Tel Aviv University
School of Management

חוכמת ההמונים, Big Data והתנהגות צרכנית ברשת

שחר רייכמן
הפקולטה לניהול, אוניברסיטת תל אביב
sr@tau.ac.il

Tel Aviv University
School of Management

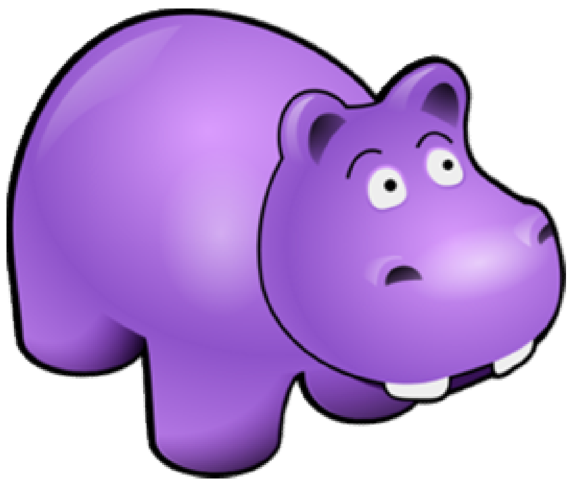
Data Driven Decision Making

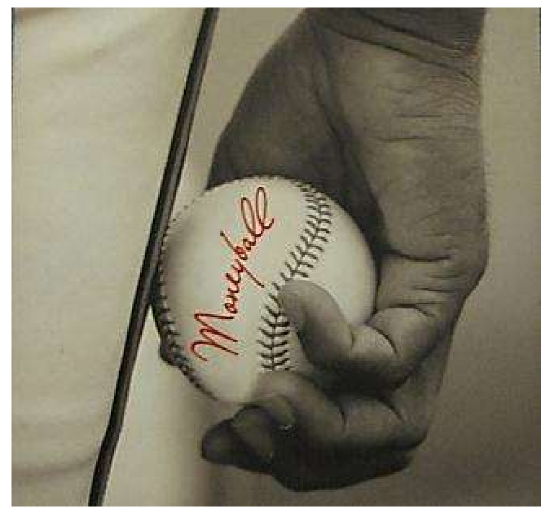
"After careful consideration of all 437 charts, graphs, and metrics, I've decided to throw up my hands, hit the liquor store, and get snockered. Who's with me?!"

Source: equest.com

Motivation

Hippo Vs. DDD





Hi**ghest **P**aid **P**erson's **O**pinion

Data **D**riven **D**ecision making

Data Science

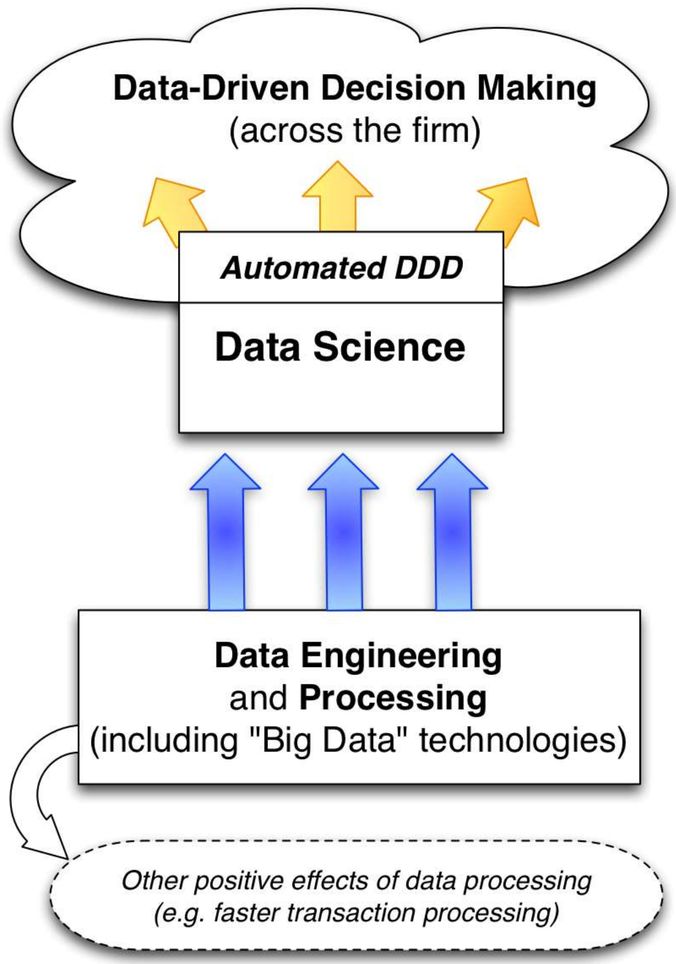


Image from "Data Science for Business", Provost & Fawcett, 2013

Data Scientist:

The Sexiest Job of the 21st Century

Harvard Business Review October 2012

By 2018, the United States alone could face a **shortage** of 140K-190K people with analytical expertise and **1.5 million managers** and analysts **with the skills to understand and make decisions based on the analysis of big data**.

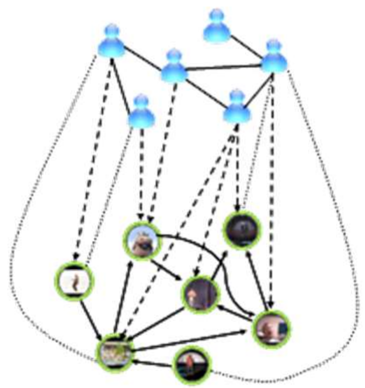
(Manyika, 2011).

Tel Aviv University
School of Management

Research Agenda

- Utilizing Big Data to improve:
  - Businesses' performances
  - Consumers' experiences

- Analyzing Networks-of-Networks:
  - Managerial decisions
  - Consumers decisions
- Online Experiments
  - Cause and effect in business decisions



Tel Aviv University
School of Management

Crowd-Squared:

Amplifying the Predictive Power of Large-Scale Crowd-Based Data

Shachar Reichman[1,2]

Erik Brynjolfsson[2], Tomer Geva[1]

[1] School of Mangement, Tel Aviv University

[2] MIT Sloan School of Management

Introduction

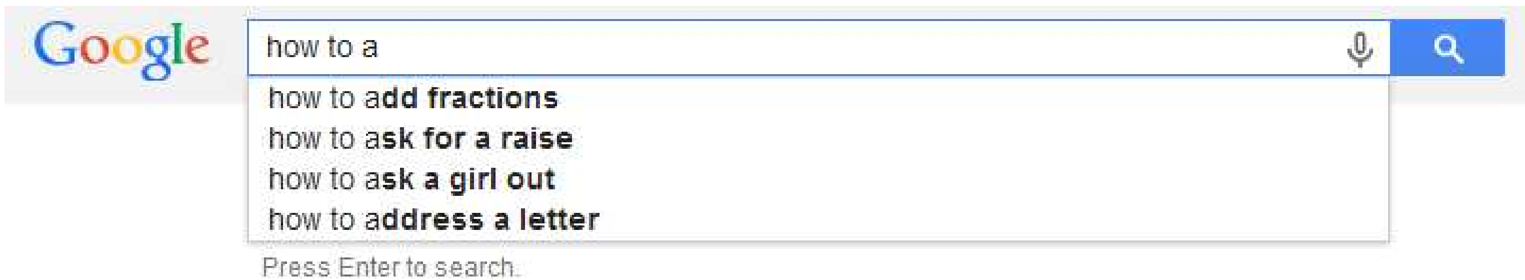
Consumers activity online

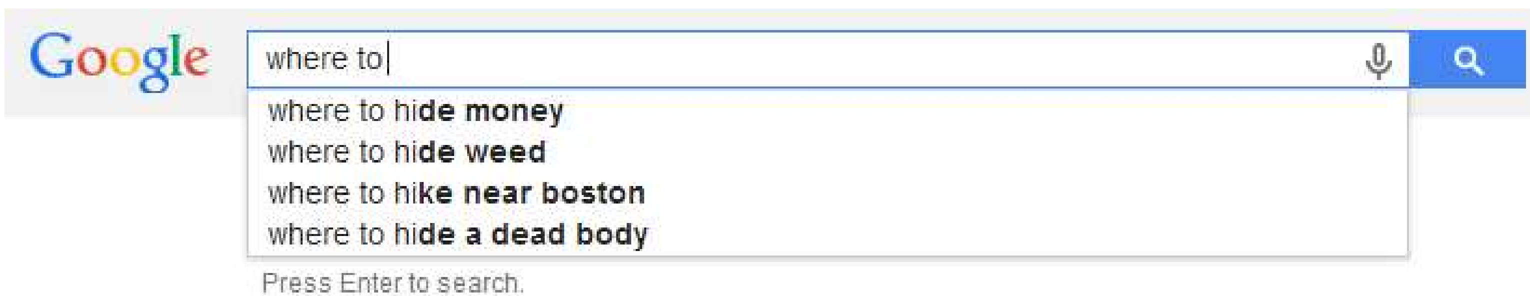
- Search
- Opinions
- Purchase

Introduction

Sources:
1. Instagram.com
2. twitter.com

Introduction





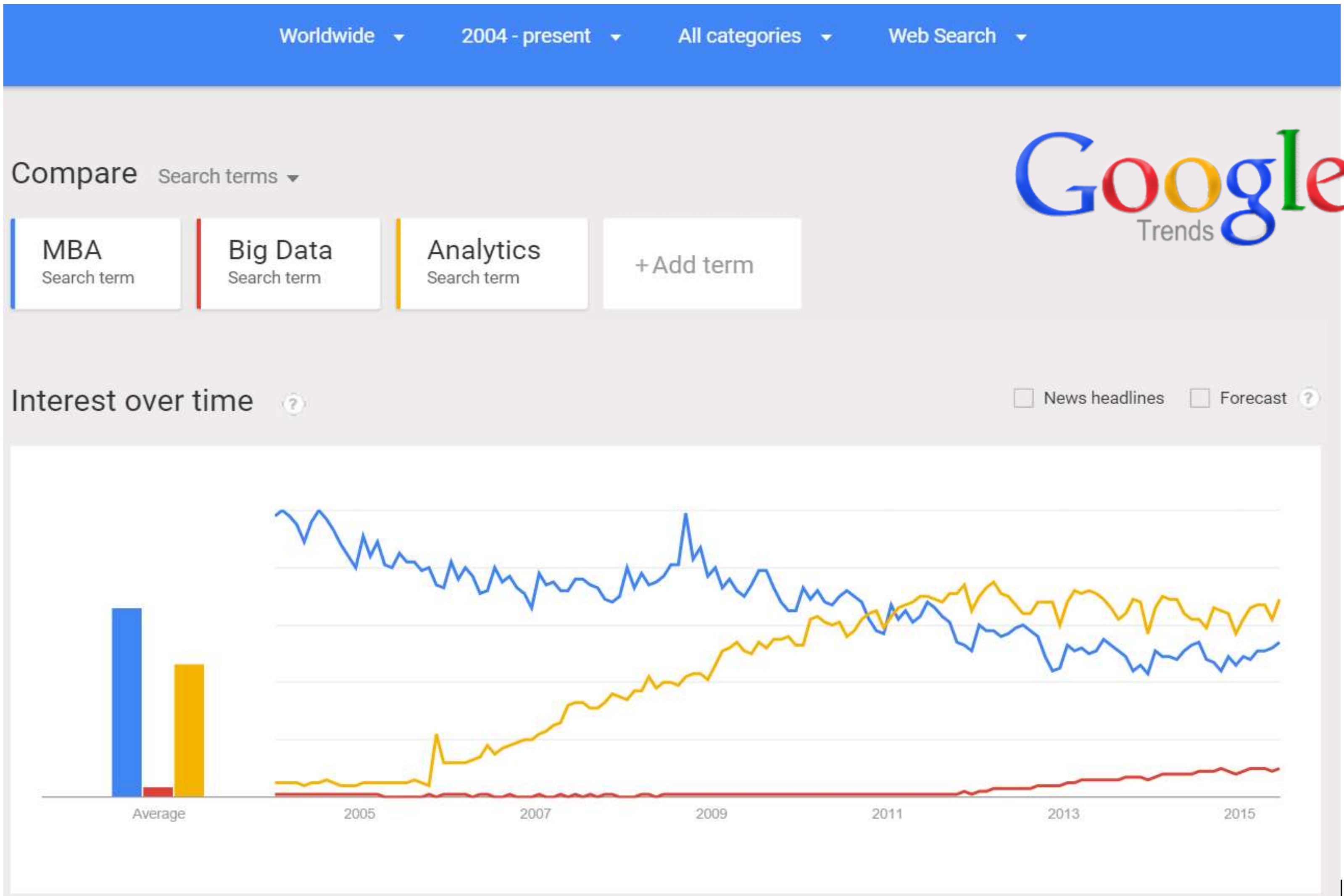
Introduction

Online activity reveals valuable information

- Consumer's intentions
- Consumer's preferences
- Purchase decisions
- Information on the intention of groups of consumers

