiency. We are able to see how much welfare loss we encounter.

Empirical Exercise: Welfare Simulation

Procedure

1. Estimating the marginal cost of the plant level firm. We don't directly have the data of the firm's marginal cost at each level of supply, but we could based on the power plant's characteristics to calculate the marginal cost based on the method in Bushnell, Mansur and Saravia (2008).
2. Simulate the best response of Function, and test if there are deviation due to market structure.
3. Use the estimated parameter from our repeated game model, we could simulate the counterfactual, when there is no collusion, and thus no market inefficiency. We are able to see how much welfare loss we encounter.

Data

- Nordic Pool Trading Data - Rich Data Environment
  - Bidding Price and Volume of Each Companies
  - Intradday Data Set and Day-Ahead Data set
- Nordic Authority Estimation Data
  - Data of Demand Estimation of the regulation authority
  - Firm that are
- Plant Level Characteristics Data
  - Ideally we shall have plant level data to estimate capacity and marginal cost for electricity for each company, I am not sure if we do have this data set yet. But professor said that marginal cost data are not difficult to get.

Reference

Ito, Koichiro and Reguant, Mar
Sequential markets, market power, and arbitrage
American Economic Review, 106(7):1921–57, 2016.

Hortaçsu, et.al.
Does Strategic Ability Affect Efficiency? Evidence from Electricity Markets
National Bureau of Economic Research 2017.

Bushnell, James B., Erin T. Mansur, and Celeste Saravia.
"Vertical arrangements, market structure, and competition: An analysis of restructured US electricity markets."
American Economic Review 98, 1 : 237-66 2008

OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

Forecasting Stock Returns with Search Ranking

Hyun Ki Kim

Research Question

Does search trend precede price change?

순위	종목명	검색비율	현재가	전일비	등락률	거래량	시가	고가	저가	PER	ROE
1	금호타이어	2.43%	7,780	⬆ 1,790	+29.88%	45,121,409	7,240	7,780	6,700	-34.12	-3.21
2	NAVER	2.25%	785,000	0	0.00%	68,102	798,000	799,000	780,000	33.48	18.50
3	인스코비	1.91%	8,450	▼ 750	-8.15%	13,795,286	8,910	8,970	8,080	-1,408.33	-1.48
4	셀트리온	1.21%	299,000	▼ 5,000	-1.64%	857,981	299,500	304,500	297,000	93.58	17.84
5	코디엠	1.04%	1,230	▲ 20	+1.65%	97,408,897	1,275	1,370	1,155	-33.24	-9.28
6	텔콘	1.04%	15,600	▲ 150	+0.97%	10,278,395	14,950	16,350	14,850	-65.82	-17.37
7	한농화성	1.04%	6,180	⬆ 1,420	+29.83%	7,524,777	4,860	6,180	4,855	11.10	8.30
8	삼성바이오로직스	0.69%	485,500	▼ 500	-0.10%	127,365	484,000	489,500	480,000	-331.17	-2.41
9	제이준코스메틱	0.69%	23,150	▲ 1,600	+7.42%	3,606,017	21,300	23,400	21,100	21.16	25.67
10	삼성전자	0.69%	2,406,000	▼ 21,000	-0.87%	253,716	2,394,000	2,407,000	2,364,000	8.88	21.01
11	SK하이닉스	0.69%	81,800	▲ 500	+0.62%	3,546,192	80,300	81,800	78,700	5.60	36.80
12	아모레퍼시픽	0.52%	345,000	▲ 12,000	+3.60%	367,448	333,000	346,500	332,500	60.43	9.81
13	LG전자	0.52%	110,500	0	0.00%	375,711	109,500	112,000	109,000	11.58	13.69
14	라온시큐어	0.52%	4,015	▲ 110	+2.82%	1,339,573	3,875	4,080	3,830	37.88	23.82
15	카페24	0.52%	129,100	▲ 2,100	+1.65%	126,323	125,100	130,000	124,700	201.40	N/A
16	파라다이스	0.52%	24,350	▲ 750	+3.18%	4,192,209	23,450	24,900	23,250	-117.07	-1.86
17	KT	0.52%	27,500	▲ 100	+0.36%	525,157	27,250	27,750	27,200	15.06	4.12
18	에이치엘비	0.52%	76,200	▲ 4,500	+6.28%	1,399,700	70,200	76,200	70,100	-166.01	-19.92
19	애경산업	0.52%	41,750	▲ 3,650	+9.58%	2,498,521	37,500	42,850	37,150	41.30	26.08
20	한국전력	0.52%	34,950	▲ 1,700	+5.11%	4,538,306	33,200	35,350	33,000	17.28	1.81

Introduction

Naver:

1. Largest search engine in Korea
2. Minute-level data

Search ranking:

1. Heuristic for choosing a stock
2. Real-time web scraping

Background

Forecasting stock returns:

1. Twitter (Bollen et al., 2011)
2. Google Trend (Preis et al., 2013)
3. Wikipedia (Moat et al., 2013)

Limitations:

1. Trading cost
2. Backtesting
3. Risk-adjusted return

Rationale

Why people search:

1. Price fluctuation
2. Limited attention
3. Good or bad news

Need to measure:

1. Price volatility, newcomer
2. Company size, popularity
3. Price trend

Rationale

Why people search:

1. Price fluctuation
2. Limited attention
3. Good or bad news

Need to measure:

1. Price volatility, newcomer
2. Company size, popularity
3. Price trend

Method and Design

Data:

1. High frequency data
2. Training period (4/3 – 4/13)
3. Validation period (4/16 – 4/27)
4. Testing period (4/30 – 5/11)

Theory:

1. Supervised learning
2. Model selection
3. Forwardtesting
4. Risk-adjustment (Fama-French 5-factor model)

Method and Design

Data:

1. High frequency data
2. Training period (4/3 – 4/13)
3. Validation period (4/16 – 4/27)
4. Testing period (4/30 – 5/11)

Theory:

1. Supervised learning
2. Model selection
3. Forwardtesting
4. Risk-adjustment (Fama-French 5-factor model)

Q&A

Questions?

References

Bollen, J., Mao, H. & Zeng, X. **Twitter mood predicts the stock market.** *Journal of Computational Science* 2, 1–8 (2011).

Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E. & Preis, T. **Quantifying Wikipedia usage patterns before stock market moves.** *Scientific Report* 3, 1801 (2013).

Preis, T., Moat, H. S. & Stanley, H. E. **Quantifying Trading Behavior in Financial Markets Using Google Trends.** *Scientific Report* 3, 1684 (2013).

OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

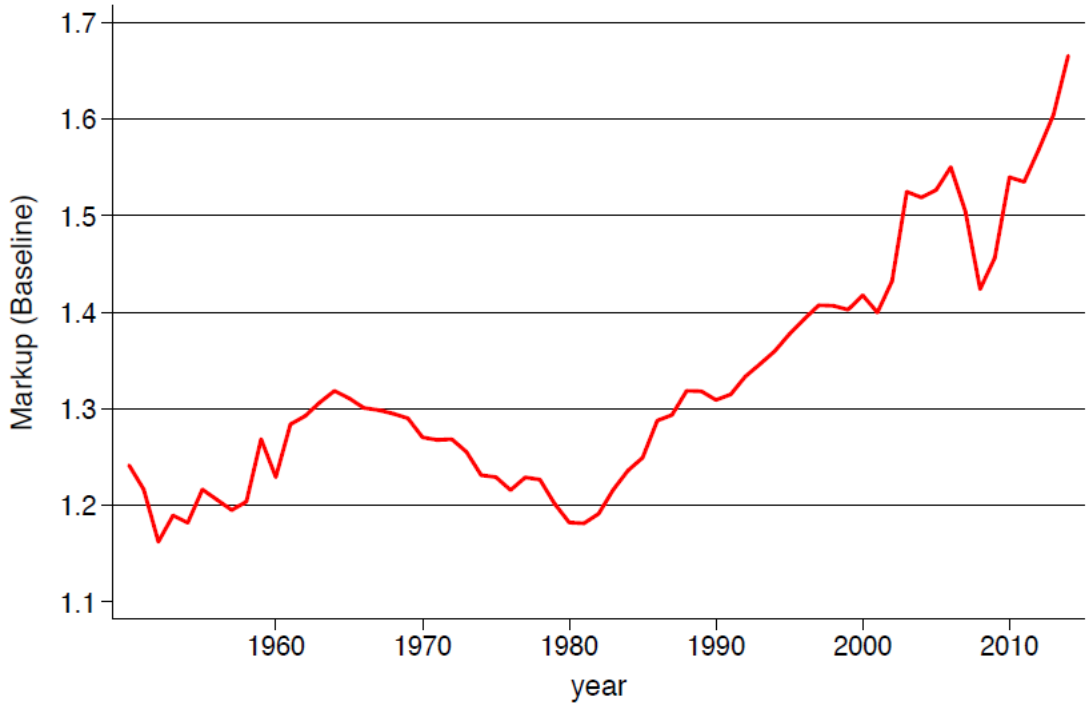
The Rise of Market Power ?

Chen Anhua, Peter

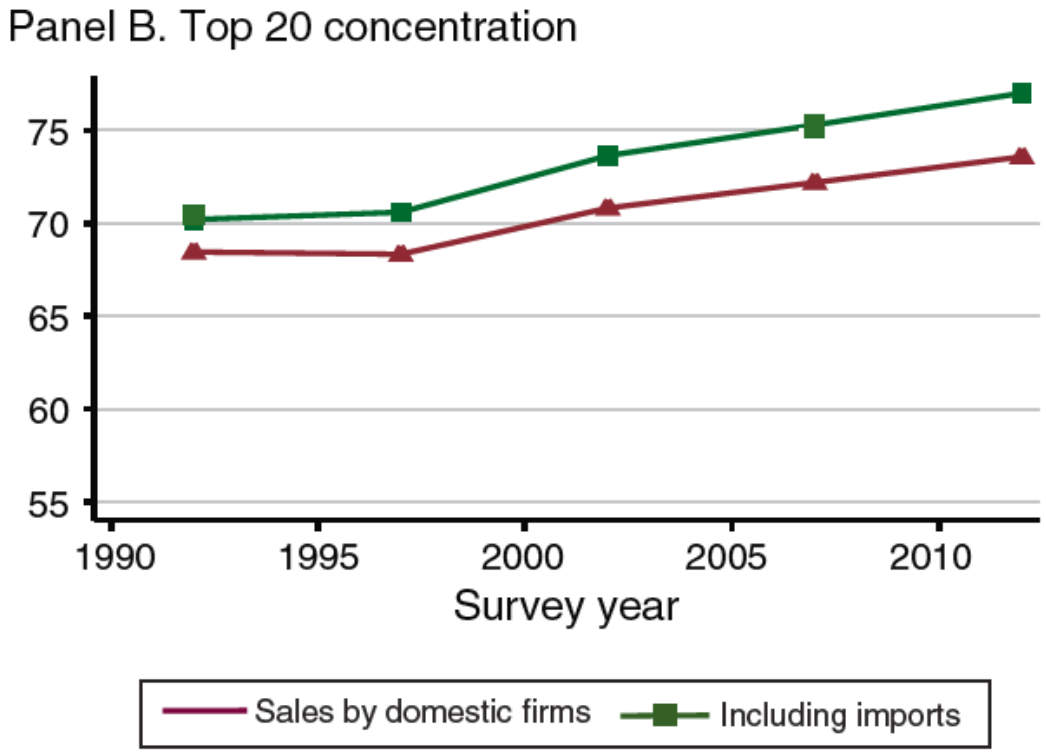
MACSS

“Evidence” on rise of market power

De Loecker & Eeckhout (2017)

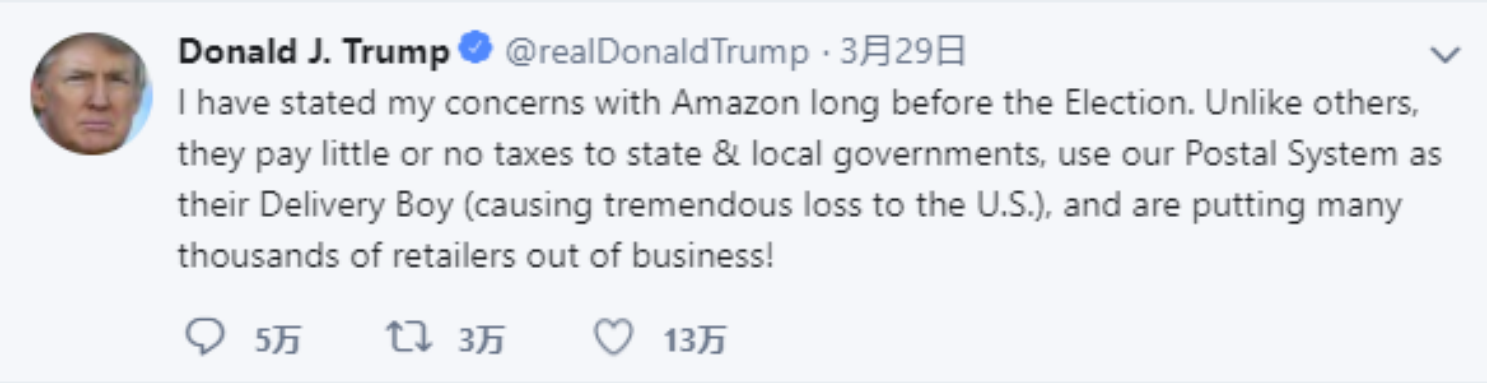


Autor et al. (2017)



Implication of change in market power

- Call for anti-trust policy



- Implication for macro inequality
  - A consensus on global decline in labor share
- Implication for change in productivity growth
  - A consensus on slowing productivity growth in recent decade(s)

Research Questions

- Are those evidence showing rising market power robust?
  ---- Can we really find evidence on rising market power?
  Or is it capturing something else?

- If there is strong evidence on rise of market power (or any other market dynamics), where does it come from?
  - "Star firms"? Specific industry? Aggregation issue? …

- A heterogeneous-firm model to explain the market dynamics

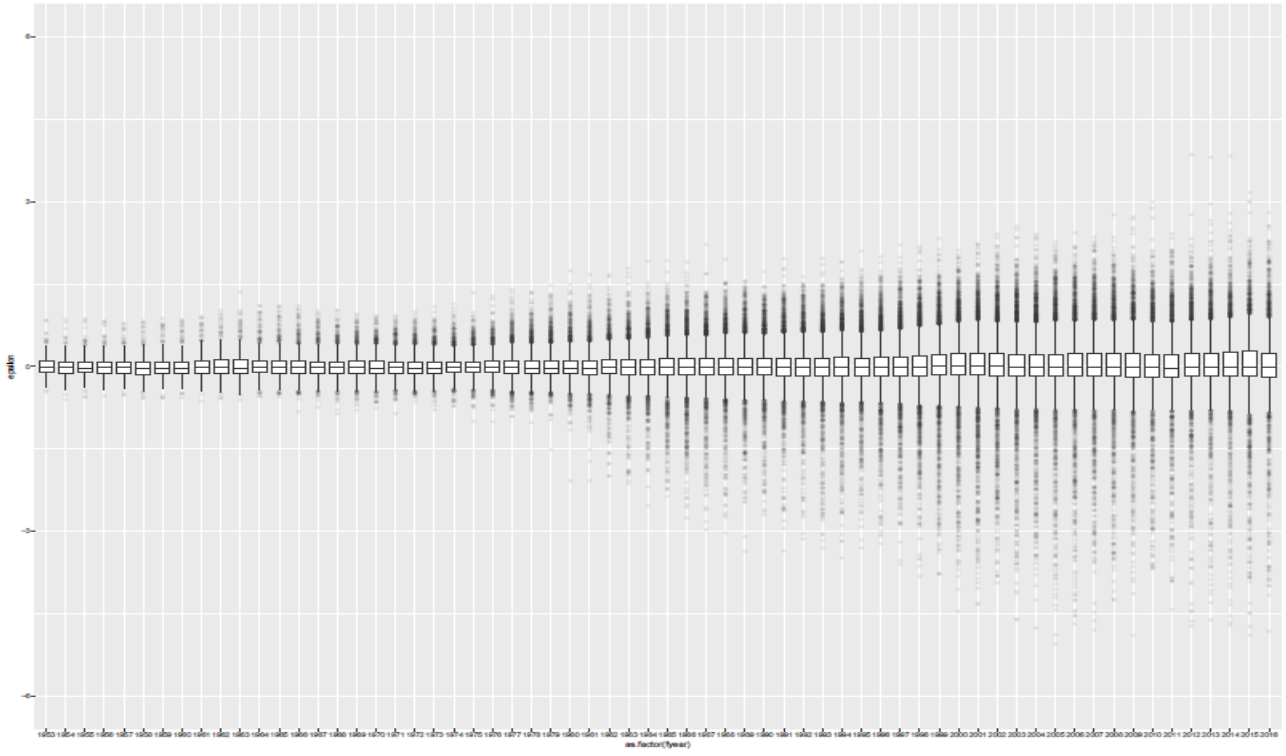
Implementation plan

- Firm producing behavior
  - Time-serial
  - From the most-nonparametric estimation to more structural estimation
- Market structure
  - Cross-sectional/Panel
  - Decomposition: is it within industry/firm or reallocation?
- Data
  - Compustat

Firm producing behavior

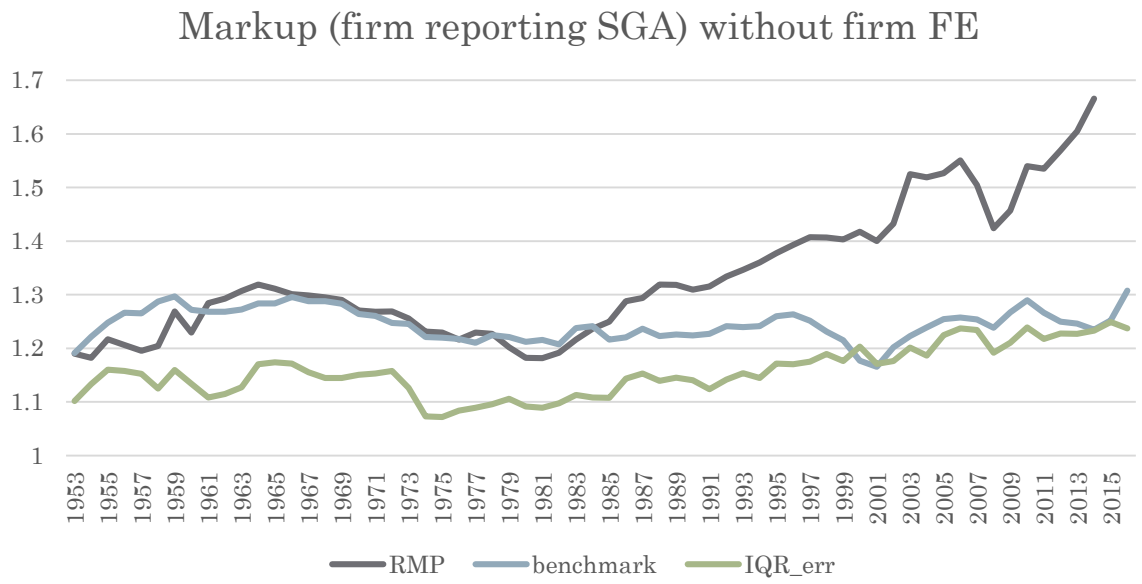
- Even using a very flexible production functional form (and controlling for time fix effect), we still witness an increase in second-moment of residuals over time.

