Social Media as an Alternative to Surveys of Opinions about the Economy

Frederick G. Conrad[1]
Johann A. Gagnon-Bartsch[1]
Robyn A. Ferg[1]
Michael F. Schober[2]
Josh Pasek[1]
Elizabeth Hou[1]

[1]University of Michigan
[2]New School for Social Research

Social Media Content and Survey Data

- Enthusiasm about exploiting social media content for social research for at least a decade
- May be timelier and less expensive than traditional survey data
- Especially intriguing as relevance of surveys called into question due to low participation rates

2

Comparability

- In general, must convert social content (open text in most social media) into information that
  1. can be used to address research questions
  2. is comparable to the survey data in units, format, etc.
- Sentiment scores (positive vs. negative) common format into which social media content is transformed
  - Opinions are, so far, more widely investigated than objective phenomena (behaviors and facts)

3

Two Visions

1. Enhance survey data with social media content
   - e.g., include data derived from social media content in statistical models otherwise based on survey results
2. Replace survey data with social media content
   - eliminate certain questions and variables they produce
   - reduce frequency of survey waves in longitudinal studies, basing estimates on social media in off-months

4

Enhancing Survey Data: Example

- Including # views of US senate candidates' Wikipedia pages in survey-based models that predict election outcomes, improves models' performance
  - Smith &Gustafson (2017) modeled US 100 senate races between 2008 and 2012 with and without pageview data
  - Assumed visiting a candidate's Wikipedia page associated with increased likelihood of voting for candidate
  - Models without pageview data (just survey data) quite accurate but models that include pageview data significantly more accurate

5

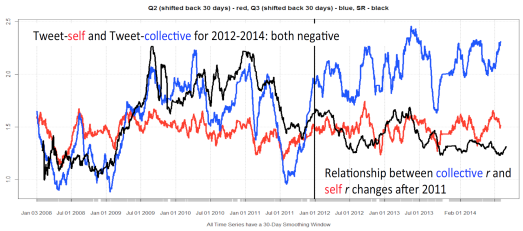
Replacing Survey Data

- If one can tell same story with social media content as survey data might be able to replace latter with former
- Daily sentiment of tweets in 2008-9 containing particular keywords shown to correlate with survey-based measures of public opinion (O' Connor, 2010)
  - Consumer confidence: sentiment of tweets containing "jobs" correlates (r=.79) with Gallup's Economic Confidence Index and (r=.64) with Michigan's Index of Consumer Confidence (ICS)
  - Presidential approval: sentiment of tweets containing "Obama" correlate (r=.75) with Gallup's Daily Tracking Poll
- Findings for consumer confidence replicated by Conrad et al. (2015) through 2011 and Daas et al. (2015) from 2010-2012

6

Replacing Survey Data (2)

- After ~2011, correlations very small and sometimes negative
  - Conrad et al. (unpublished)
  - Antonucci et al. (2014)
  - Pasek et al. (in press)



- This pattern raises serious questions about the viability of using social media content (at least tweets) in place of survey data

7

Current Study

- Why might relationship between sentiment of "jobs" tweets and ICS have weakened over time?
- Might it be restored through different analytical methods?
- If not, might original relationship have been spurious?

8

Approach

1. Reproduce key findings of O'Connor et al. (tweets from 2008-9) with our corpus
2. Test impact of different analytic decisions in original time period
3. Test whether different analytic decisions restore relationship after 2011

9

Analytic Decisions

- Investigators presented with many choices in how to process and analyze both data types of data
- Possible particular choices can affect relationship between sentiment in survey responses and tweets

10

Analytic Decisions

- Classification of tweets
- Smoothing and Lag intervals
- Sentiment scoring tool
- Calculation of daily sentiment
- Measure of association

11

Classifying Tweets

- Assigned "jobs" tweets to five broad content categories:
  - *news/politics*
  - *personal*