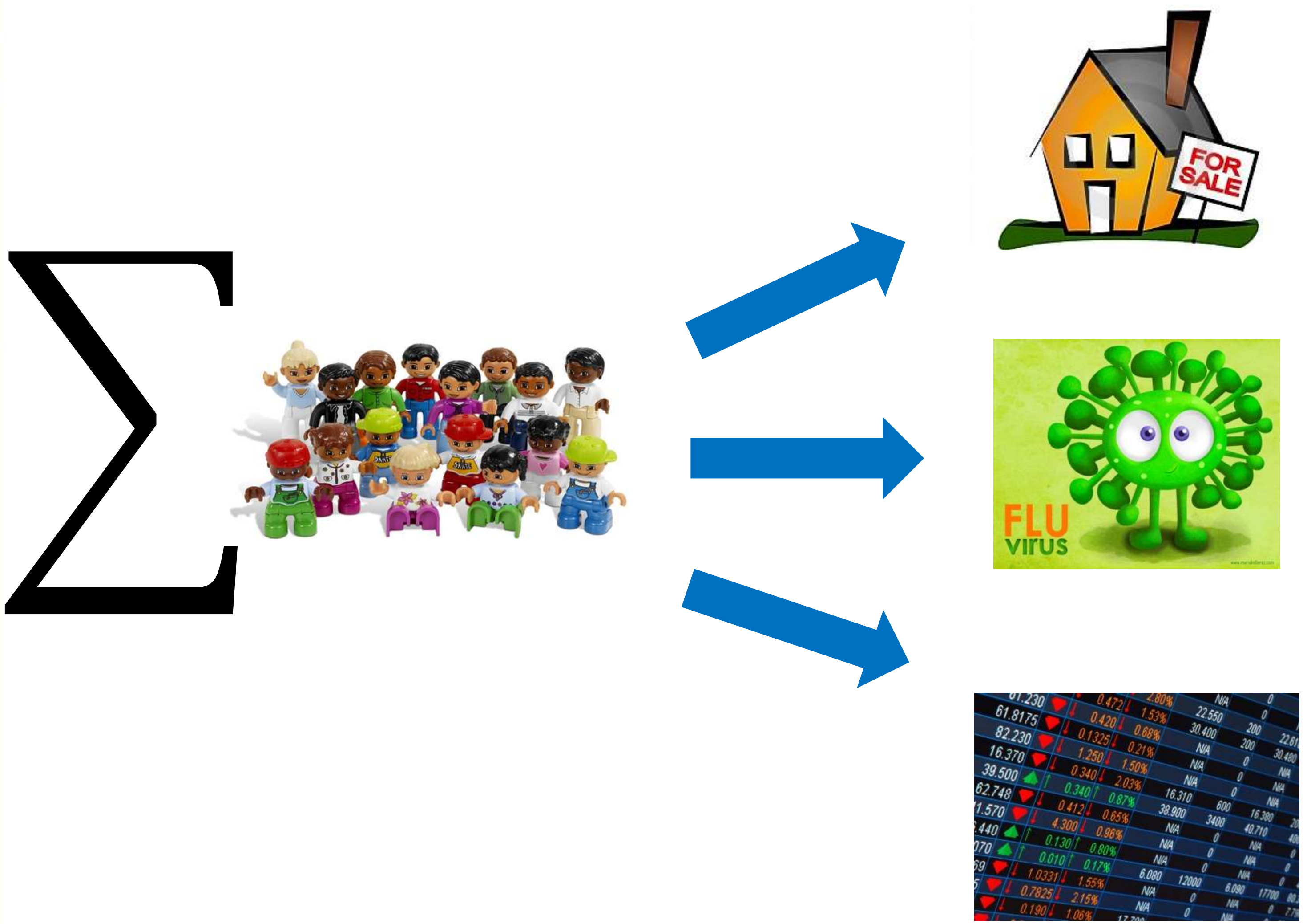
Tel Aviv University
School of Management

Predictions using online activity

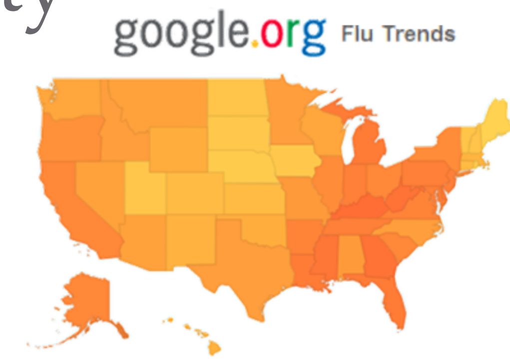


http://www.google.com/trends/

Predictions using online activity



Predictions using online activity



- Flu trends (Ginsberg et al. 2009)
- Real estate market(Wu & Brynjolfsson 2009)
- Unemployment (Choi & Varian 2012)
- Financial Trading (Preis et al. 2013)

Predictions using online activity

N – a set of all activities

D activities

L activities

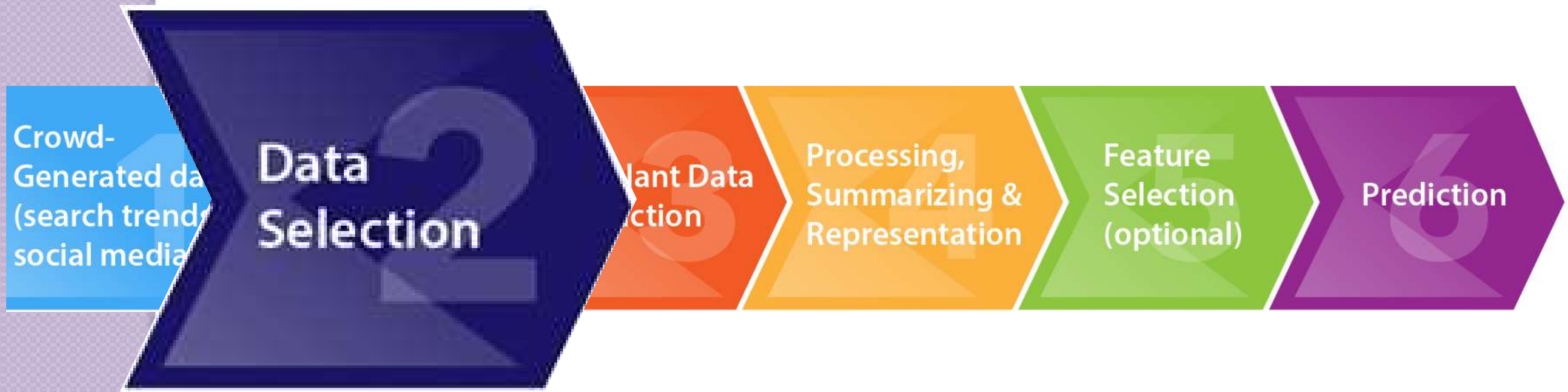
G activities

M activities

$$\sum\sum$$

?

Predictions using online activity



Tel Aviv University School of Management

Motivation

How to choose the "right" data?

"Right" → best reflect the phenomenon

Goal:

Develop a structured and practical data selection method

Tel Aviv University
School of Management

Predictions using online activity

<u>Current keywords selection methods</u>

- Prior knowledge and Intuition

(Seebach et al. 2011, D'Amuri and Marcucci 2012)

London calling: NFL wants UK team

Predictions using online activity

<u>Current keywords selection methods</u>

- Prior knowledge and Intuition

  (Seebach et al. 2011, D'Amuri and Marcucci 2012)

- Comprehensive scan of search engine data

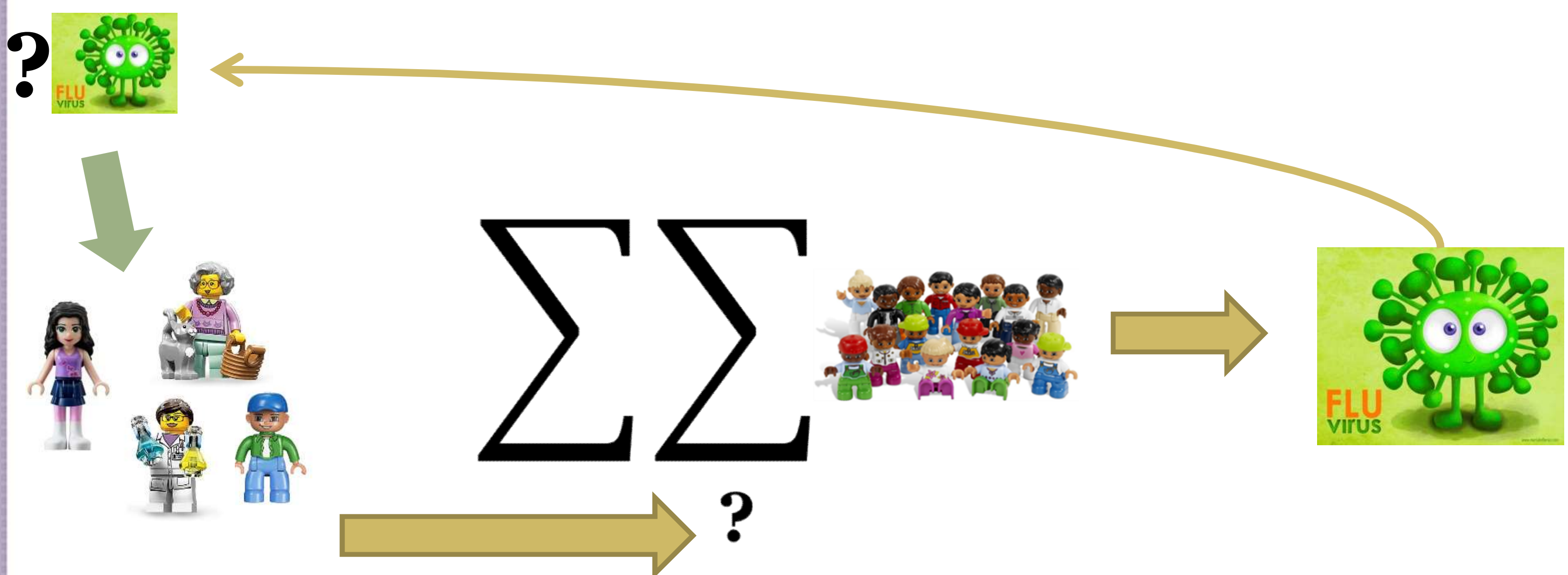
  (Ginsberg et al. 2009)

- Search engine categories (automated classifier)

  (Wu & Brynjolfsson 2009, Choi and Varian, 2012)

Crowd-Squared

Crowdsource keywords selection



Crowd-Squared

<u>Crowdsourcing</u>:

"The practice of obtaining needed services, **ideas, or content** by soliciting contributions **from a large group of people…"**

