When can Twitter content really be used for social statistics?

Fred Conrad (ISR)
Michael Schober (New School for Social Research)
Cliff Lampe (SI)
Josh Pasek (Comm. Studies/ISR)
Daniel Romero (SI/EECS)

Team

- Started to work together in Jan. 2013

- MCubed support in 2014

- We have explored alignment between Twitter sentiment and data from SCA* leading, so far, to ~8 conference talks

- Research synthesis in *POQ* (2016): Social media analyses for social measurement

*Surveys of Consumers, ISR/SRC monthly survey producing Index of Consumer Sentiment

Mining Social Media Content

- Promise is that this can supplement – even replace – traditional surveys, providing accurate estimates much more cheaply and quickly than conventional methods
  - Maybe even for official statistics
- Hybrid approach:
  - May enable collecting survey data every other month instead of every month, adjusting the prior survey estimates with social media content

Track Record

- Successes replicating survey data with social media data:
  - O'Connor et al. (2010) found that content of tweets correlated highly with Gallup's Economic Confidence Index based on answers to several survey questions
  - Tumasian et al. (2011) found that number of tweets mentioning a political party predicted German election outcomes about as accurately as traditional pre-election polls
  - Others: Ceron et al. (2014), Fu and Chan (2013), Jensen and Anstead (2013), and Sang and Bos (2012)

- Failures replicating survey data with social media data:
  - O'Connor et al. (2010) found that content of tweets only weakly correlated with answers to survey questions contributing to SCA's ICS
  - Jungherr, Jürgens, and Schoen (2012) removed tweets mentioning "pirate party" from corpus analyzed by Tumasian et al. (2011)
  - Unpublished studies?

Why does social media content sometimes align with survey results and sometimes not?

- Topic coverage
  - When there is alignment, range of sentiment and experiences in posts is proportional to those in the population
  - despite demographic differences between between users and population
- Particular survey questions
  - Three hypotheses
- Data extraction/wrangling tools
  - Sentiment (positive/negative) vs. semantic analyses
  - Machine learning algorithms, particular models

Three hypotheses about why alignment* may (not) occur

1. *Collective vs. self hypothesis*. Alignment is more likely when survey *Q*s to which social media content compared concern groups larger than individuals or their households

2. *Nowcasting hypothesis*. Alignment is more likely when the survey estimates are based on *Q*s about the present, as opposed to the past or future.

3. *Stigmatized topics hypothesis*. Tweets and survey responses should be less well aligned for more stigmatized topics.