$$q_{it} = \phi_t(v_{it}, k_{it}) + \epsilon_{it}$$



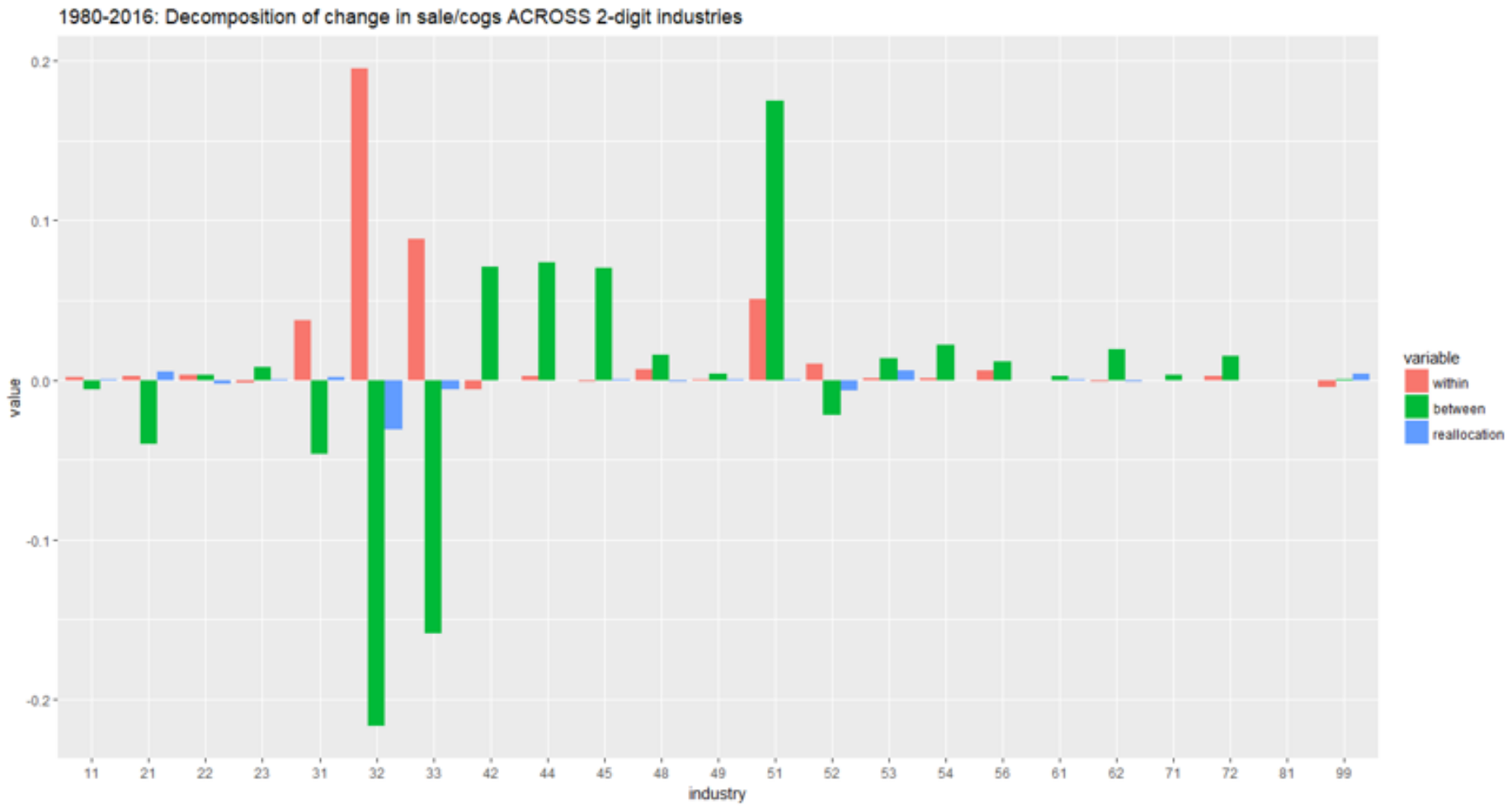
Firm producing behavior (continued)

- Outliers are important
  - If been corrected or excluded, the "rise of markup" is gone. These outliers/unexplained behavior, really drive the "rise of markup" (market power)



Market Structure: decomposition

- Within firm/industry change or reallocation ?



Next step

- A more detailed probe into firm producing behavior
  - Structural model
- Link between unexplained producing behavior with market structure (reallocation)
- Extension to other countries

Reference

- De Loecker, Jan and Jan Eeckhout, *"The Rise of Market Power and the Macroeconomic Implications,"* Working Paper 23687, National Bureau of Economic Research, August 2017.
- Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2017): *"Concentrating on the Fall of the Labor Share,"* National Bureau of Economic Research Working Paper.

OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

The effect of immigrants on American labor market outcomes

Presentation of Research Proposal

Ruxin Chen

The University of Chicago, Master of Computational Social Science

Spring 2018

Research Question

- Identify the effect of immigrants on American labor market outcomes
  - Hypothesis: inflow of immigrants (presumably less-educated) might decrease the wage and employment of less-skilled American workers
- Capture heterogeneous effect on different subgroups of American workers

Previous Literature

- LaLonde, Robert J., and Robert H. Topel. "Immigrants in the American labor market: Quality, assimilation, and distributional effects." The American economic review 81, no. 2 (1991): 297-302.
    - Policy concerns were raised in 1980s when there was substantial inflow of immigrants. The empirical results showed that these immigrants assimilate rapidly and have insignificant/trivial effect on the non-immigrants (wage, unemployment). The long-term earning potential of the immigrants will be much like their ethnically similar natives.

Previous Literature

- Foged, Mette, and Giovanni Peri. "Immigrants' effect on native workers: New analysis on longitudinal data." American Economic Journal: Applied Economics 8, no. 2 (2016): 1-34.
  - An increase in the supply of refugee-country immigrants pushed less educated native workers to pursue less manual-intensive occupations, hence the wages for low-skilled natives are increased.

Motivation

- Why is it interesting to study?
- Donald Trump announced his reform in immigration policy
  1. Build "wall" against Mexico immigrants: increase fees for worker visa from Mexico
  2. Enforcement of immigration law: triple the number of ICE officers, return of all criminal aliens, enhanced penalties for overstaying a visa
  3. Put American workers first: increase barriers for H1-B visa, the termination of J-1 visa.

Data

- Dataset: IPUMS-USA 2001-2016
- Exogenous policy shock: George W. Bush signed the Secure Fence Act of 2006
  - Build a wall of hundreds of miles of additional fencing along the Southern border (reported completion in 2015)
  - Authorize vehicle barriers, checkpoints and advanced technology to prevent people from entering illegally

Model Setup: Conventional D-in-D model

- Difference-in-difference model

$$y_{ist} = \gamma_s + \lambda_t + \delta D + \beta X_{ist} + \epsilon_{ist} \tag{1}$$

  where y_{ist} is the labor market outcomes including wage and employment status. D is a dummy variable indicating the policy shock, and X_{ist} is a set of covariates (demographic characteristics: age, marital status, education attainment and occupation etc.).
- δ is the ATE (average treatment effect) of a tightened immigration policy on the labor market outcomes.

Model Setup: Spline Regression

- Question: does the policy have the same effect on different subgroups of American labor force?
- Hypothesis
  - heterogeneous effect on labor force with different education attainment
  - Spline Regression

$$y_{ist} = \gamma_s + \lambda_t + \delta D + \beta X_{ist} + \theta b(edu_{ist}, \kappa) + \epsilon_{ist} \tag{2}$$

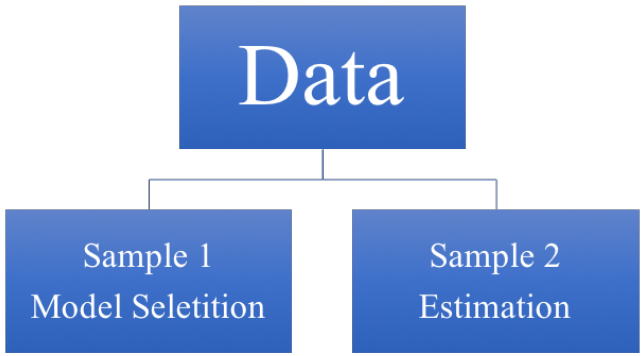
where $b(edu_{ist}, \kappa)$ is the basis of spline.

Model Setup: 'honest approach'

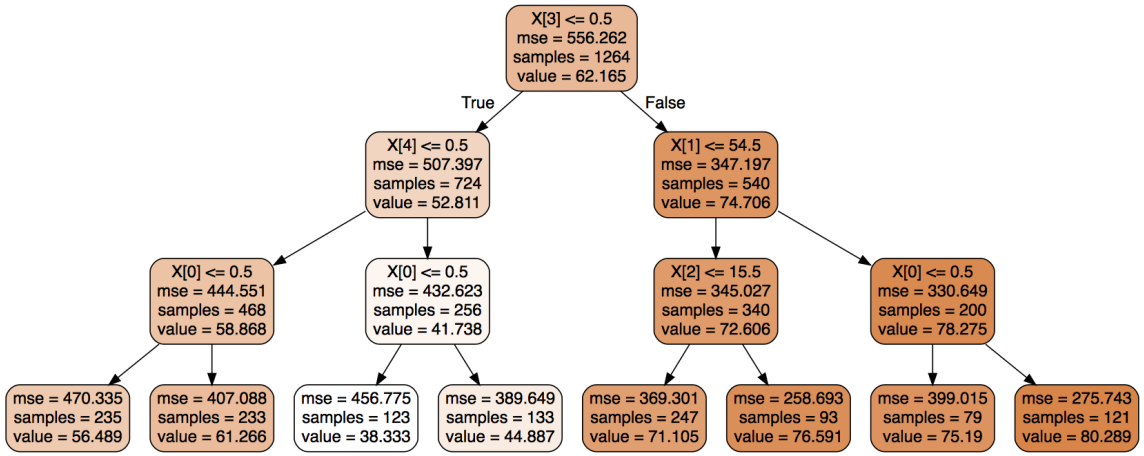
- Susan Athey: combine supervised machine learning techniques to estimate more precisely average treatment effects in different subpopulations
- Two-step
  - Model selection: find the partition using regression tree method
  - Estimation: estimate treatment effects within leaves of the tree
  - Use different samples in the two steps so the estimation is unbiased

Honest Approach: Flow Chart

- Sample-splitting



- Expected result of regression tree



OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

Forecasting U.S. recessions with a large number of predictors

Shuting Chen

MACSS Project Proposal
April 4, 2018

Introduction - Motivation of the project

- **Predicting recessions using probit models**
  - Estrella and Mishkin (1998) predict the recession indicator y_t with a static probit model
  - Kauppi and Saikkonen (2008) propose the dynamic autoregressive specification
    - Improve the forecasting performance by including lags of π_t, the usual probit latent variable
    - Only use interest rate spread as the driving predictor
    - Use the same variables for different forecasting horizons
- **Using a large panel of predictors**
  - Stock and Watson (2002) forecast real-valued economic activities with a large set of predictors
  - Chen, Iqbal, and Li (2011) demonstrate that including principal components improves dynamic probit models for predicting recessions

Introduction - Motivation of the project

- **Predicting recessions using probit models**
  - Estrella and Mishkin (1998) predict the recession indicator y_t with a static probit model
  - Kauppi and Saikkonen (2008) propose the dynamic autoregressive specification
    - Improve the forecasting performance by including lags of π_t, the usual probit latent variable
    - Only use interest rate spread as the driving predictor
    - Use the same variables for different forecasting horizons
- **Using a large panel of predictors**
  - Stock and Watson (2002) forecast real-valued economic activities with a large set of predictors
  - Chen, Iqbal, and Li (2011) demonstrate that including principal components improves dynamic probit models for predicting recessions

Introduction - Purpose of the project

- Extend Kauppi and Saikkonen's (2008) dynamic autoregressive probit model by
  - replacing the interest rate spread with a few factors or predictors selected from a large panel of predictors
  - augmenting nonlinearity with factors selected from a pool set of predictors and their squares
- Ideal contributions:
  - Identify most informative predictors/factors
  - Understand whether the selected predictors/factors are horizon specific

Models

- **How to predict the probability of recessions occurring at time t?**
  - y_t - recession indicator, having a Bernoulli distribution with conditional probability p_t
  - Using probit model: $\Phi(\pi_t) = p_t$
  - Model specification:

$$\pi_t = \omega + y_{t-1}\alpha + \pi_{t-1}\delta + f'_{t-1}\beta$$

  - How to determine f_{t-1}?
    - Option 1: factors extracted from a large set of predictors by principal component analysis
    - Option 2: predictors selected from the same set of predictors by Adaboost (Ng, 2014)

Models

- **How to predict the probability of recessions occurring at time t?**
  - y_t - recession indicator, having a Bernoulli distribution with conditional probability p_t
  - Using probit model: $\Phi(\pi_t) = p_t$
  - Model specification:

$$\pi_t = \omega + y_{t-1}\alpha + \pi_{t-1}\delta + f'_{t-1}\beta$$

  - How to determine f_{t-1}?
    - Option 1: factors extracted from a large set of predictors by principal component analysis
    - Option 2: predictors selected from the same set of predictors by Adaboost (Ng, 2014)

Estimation and forecasting

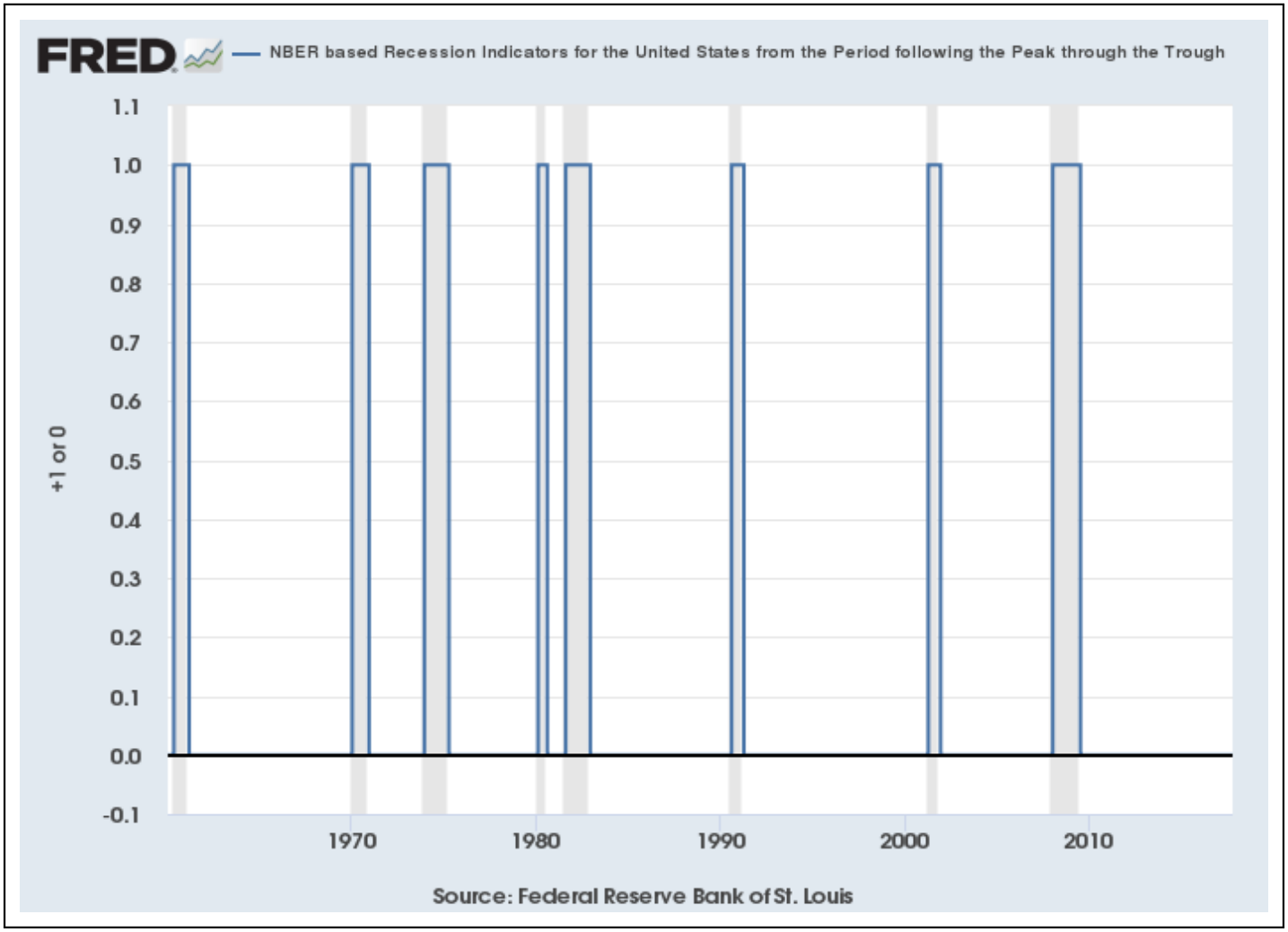
- Estimate parameters, $\theta = \left(\omega, \alpha, \delta, \beta', \pi_0\right)'$ in the specification of π_t, by maximum likelihood estimation
- Forecasting procedures
  - Forecasting horizons: h = 3, 6, 12 months ahead
  - Direct approach vs iterative approach
    - Iterative approach: need to consider every possible path of y_t through the $h-1$ months
    - Potential drawbacks

Data description

Use data from two different sources

- U.S. business cycle expansion and contraction dates announced by NBER
  - The first month following a peak month defines the first recession month
  - The last month of a trough defines the last recession month
- Monthly frequency macroeconomic series from FRED-MD
  - disposed by McCracken and Ng with the data desk at the Federal Reserve Bank of St. Louis
  - including 123 variables over 1960:M1 to 2017:M10
  - covering macroeconomic and financial series such as real activity indicators, interest rate indices and price indices

Data Figure - U.S. Business Cycle Dates by NBER



Data Table - Interest and exchange rates

Table 1: Summary Statistics for 10 Predictors
Classified as Interest and Exchange Rates

Fred	Description	Mean	S. D.
FEDFUNDS	Effective Federal Funds Rate	5.1085	3.6843
CP3Mx	3-Month AA Financial Commercial Paper Rate	5.2132	3.4627
TB3MS	3-Month Treasury Bill	4.6322	3.1724
TB6MS	6-Month Treasury Bill	4.7704	3.1582
GS1	1-Year Treasury Rate	5.1491	3.3706
GS5	5-Year Treasury Rate	5.8484	3.0684
GS10	10-Year Treasury Rate	6.1815	2.8436
AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS	2.0696	1.9733
BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS	3.0836	2.0758
TB3SMFFM	3-Month Treasury C Minus FEDFUNDS	-0.4764	0.7146

OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

The Trade Between Money and Time
Model Selection for Intertemporal Choice

Xi Chen

MACSS Project Proposal

April 4, 2018

Introduction

Decisions involving consequences at different time points are referred to as *intertemporal choice* (Frederick, Loewenstein, & O'Donoghue, 2002).

Intertemporal choice is also known as *temporal discounting*, *delay discounting*, *time preference*, or *time discounting*.

What model has the best prediction for money-earlier-or-later (MEL) decisions / intertemporal choices?

Literature

- Exponential model
- Hyperbolic model
- Quasi-Hyperbolic model / $\beta - \delta$ discounting model
- Fixed cost model (Benhabib, Bisin, & Schotter, 2009)
- Exponential time (Roelofsman, 1996)
- Discounting-by-intervals (DBI) (Scholten and Read, 2006)
- Constant sensitivity (Ebert and Prelec, 2007)
- Discounting fractions (Read, 2001)
- Hyperboloid model (Green & Myerson, 1995)
- Generalized Hyperboloid model (Loewenstein & Prelec, 1992)
- Constant absolute decreasing impatience (Bleichrodt et al., 2009)

...

Data

Marzilli Ericson, K. M., White, J. M., Laibson, D., & Cohen, J. D. (2015). Money earlier or later? Simple heuristics explain intertemporal choices better than delay discounting does. Psychological science, 26(6), 826-833.

- The amount of data: 23500
- 940 participants; 25 questions for each participants
- Money range: \$0.01 to \$100,000.00
- Time range: 0 weeks to 6 weeks

Experiment

- Standard *money-earlier-or-later*(MEL) format
(e.g., $5 today vs. $10 in 4 weeks)

- Absolute money value, delay framing
(e.g., $5 today vs. $5 plus an additional $5 in 4 weeks)

- Relative money value, delay framing
(e.g., $5 today vs. $5 plus an additional 100% in 4 weeks)

- Absolute money value, speedup framing
(e.g., $10 in 4 weeks vs. $10 minus $5 today)

- Relative money value, speedup framing
(e.g., $10 in 4 weeks vs. $10 minus 50% an additional 100 % today)

ITCH Model

A simple heuristics model, *intertemporal choice heuristic* (ITCH) model outperformed the traditional utility-discounting models. (Ericson, White, Laibson, & Cohen, 2015)

The outcome is binary variable:
0 represents smaller sooner option, 1 represents larger later option.