  Expect to be more relevant to survey measures of consumer sentiment
  - *advertisement*
  - *junk*
  - *other*

  Expect to be less relevant to survey measures of consumer sentiment

12

Smoothing and Lag

- Smoothing (moving average) over past *K* days used to reduce noise in time series
  - We varied between 1 and 100
- Lag (lead) is number of days (*L*) shifted to offset the two measures
  - Positive *L* implies that tweets lag survey responses
  - We varied between -100 (survey data precedes tweets) and +100 (tweets precede survey data)

13

Sentiment Scoring Tools

- Dictionary methods used to assign sentiment to individual words (-1, 0, +1) but not entire tweet
  - Insensitive to irony and sarcasm, negation, context
  - Developed for long texts like newspaper articles
- Machine learning methods used to assign sentiment to entire tweet (continuous between -1 and +1)
  - Developed for tweets and similar texts
- We test 3 dictionary methods (Opinion Finder, Lexicoder, and Liu Hu) and two machine learning methods (Vader and Tetblob)
  - Vader developed for use with tweets
  - Textblob developed with movie reviews
  - Both use less formal/professional language than the dictionary methods

14

Daily Sentiment

- Overall sentiment for words or tweets has been calculated by:

1. $\frac{positives}{negatives}$
2. $\frac{positives\ -\ negatives}{total}$
3. $\frac{postives}{positives\ +\ negatives}$

- Only needed for dictionary approach

15

Measures of Association

- Pearson's correlation commonly used for assessing relationships between survey responses and Twitter sentiment
  - Weaknesses include sensitivity to outliers and overall linear trends (e.g., if both series increase over time)
- Comovement: how often two time series move in the same direction from one time period to the next
  - Not sensitive to outliers or overall linear trends
  - Easy to interpret: if comovement is 0.9, then time series move in same direction 90% of the time

16

Data Sources

- Corpus of tweets containing "jobs" from 1-Jan-08 through 27-Jun-14
  - Acquired from Topsy API (spam tweets removed via proprietary process)
  - Randomly sampled 500 tweets per day to reduce computational burden
  - Calculated sentiment scores; values between -1 and +1
- Survey data
  - Daily responses to 5 survey questions from University of Michigan's Surveys of Consumers (SCA) collected mostly by telephone
  - Questions concern personal finances and national economy: all have positive, negative and neutral response options
  - Data mostly collected via telephone
  - ICS based on responses to all 5 questions (historical range from 59.5 to 112.0)

17

1. Replication of O'Connor et al.

- Compared sentiment of "jobs" tweets from 2008-2009 to consumer confidence, as measured by Michigan ICS
- Used same "settings" as in original study:
  - daily sentiment calculated as ratio of positive to negative tweets
  - Scores assigned to words with the OpinionFinder dictionary
  - Twitter sentiment smoothed by K = 30 days; shifted by L = -50 days
  - Association measured by Pearson correlation
- Differences
  - Original corpus obtained from Twitter API; current corpus from Topsy
  - Original survey data (ICS) *monthly*; current survey data (ICS) *daily*
- Result: O'Connor et al. find r = 0.64 and we find r = 0.65 -- replication successful

18

2. Impact of Analytic Decisions, Original Years

- Classifying the "jobs" tweets does not strengthen the relationship
  - Categories of "jobs" tweets that do not concern employment correlate more highly than those that do

19

Correlation by "jobs" Tweet Category and Daily Sentiment Formula

	$\frac{postive\ tweets}{negative\ tweets}$	$\frac{positive\ tweets - negative\ tweets}{total\ tweets}$	$\frac{positive\ tweets}{positive\ tweets + negative\ tweets}$
All tweets	0.65	0.00	0.48
News/politics	0.17	0.30	0.19
Personal	-0.23	-0.30	-0.26
Advertisements	0.71	-0.24	0.32
Junk	0.42	0.16	0.32
Other	0.19	0.43	0.52