*between tweets and survey responses

Questions from SCA

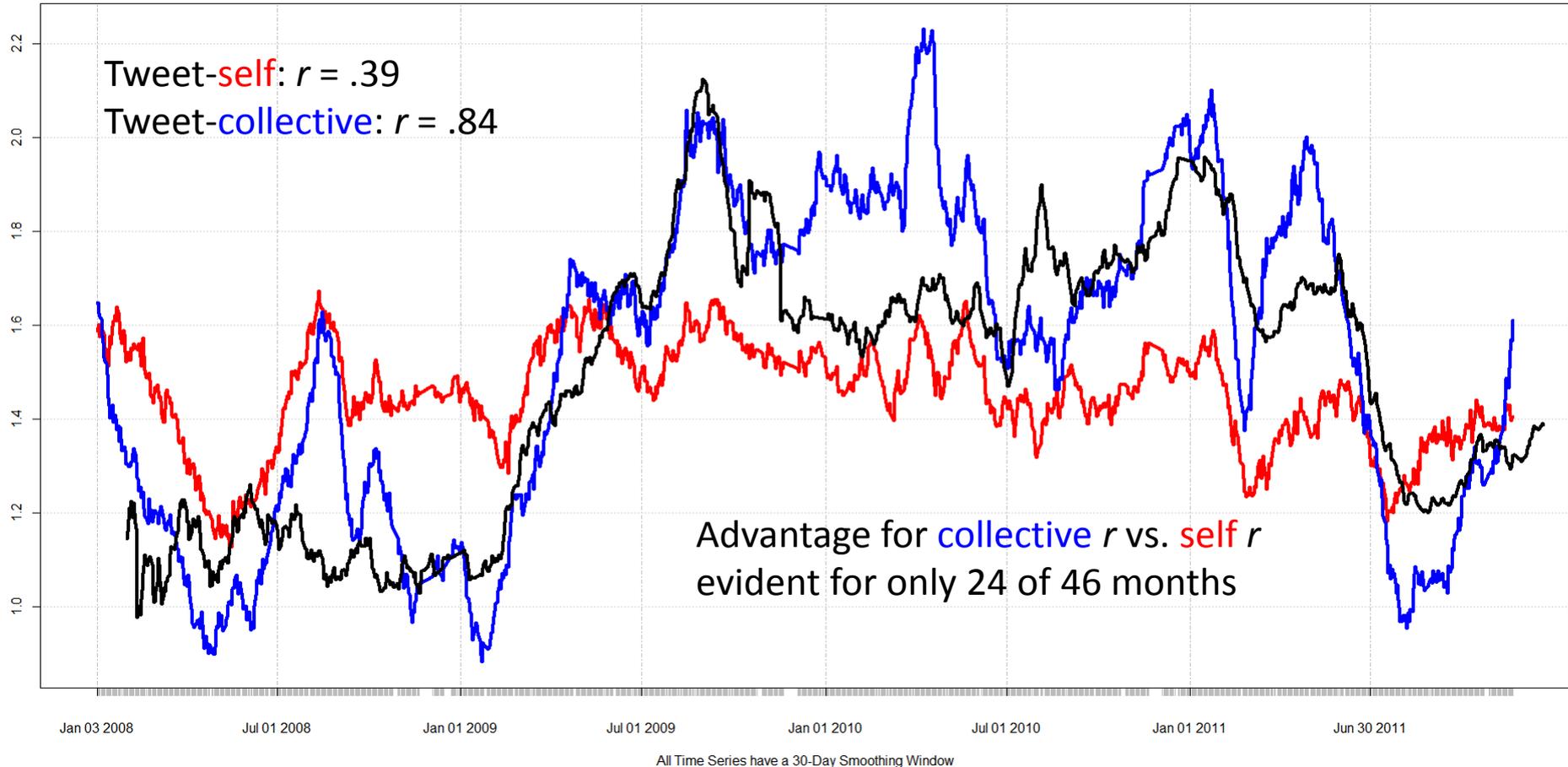## Self

- *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?*

Collective

- *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?*

Q2 (shifted back 30 days) - red, Q3 (shifted back 30 days) - blue, SR - black

Tweet-self: *r* = .39
Tweet-collective: *r* = .84

Advantage for collective *r* vs. self *r*
evident for only 24 of 46 months

All Time Series have a 30-Day Smoothing Window

Tweet sentiment ratio
Positivity of responses to self question (30 days later)
Positivity of responses to collective question (30 days later)

Proposed Studies

A. Test the three hypotheses in public survey data:
   – Consumer Sentiment, Electoral Politics, Labor Force Activities, Health Behaviors

   – Compare results using
     • Different lexical and NLP tools and different machine learning approaches
     • Different methods of sampling tweets, e.g., all vs. 1 per user

   – Will test alignment with all combinations and evaluate in multilevel Bayesian model
     • which will allow us to see which combinations lead to the best prediction under which circumstances

Proposed Studies (2)

B.   Does alignment occur at individual level?

- Recruit users who have tweeted on topics in A
- Recruit into online survey where will be asked about those topics for which alignment high and low
- If within-individual alignment is high suggests (1) survey response process similar to tweeting and (2) no coverage bias
- If low, supra-individual alignment could occur if
  - tweets reflect a general mood, possibly distilling media reports and conversations with friends
  - personal circumstances
  - But not a one-to-one correspondence between tweets and survey responses

Questions?