L(x) represents the inverse logistic function of x:
$L(x) = (1 + e^{-x})^{-1}$

ITCH model:
$P(LL) = L(\beta_I + \beta_{xA}(x_2 - x_1) + \beta_{xR}\frac{x_2 - x_1}{x^*} + \beta_{tA}(t_2 - t_1) + \beta_{tR}\frac{t_2 - t_1}{t^*})$

Models

- Exponential model: $P(LL) = L(a(x_2\delta^{t_2} - x_1\delta^{t_1}))$

- Hyperbolic model: $P(LL) = L(a(x_2(1+\alpha t_2)^{-1} - x_1(1+\alpha t_1)^{-1}))$

- Quasi-hyperbolic model:
$P(LL) = L(a(x_2\beta^{I(t_2>0)}\delta^{t_2} - x_1\beta^{I(t_1>0)}\delta^{t_1}))$

- Tradeoff model:
$P(LL) = L(a((log(1+\gamma_x x_2)/\gamma_x - log(1+\gamma_x x_1))/\gamma_x - k(log(1+\gamma_t t_2 - log(1+\gamma_t t_1)/\gamma_t))))$

- DRIFT model:
$P(LL) =$
$L(\beta_0 + \beta_1(x_2 - x_1) + \beta_2\frac{x_2-x_1}{x_1} + \beta_3((\frac{x_2}{x_1})^{\frac{1}{t_2-t_1}} - 1) + \beta_4(t_2 - t_1))$

- ITCH model:
$P(LL) = L(\beta_I + \beta_{xA}(x_2 - x_1) + \beta_{xR}\frac{x_2-x_1}{x^*} + \beta_{tA}(t_2 - t_1) + \beta_{tR}\frac{t_2-t_1}{t^*})$

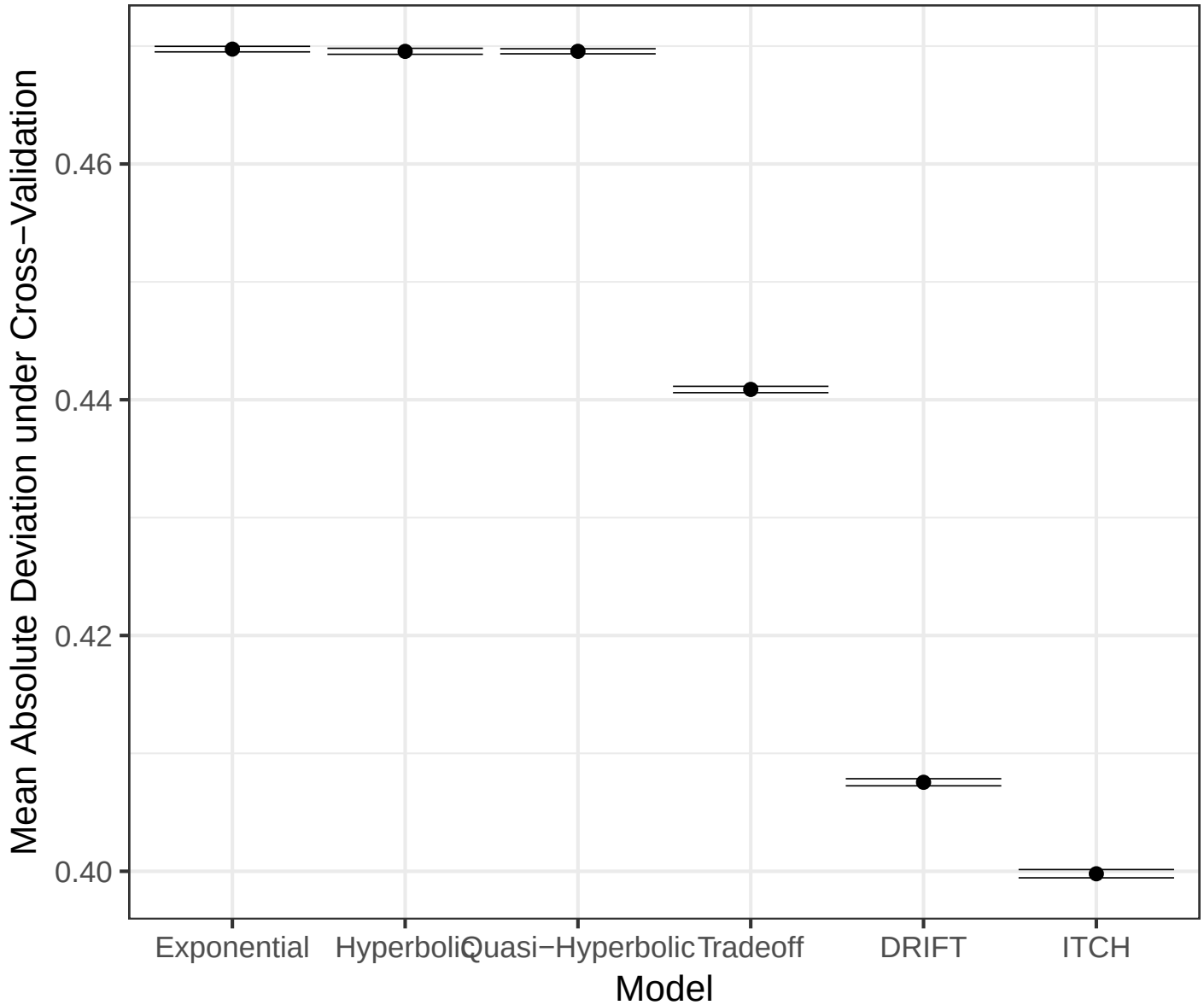
Method

- What are the features in each of these models?
- Why heuristic models outperform traditional utility-discounting models?
- Why ITCH model outperforms the other models?
- What are the most important terms?
- Develop a new model which is nested in the traditional utility-discounting models and the heuristics model

- Cross-validation analysis
- Maximum likelihood estimation techniques

...

Results


Mean Absolute Deviation under Cross−Validation
0.46
0.44
0.42
0.40
Exponential
Hyperboli
Quasi−Hyperbolic
Tradeoff
DRIFT
ITCH
Model

Future directions

- May design new experiments based on specific conditions to compare the models' performance across conditions
- May design new experiments to examine the indifference point between the larger-later option and smaller-sooner option

...

OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

Endogenous Health Care in Overlapping Generations Model:

Simulation for Health Care and Economy

Fiona Fan

MACSS Project Proposal
April 4, 2018

Motivation

- **Health is an overlapping generations thing – Grossman Model (Grossman, 1972)**

$$H_{t+1} = (1 - \delta)(H_t + I_t)$$

- **Key Features of Model**
  - Agents in the model choose health care to consume, in addition to consumption and savings.
  - Consumption of health care at time t boosts labor productivity at time t+1 (consistent with Grossman).
  - Insurance in forms of Medicare (young people pay for old people's health insurance). Other kinda insurance could be added (Hashimoto and Tabata, 2010)
  - Production of health care VS non-health-care good.

Motivation

- **Health is an overlapping generations thing – Grossman Model (Grossman, 1972)**

$$H_{t+1} = (1 - \delta)(H_t + I_t)$$

- **Key Features of Model**
  - Agents in the model choose health care to consume, in addition to consumption and savings.
  - Consumption of health care at time t boosts labor productivity at time t+1 (consistent with Grossman).
  - Insurance in forms of Medicare (young people pay for old people's health insurance). Other kinda insurance could be added (Hashimoto and Tabata, 2010)
  - Production of health care VS non-health-care good.

What we can learn from simulation

- With the repeal of mandate (decreased insurance), what will happen to economic growth/ labor participation rate in healthcare VS non-health-care/ consumption of health care/ consumption of non-health-care, etc?
- What effect of aging/ decreased mortality rate affect health care consumption/ spending?

Demographics

$$\omega_{1,t+1} = (1-\rho_o)\sum_{s=1}^{E+S} f_s\omega_{s,t} + i_1\omega_{1,t}, \quad \forall t \tag{1}$$

$$\omega_{s+1,t+1} = (1-\rho_s)\omega_{s,t}(1+\zeta_s(h_{s,t})) + i_{s+1}\omega_{s+1,t}, \quad \forall t \quad \text{and} \quad 1 \leq s \leq E+S-1 \tag{2}$$

$$N_t = \sum_{s=1}^{E+S} \omega_{s,t} \qquad \tilde{N}_t = \sum_{s=E+1}^{E+S} \omega_{s,t} \tag{3}$$

$$n_{s,t} = \begin{cases} 1, & E+1 \leq s \leq E + round(\frac{2S}{3}) \\ 0.2, & s \geq E + round(\frac{2S}{3}) \end{cases} \tag{4}$$

Households

- Budget Constraints

$$c_{s,t} + b_{s+1,t+1} + h_{s,t} + m_{s,t} = (1 + r_t)b_{s,t} + w_t n_s, \quad \forall E + 1 \leq s < round(\frac{2S}{3}) \tag{5}$$

$$c_{s,t} + b_{s+1,t+1} + h_{s,t} = (1 + r_t)b_{s,t} + w_t n_s + \frac{\sum_{s=E+1}^{round(\frac{2S}{3})} m_{s,t}}{round(\frac{S}{3})}, \quad \forall s \geq round(\frac{2S}{3}) \tag{6}$$

Households

- Budget Constraints

$$c_{s,t} + b_{s+1,t+1} + h_{s,t} + m_{s,t} = (1 + r_t)b_{s,t} + w_t n_s, \quad \forall E+1 \leq s < round(\frac{2S}{3}) \tag{5}$$

$$c_{s,t} + b_{s+1,t+1} + h_{s,t} = (1 + r_t)b_{s,t} + w_t n_s + \frac{\sum_{s=E+1}^{round(\frac{2S}{3})} m_{s,t}}{round(\frac{S}{3})}, \quad \forall s \geq round(\frac{2S}{3}) \tag{6}$$

- Utility Maximization

$$\max_{\substack{\{c_{s,t+s-1}, h_{s,t+s-1}\}_{s=E+1}^{E+S}, \\ \{b_{s+1,t+s}\}_{s=E+1}^{E+S-1}}} \sum_{s=E+1}^{E+S} \beta^{s-E-1} [\Pi_{n=E}^{s-1}(1-\rho_n)] U(c_{s,t+s-E-1}, h_{s,t+s-E-1}) \quad \forall s,$$

$s.t.$ 5 and 6, and $b_{E+1,t}, b_{E+S+1,t} = 0 \quad \forall t$ and $c_{s,t} \geq 0 \quad \forall s, t,$

where $U = ln(c_{s,t+s-E-1}^{\gamma} h_{t+1}^{1-\gamma})$

Firm

$$Y_t^H = A^H L_t^H \tag{7}$$

$$Y_t^N = F(K_t, L_t^N) = A^N K_t^{\alpha} (e^{g_y t} L_t)^{1-\alpha} \tag{8}$$

$$r_t = \alpha(\frac{Y_t}{K_t}) - \delta \tag{9}$$

$$w_t = (1-\alpha)(\frac{Y_t}{K_t}) \tag{10}$$

Market Clearing

$$L_t^H + L_t^N = \sum_{s=E+1}^{E+S} \omega_{s,t} n_s (1 + \zeta_s(h_{s,t})) \tag{11}$$

$$K_t = \sum_{s=E+2}^{E+S} (\omega_{s-1,t-1} b_{s,t} + i_s \omega_{s,t-1} b_{s,t}) \tag{12}$$

Market Clearing

$$L_t^H + L_t^N = \sum_{s=E+1}^{E+S} \omega_{s,t} n_s (1 + \zeta_s(h_{s,t})) \tag{11}$$

$$K_t = \sum_{s=E+2}^{E+S} (\omega_{s-1,t-1} b_{s,t} + i_s \omega_{s,t-1} b_{s,t}) \tag{12}$$

$$Y_t^N = C_t + I_t - \sum_{s=E+2}^{E+S} i_s \omega_{s,t} b_{s,t+1} \quad \text{where} \tag{13}$$

$$I_t = K_{t+1} - (1 - \delta) K_t \quad \text{and,} \tag{14}$$

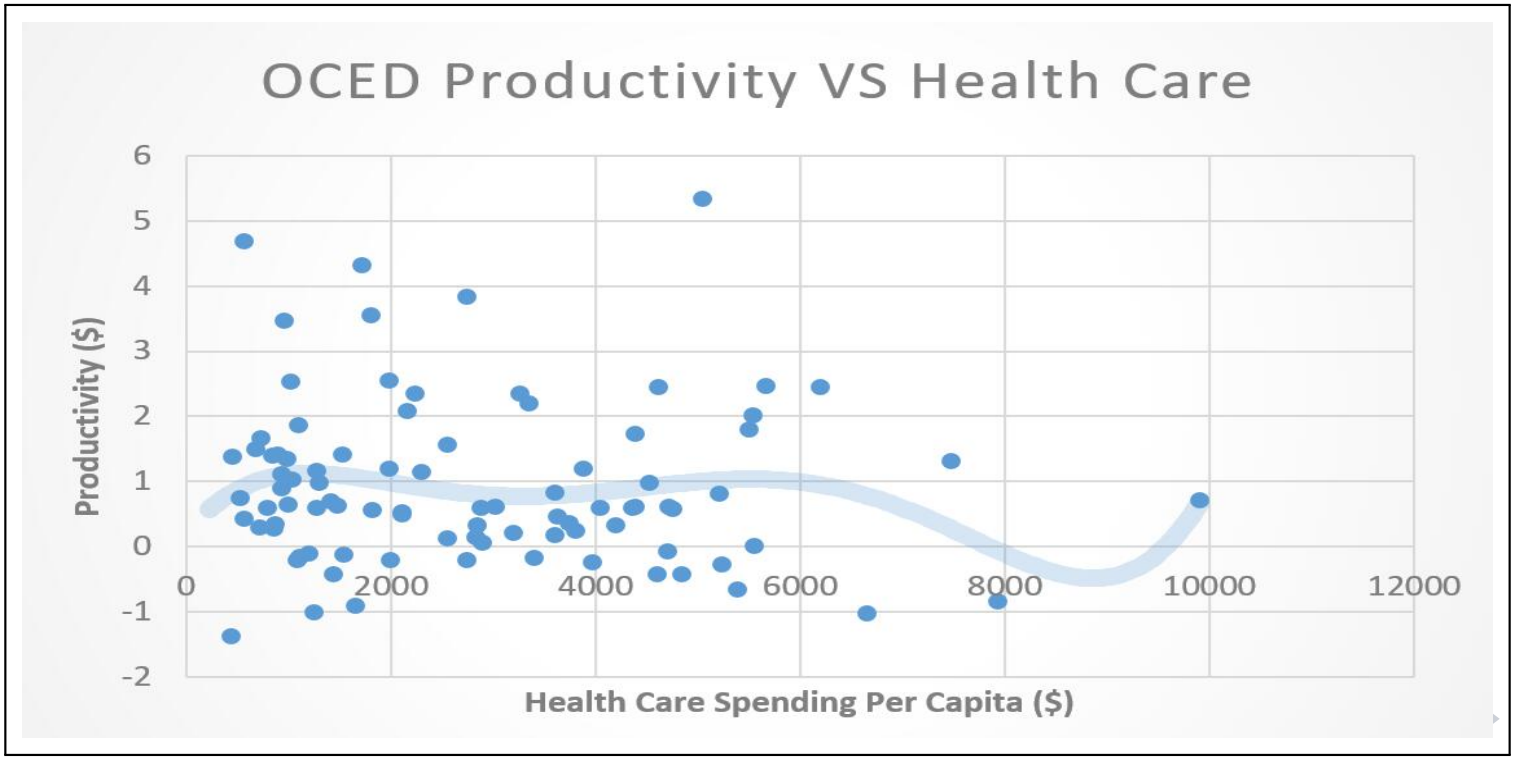
$$C_t = \sum_{s=E+1}^{E+S} \omega_{s,t} c_{s,t} \tag{15}$$

$$Y_t^H = \sum_{s=E+1}^{E+S} \omega_{s,t} h_{s,t} \tag{16}$$

Calibration and Simulation

Calibrations

- ζ_s: OCED data on increase in productivity VS health care spending/capita
- ρ, f, i: US Census data



OK. You're done.

Let's get the next person up.

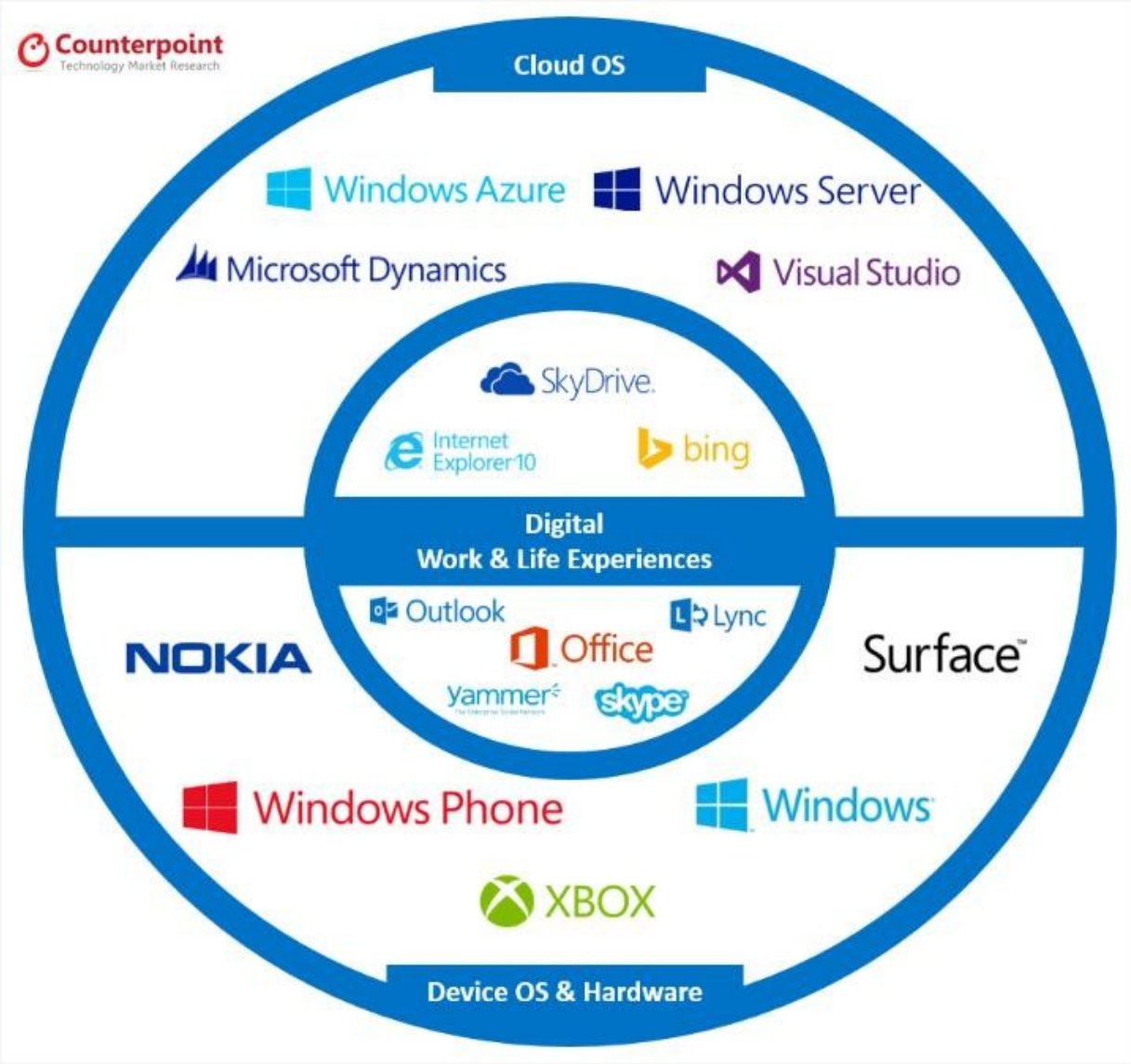
Your presentation starts in 30 seconds

Firm Expansion through Innovation Network

Zhiyu Fu

M.A. in Computational Social Science

April 3, 2018



https://www.linkedin.com/pulse/20140721180246-21534059-microsoft-mobile-first-nokia-last

Questions

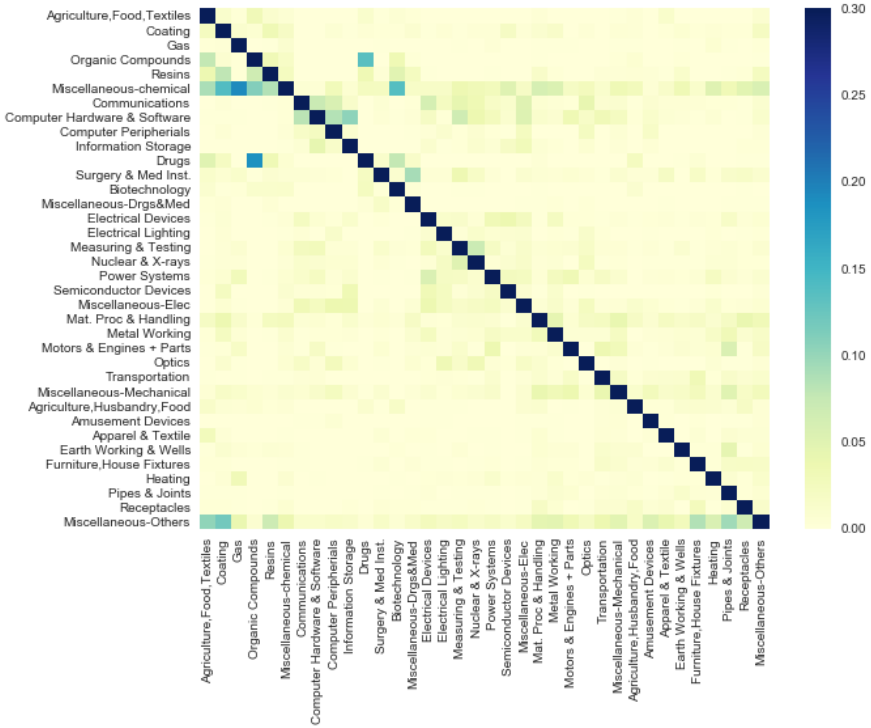
- Why are firms diversifying their portfolio?
  - Agency theory: e.g., Jensen & Meckling (1976); Williamson (1975);
  - Resource-Based theory: See Wan et al. (2011) for a review;
  - Within-firm R&D spillover: Klette (1996); MacDonald (1985); Jovanovic and Gilbert (1993); .
- Where do they expand?
  - Related industries: e.g., product similarity (Berry, 1975); similarity of customers (Lemelin, 1982); similar intensity of R&D (MacDonald, 1985).