20

2. Impact of Analytic Decisions, Original Years

- Classifying the "jobs" tweets does not strengthen the relationship
  - Categories of "jobs" tweets that do not concern employment correlate more highly than those that do
  - Suggests original correlation might have been spurious
- Formula for calculating daily sentiment has a large impact on correlation

21

Correlation by "jobs" Tweet Category and Daily Sentiment Formula

	$\frac{postive\ tweets}{negative\ tweets}$	$\frac{positive\ tweets - negative\ tweets}{total\ tweets}$	$\frac{positive\ tweets}{positive\ tweets + negative\ tweets}$
All tweets	0.65	0.00	0.48
News/politics	0.17	0.30	0.19
Personal	-0.23	-0.30	-0.26
Advertisements	0.71	-0.24	0.32
Junk	0.42	0.16	0.32
Other	0.19	0.43	0.52

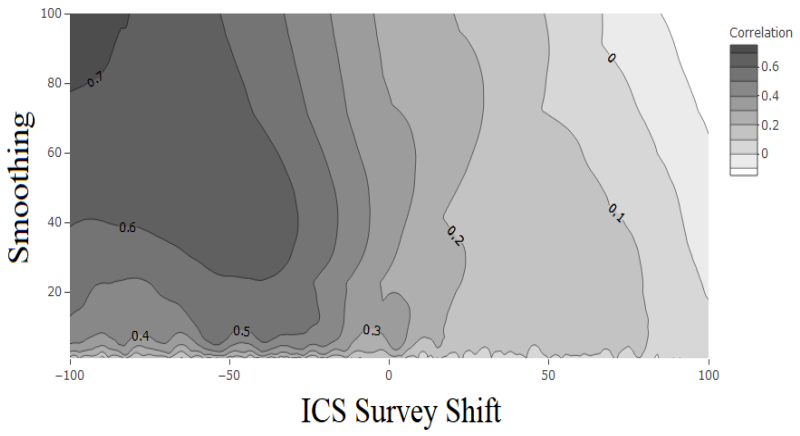
22

2. Impact of Analytic Decisions, Original Years

- Classifying the "jobs" tweets does not strengthen the relationship
  - Categories of "jobs" tweets that do not concern employment correlate more highly than those that do
  - Suggests original correlation might have been spurious
- Formula for calculating daily sentiment has a large impact on correlation
  - Suggests results not very robust
- Smoothing and Lag intervals have large effect on *r*

23

Impact of Smoothing and Lag Interval



24

2. Impact of Analytic Decisions, Original Years

- Classifying the "jobs" tweets does not strengthen the relationship
  - Categories of "jobs" tweets that do not concern employment correlate more highly than those that do
  - Suggests original correlation might have been spurious
- Formula for calculating daily sentiment has a large impact on correlation
  - Suggests results not very robust
- Smoothing and Lag intervals have large effect on *r*
  - Suggests results may depend on arbitrary choices
- Comovement: relationships are close to chance (.5) and not significant; sensitive to starting date
  - If genuine relationship between survey responses and Twitter sentiment, we would expect comovement to be large and robust to starting date

25

3. Impact of analytic decisions after 2011

- Does relationship between Twitter sentiment and survey responses actually weaken over time?
- Or did it simply start and remain volatile?
- Examining the correlations under the original "settings" for each year (2008-2014) indicates considerable volatility from year to year
  - Sentiment scoring tool developed for tweets (Vader) and movie reviews (Textblob) do not increase correlations

26

Correlations by Year

	2008	2009	2010	2011	2012	2013	2014
All tweets	0.21	0.66	-0.03	0.54	0.02	0.28	0.41
News/politics	-0.05	0.18	0.22	0.37	-0.02	0.02	-0.61
Personal	-0.10	0.36	0.08	0.23	-0.07	0.09	-0.24
Advertisements	-0.02	0.64	0.01	0.59	-0.17	0.29	0.84
Junk	0.06	0.29	-0.21	-0.21	-0.16	-0.16	0.16
Other	-0.38	0.46	-0.57	0.67	0.02	0.53	-0.25