(Source: Merriam-Webster Dictionary)

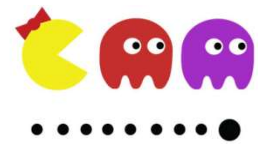
Tel Aviv University
School of Management

Crowd-Squared

Crowdsourcing advantages:

- A variety of users
  - Backgrounds
  - level of expertise
  - Demographics
- Low cost

Methodology



- Online game environment
  - Effective technique to capture crowd knowledge
  - Provides reliable information without any supplementary verification of the users' answers

(Von Ahn 2006)

  - May generate results at the same level as experts

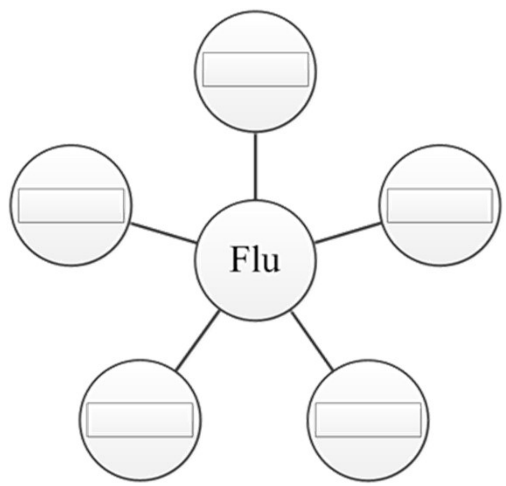
(Snow et al. 2008)

Word Association Game

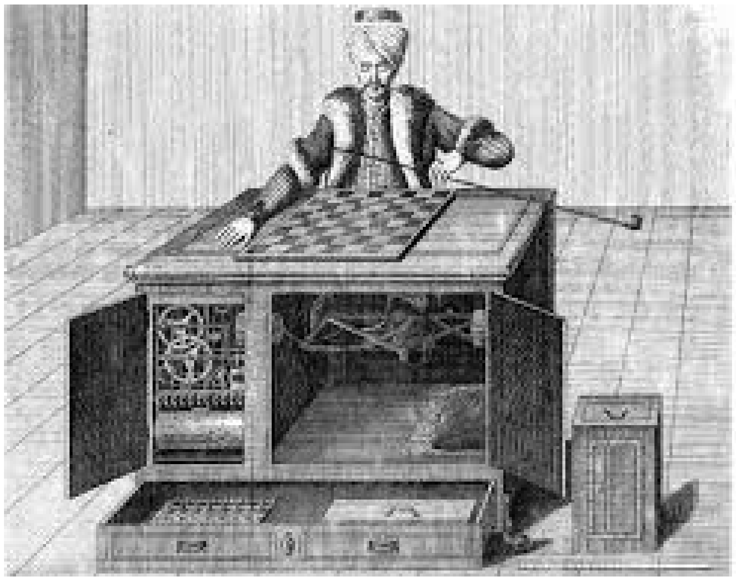
- Word association game website on the

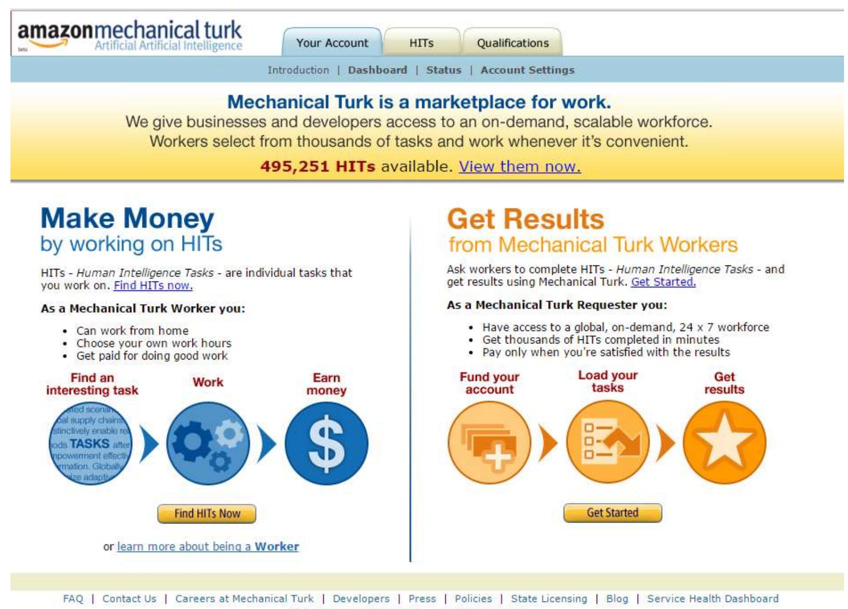
amazon mechanical turk
Artificial Artificial Intelligence
platform

Please provide 5 terms that come to mind when seeing the word:



Amazon Mechanical Turk

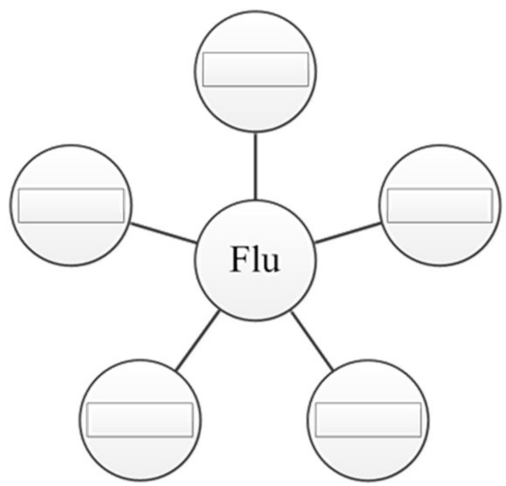




Word Association Game

- Word association game website on the amazon mechanical turk (Artificial Artificial Intelligence) platform

Please provide 5 terms that come to mind when seeing the word:



Why Word Association?

- People are using the web as external memory

  (Sparrow, Liu & Wegner 2011)

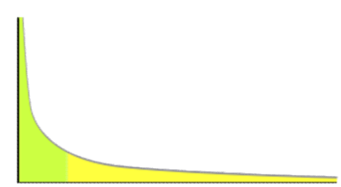
- Accessing the “web memory” will involve similar processes of retrieving from the human memory

Tel Aviv University School of Management

Why Word Association?

Words that come to mind when seeing a <u>term</u>

- Used in everyday activities for "collecting thoughts" (Nelson et al. 2000)
- Taps into lexical knowledge that is based on real-world experience (Nelson et al. 2004)
- Consistent across different people in the same culture (Nelson et al. 1998)
- Provides a power law distribution of term associations (Steyvers and Tenenbaum 2005)



Word Association Game

- 1100 participants (550 in each domain)
- Each participant was paid $0.05-$0.07
- Average duration - 53 seconds, including 4 demographic questions

Tel Aviv University
School of Management

Predicted Variables

- Influenza epidemics - ILI
- Unemployment – Initial claims for unemployment

Empirical Analysis

- Comparison to well known studies using search trends data:
  - Ginsberg et al., 2009 – Flu outbreak prediction
  - Choi and Varian, 2012 – Unemployment claim prediction

- **Only difference - data selection procedure**

Tel Aviv University
School of Management

Compared Forecast Models

	Influenza epidemics	**Unemployment**
Compared Model	Ginsberg et al. ,2008 (Google Flu Trends)	Choi and Varian, 2012
Keyword selection method	50 million most popular search terms	Google's categoris: "Jobs" and "Welfare & Unemployment"
Training Period	Jan 2004 - Mar 2007 (167 Weeks)	Jan 2004 - July 2011 (a one-week-ahead rolling prediction)
Validation Period	Mar 2007 - May 2008 (61 Weeks)	
Evaluation Index	Out-of-Sample Correlation	Mean Absolute Error (MAE)

Tel Aviv University
School of Management

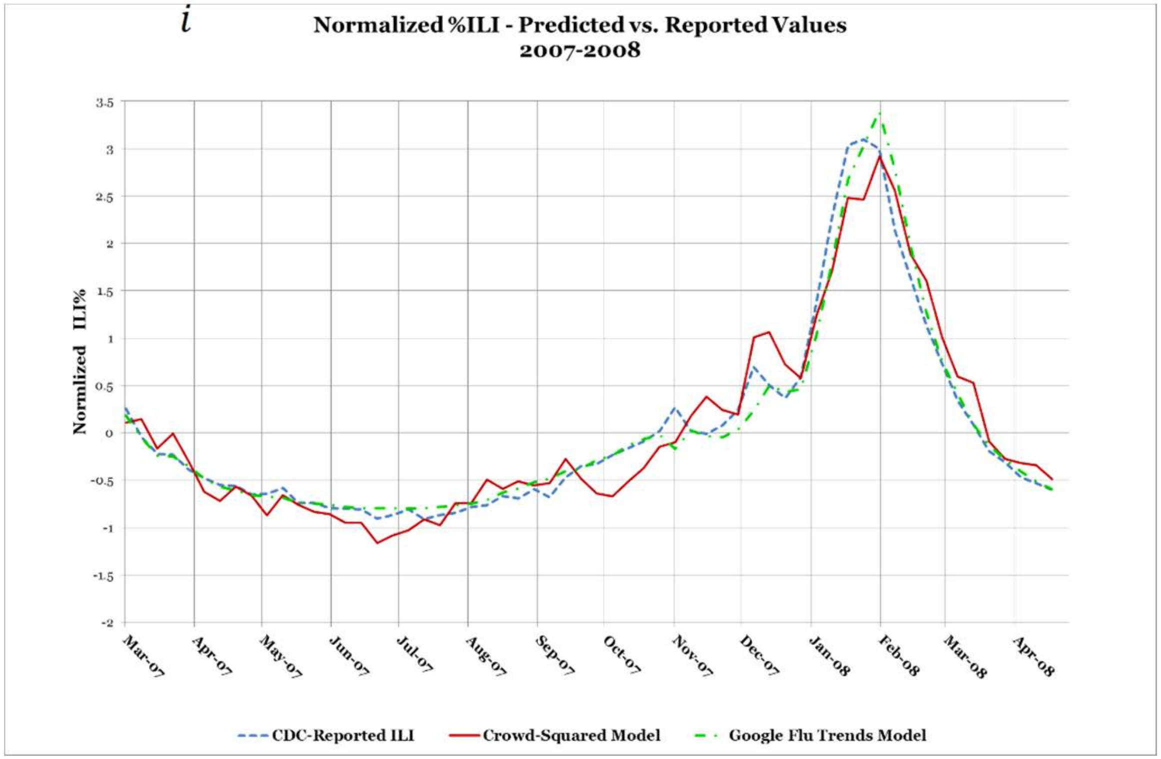
Results

We generated a list of the terms associated with each domain

Results - Influenza Epidemics

$$ILI(t) = \alpha + \sum_i \beta_i AssociatedTerm_i\,(t) + \varepsilon_i$$

2007-2008



Out-of-Sample Correlation (2007-2008)	
Crowed-Squared	Google Flu-Trends
0.97	0.97

nature International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For

Archive > Volume 494 > Issue 7436 > News > Article

NATURE | NEWS

عربي

When Google got flu wrong

US outbreak foxes a leading web-based method for tracking seasonal flu.