My answer: Firms expand into new industries due to the R&D spillovers through innovation network.

Innovation Network

Innovation network captures how knowledge is shared and transferred across knowledge fields (Acemoglu et al., 2016).

Figure: Citation Matrix ($\alpha_{ij} = \frac{\text{Citation}_{i \to j}}{\sum_h \text{Citation}_{i \to h}}$)



Adapted from Acemoglu et al. (2016)

Results and Contributions

- Result:
  - Firms tend to expand into indutries that are closely connected in innovation networks, both upstream and downstream.
- Contribution:
  - Provide a new explanation and evidence of corporate diversification in a dynamic perspective with detailed mechanism;
  - Propose a framework for endogenous growth models that incorporates hetergeneous industries and innovation networks;
  - Give policy suggestions under heterogeneous industries.

Empirical Evidence

Strategy

- Data: NBER patent and citation data 1963-1997 (U.S. non-government organization only)
- Identify "expansion":
  - Accumulation Period: Firms built their portfolios $\mathbf{m}_f$ during this period;
  - Expansion period: Once a firm holds patents in a new industry above certain thesholds (e.g., 30% of current portfolio), it is counted as an expansion;

Empirical Evidence

Index Construction

Network Strength:

- Upstream$_j = \frac{1}{N} \sum_{i \in \mathbf{m}_f} n_i \alpha_{ij}$
- Downstream$_j = \frac{1}{N} \sum_{i \in \mathbf{m}_f} n_i \alpha_{ji}$

$$V_{fji} = \beta_1 \text{Upstream}_{fj} + \beta_2 \text{Downstream}_{fj} + \beta_j \mathbf{x}_{fj} + \epsilon_{fji}$$

Empirical Evidence

Results

Table: Network Strengths and Probability of Expanding into a Certain Industry

	Probability of Expanding into a certain industry			
	(1)	(2)	(3)	(4)
Upstream	14.009***	17.308***	11.625***	20.571***
	(1.364)	(3.215)	(4.248)	(4.466)
Downstream	16.311***	20.347***	20.437***	17.352***
	(1.360)	(3.235)	(4.473)	(4.354)
Avg. Marginal Effect: Up	0.367	0.45	0.3	0.532
Avg. Marginal Effect: Down	0.428	0.529	0.527	0.449
Min. Percentage	30%	30%	50%	30%
Min. Num	1	3	3	3
Accumulation (yr)	10	10	10	15
Expansion (yr)	10	10	10	5
Observations	2,641	585	296	270

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Theoretical Model

Main Features

Firms' R&D function:

$$x_i = [\prod_h s_h^{\alpha_{ih}}]^{\alpha}, \sum_{\alpha_{ih}} = 1$$

Consider a firm targeting in innovating in industry i:

- Upstream effect: It will invest higher in industry j where α_{ij} is larger;
- Downstream effect: It will invest heavily in i, generating higher innovation rate in industry j where α_{ji} is higher.

Theoretical Model

Simulation (Partial Equilibrium)

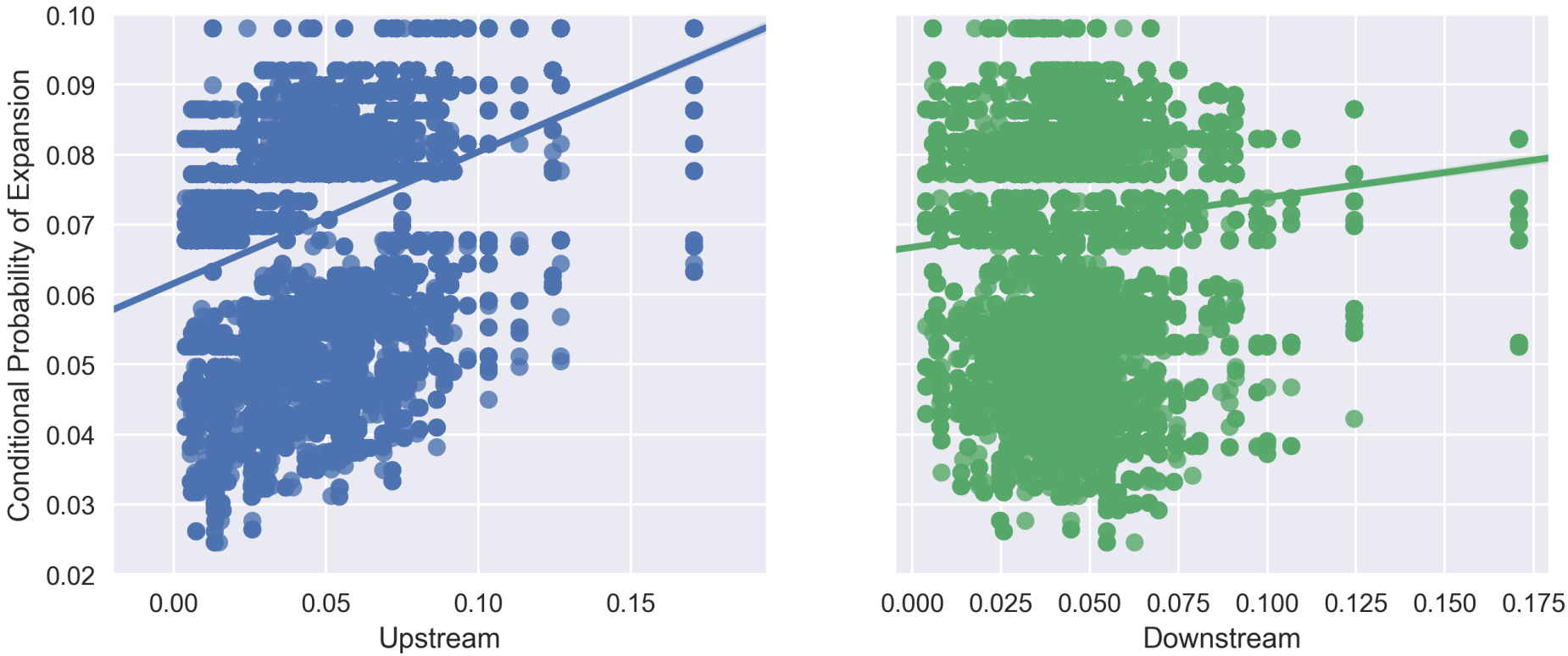


Figure: Simulated Results in Partial Equilibrium

Further Steps

- Derive general equilibria;
- Use Simulated Moment Method (SMM) to calibrate the model;
- Analyze policy instruments.

OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

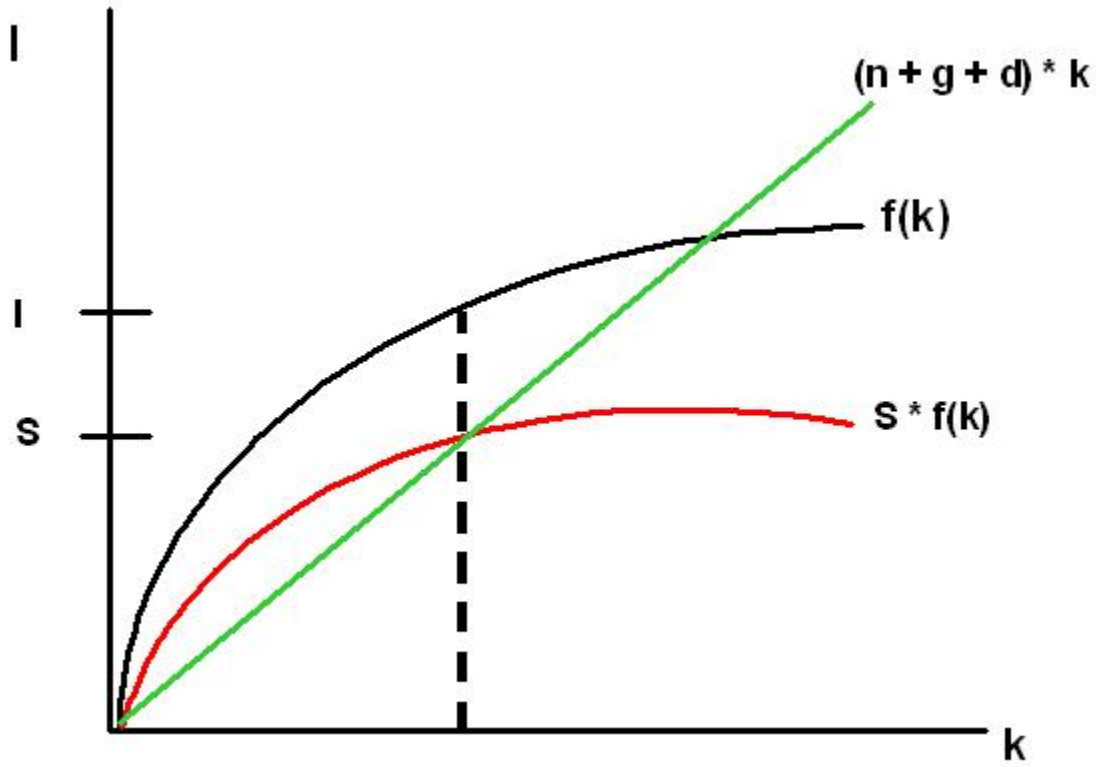
How do international economic activities between two countries impact a nation's innovation over time?

MACSS Student
Yuqian Gong (Nancy)

- National Innovation level before and after joining WTO?
- Heterogeneous impacts on emerging innovator nation and leading innovator nation?
- More trade and FDI lead to more innovation?

Background

Solow Growth Model

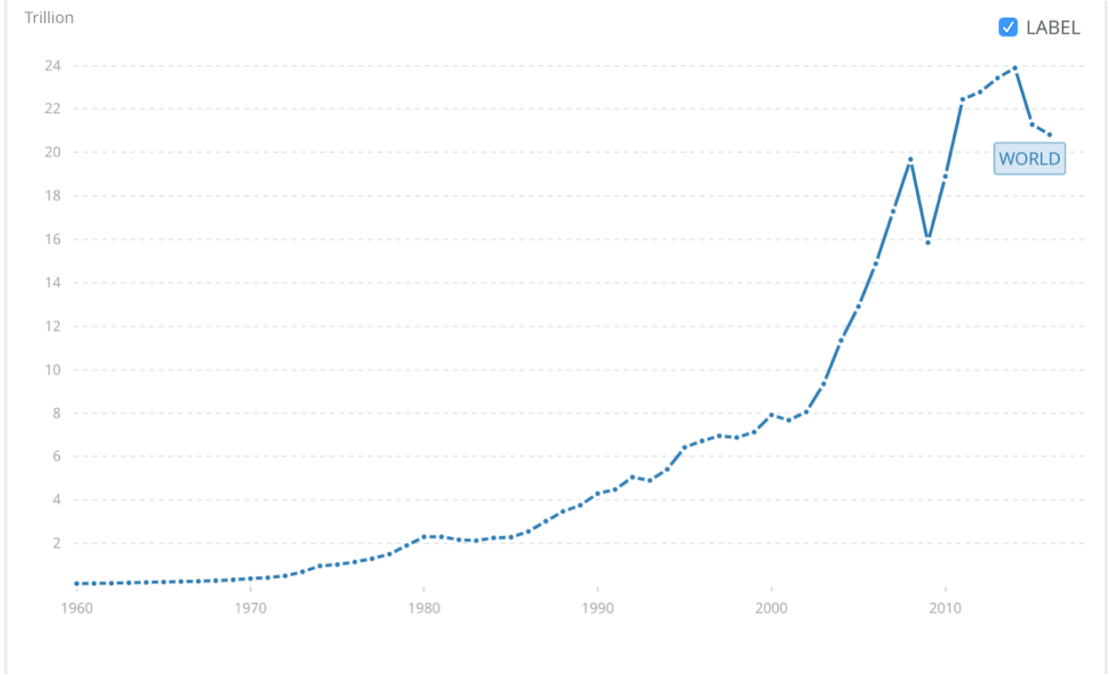


Previous literature on drivers of innovation process

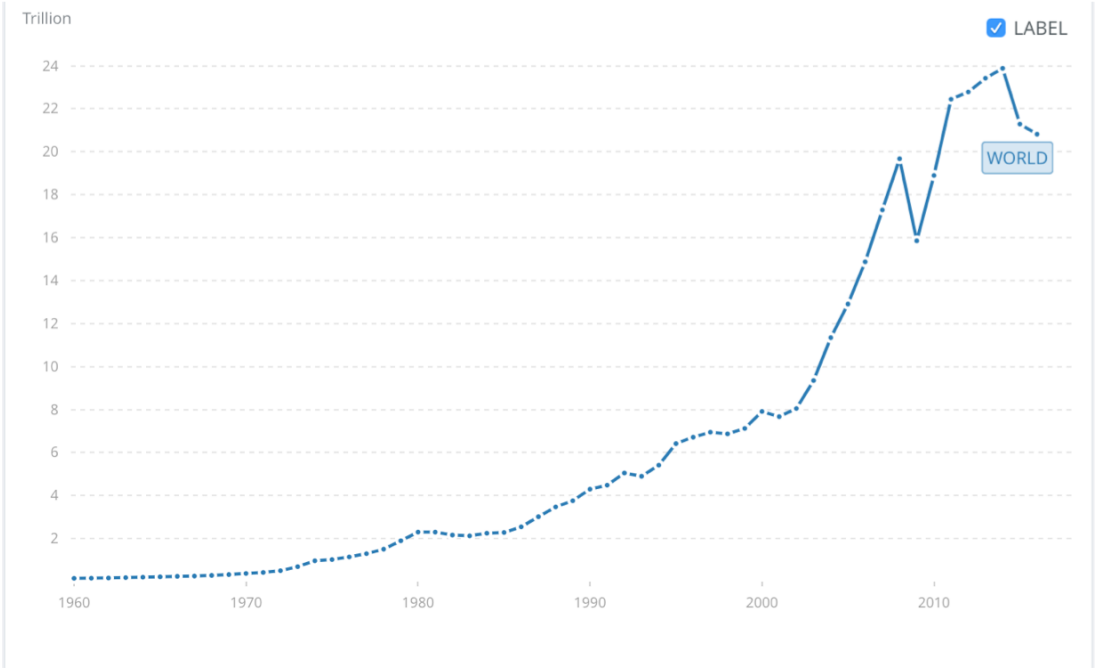
- National investments
- Other factors play a role:
  - Intensity of a nation's financial resources
  - Human capital to innovation activities
  - Accumulated technological capital
  - Supportive innovation environment in a nation's industrial clusters
  - …….

From closed economy to open economy

International Trade Flows (US$)



Foreign Direct Investment, net inflows (US$)



Model by Furman & Hayes

$$\bar{A}_{j,t} = (X_{j,t}^{INF}, Y_{j,t}^{CLUS}, Z_{j,t}^{LINK}) H_{j,t} A_{j,t}$$

Determinants of national innovative capacity:

(a) Common innovation infrastructure
(e.g. Patent Stock, GDP, Education expenditure)
(b) Cluster-specific innovation environment
(c) Quality linkage between the two above

Variables of Interest:

$\bar{A}$: Flow of innovations
X: Level of resource commitments and policy choices that constitute the innovation structure
Y: Environments for innovation in a country‘s industrial clusters
Z: Strength of linkages between common infrastructure and nation’s industrial clusters
H: Total level of human capital and labor recourses
A: Stock of knowledge

Source: Furman, J. L., & Hayes, R. (2004). Catching up or standing still: National innovative productivity among ‘follower’ countries, 1978–1999. Research Policy, 33(9), 1329–1354

Model by Furman & Hayes

$$\bar{A}_{j,t} = (X_{j,t}^{INF}, Y_{j,t}^{CLUS}, Z_{j,t}^{LINK}) H_{j,t} A_{j,t}$$

Determinants of national innovative capacity:

(a) Common innovation infrastructure
(e.g. Patent Stock, GDP, Education expenditure)
(b) Cluster-specific innovation environment
(c) Quality linkage between the two above

Variables of Interest:

$\bar{A}$: Flow of innovations
X: Level of resource commitments and policy choices that constitute the innovation structure
Y: Environments for innovation in a country‘s industrial clusters
Z: Strength of linkages between common infrastructure and nation’s industrial clusters
H: Total level of human capital and labor recourses
A: Stock of knowledge

Source: Furman, J. L., & Hayes, R. (2004). Catching up or standing still: National innovative productivity among ‘follower’ countries, 1978–1999. Research Policy, 33(9), 1329–1354

New Model

$$\overline{A}_{j,t} = (X_{j,t}^{INF}, Y_{j,t}^{CLUS}, Z_{j,t}^{LINK}) M_{j,t} F_{j,t} H_{j,t} A_{j,t} C_{j,t}$$

New variables to be incorporated:

M: trade flows
F: foreign direct investment
C: control variables

Data and Methods

- **Innovative Output**
  variable:
  Patent Granted for each country each year
  source:
  WIPO

- **Quality of innovation**
  variable:
  GDP, National education expenditure, number of educational institutions…..
  source:
  WDI

- **Cluster-specific innovation environment**
  variable:
  R&D expenditure by private industry
  source:
  OECD

- **Quality of linkage**
  variable:
  national R&D expenditure(not industry)
  source:
  OECD

- **Trade flows/FDI**
  variable:
  imports/exports of goods and services, high- technology imports and exports, foreign direct investment net inflows
  source:
  WDI
- **Other control variable**

Data and Methods

- **Innovative Output**
  variable:
  Patent Granted for each country each year
  source:
  WIPO

- **Quality of innovation**
  variable:
  GDP, National education expenditure, number of educational institutions…..
  source:
  WDI

- **Cluster-specific innovation environment**
  variable:
  R&D expenditure by private industry
  source:
  OECD

- **Quality of linkage**
  variable:
  national R&D expenditure(not industry)
  source:
  OECD

- **Trade flows/FDI**
  variable:
  imports/exports of goods and services, high- technology imports and exports, foreign direct investment net inflows
  source:
  WDI

- **Other control variable**

Methods

- Time Series Method/Autoregression
- PCA
- Neural Networks

Further Discussion and Challenges

- Alternative model
- Better operationalize my model
  - Variable to measure innovation output
  - Variable to measure quality of linkage
  - Impacts of trade flows
  - Control variables

OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

LI RUIXUE

DOES SPECIALIZATION INCREASES RESEARCHERS' PRODUCTIVITY

MOTIVATION

- "Well then, how will our state supply these needs? It will need a farmer, a builder, and a weaver, and also, I think, a shoemaker and..."