27

Challenges of Future Work

- In retrospect, we should not be surprised about the lack of correspondence between survey responses and Twitter sentiment

1. Representation: Users who create social media are unlikely to look like the population of interest
2. Measurement: Process of posting Twitter content fundamentally different than answering survey questions

28

Representation

- The sentiment expressed in any Twitter corpus may represent the sentiment of Twitter users but cannot be assumed to represent the mood in any other population (e.g., Baker, 2017)
  - It is possible that both data sources can tell the same story suggesting not that the Twitter users represent the population but that the topics are similarly covered (Schober et al., 2016)
- If goal is to generalize sentiment in a Twitter corpus to a national population, then worth exploring whether methods to produce estimates from nonprobability survey samples can be applied to Twitter content

29

Weighting Twitter Content

1. Infer covariates from content of posts (plus metadata)
   - e.g., location, political affiliation, income, computer use
   - Results so far are mixed but promising
2. Weight survey respondents to look like Twitter user base
   - Allows one to correlate survey data with Twitter population
   - Pasek et al. (in press) first to do this; little improvement but promising
   - However, will not support population estimates
3. Recruit Twitter users who have tweeted on topics of interest to participate in survey; simultaneously conduct calibration survey
   - Weight tweets based on demographics of Twitter survey sample adjusted to match calibration sample

30

Measurement

- Differences in why and how respondents and users create data
- Topic:
  - Respondents provide information about a topic chosen by researchers
  - Social media users determine the content about which they post
- Stimulus
  - Survey researchers present exactly the same stimulus (the question) to all respondents so the data are comparable
  - Social media users, post content in response to unknown – but presumably highly varied – stimuli
- Audience
  - Survey respondents, especially when questionnaire self-administered, do not seem to respond with particular audience in mind
  - Twitter users can have very specific audiences in mind

31

Why and for Whom Tweets Are Created

- Twitter users post about what is currently in mind (Naaman et al., 2010):
  - "Me now" (41%) e.g., "tired and upset"
  - "Random Thoughts" (25%), e.g., "I miss New York but I love LA …",
  - "Opinions/Complaints" (24%), e.g., "Illmatic = greatest rap album ever"
- And often for specific imagined audiences (Marwick & Boyd, 2011)
  - "I think of a room filled with friends when I tweet. I assume people like me that are reading my tweets."
  - "i'm very conscious that twitter is public. i wouldn't tweet anything i didn't want my mother/employer/professor to see".
- Possible for some topics on some occasions differences in why and for whom content is created does not affect comparability of two data sources
- But we are not aware of any research about when this might be the case

32

Conclusion

- We have not uncovered any evidence that there is a credible relationship between sentiment expressed in answers to ICS questions and "jobs" tweets
- It may be possible under some circumstances to capture sentiment from social media data that is genuinely associated with the sentiment of the entire population.
- Our concern is that, as Groves (2011) said about non-probability surveys, "such designs work well until they don't; there is little theory undergirding their key features."
- Developing such theory, if it can be developed, should be a priority

33

Thank you!

34

Five SCA Questions used to calculate ICS

1. "We are interested in how people are getting along financially these days. Would you say that you (and your family living there) are better off or worse off financially than you were a year ago?"
2. "Now looking ahead--do you think that a year from now you (and your family living there) will be better off financially, or worse off, or just about the same as now?"
3. "Now turning to business conditions in the country as a whole--do you think that during the next twelve months we'll have good times financially, or bad times, or what?"
4. "Looking ahead, which would you say is more likely--that in the country as a whole we'll have continuous good times during the next five years or so, or that we will have periods of widespread unemployment or depression, or what?"
5. "About the big things people buy for their homes--such as furniture, a refrigerator, stove, television, and things like that. Generally speaking, do you think now is a good or bad time for people to buy major household items?"

35