Declan Butler

13 February 2013

PDF | Rights & Permissions

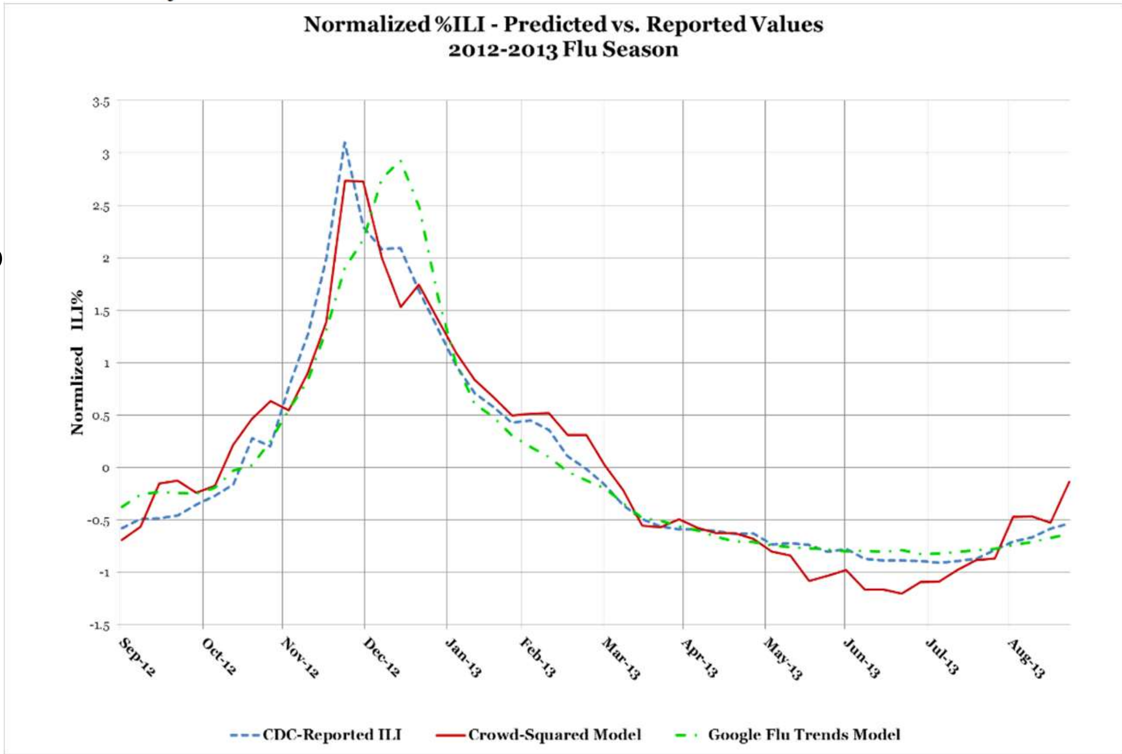
Tel Aviv University
School of Management

Tel Aviv University
School of Management

Results - Influenza Epidemics

$$ILI(t) = \alpha + \sum_i \beta_i AssociatedTerm_i\,(t) + \varepsilon_i$$

2012-2013



Out-of-Sample Correlation (2012-2013)	
Crowed-Squared	Google Flu-Trends
0.971	0.955

Tel Aviv University
School of Management

Influenza Epidemics

BIG DATA

The Parable of Google Flu: Traps in Big Data Analysis

Science
AAAS

David Lazer,[1,2*] Ryan Kennedy,[1,3,4] Gary King,[3] Alessandro Vespignani[3,5,6]

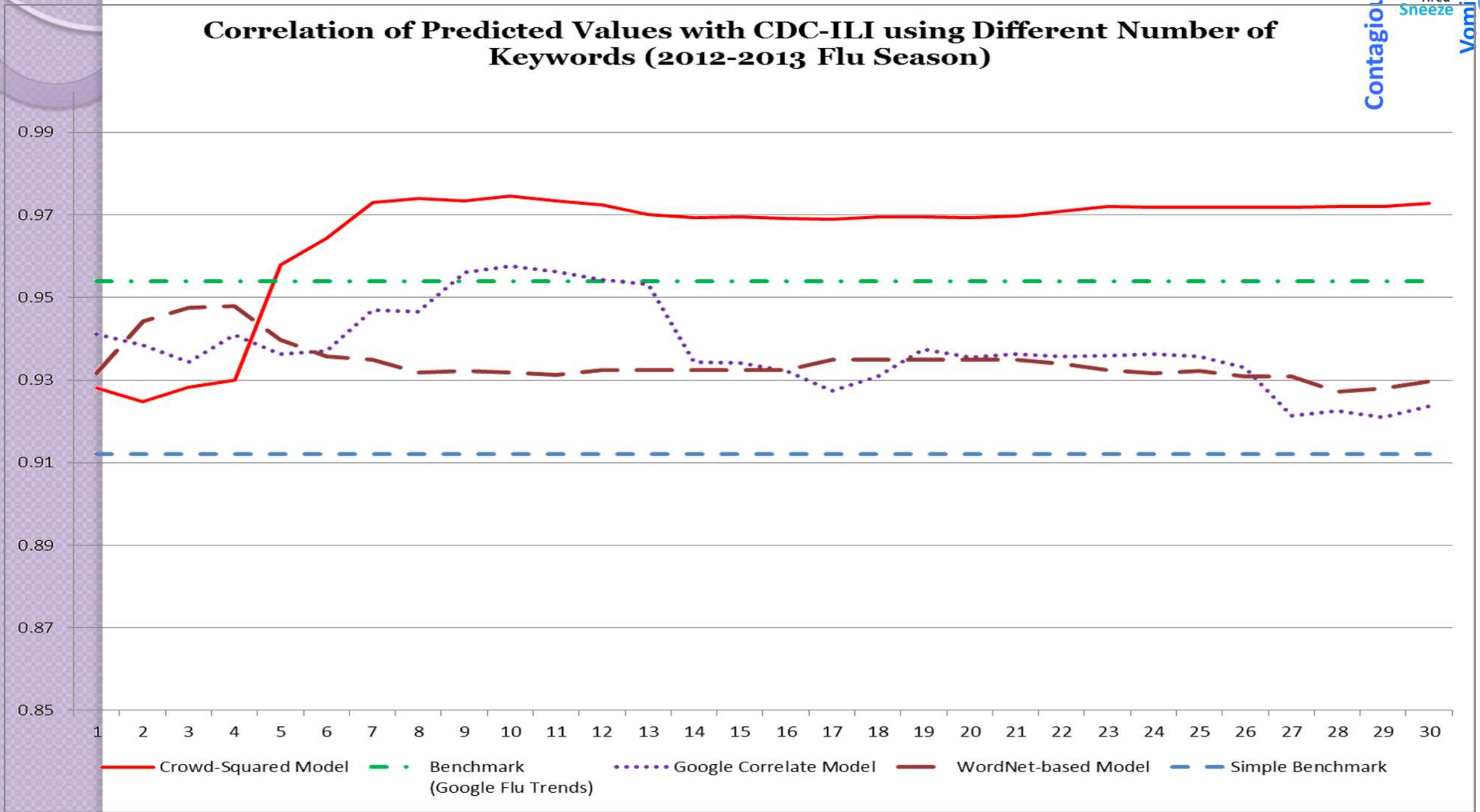
In February 2013, Google Flu Trends (GFT) made headlines but not for a reason that Google executives or the creators of the flu tracking system would have hoped. *Nature* reported that GFT was predicting more than double the proportion of doctor visits for influenza-like illness (ILI) than the Centers for Disease Control and Prevention (CDC), which bases its estimates on surveillance reports from laboratories across the United States (*1*, *2*). This happened despite the fact that GFT was built to predict CDC reports. Given that GFT is often held up as an exemplary use of big data (*3*, *4*), what lessons can we draw from this error?

The problems we identify are not limited to GFT. Research on whether search or social media can

Out-of-Sample MAE		
Crowd - Squared	Lazer et al.	Google Flu Trends
0.209	0.232	0.486

* Lazer et al., "The Parable of Google Flu: Traps in Big Data Analysis," 2014, *Science*

Influenza Epidemics: Sensitivity Analysis



Initial Claims for Unemployment Benefits

- Choi and Varian, 2012
- Categories selected by expert researchers
  - Google Trends categories : "Jobs" and "Welfare & Unemployment"
- Benchmarks: AR(1), Google Correlate, WordNet Lexicon

$$UIC(t) = \alpha + \delta_1 UIC(t-1) + \sum_i \beta_i AssociatedTerm_i\,(t)$$

Initial Claims for Unemployment Benefits

- Jan 2004 - July 2011

  (one-week-ahead expanding window prediction)

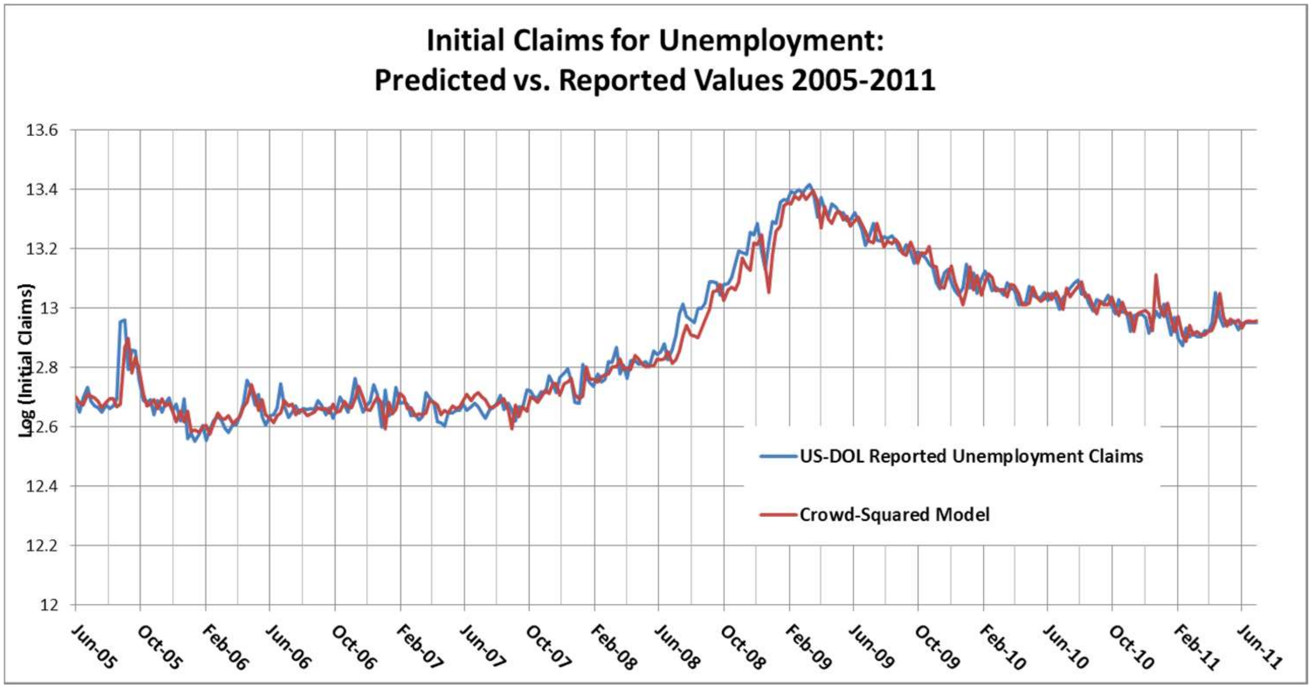
- Performance Measure: MAE (Out-of-Sample)

Out-of-Sample Mean Absolute Error (MAE)				
Crowd-Squared	Choi and Varian, 2012	AR (1)	Google Correlate	WordNet
3.24%	3.68%	3.36%	3.45%	3.69%