  - Plato, *Republic*

- " The greatest 'improvement' in the productive powers of labour, and the greater part of the skill, dexterity, and judgment with which it is any where directed, or applied, seem to have been the effects of the division of labour.

  - Adam Smith, *Wealth of Nations*

- Specialization increases productivity.

EMPIRICAL EVIDENCE FROM OTHER FIELDS

- Highly specialized medical doctors make better diagnoses and fewer mistakes, and are paid more. (Lovinger 2003)
- High school sports players who specialize perform better. (Mulligan 2018)
- Venture capitals that specialize are more successful. (Gompers, Kovner and Lerner 2009)

WHEN IT COMES TO RESEARCH…?

- It's not clear what's the direction the effect of specialization on researchers' productivity:
  - ↑: Researchers become better and more influential in their field.
  - ↓: Less likely to gain insights from other fields or come up with creative ideas.

RESEARCH DESIGN / METHODOLOGY

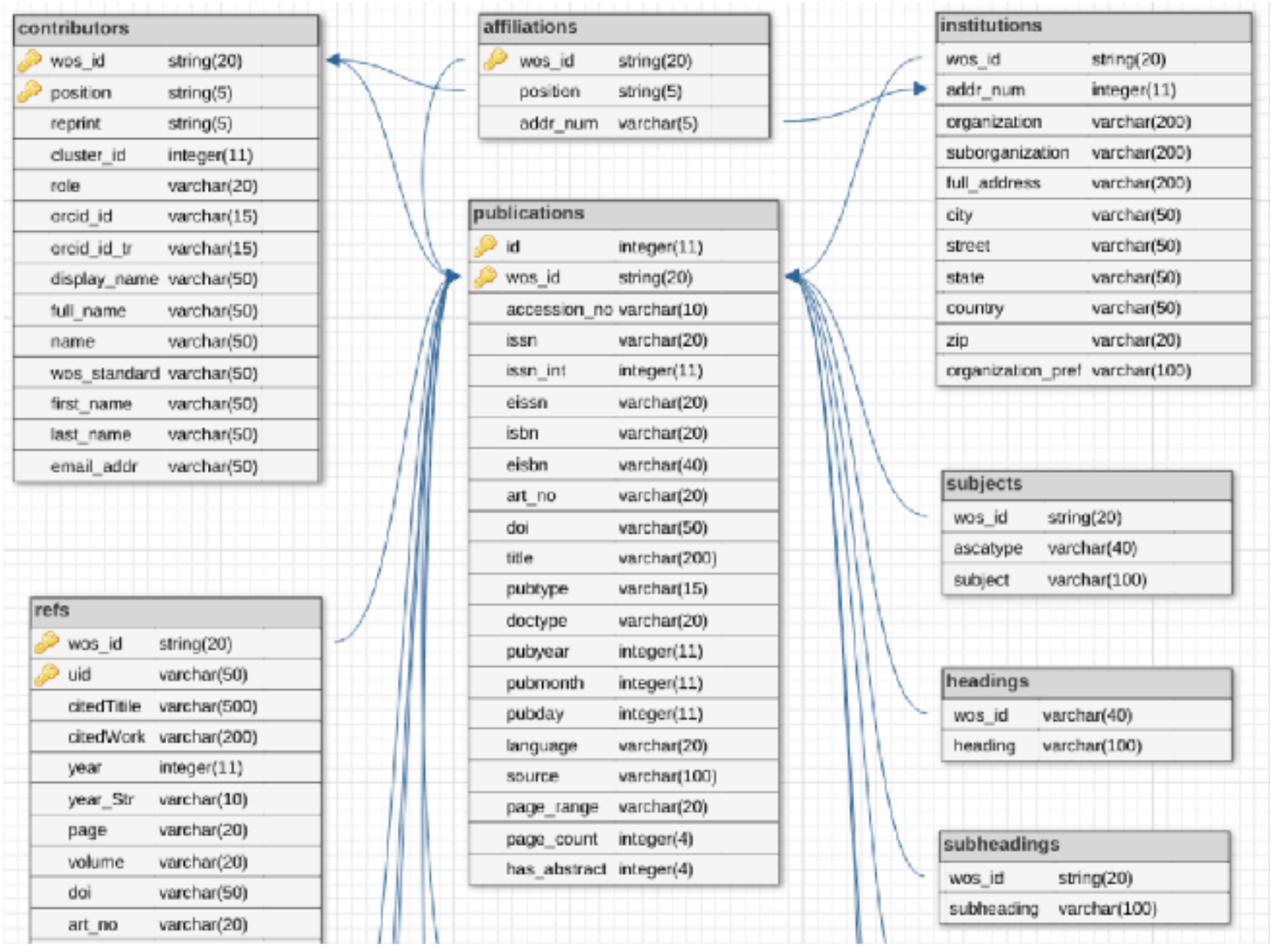
- Construct a measure of researcher productivity as endogenous variable
  - Research paper quantity and quality (citations, downloads, views, journal impact factors, etc. (Seglen 1997))
- Construct a measure of the level of specialization as exogenous variable
  - Using **computational text analysis** to measure how similar a researcher's papers are
  - Fields of the papers citing or cited by a researcher
- Fit a (or a few) **computational model**(s): linear regression, GMM, etc.
- Control for researcher's characteristics and other fixed effects

RESEARCH DESIGN / METHODOLOGY (CTD.)

- Construct a measure of researcher productivity as endogenous variable
  - Research paper quantity and quality (citations, downloads, views, journal impact factors, etc. (Seglen 1997))
- Construct a measure of the level of specialization as exogenous variable
  - Using **computational text analysis** to measure how similar a researcher's papers are
  - Fields of the papers citing or cited by a researcher
- Fit a (or a few) **computational model**(s): linear regression, GMM, etc.
- Control for researcher's characteristics and other fixed effects

DATA

- Web of Science (information on authors, institutions, papers, etc.)



CHALLENGES

- Identifying causal effect
- Many confounding factors
- Quality of the constructed measures

CONTRIBUTIONS

- Provide the earliest empirical evidence on this topic.
- Create comprehensive measure of researcher productivity.
- May be able to identify heterogeneity across fields since dataset is large.
- Findings will be of interest to researchers.
- Possible policy implications for institutions.

REFERENCES

"David Meltzer, Physician And Economist, Discusses The New Hospitalist Movement". 2018. *Jamanetwork-Com.Proxy.Uchicago.Edu*. https://jamanetwork-com.proxy.uchicago.edu/journals/jama/article-abstract/195836.

Gompers, Paul, Anna Kovner, and Josh Lerner. 2009. "Specialization And Success: Evidence From Venture Capital". *Journal Of Economics & Management Strategy* 18 (3): 817-844. doi:10.1111/j.1530-9134.2009.00230.x.

Mulligan, Casey. 2018. "Reaping The Gains From Specialization". *Economix Blog*. https://economix.blogs.nytimes.com/2009/03/18/reaping-the-gains-from-specialization/.

Plato., and C. J Rowe. 2012. *Republic*. London: Penguin.

Seglen, P. O. 1997. "Why The Impact Factor Of Journals Should Not Be Used For Evaluating Research". *BMJ* 314 (7079): 497-497. doi:10.1136/bmj.314.7079.497.

Smith, Adam. 1955. An Inquiry Into The Nature And Causes Of The Wealth Of Nations. Chicago: Encyclopædia Britannica.

OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

ESTIMATING ECONOMIC TIME SERIES USING HIGH RESOLUTION SATELLITE IMAGES AND NEURAL NETWORKS

By Cooper Nederhood

DATA POOR DEVELOPING ECONOMIES

- Developing economies lacking official economic measures
  - Higher resolution than other economic statistics
  - Worldwide coverage
  - Low marginal cost
- Modern computer vision techniques to analyze images
- Jean et al (2016) "Combining satellite imagery and machine learning to predict poverty"

DATA POOR DEVELOPING ECONOMIES

- Developing economies lacking official economic measures
  - Higher resolution than other economic statistics
  - Worldwide coverage
  - Low marginal cost
- Modern computer vision techniques to analyze images
- Jean et al (2016) "Combining satellite imagery and machine learning to predict poverty"

SATELLITE DATA – LANDSAT

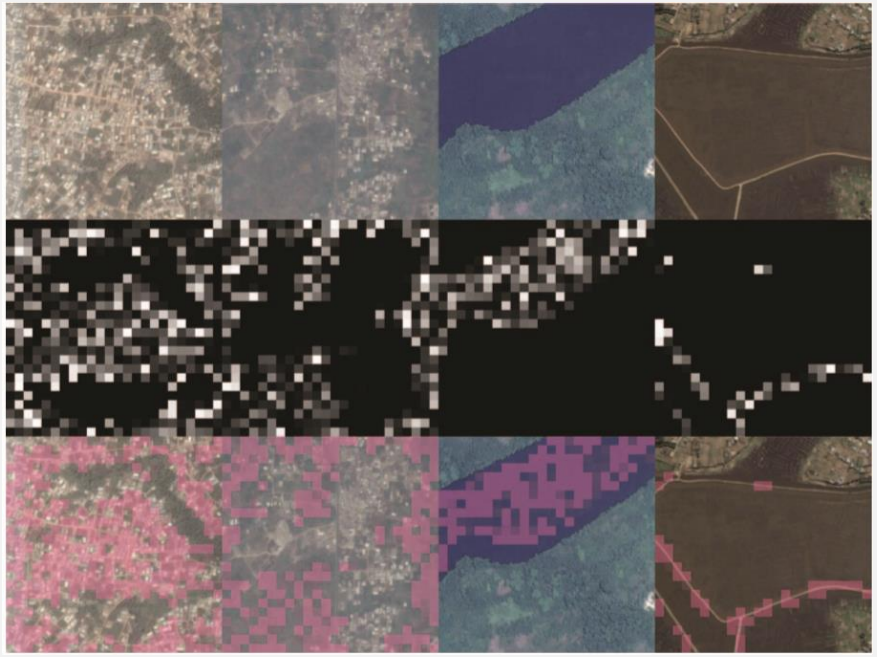
- Available from 1972 to today
- 30 meter resolution
- Cover entire Earth's surface every two weeks
- Access through Google Earth Engine API
  - Goldblatt et al, 2016

GROUND TRUTH DATA – DHS SURVEY

- "Demographic and Health Survey"
- Includes Wealth Index, composite index
- Uganda (2016, 2011, 2006, 2000-01, 1995, 1988-89)

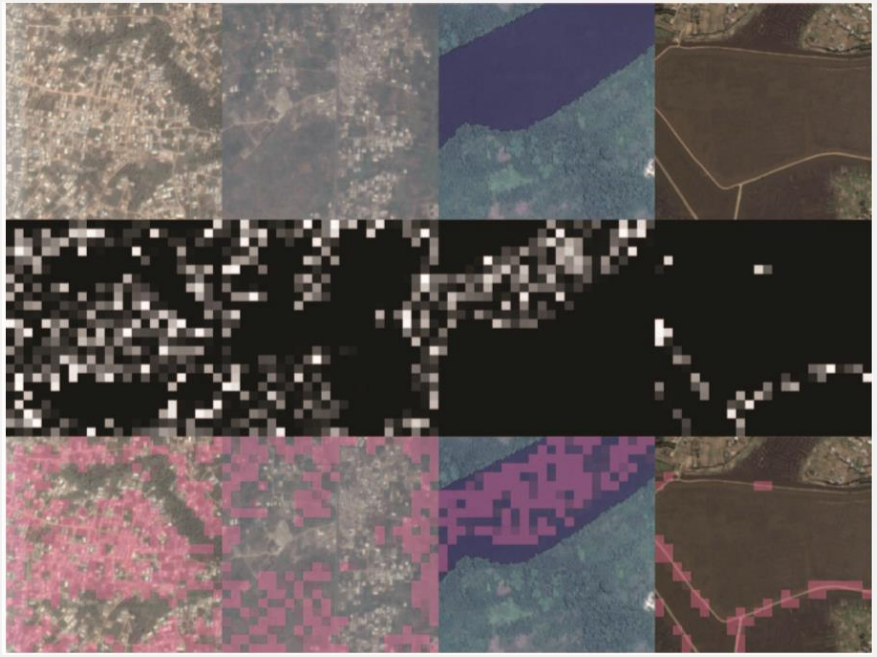
CONVOLUTIONAL NEURAL NETWORK

- Deep neural network used in computer vision
- Learns to identify image features relevant to economic activity
- Transfer learning from ImageNet
- Jean et al (2016) "Combining satellite imagery and machine learning to predict poverty"
- Banerjee et al (2017) "On monitoring development using high resolution satellite images"



CONVOLUTIONAL NEURAL NETWORK

- Deep neural network used in computer vision
- Learns to identify image features relevant to economic activity
- Transfer learning from ImageNet
- Jean et al (2016) "Combining satellite imagery and machine learning to predict poverty"
- Banerjee et al (2017) "On monitoring development using high resolution satellite images"



POTENTIAL RESULTS

- Much research in 'night lights' as a proxy for economic activity
- Convolutional Neural Nets much richer, complex
  - Success cross-sectionally
- "We however find poor prediction power of [nightlights and Landsat images] in the time-series which severely limits the usage of remote sensing for predicting economic changes over time at small geographies" (Goldblatt et al, 2017)
  - Simple model?
  - Or physical features not changing?

POTENTIAL RESULTS

- Much research in 'night lights' as a proxy for economic activity
- Convolutional Neural Nets much richer, complex
  - Success cross-sectionally
- "We however find poor prediction power of [nightlights and Landsat images] in the time-series which severely limits the usage of remote sensing for predicting economic changes over time at small geographies" (Goldblatt et al, 2017)
  - Simple model?
  - Or physical features not changing?

VAST POSSIBILITIES IN USING SATELLITE DATA IN ECONOMICS

- "The View from Above: Applications of Satellite Data in Economics", Donaldson and Storeygard
- Questions?

OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

Which method of delay discounting is the best method of measuring impulsive behavior?

John-Henry Pezzuto

Perspectives on Computational Research

April 4, 2018

Monetary Delay Discounting is a psychological technique used in measuring patience.

It is solicited by finding an individuals **indifference point** by asking them series of now or later questions about money

	Option 1	Option 2	
$0.01 today	○	○	$30 in six months
$2 today	○	○	$30 in six months
$4 today	○	○	$30 in six months
$6 today	○	○	$30 in six months
$8 today	○	○	$30 in six months
$10 today	○	○	$30 in six months
$12 today	○	○	$30 in six months
$14 today	○	○	$30 in six months
$16 today	○	○	$30 in six months
$18 today	○	○	$30 in six months

Method 1: Time Discounting
(small, present)

Monetary Delay Discounting is a psychological technique used in measuring patience.

It is solicited by finding an individuals **indifference point** by asking them series of now or later questions about money

	Option 1	Option 2	
$0.01 today	○	○	$30 in six months
$2 today	○	○	$30 in six months
$4 today	○	○	$30 in six months
$6 today	○	○	$30 in six months
$8 today	○	○	$30 in six months
$10 today	○	○	$30 in six months
$12 today	○	○	$30 in six months
$14 today	○	○	$30 in six months
$16 today	○	○	$30 in six months
$18 today	○	○	$30 in six months

Method 1: Time Discounting
(small, present)

But there are different ways to measure delay discounting. These can also vary across **size** and **time**

	Yes	No
Would you pay $0.01 to get the $300 today?	○	○
Would you pay $20 to get the $300 today?	○	○
Would you pay $40 to get the $300 today?	○	○
Would you pay $60 to get the $300 today?	○	○
Would you pay $80 to get the $300 today?	○	○
Would you pay $100 to get the $300 today?	○	○
Would you pay $120 to get the $300 today?	○	○
Would you pay $140 to get the $300 today?	○	○
Would you pay $160 to get the $300 today?	○	○

Method 2: Speedup Method (large, present)

In this study we will measure delay discounting **8 (2x4) different ways**

Time Discounting (either or small now or larger later)	Speed Up Method (*pay* to get money now or wait for full amount later)	
Small present	Small present	30 now
Large present	Large present	300 now
Small future	Small future	30 in one month
Large future	Large future	300 in one month

All of these methods involve indifference points that can be standardized and then compared via AUC

$x_2 - x_1 [(y_1 + y_2)/2]$, where x1 and x2 are successive delays and y1 and y2 are the indifference points associated with those delays

Previously, delay discounting has been linked to many impulsive and patient behaviors…*

- Risky Sexual Behavior (Chesson, 2006)
- Debt and poor financial planning (Chabris et al., 2008)
- Not finishing prescriptions (Chabris et al., 2008)
- Poor dental status (Kang & Ikeda, 2015)
- Interest in the political campaigns (Fowler & Kam, 2006)

Which method of the 8 time discounting methods is the best at measuring impulsive behavior?

*Full references available upon request

The Proposed Study:

- A large mTurk survey establishing a rate with all 8 methods AND questions about various categories of impulsive behaviors (n ≈ 600)
- Survey will cover substance use, political interest, financial behavior, health behavior, & sexual behavior (approx. 100 questions total)
- Already IRB approved & will be paid for by our friends at Chicago Booth
- Useful for finding most accurate predictor of behavior for future research or supports the idea that there is no best predictor
- Multivariate Regression

Why is this important?

- Discounting is useful for summarizing the results of experiments on sensitivity to delayed rewards
- Has powerful cross-species, cross-population, as well as intraindividual replicability
- If we provide more evidence of that humans either treat the models differently or not, this can be useful for researchers

Predictions

- Based on findings from behavioral economics, I think we can expect to see large, present condition willingness to pay will be the best overall predictor
- Most discounting research overlooks the question of stakes instead opting to use a standardized questionnaire

OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

News and Expected Volatility in the Stock Market

Mengchen Shi

MACSS Project Proposal
April 4, 2018

Introduction

Research Question:

- What is the relationship between news and expected volatility in stock market?

Background:

- VIX: measure of the stock market's expectation of volatility
- Summarize real-time data from news
- Both are utilized to design investment strategy

Introduction

Lit Review:

- Antweiler and Frank (2004) used document length and sentiment analysis to fuel NB (Rainbow) and SVM classification algorithms, showing that financial news correlate with stock returns, trading volume and stock volatility

- Manela and Moreira (2013) constructed news implied volatility index to predict disasters and returns

- Smales (2014) constructed sentiment indicator to confirm the ‘significant negative relationship’ between 2000–2010 news releases and market volatility.