Tel Aviv University
School of Management

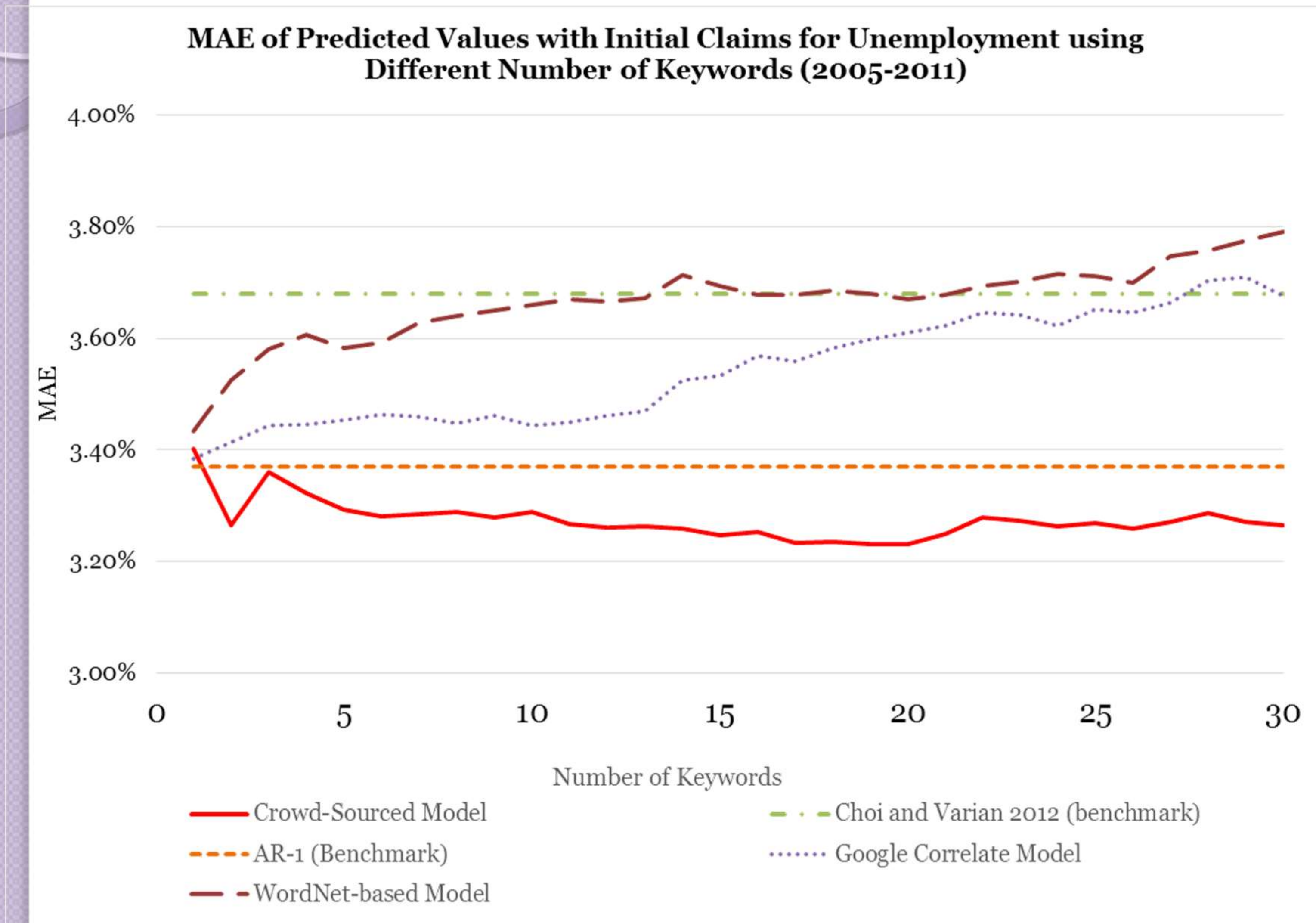
Results - Unemployment

$$UIC(t) = \alpha + \delta_1 UIC(t-1) + \sum_i \beta_i AssociatedTerm_i\,(t)$$



Out-of-Sample Mean Absolute Error (MAE)	
Crowed-Squared	Choi-Varian
3.23%	3.68%

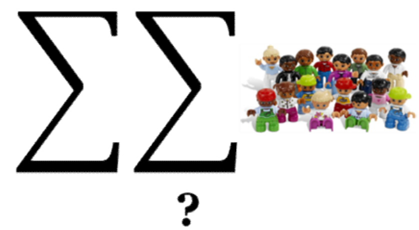
Unemployment - Sensitivity Analysis



Tel Aviv University
School of Management

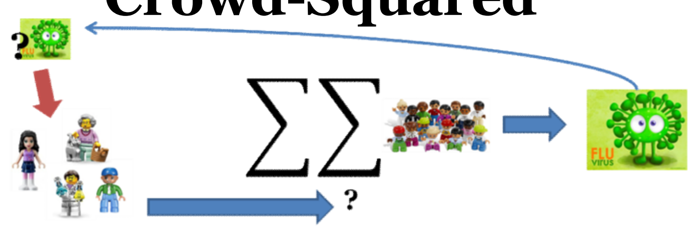
Summary

- Data selection is a critical aspect in predictions using crowd data



- We present a the **crowd-squared**, a concept for using the crowd to identify relevant keywords

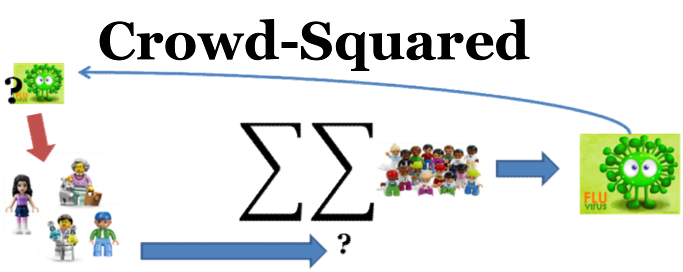
Crowd-Squared



Summary

Crowd-Squared provides:

- Low cost
- Low computational requirements
- Structured and transparent approach
- Easy to implement
- Good predictive performance



Tel Aviv University
School of Management

When Online Engagement Gets in the Way of Offline Sales

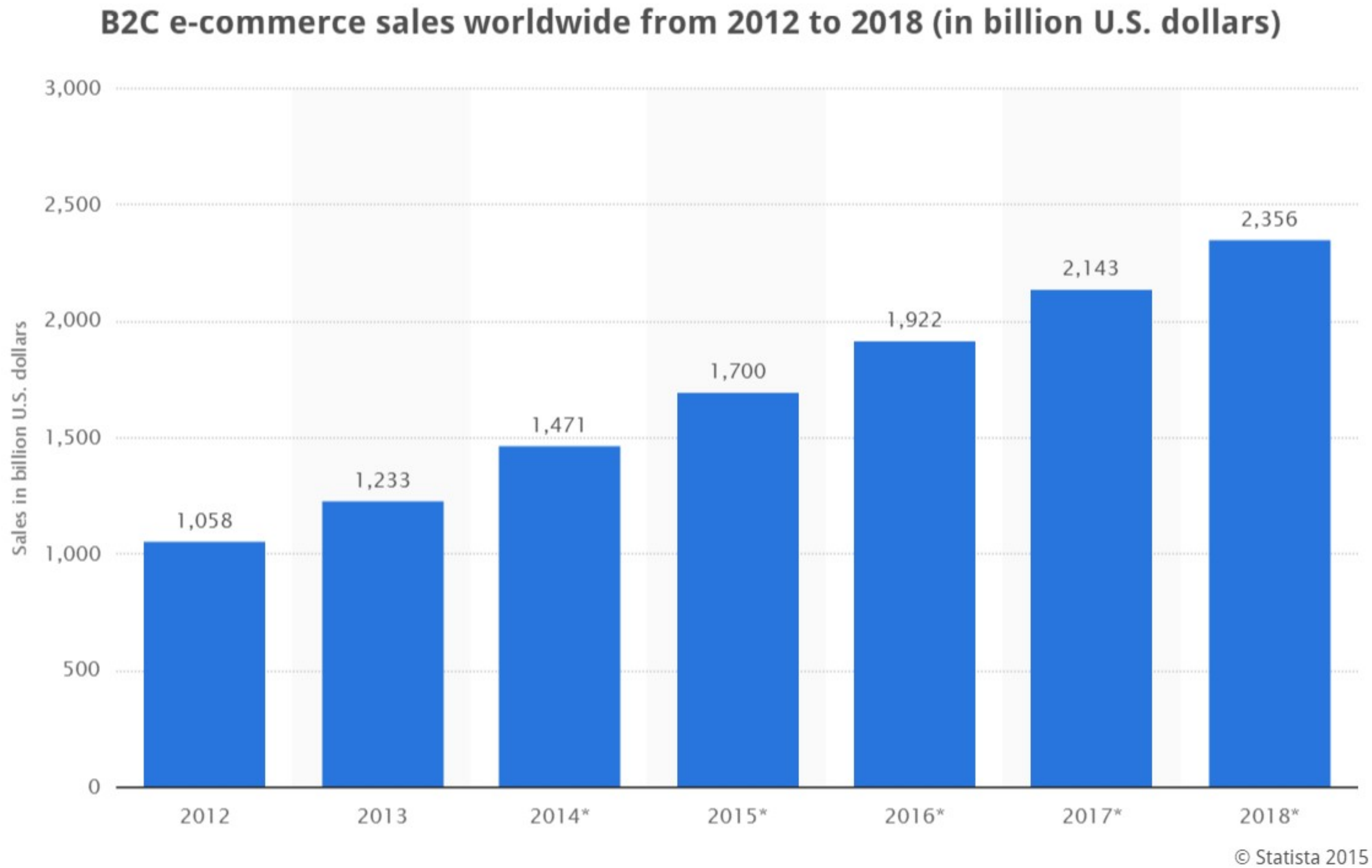
Sagit Bar-Gill and Shachar Reichman

Tel Aviv University
School of Management

E-Commerce is Huge

Source: http://www.executionists.com/

E-Commerce is Growing



Online Engagement

- Consumers' interactive brand-related dynamics

(Brodie et al. 2011)

- High relevance of brands to consumers
- The development of an emotional connection between consumers and brands

(Rappaport , 2007)

Tel Aviv University
School of Management

Online Engagement – The Holy Grail?

- "Online mechanism that delivers competitive advantage"

  (Mollen & Wilson 2010)

- "Expected to provide enhanced predictive and explanatory power of focal consumer behavior outcomes"

  (Hollebeek et al. 2014)

Tel Aviv University School of Management

Brick-and-mortar is not dead yet!

* Source: Forrester, 2015

Tel Aviv University
School of Management

The Case of Amazon.com

theguardian

UK world sport football opinion culture business lifestyle fashion environment tech travel

home › tech

Amazon.com

Graham Ruddick

Tuesday 3 November 2015 12.20 GMT

Amazon begins a new chapter with opening of first physical bookstore

World's biggest online retailer opens shop in Seattle's University Village stocked with 6,000 books at same price as on its website

E-Commerce Moves Offline?

- Amazon stores

- Google opened first-ever shop in London:

"With the Google shop, we want to offer people a place where they can play, experiment and learn about all of what Google has to offer" (James Elias, Google UK marketing director)

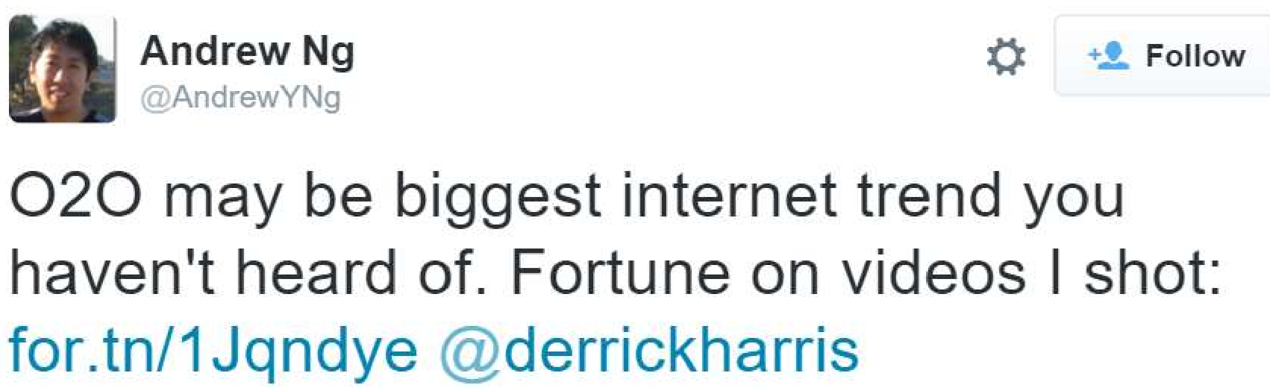
- Frank & Oak (online menswear start-up) now opening offline locations:

".. physical space that we can leverage to communicate our brand value"

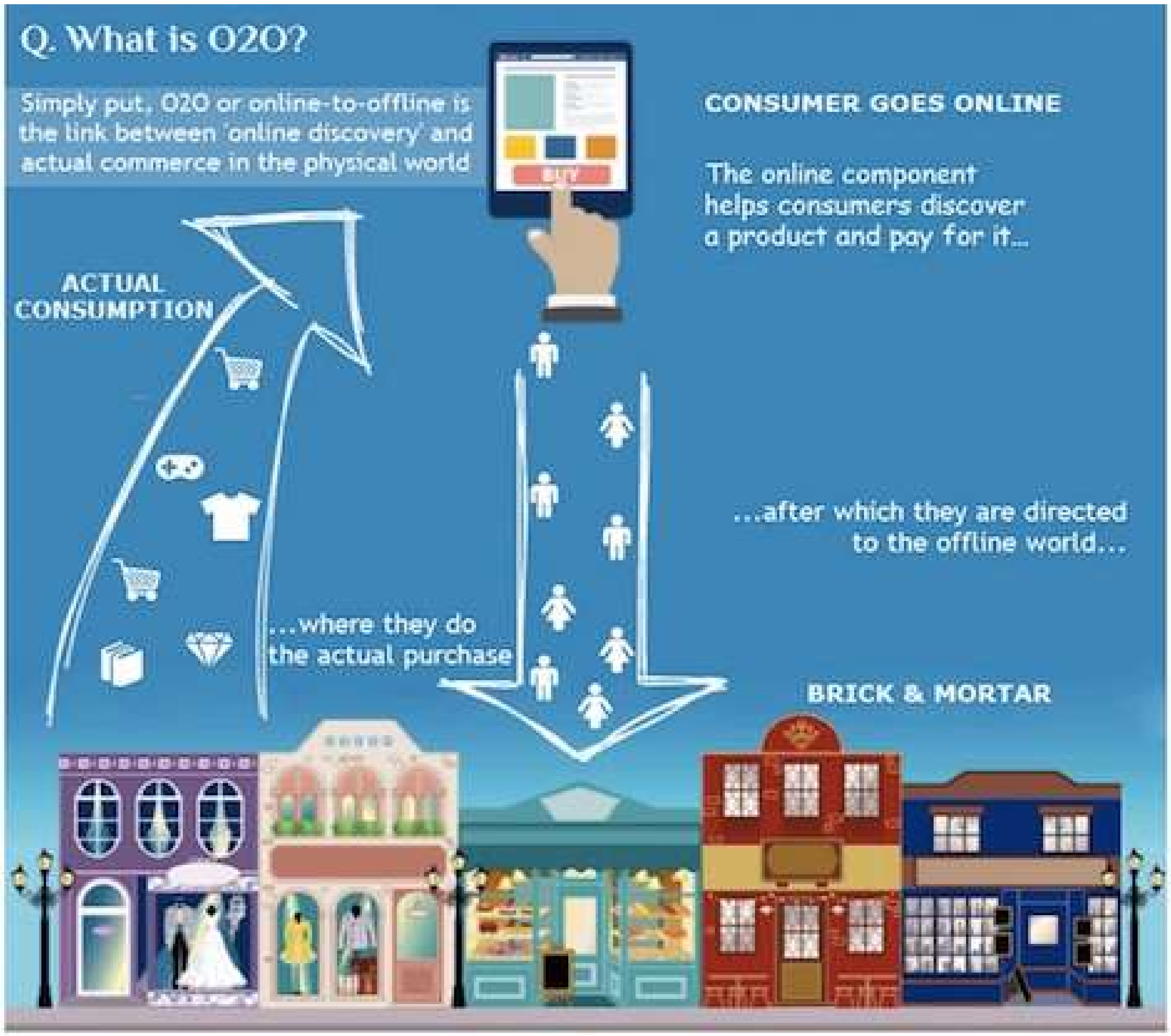
Online-to-Offline – O2O

- Anything digital which brings people to shop offline

Baidu 百度



Online-to-Offline – O2O



Online Engagement – The Holy Grail?

- Online engagement → e-WOM, online sales

  (e.g. Chevalier and Mayzlin 2006; Godes and Mayzlin 2009; Gu et al. 2012)

- Online and offline retailers maximize traffic and online engagement
  - Shown to increase online purchase probability

    (Agarwal and Venkatesh 2002 and 2006)

Tel Aviv University
School of Management

Online vs. Offline

See and feel at brick-and-mortar → Physical stores as showrooms for online shoppers? → Reverse is more common!

- Offline → online: 46%
- Online → offline: 69%

(recent poll, US shoppers)

- Online-offline strategies studied:
  - Connections between websites and physical stores are strategic tools (Ghose et al. 2007)
  - Firm responses to online leads → conversions (Oldroyd et al. 2011)
  - Omni-channel retailing for products sold both online and offline (Brynjolfsson et al. 2013)

Online vs. Offline

- Online and offline activity are substitutes

(Brynjolfsson et al. 2009, Forman et al. 2009)

- Complementing cross-channel effects

(Wiesel et al. 2011)

- Complementarities and substitutes

Goldfarb & Wang (2014)

Tel Aviv University
School of Management

Research Objective

How online engagement affects offline sales

- Pure offline products – products sold only offline

Tel Aviv University
School of Management

New Brand Website → Increase Engagement

- Leading automobile brand launched a new interactive website in 9 out of 15 markets

- Upgraded website aimed at "increasing user engagement"
- Launch times exogenous (according to manufacturer).

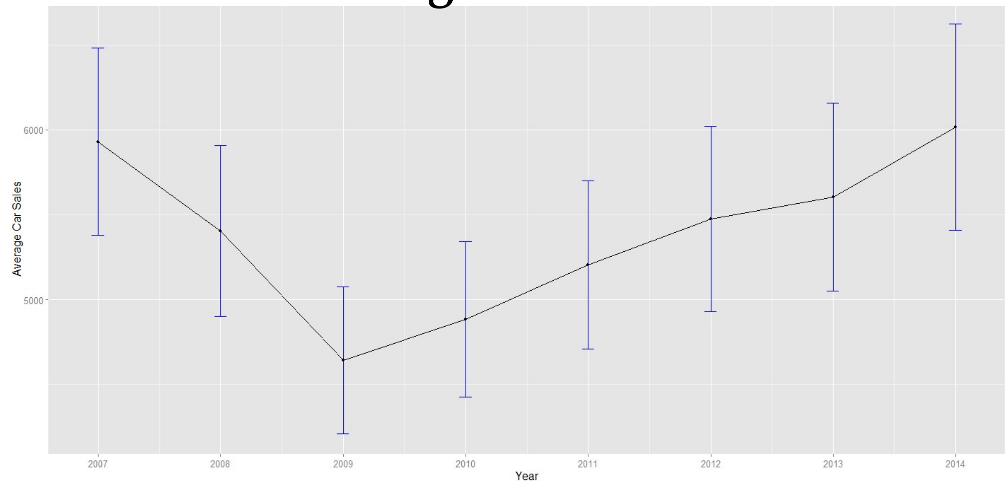
Tel Aviv University
School of Management

Data

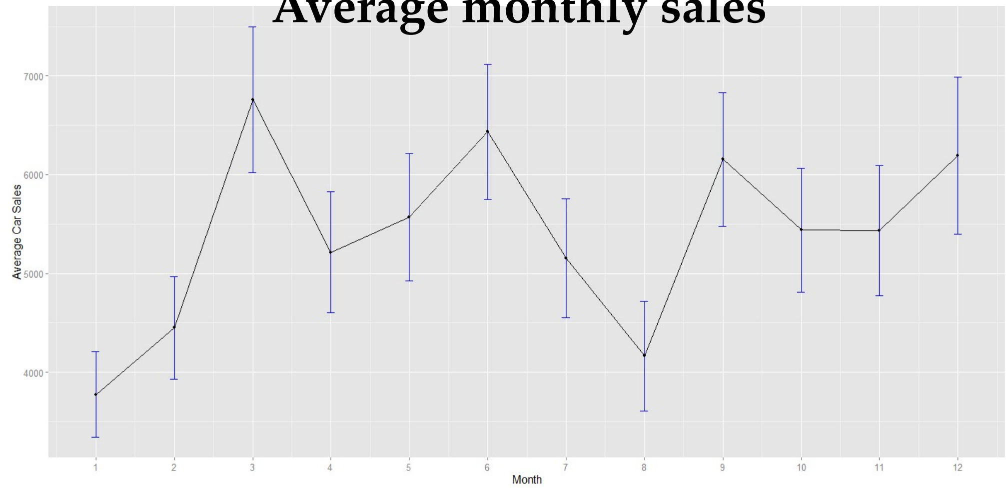
- Monthly sales 2007-2014 (15 markets)
- Web activity variables (9 treatment markets):
  - Visits
  - RFI - Requests For information
  - RFO - Requests For Offer
  - TDA – Test Drive Application
- Alexa.com: time on site, traffic rank
- Web activity data – Google Trends, Wikipedia visits