Data

1. News articles:
    - Financial Times API
    - Scrape from websites: Wall Street Journal, New York Time
2. VIX: Chicago Board Options Exchange website

Computational Tools & Methods

1. Sentiment analysis (positive & negative attitudes)
2. Topic modeling
3. Bag-of-Words & Word2Vec

Computational Tools & Methods

1. Sentiment analysis (positive & negative attitudes)
2. Topic modeling
3. Bag-of-Words & Word2Vec

Models

- VIX and news sentiment:
  - OLS:

$$\Delta VIX_t = \beta_0 + \beta_1 News_t + \varepsilon_i$$

$$\Delta VIX_t = \beta_0 + \beta_1 News_t^+ + \beta_1 News_t^- + \varepsilon_i$$

- VIX and news topics:
  - OLS:

$$\Delta VIX_t = \beta_0 + \sum_{i=1}^{N} \beta_i Topic_{it} + \varepsilon_i$$

- VIX and news content (Bag-of-Words and Word2Vec):
  - Lasso Regression:

$$\Delta VIX_t = \beta X_t + \varepsilon_i$$

Models

- VIX and news sentiment:
  - OLS:

$$\Delta VIX_t = \beta_0 + \beta_1 News_t + \varepsilon_i$$

$$\Delta VIX_t = \beta_0 + \beta_1 News_t^+ + \beta_1 News_t^- + \varepsilon_i$$

- VIX and news topics:
  - OLS:

$$\Delta VIX_t = \beta_0 + \sum_{i=1}^{N} \beta_i Topic_{it} + \varepsilon_i$$

- VIX and news content (Bag-of-Words and Word2Vec):
  - Lasso Regression:

$$\Delta VIX_t = \beta X_t + \varepsilon_i$$

Potential Contributions:

1. Analyzing both news sentiment and detailed content
2. Predict VIX in the future:
    - Improve investment strategy in the financial market
    - Improve investment strategy in VIX futures and options market
3. Estimate a measure of implied volatility that extends backwards (before 1980’s)

Potential Contributions:

1. Analyzing both news sentiment and detailed content
2. Predict VIX in the future:
    - Improve investment strategy in the financial market
    - Improve investment strategy in VIX futures and options market
3. Estimate a measure of implied volatility that extends backwards (before 1980’s)

OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

Parallel Optimization for Simulation Data Based on "DASK"

Jiang Wang

MACSS Project Proposal

Apr 4th, 2018

Github: https://github.com/Otamio/DASKopt

Introduction

- Why do we need parallel computing?
- Why choose to parallel optimization based on “DASK”?
- Are there any needs to improve the “DASK” Module?
- Goal: Improve parallel performance for large scale simulation data by balancing and optimizing the “DASK” scheduler

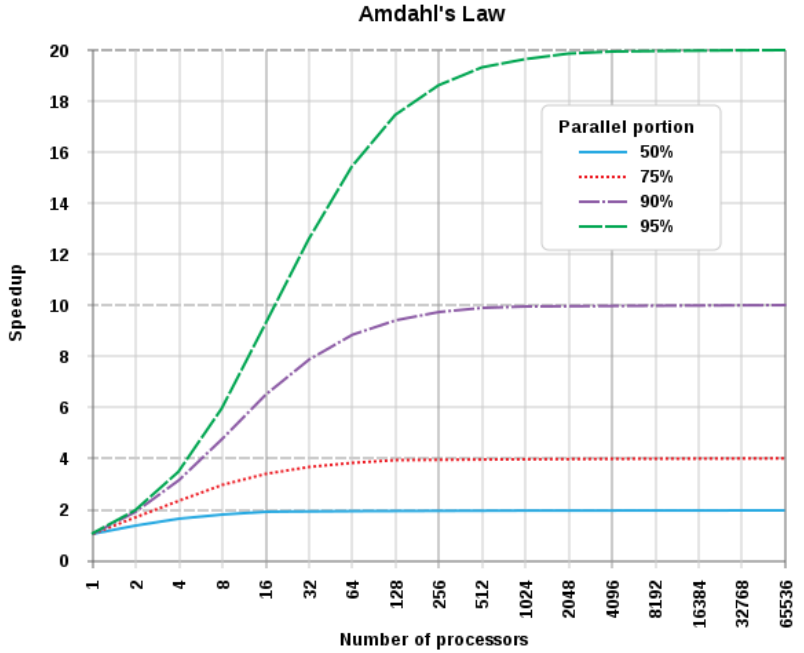
Basic Theory of Parallel Computing

- Amdahl's Law?

$$S_{\text{latency}}(s) = \frac{1}{1 - p + \frac{p}{s}}$$

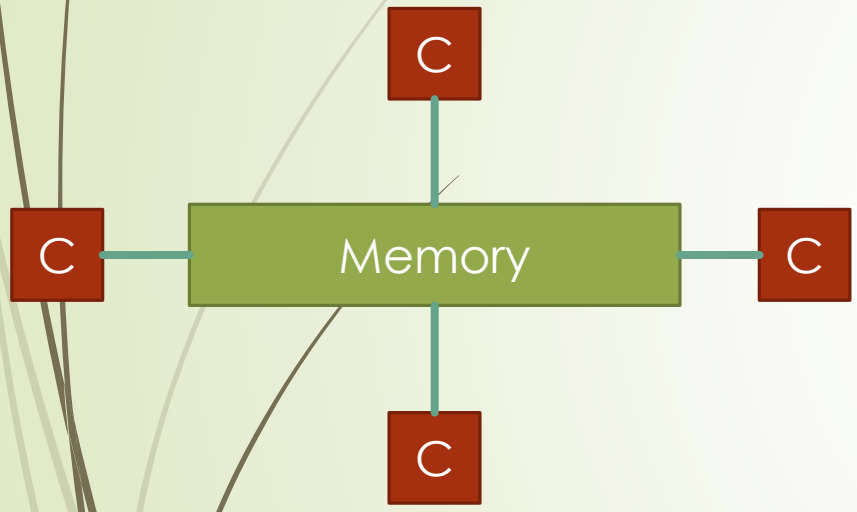
  - p: Execution time of parallelizable part
  - s: Speed up of the parallelized part
  - S: Speed up of the whole part

  - Let p=.8, s=4, we would have S=2.5
  - Let p=0.9, s =16, we would have S=6.4
  - Normally, S would reach a ceiling of 20x

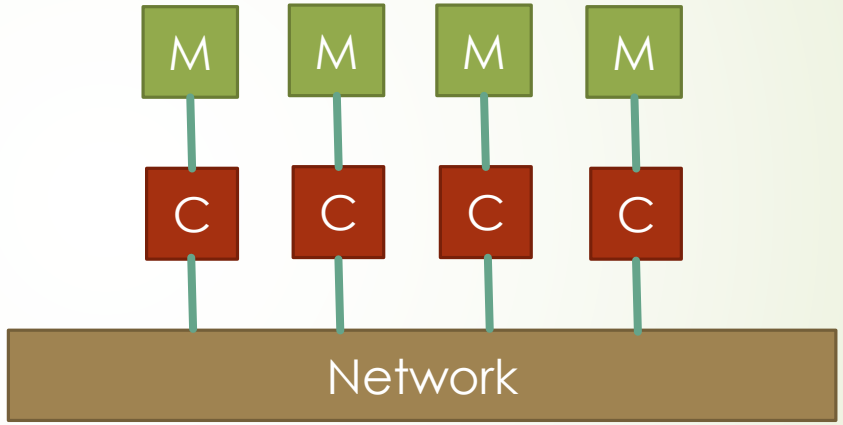


Basic Theory of Parallel Computing

- Memory Management and Communicating Cost



Shared Memory



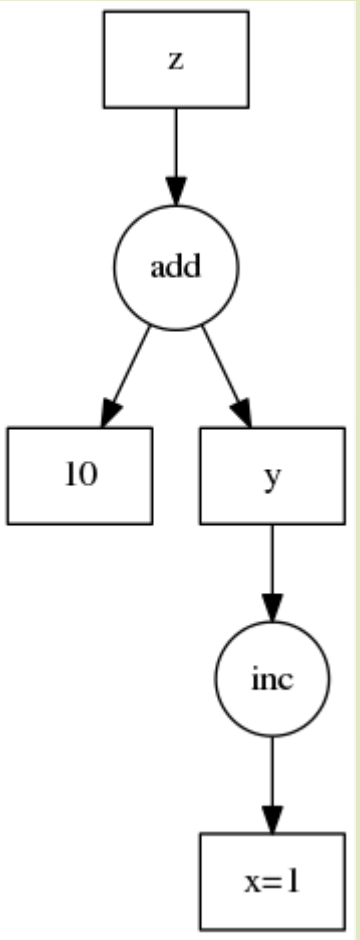
Distributed Memory

- Modern architecture takes a blend of two
- Transferring data between processes is costly!
- We should prevent unnecessary data transferring, especially when doing cheap operations

Methods: The "DASK" Scheduler

- Scheduler: The work needs to be serialized and distributed to different processes before they are executed
- DASK Scheduler is a graph
  - E.g.

```
def inc(i):
    return i + 1

def add(a, b):
    return a + b

x = 1
y = inc(x)
z = add(y, 10)
```



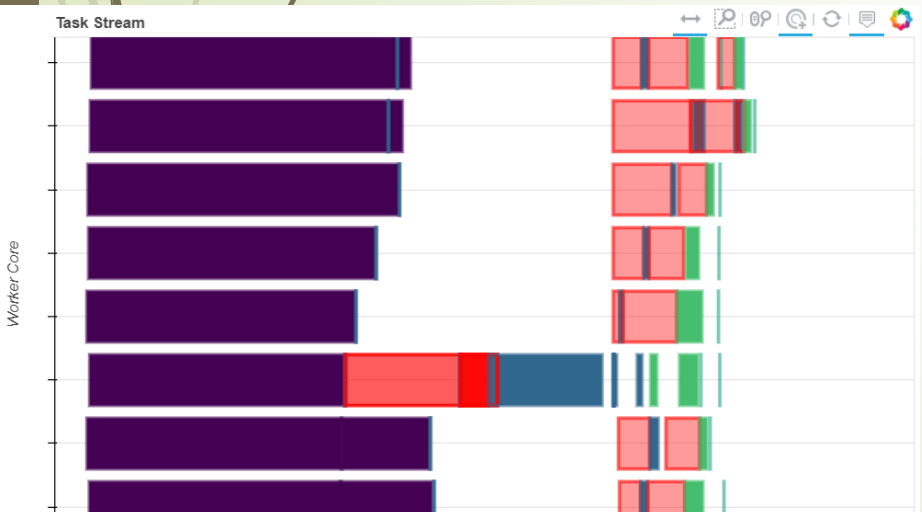
  - Which is encoded as a Python Dictionary:

```
d = {'x': 1,
     'y': (inc, 'x'),
     'z': (add, 'y', 10)}
```

Methods: The "DASK" Scheduler

- Some Issues with the default "DASK" Scheduler
  - Python GIL (Global Interpreter Lock)
    - A "Pain" of CPython when computing with Python data structures
    - Consider the following code:

```
a = da.random.random(size=(10000, 1000), chunks=(1000, 1000))
q, r = da.linalg.qr(a)
a2 = q.dot(r)

out = a2.compute()
```



- We have the following performance analysis:
  - Purple: da.random.random()
  - Red: Transferring Cost
  - Blue: da.linalg.qr()
  - Green: da.array.dot()

Methods: The "DASK" Scheduler

- Data Simulation: Get annual income 40 years later
  - Compare "DASK" and numpy. Which is faster? (Use %time callable)

```
def simulation_dask(init, years):
    y = da.from_array(np.log(init), chunks=(1000000, 1))
    for i in range(years+1):
        n_errors = da.random.normal(0, 0.1, size=(1000000,1), chunks=(1000000, 1))
        y = (1 - 0.2) * (np.log(init) + 0.03 * i) + 0.2 * y + n_errors
    y = da.exp(y)
    return y
```

```
def simulation_numpy(init, years):
    y = np.full((1000000, 1), np.log(init))
    for i in range(years+1):
        n_errors = np.random.normal(0, 0.1, (1000000, 1))
        y = (1 - 0.2) * (np.log(init) + 0.03 * i) + 0.2 * y + n_errors
    y = np.exp(y)
    return y
```

Methods: The "DASK" Scheduler

- Result:

```
%time simulation_numpy(80000, 40).mean()
```

```
Wall time: 3.03 s

264957.69206603663
```

```
%time simulation_dask(80000, 40).mean().compute()
```

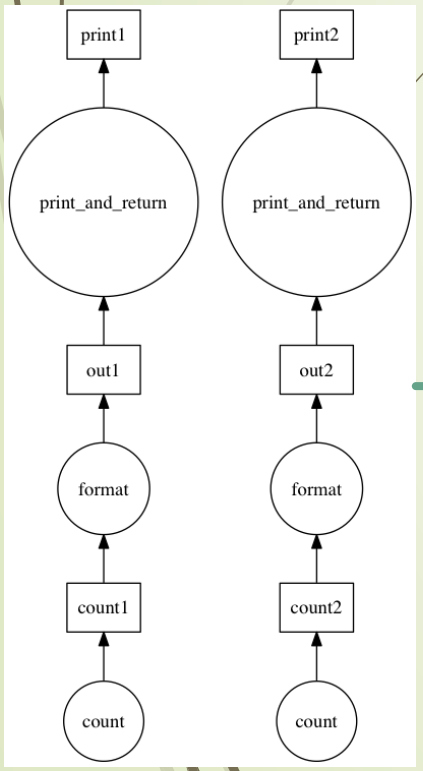
```
Wall time: 5.23 s

264970.07779333438
```

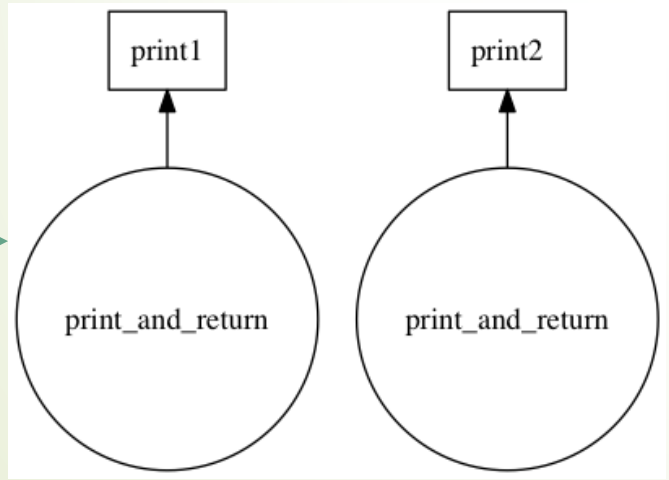
- Reasons?
  - Dependency

```
def simulation_dask(init, years):
    y = da.from_array(np.log(init), chunks=(1000000, 1))
    for i in range(years+1):
        n_errors = da.random.normal(0, 0.1, size=(1000000,1), chunks=(1000000, 1))
        y = (1 - 0.2) * (np.log(init) + 0.03 * i) + 0.2 * y + n_errors
    y = da.exp(y)
    return y
```

Methods: Graph Optimization

- Some possible solutions:
  - Reduce Dependency (Avoid recursion and iteration).
  - Fuse






```
from dask.optimization import fuse
dsk4, dependencies = fuse(dsk3)
results = get(dsk4, outputs)
```

- Advanced: User defined graph (using dictionaries)

Conclusion

- Research Question: Improve DASK's efficiency for large scale data analysis in simulation research
  - Most emphasis will be placed on optimizing the DASK scheduler
- Schedule:
  - A mini Python project to be completed reducing the data transferring costs between processes
  - A project (mostly Python) to deal with schedulers, dependencies, GIL, and possibly, code rewrite and compilers
- References:
  - http://dask.pydata.org/en/latest/docs.html
- Slides and some example codes will be posted on https://github.com/Otamio/DASKopt
- Thanks for your attention!

OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

The Relationship between Macroeconomic Uncertainty and Interest Rates in China

MACS 30200 Project Proposal

Presenter: Zunda Xu

Motivation

Uncertainty

- Uncertainty about future economic growth is thought to have a broad impact on the economy

- Volatility —> Uncertainty: Using conditional variance to present uncertainty

Interest Rates

- The real risk-free interest rate is determined by two economic forces

- Inter-temporal smoothing— a positive relation between the interest rate and expectations of economic growth

- precautionary savings — a negative relation between the interest rate and uncertainty, the conditional variance of growth

Motivation

Uncertainty

- Uncertainty about future economic growth is thought to have a broad impact on the economy

- Volatility —> Uncertainty: Using conditional variance to present uncertainty

Interest Rates

- The real risk-free interest rate is determined by two economic forces

- Inter-temporal smoothing— a positive relation between the interest rate and expectations of economic growth

- precautionary savings — a negative relation between the interest rate and uncertainty, the conditional variance of growth

Model

Basic Model:

$$r_t = \beta_0 + \beta_1 E_t[g_{t+1}] + \beta_2 Var_t[g_{t+1}]$$

r_t is the log of the time t to t+1 risk-free rate,

g_{t+1} is the log economic growth rate in the subsequent period,

E_t is the expectation conditioning on in- formation at time t,

Var_t is the variance conditioning on information at time t.

Data Source

Measures of Economic Growth Rate:

Consumption Expenditure:

Private Final Consumption Expenditure in China
(https://fred.stlouisfed.org/series/CHNPFCEADSMEI)

GDP:

Real GDP at Constant National Prices for China
(https://fred.stlouisfed.org/series/RGDPNACNA666NRUG)

Industrial Production:

Total Industry Production Excluding Construction for China
(https://fred.stlouisfed.org/series/RGDPNACNA666NRUG)

Data Source

Main Interest Rate Data:

3-Month or 90-day Rates and Yields: Treasury Securities for China

Federal Reserve Bank of St. Louis
(https://fred.stlouisfed.org/series/IR3TTS01CNM156N)

China Interest Rate

Trading Economics

Interest Rates for China:

Federal Reserve Bank of St. Louis

Method

Estimates of Expected economic growth and the variance of economic growth:

Model growth as ARMA(1,1):

$$g_{t+1} = \phi g_t + \theta \epsilon_t + \mu + \epsilon_{t+1}$$

Two methods to estimate variance:

(1) the square of the residuals in period t: $(\hat{g}_t - g_t)^2$

(2) using a GARCH(1,1) model (Engle 1982; Bollerslev 1986) :

$$g_{t+1} = \phi g_t + \theta \epsilon_t + \mu + \epsilon_{t+1}$$

$$Var_t(\epsilon_{t+1}) = \gamma + \alpha_1 \epsilon_t^2 + \alpha_2 \sigma_t^2$$

Method

Estimates of Expected economic growth and the variance of economic growth:

Model growth as ARMA(1,1):

$$g_{t+1} = \phi g_t + \theta \epsilon_t + \mu + \epsilon_{t+1}$$

Two methods to estimate variance:

(1) the square of the residuals in period t: $(\hat{g}_t - g_t)^2$

(2) using a GARCH(1,1) model (Engle 1982; Bollerslev 1986) :

$$g_{t+1} = \phi g_t + \theta \epsilon_t + \mu + \epsilon_{t+1}$$

$$Var_t(\epsilon_{t+1}) = \gamma + \alpha_1 \epsilon_t^2 + \alpha_2 \sigma_t^2$$

Method

Using Alternative Forecasts:

Professional Forecasts:

China Economic Forecasts
(http://www.consensuseconomics.com/China_Economic_Forecasts.htm)

Data of other Developing Countries:

India, Brazil, South Africa and etc.

OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

The Effects of Health Shock on Household Income Mobility: Evidence from China

Xiang Zhang

The University of Chicago

snzhang@uchicago.edu

April 4, 2018

Motivation

- High income inequality in many countries
  - China's official statistics: less than 0.3 in 1980 to 0.465 in 2016
  - The United States: 0.45 in 2016
- Intra-generational Income Mobility can serve as an equalizer of long-term income inequality (Shorrocks [1978])
  - GINI coefficient is a static measure of income inequality
  - Two societies with identical GINI coefficient, the one with higher income mobility will have relatively small overall income inequality

Motivation

- High income inequality in many countries
  - China's official statistics: less than 0.3 in 1980 to 0.465 in 2016
  - The United States: 0.45 in 2016
- Intra-generational Income Mobility can serve as an equalizer of long-term income inequality (Shorrocks [1978])
  - GINI coefficient is a static measure of income inequality
  - Two societies with identical GINI coefficient, the one with higher income mobility will have relatively small overall income inequality

Motivation

- How can a household maintain high income mobility, or specifically, high upward mobility?

- Perspective from health human capital
  - One central pillars of human capital (the other is education)
  - Poor stay poor due to bad health condition of household members (no hope of achieving "American Dream")
  - Impoverished again due to health shock

Motivation

- How can a household maintain high income mobility, or specifically, high upward mobility?

- Perspective from health human capital
  - One central pillars of human capital (the other is education)
  - Poor stay poor due to bad health condition of household members (no hope of achieving "American Dream")
  - Impoverished again due to health shock

This Paper

- Present stylized facts on the relationship between health human capital and income mobility

- Estimate the effects of health shock on household income mobility

Data

- China Health and Nutrition Survey (CHNS)
  - From 1991 to 2011
  - Not a nationally representative data
  - Target sample: 4400 households
- China Family Panel Studies (CFPS)
  - A bi-annual survey starting from 2010
  - Nationally representative
  - Target sample: 16000 households

Theories behind Health Human Capital and Income Mobility

- General health human capital and income productivity theory
  - Becker [1962], Grossman [1972]

- Conceptual framework
  - Gertler and Gruber [2002], Dobkin et al. [2018]

Theories and Empirical Work on Measurement of Income Mobility

- Parametric estimation: Solon [1992], Chetty et al. [2014]
  - Inter(Intra)-generational income elasticity
  - Transition matrix
- Non-parametric estimation: Chetty et al. [2014, 2017], Bhattacharya and Mazumder [2011]

Potential Computational Methods in Analysis

- Sample selection:
  - K-NN
- Correlational study:
  - Generalized method of moments (GMM)
  - Non-parametric kernel density estimation
- Econometric analysis:
  - LASSO (least absolute shrinkage and selection operator)
  - Linear regression (DID and event study specification should be plausible)
  - Principal Component Analysis (PCA) and Factor Analysis

OK. You're done.