Descriptive Analysis

Average annual sales



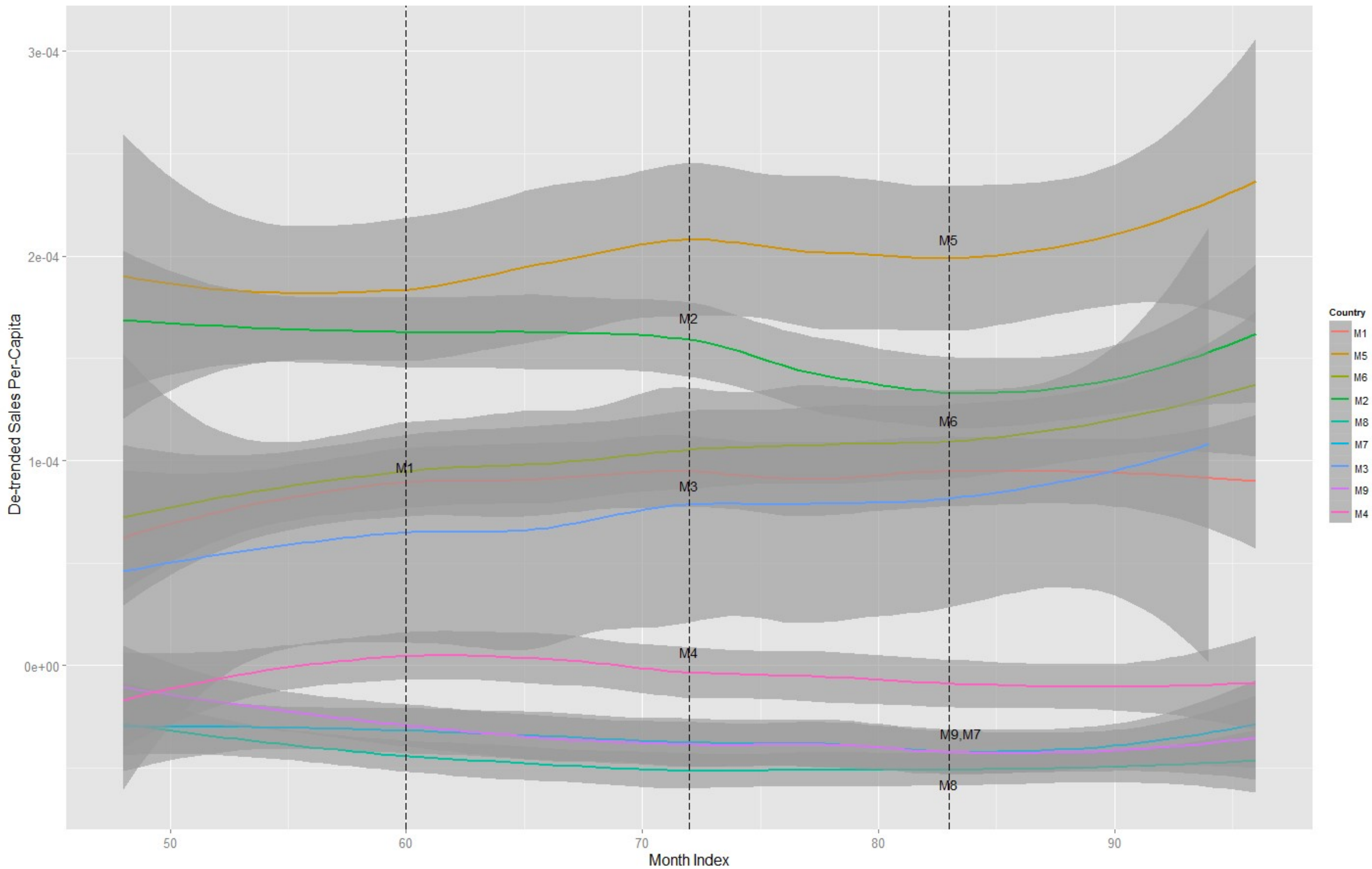
Average monthly sales



Tel Aviv University
School of Management

Descriptive Analysis

De-trended Sales Per Capita for the 9 Digital Markets



Descriptive Analysis

Online KPIs					
Statistic	N	Mean	St. Dev.	Min	Max
Visits	259	761,788.80	814,716.70	66,762	3,680,170
RFI	189	1,121.29	1,601.41	8	6,863
RFO	225	143.36	175.49	15	814
TDA	225	214.90	245.60	20	1,072

Initial Analysis – Sales Predictions

- Predicted Variable:
  - Car sales (all models) at month *t* in market *i*
- Predictors
  - Number of visits
  - Website requests KPIs: TDA, RFI, RFO, VCO
  - Google search volume (from Google Trends)
  - Wikipedia page visits
  - Lagged data (previous month)

Tel Aviv University
School of Management

Sales Predictions - Model

- **Linear regression**

$$Sales_i\,(t) =$$

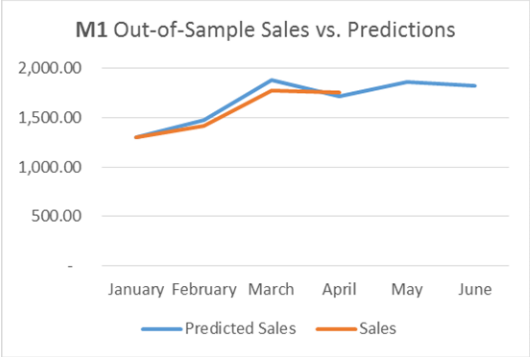
$$\alpha_0 + \varphi Sales_i(t-1) + \rho Visits(t-1) + \sum_{l=1}^{L} \gamma_l KPI_{il}(t-1)$$
$$+\,\delta GoogleTrends_i(t-1) + \beta LocalWikipedia_{\,i}\,(t-1)$$
$$+\ \tau EnglishWikipedia_{\,i}\,(t-1)$$

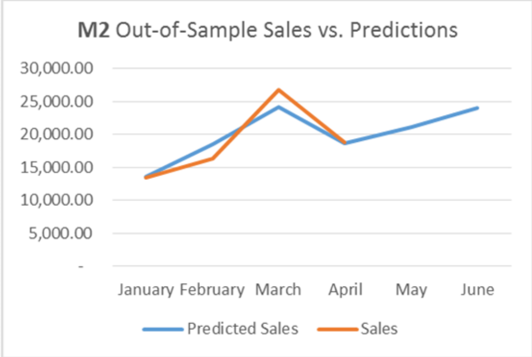
Tel Aviv University
School of Management

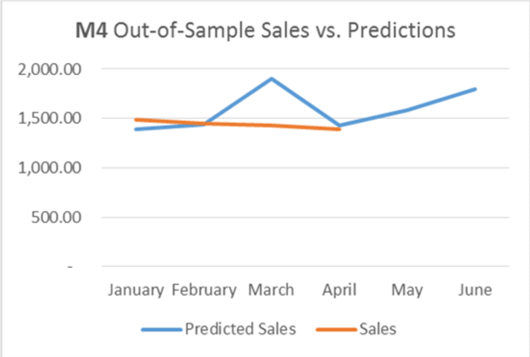
Sales Predictions - Results

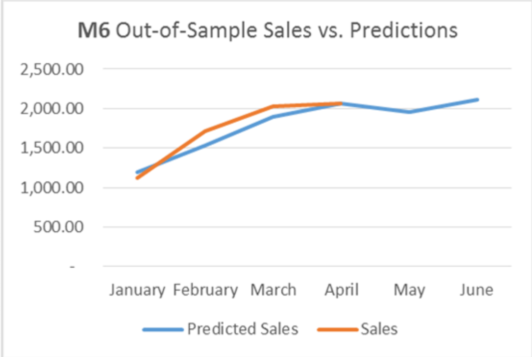
- Training: 2013-2014
- Out-of-Sample: Jan-Apr 2015

Market	MAPE
M1	7.69%
M2	4.60%
M4	13.42%
M6	6.66%







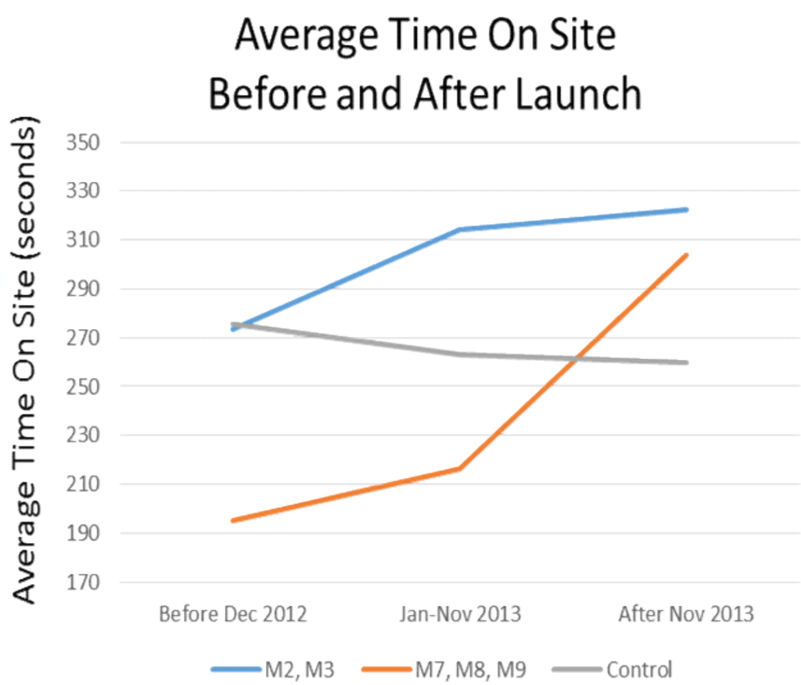


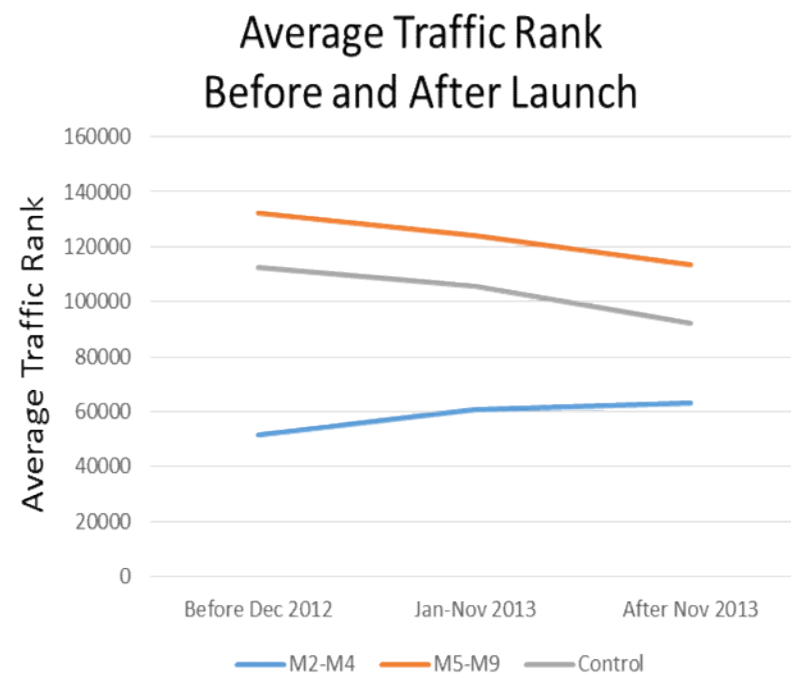
Tel Aviv University
School of Management

Natural Experiment Analysis

- What was the effect of the new website:
  - Effect on online engagement
  - Effect on online requests
  - Effect on offline sales

Engagement Indeed Increased





Engagement Indeed Increased

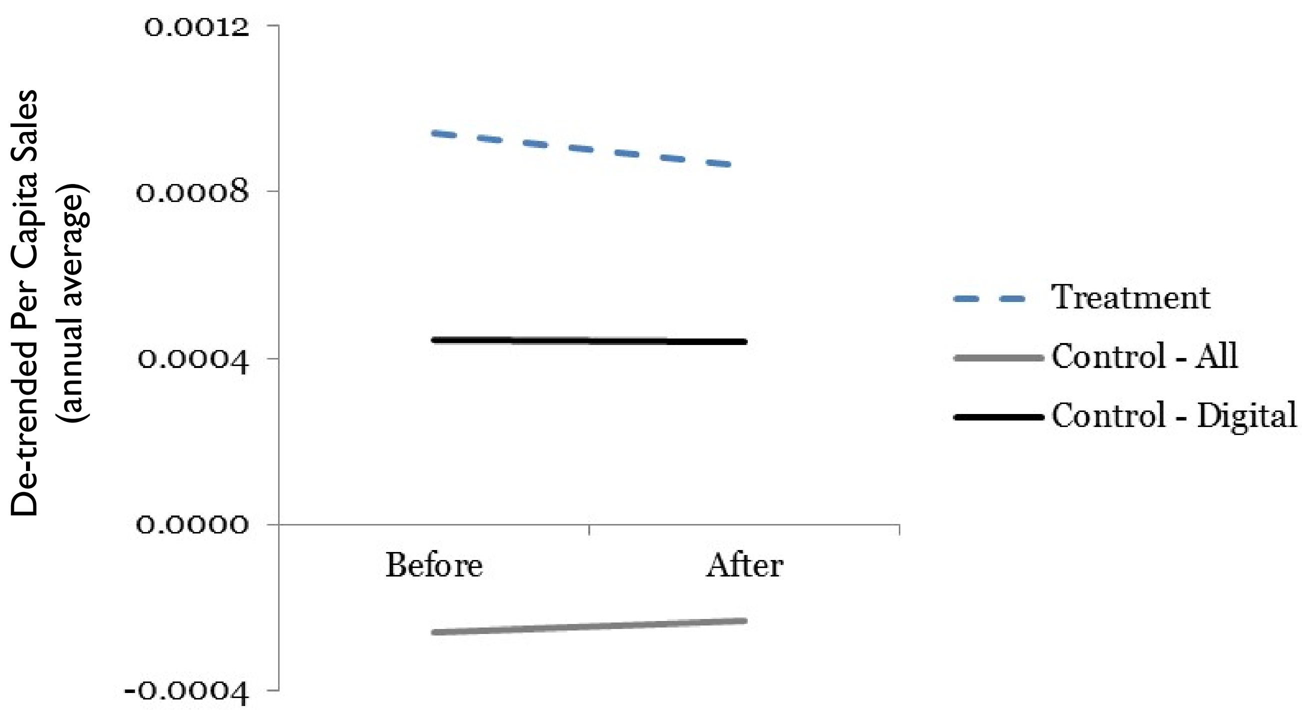
- Difference in Differences:
  - Alexa engagement and traffic data