Let's get the next person up.

Your presentation starts in 30 seconds

How Do We Learn about some New: Interpreting Generic Statements Using Bayesian Inference

Just in case you don't know it already: $$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Presenter: Flora Zhang
Principle Investigator: Dr. Daniel Yurovsky

MACS Github Repo: https://github.com/xiuyuanzhang/MACS30200proj
Research Project Repo: https://github.com/xiuyuanzhang/generic-statement

Communication and Learning Lab

"[Puppies] are [friendly.]"

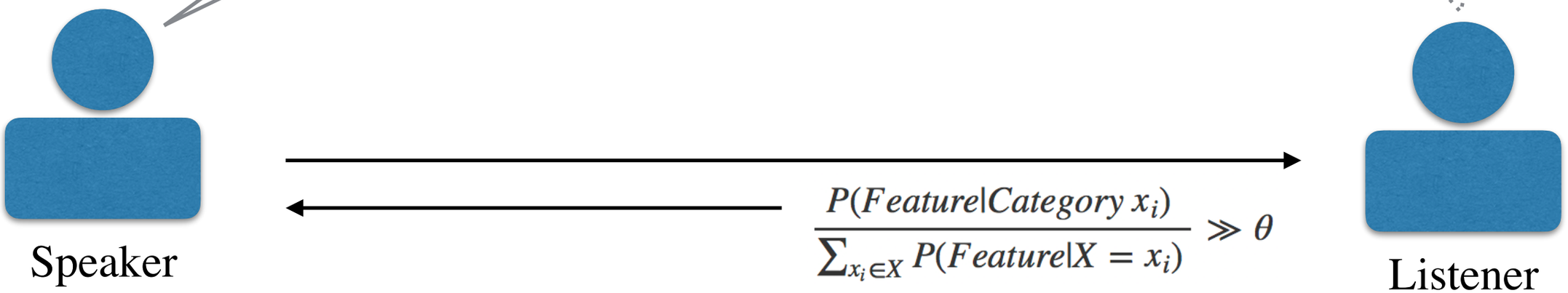
"[Squirrels] are [friendly.]"

"[Tigers] are [friendly.]"

True

True, but maybe less so

Eh, no



Speaker

produce an utterance

Listener

process the utterance

Informally defined, generic statements are blanket statements about members of a category and their diagnostic features.

Communication and Learning Lab

"[Puppies] are [friendly.]"

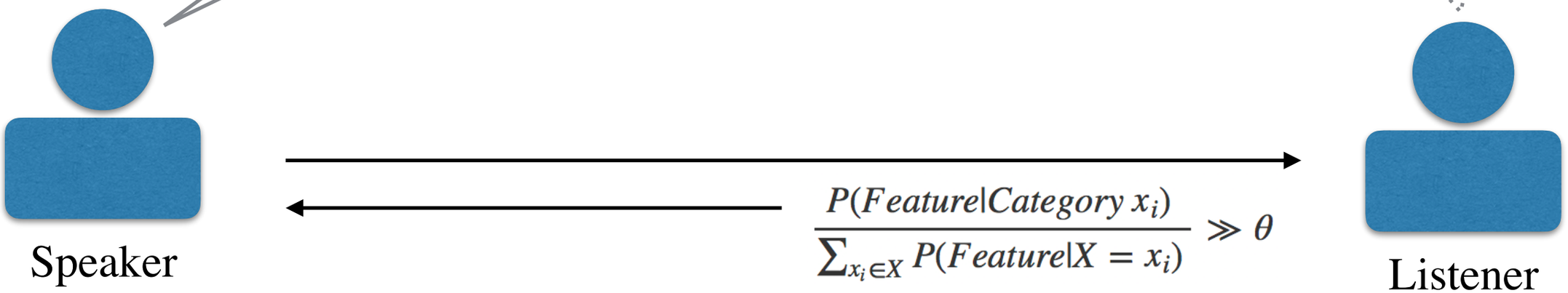
"[Squirrels] are [friendly.]"

"[Tigers] are [friendly.]"

True

True, but maybe less so

Eh, no



Speaker

produce an utterance

Listener

process the utterance

Informally defined, generic statements are blanket statements about members of a category and their diagnostic features.

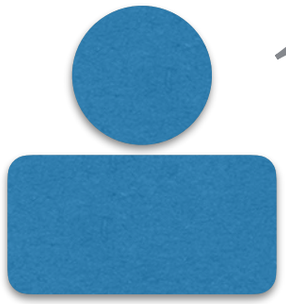
Communication and Learning Lab

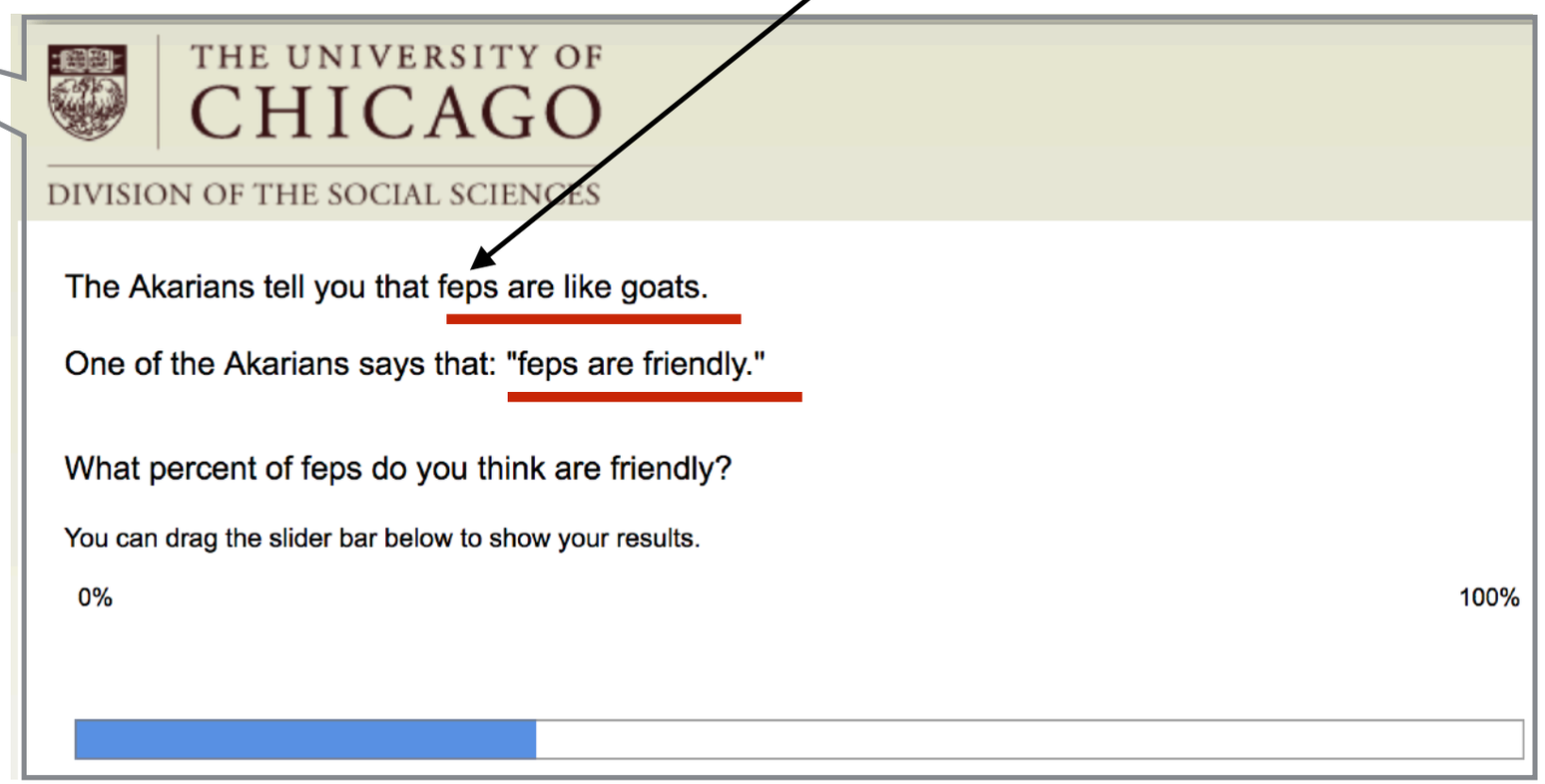
"[Puppies] are [friendly.]"

"[Squirrels] are [friendly.]"

"[Tigers] are [friendly.]"

"feps" is a made-up word,
a novel category



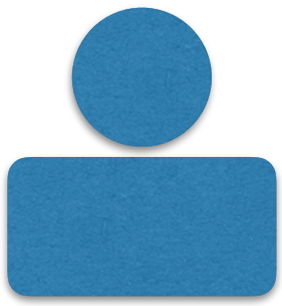


Communication and Learning Lab

A Speaker-Listener Interaction Model

$$\frac{P(Feature|Category\ x_i)}{\sum_{x_i \in X} P(Feature|X = x_i)} \gg \theta$$

$$\frac{P(Feature|Novel\ Category)}{\sum_{x_i \in X} P(Feature|X = x_i)} \gg \theta$$



Listener

process the utterance

$$P(Feps\ are\ friendly|speaker\ uttered\ "Feps\ are\ friendly")$$

$$= \frac{P(speaker\ uttered\ "Feps\ are\ friendly"|Feps\ are\ friendly)P(Feps\ are\ friendly)}{P(speaker\ uttered\ "Feps\ are\ friendly")}$$

$$= \mathbf{E}(P(speaker\ uttered\ "Feps\ are\ friendly"|Feps\ are\ friendly)P(Feps\ are\ friendly))$$

$$= \mathbf{E}(P(speaker\ uttered\ "Feps\ are\ friendly"|Feps\ are\ friendly)) \cdot 0.5$$

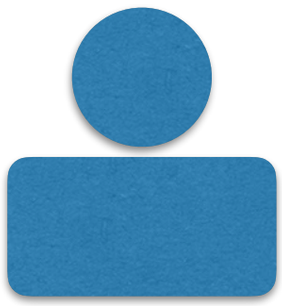
$$= \mathbf{E}(\int_{P(Friendly|Feps)} \frac{e^{\alpha P(Friendly|Feps)}}{e^{\alpha P(Friendly|Feps)} + e^{\alpha P(Friendly)}} \delta P(Friendly|Feps)) \cdot 0.5$$

Communication and Learning Lab

A Speaker-Listener Interaction Model

$$\frac{P(Feature|Category\ x_i)}{\sum_{x_i \in X} P(Feature|X = x_i)} \gg \theta$$

$$\frac{P(Feature|Novel\ Category)}{\sum_{x_i \in X} P(Feature|X = x_i)} \gg \theta$$



Listener

process the utterance

$$P(Feps\ are\ friendly|speaker\ uttered\ "Feps\ are\ friendly")$$

$$= \frac{P(speaker\ uttered\ "Feps\ are\ friendly"|Feps\ are\ friendly)P(Feps\ are\ friendly)}{P(speaker\ uttered\ "Feps\ are\ friendly")}$$

$$= \mathbf{E}(P(speaker\ uttered\ "Feps\ are\ friendly"|Feps\ are\ friendly)P(Feps\ are\ friendly))$$

$$= \mathbf{E}(P(speaker\ uttered\ "Feps\ are\ friendly"|Feps\ are\ friendly)) \cdot 0.5$$

$$= \mathbf{E}(\int_{P(Friendly|Feps)} \frac{e^{\alpha P(Friendly|Feps)}}{e^{\alpha P(Friendly|Feps)} + e^{\alpha P(Friendly)}} \delta P(Friendly|Feps)) \cdot 0.5$$

Method

Data:

- Involves Human Subjects (SBS IRB Approval No.: IRB16-1118) - Online Survey using Amazon MTurk
- Pew Research Center's American Trends Panel Wave 26 (n = 5155)

Analysis and Modeling in R & Stan:

- Random Forest - finding distinct demographic trends on social and political issues (for next step surveys)
- Linear Regression - evaluating estimate probability for novel category
- Logistic Regression - evaluating binary True/False var
- Stan(probabilistic programming language) - build Bayesian statistical model for our hypothesis

Previous Related Work

- Ward, Andrew, L. Ross, E. Reed, E. Turiel, and T. Brown. "Naive realism in everyday life: Implications for social conflict and misunderstanding." *Values and knowledge* (1997): 103-135.
- Rhodes, Marjorie, Sarah-Jane Leslie, and Christina M. Tworek. "Cultural transmission of social essentialism." Proceedings of the National Academy of Sciences 109, no. 34 (2012): 13526-13531.
- Tessler, Michael Henry, and Noah D. Goodman. "A pragmatic theory of generic language." *arXiv preprint arXiv:1608.02926*(2016).

Method

Data:

- Involves Human Subjects (SBS IRB Approval No.: IRB16-1118) - Online Survey using Amazon MTurk
- Pew Research Center's American Trends Panel Wave 26 (n = 5155)

Analysis and Modeling in R & Stan:

- Random Forest - finding distinct demographic trends on social and political issues (for next step surveys)
- Linear Regression - evaluating estimate probability for novel category
- Logistic Regression - evaluating binary True/False var
- Stan(probabilistic programming language) - build Bayesian statistical model for our hypothesis

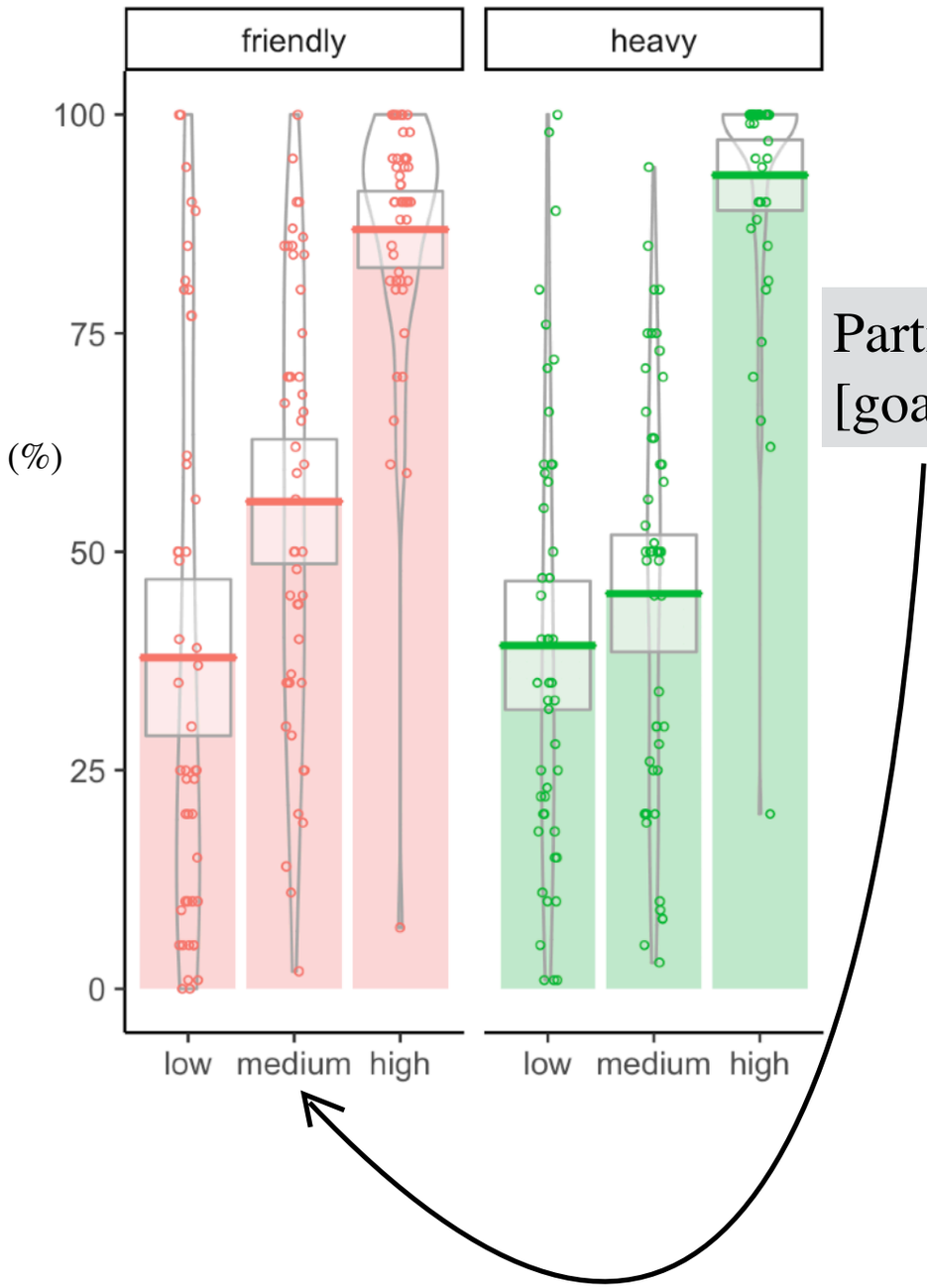
Previous Related Work

- Ward, Andrew, L. Ross, E. Reed, E. Turiel, and T. Brown. "Naive realism in everyday life: Implications for social conflict and misunderstanding." *Values and knowledge* (1997): 103-135.
- Rhodes, Marjorie, Sarah-Jane Leslie, and Christina M. Tworek. "Cultural transmission of social essentialism." Proceedings of the National Academy of Sciences 109, no. 34 (2012): 13526-13531.
- Tessler, Michael Henry, and Noah D. Goodman. "A pragmatic theory of generic language." *arXiv preprint arXiv:1608.02926*(2016).

Communication and Learning Lab

Preliminary Results

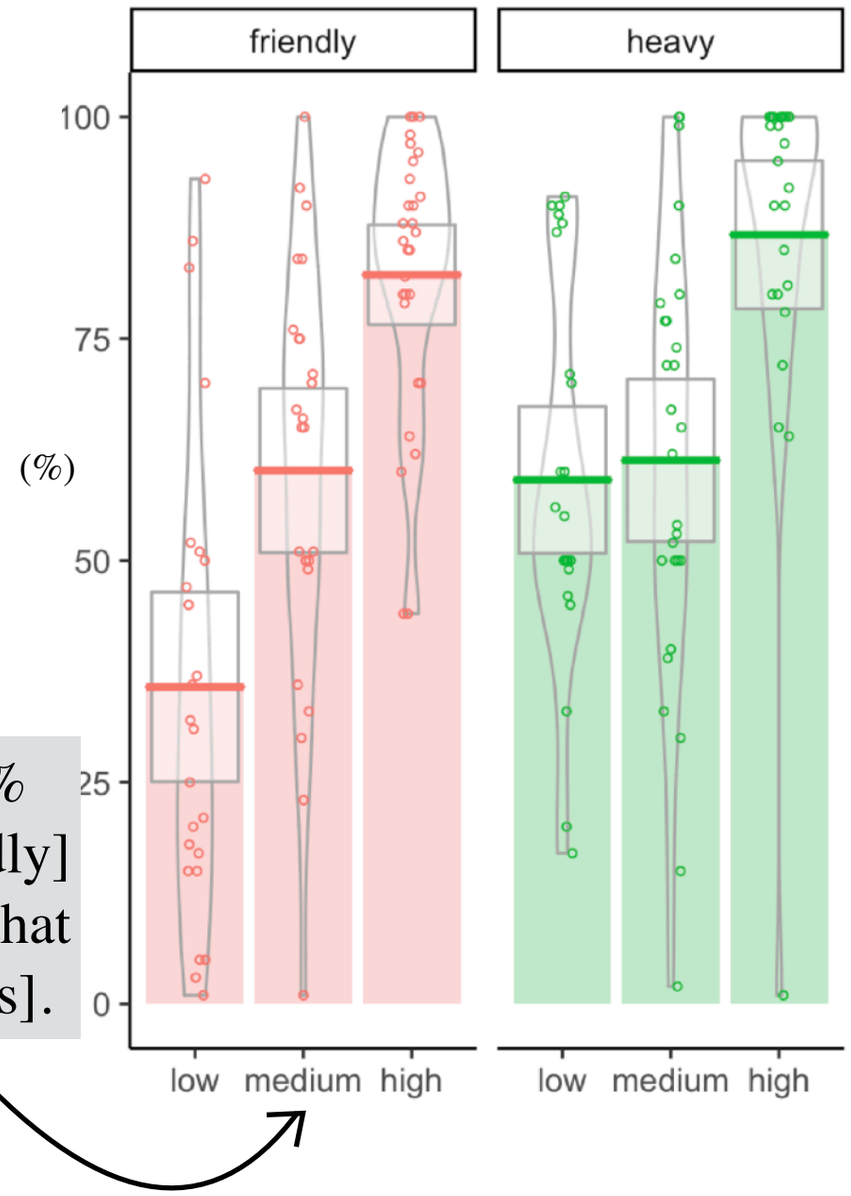
Baseline Generic Estimation by Features (n = 145)



Participants say __% of [goats] are [friendly].

Participants say __% of [feps] are [friendly] when we tell them that [feps] are like [goats].

Novel Generic Estimation by Features (n = 78)



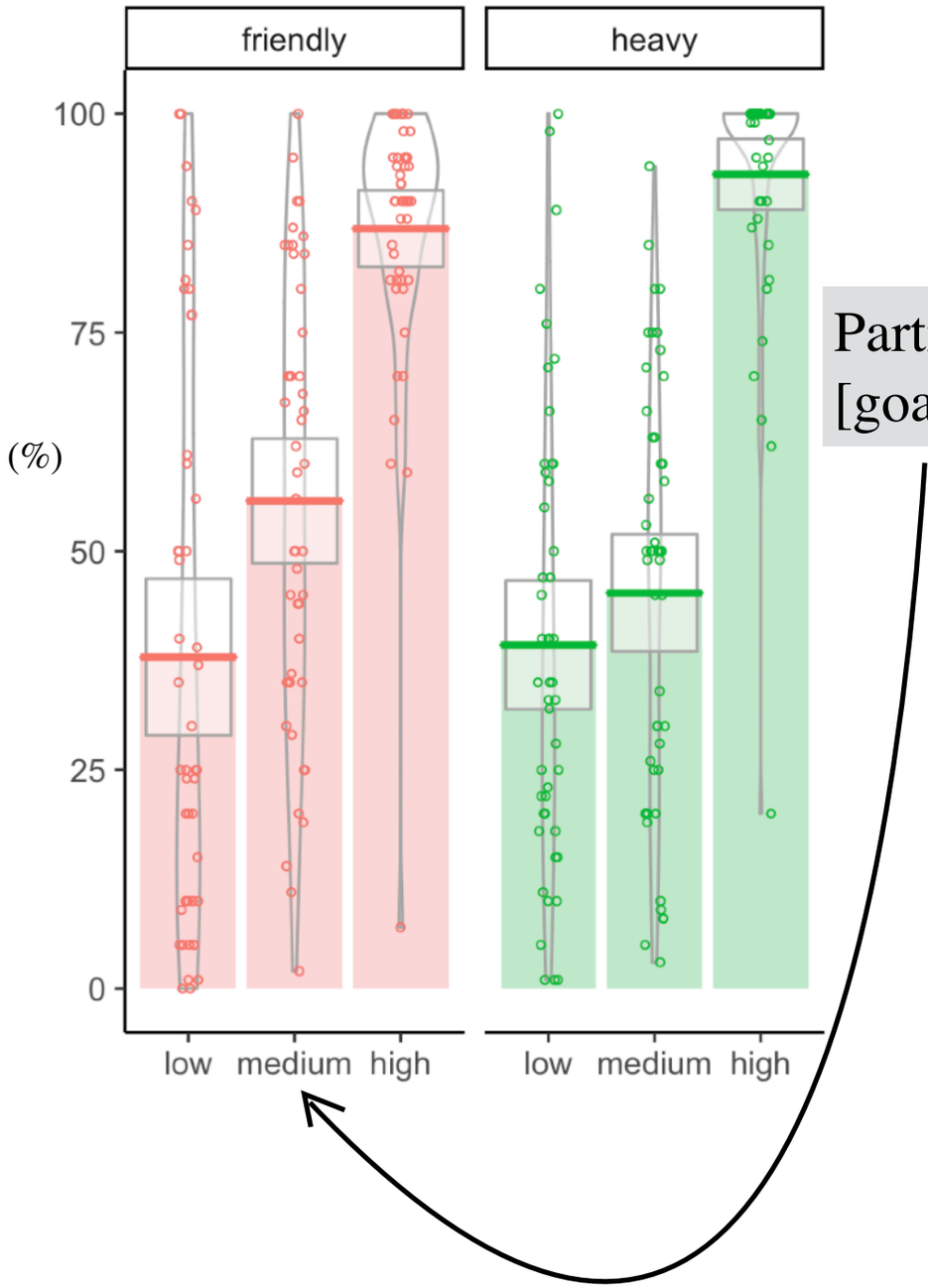
Experiment Conditions for Referenced Familiar Categories by Levels of Probability Intensity

Survey participants: n = 300 (150 each)

Communication and Learning Lab

Preliminary Results

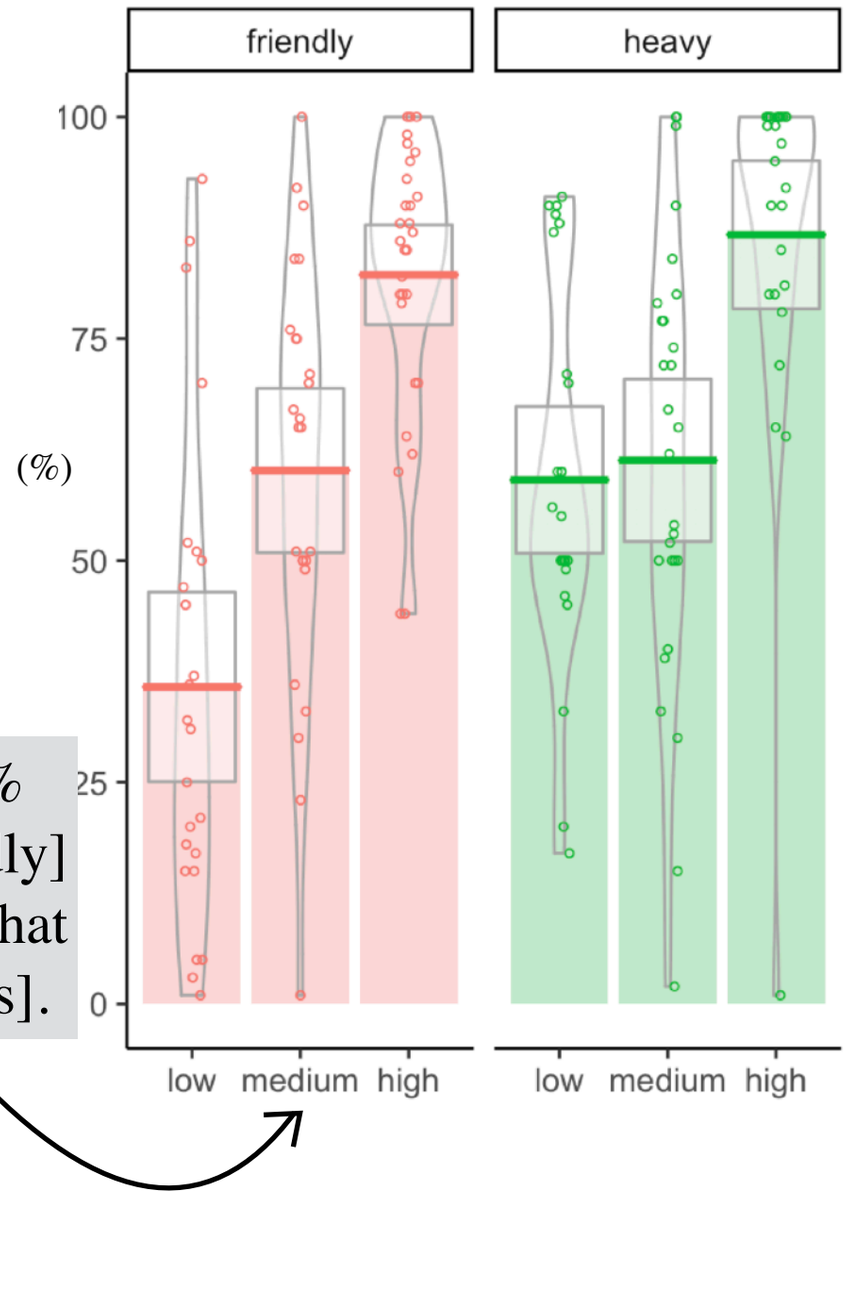
Baseline Generic Estimation by Features (n = 145)



Participants say __% of [goats] are [friendly].

Participants say __% of [feps] are [friendly] when we tell them that [feps] are like [goats].

Novel Generic Estimation by Features (n = 78)



Experiment Conditions for Referenced Familiar Categories by Levels of Probability Intensity

Survey participants: n = 300 (150 each)

OK. You're done.

Let's get the next person up.

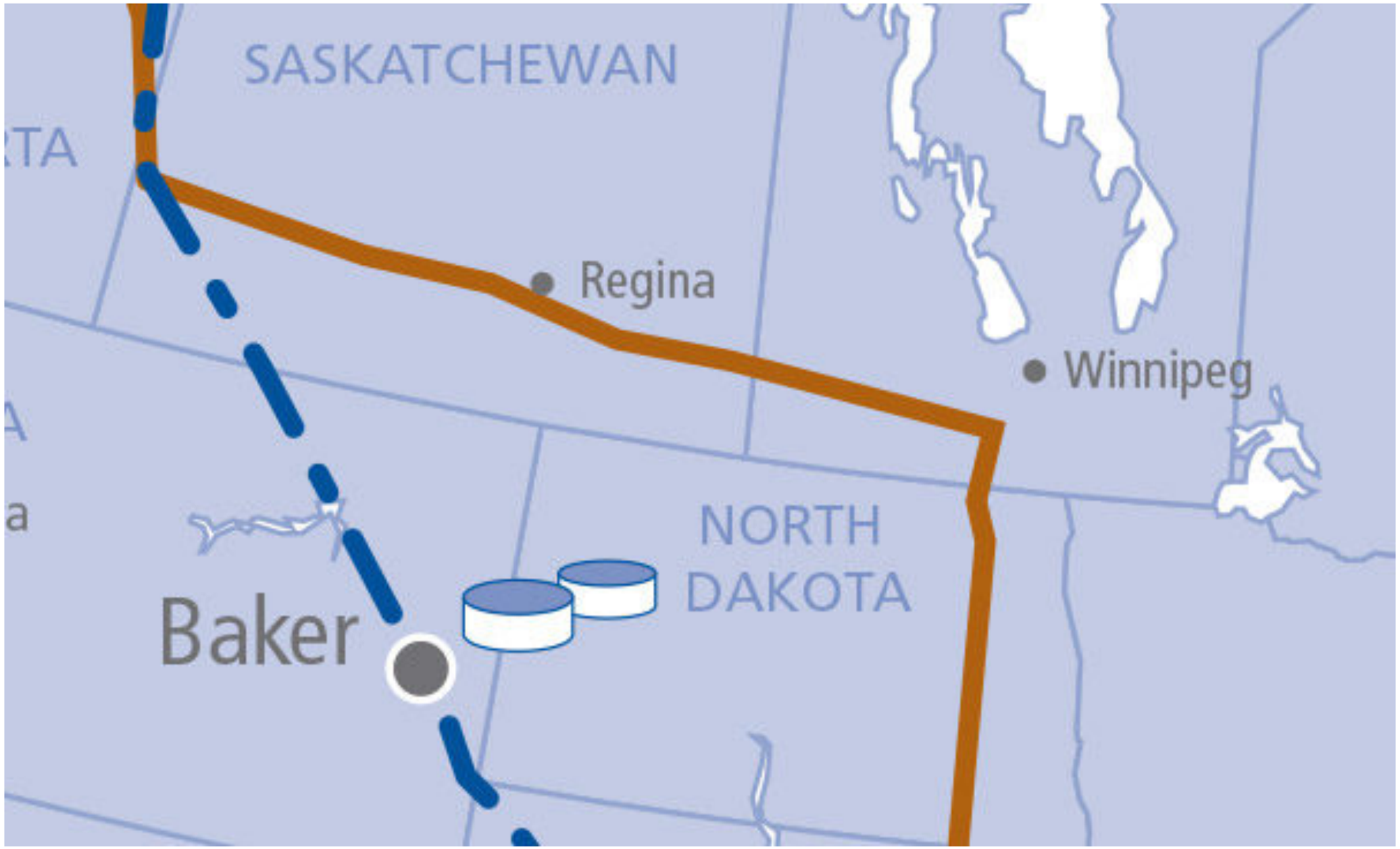
Your presentation starts in 30 seconds

Exploring the public concerns about Keystone XL Pipeline Project

A Topic Analysis Approach

Liqiang Yu
MACSS

Background



Background

- Crude Oil Pipe
- From Canada to US
- Beginning in Hardisty, Alta and ending at Steele City, Neb
- Public comments were collected

Background

regulations.gov
Your Voice in Federal Decision-Making

Home Help ▾ Resources ▾ Contact Us

DOS-2014-0003

Advanced Search

Presidential Permit Applications: TransCanada Keystone Pipeline, L.P, National Interest Determination

This Notice document was issued by the **U.S. Department of State** (DOS)

For related information, **Open Docket Folder**

Comment Period Closed
Mar 7 2014, at 11:59 PM ET

Action

Notice; Solicitation of Comments.

Summary

TransCanada Keystone Pipeline, L.P. applied on May 4, 2012, to the U.S. Department of State ("State Department") for a Presidential Permit that would authorize construction, connection, operation, and maintenance of pipeline facilities on the U.S./Canadian border in Phillips County, Montana for the importation of crude oil. The border facilities would be part of a proposed 875-mile pipeline and related facilities (the Keystone XL project) that is designed to transport up to 830,000 barrels per day of crude oil from Alberta, Canada and the Bakken shale formation in North Dakota and Montana. The pipeline would cross the U.S. border near Morgan, Montana and continue through Montana, South Dakota, and Nebraska, where it would connect to existing pipeline facilities near Steele City, Nebraska for onward delivery to Cushing, Oklahoma and the U.S. Gulf Coast region.

Background information related to the application may be found at *http://www.keystonepipeline-xl.state.gov/*. On January 31, 2014, the State Department released the Final Supplemental Environmental Impact Statement ("Final SEIS") for the proposed Keystone XL project. The application and the Final SEIS, along with other documents, are available through the State Department's web address for the project shown above.

Executive Order 13337 (69 FR 25299) calls on the Secretary of State, or his designee, to determine if issuance of a Presidential Permit would serve the national interest. This decision will take into account a wide range of factors, including energy security; environmental, cultural, and economic impacts; foreign policy; and compliance with relevant federal regulations and issues.

The State Department invites members of the public to comment on any factor they deem relevant to the national interest determination that will be made for this permit application. Along with other factors such as those listed above, these comments will be considered in the final national interest determination. The public

ID: DOS-2014-0003-0001

View original printed format: PDF

Tweet Share Email

Document Information

Date Posted:
Feb 5, 2014

Federal Register Number:
2014-02420

Show More Details

Comments

127,230
Comments Received*

Question

- What are the main topics that the public is worried about Keystone Pipeline Project?
- How are these topics different from experts' opinions?

Data

Comment

Dear Secretary Kerry:

I'm writing to encourage you to approve the construction of the Keystone XL Pipeline as quickly as possible. The pipeline will provide much needed, good paying jobs for our economy while ensuring our country's energy and national security.

Please act now. We cannot afford to wait any longer.

Comment

This is the worst possible project proposed for this or any other country.
You MUST say no.

Data

Regulations.gov /developers

regulations.gov
Your Voice in Federal Decision-Making

Regulations.gov API docs

- Overview
- API Basics
- Interactive API
- Field References
- Blog
- Contribute
- Terms of Service

Examples

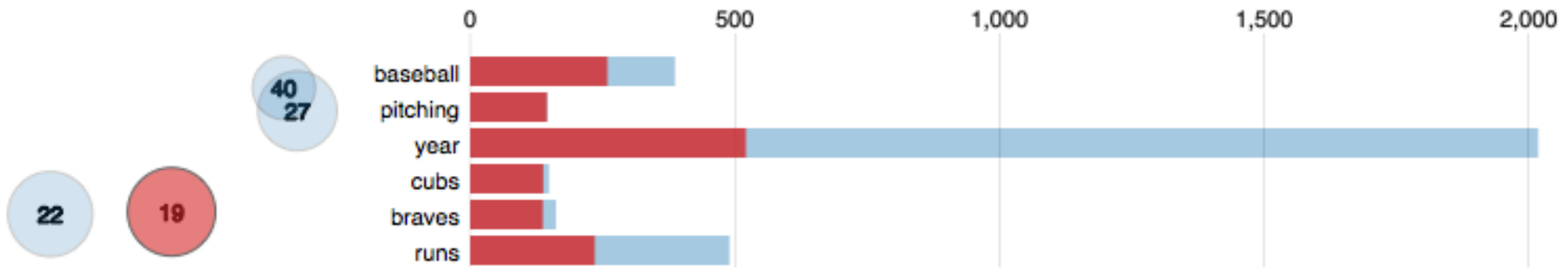
Return a list of all Rules and Proposed Rules posted in the month of September 2014 using this example API Call.

An example regulation showcasing various different fields:


```
{
  "allowLateComment": false,
  "commentDueDate": "2014-12-01T23:59:59-05:00",
  "commentStartDate": "2014-05-18T00:00:00-04:00",
  "openForComment": true,
  "postedDate": "2014-06-18T00:00:00-04:00",
  "receiveDate": "2014-06-18T00:00:00-04:00",
  "status": "Posted",
  "topics": [
    "Administrative Practices and Procedures",
    "Air Pollution Control",
    "Environmental Protection",
    "Intergovernmental Relations",
    "Reporting and Recordkeeping Requirements"
  ],
  "docketTitle": {
    "label": "Docket Title",
    "value": "Standards of Performance for Greenhouse Gas Emissions from Existing Sources: Electric Utility
Generating Units"
  },
  "abstract": {
    "label": "Abstract",
    "value": "Federal Register of June 18, 2014 (79 FR 34830) (FRL-9911-86-OAR)"
  },
  "pageCount": {
    "label": "Page Count",
    "value": "130"
  },
  "docketType": {
    "label": "Docket Type",
    "value": "Rulemaking"
```

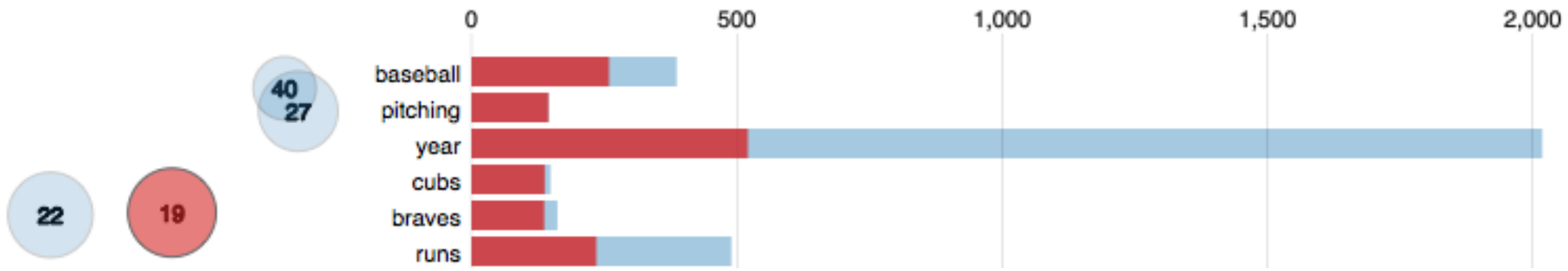
Method

- Topic Modeling



Method

- Topic Modeling



https://ldavis.cpsievert.me/reviews/vis/#topic=7&lambda=0.6&term=

Reference

- Paul W. Parfomak, Robert Pirog, Linda Luther and Adam Vann, “Keystone XL Pipeline Project: Key Issues”

- David M. Blei and John D. Lafferty, “Topic Models”

- Peter Erickson and Michael Lazarus, “Impact of the Keystone XL pipeline on global oil markets and greenhouse gas emissions”