	Dependent variable:	
	TimeOnSite	TrafficRank
Launch	**84.99***** (9.37)	1,659.84 (10,519.83)
Constant	241.98*** (14.99)	166,265.90*** (18,995.12)
Observations	221	446
R^2	0.60	0.82
Adjusted R^2	0.56	0.81
Residual Std. Error	36.76 (df = 198)	56,699.69 (df = 415)
F Statistic	13.68*** (df = 22; 198)	64.88*** (df = 30; 415)
Note:		*p<0.1; **p<0.05; ***p<0.01

* country, year, month FE not reported

But sales did not…

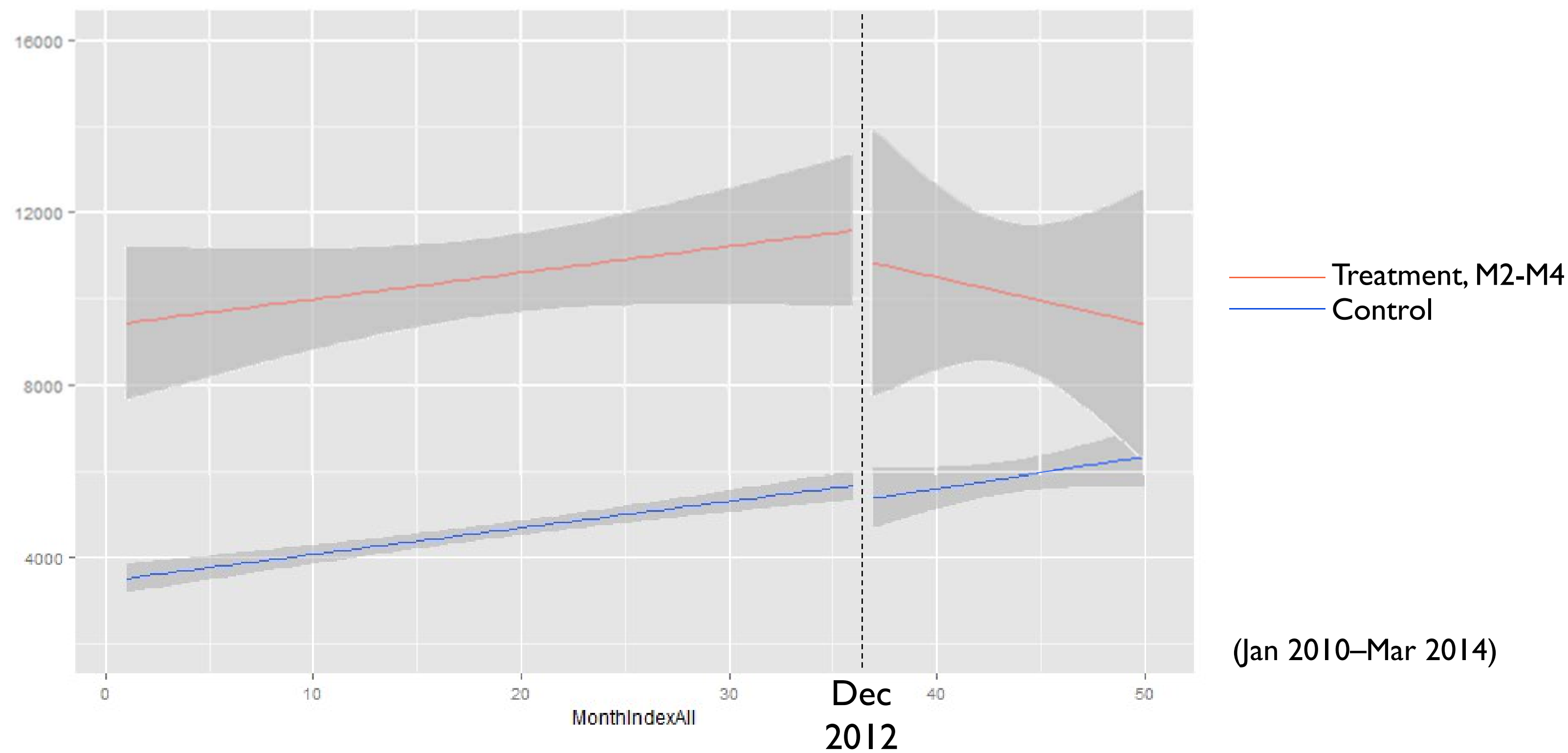
Average Sales Per Capita Before and After Launch



Tel Aviv University
School of Management

But sales did not…

Average Monthly Sales - Treatment and Control Groups



Sales Analysis

- Diff-in-Diffs model

$$Sales_{cym} = \alpha_c + \beta_y + \gamma_m + \lambda_{ym} SalesLag_{cym} + \boldsymbol{\delta} \cdot Launch_{cym} + \epsilon_{cym}$$

c - Country

y - Year

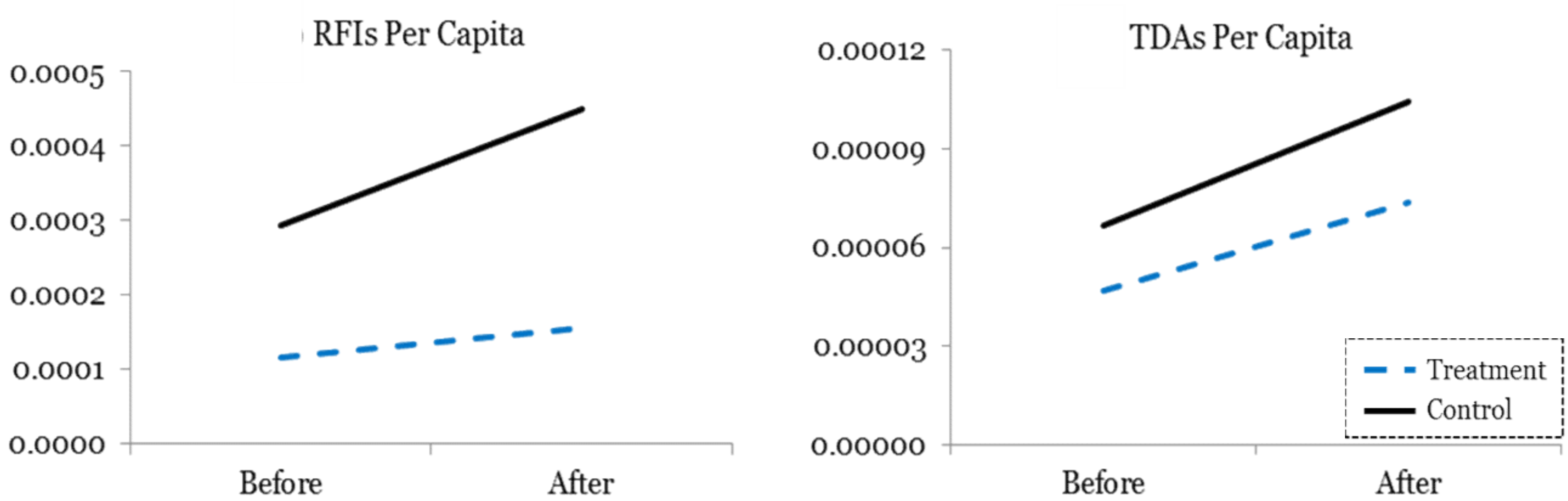
m - Month

Sales Analysis

DID Estimation of Launch Effect on Sales

	All Markets	Large 5	Small 4	no M2, M3	no M2
	(1)	(2)	(3)	(4)	(5)
Launch	**-708.99*****	**-696.14****	**-724.47*****	**-515.52*****	**-602.15****
	(246.86)	(324.05)	(259.35)	(179.31)	(251.81)
Sales(t-1)	0.13***	0.09***	0.54***	0.54***	0.13***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Constant	1.07	-509.85	-653.35**	-510.32**	351.72
	(362.42)	(448.40)	(300.94)	(251.96)	(360.30)
Observations	1,423	1,044	949	1,233	1,328
Adjusted R^2	0.91	0.91	0.94	0.94	0.88
F Statistic	409.57*** (df = 34; 1388)	351.92*** (df = 30; 1013)	519.37*** (df = 29; 919)	618.75*** (df = 32; 1200)	287.99*** (df = 33; 1294)

Online engagement↑ Sales leads ↓



- Data on online activity variables only available for treated markets.
- M5-M9 (Nov. 2013) are now "treatment" and M1-M4 "control".
- H to L engagement → smaller decrease in RFI, TDA compared to control.

Requests for Contact - Post Launch

	Visits	RFI	RFO	TDA
	(1)	(2)	(3)	(4)
Launch	-35,348.55	**-565.94******	9.13	**-65.39******
	(36,094.07)	(216.77)	(19.25)	(23.79)
Constant	270,116.90***	-30.22	61.57	-12.28
	(52,752.26)	(421.46)	(39.90)	(49.30)
Observations	259	189	225	225
Adjusted R^2	0.97	0.79	0.81	0.85
F Statistic	324.81***	32.45***	43.67***	58.63***
	(df = 23; 235)	(df = 22; 166)	(df = 22; 202)	(df = 22; 202)

Note: *p<0.1; **p<0.05; ***p<0.01

A Simple Model for Online-to-Offline

- **Customer Funnel**

A Simple Model for Online-to-Offline

- Consumer and brand match value = 0 or 1 (share τ of population match)
- Consumer uncertainty about match with brand:

$$\sigma \sim U[0, \overline{\sigma}];\ \tilde{t}_0 = \sigma,\ \tilde{t}_1 = 1 - \sigma$$

Online engagement

High Online Engagement -
- Reduces uncertainty: $\sigma_H = 0$
- Creates brand bias: $b_H \sim U\left[0, \overline{b}\right]$

Low Online Engagement -
- Uncertainty remains: $\sigma_L = \sigma$
- No bias introduced: $b_L = 0$

Offline contact at dealership

- Engage offline if updated expected match value > threshold.
- Purchase probability = expected match value.

Tel Aviv University School of Management

The Effect on Conversion Rates

- Conversion rates (at dealership) may be higher under high online engagement
- Evidence from user-level data (one treatment market):

	Before (L)	After (H)	
RFI	7.6%	8.6%	↑13.2%
TDA	13.7%	16.1%	↑17.5%

Robustness Check – Placebo Effect

	Placebo Treatment Markets		
	All markets	Dec. 2012 Launch Markets	Nov. 2013 Launch Markets
	(1)	(3)	(4)
PlaceboAllTdec09	-164.49 (251.81)		
Placebo1Dec09		115.95 (317.68)	
Placebo2Dec09			-413.26 (267.85)
SalesLag	0.04 (0.03)	0.04 (0.03)	0.04 (0.03)
Constant	309.05 (385.25)	295.02 (384.59)	229.80 (384.52)
Observations	870	870	870
Adjusted R^2	0.92	0.92	0.92
F Statistic	322.25^{***}	322.12^{***}	323.06^{***}
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Robustness Check – Other Brands

Tel Aviv University
School of Management

Discussion

- Increased online engagement:
  - Reduce consumer's uncertainty about the product/match
  - Decrease quantity of request for offline contacts
  - Increase quality of requests for offline contacts
  - Might reduce offline sales

Tel Aviv University
School of Management

Managerial Implications

- Determine the levels of online engagement:
  - Low engagement – online activity as contact channel
  - Dynamic website - based on consumers "match" level
- Operational Aspects:
  - Cost of offline stores
  - Cost of sales leads