Things Fall Apart:
Learning about a changing economy from instability in a social media nowcaster

Matthew D. Shapiro
University of Michigan

Columbia Business School conference
Program for Financial Studies
News and Finance Initiative
March 8, 2018
Acknowledgements

University of Michigan

Dolan Antenucci, Electrical Engineering and Computer Science
Mike Cafarella, Electrical Engineering and Computer Science
Margaret C. Levenstein, Institute for Social Research, and Information School and Ross School

Stanford University

Christopher Ré, Computer Science

Michigan NSF-Census Research Network (NSF SES 1131500)
Michigan Institute for Data Science (MIDAS)
References

“Using Social Media to Measure Labor Market Flows” (NBER WP 20010)

“Things Fall Apart” (this presentation, WIP)
Nowcasting job flows

Development of Twitter-based nowcaster for initial claims for unemployment insurance

• Factor model based on job loss Tweets
• Initial excellent performance
Job loss indicator

Benchmark: weekly Unemployment Insurance (UI) claims

• Most frequent measure of labor market activity
Why use Tweets?

Valuable information freely provided by the Tweeter in real time

- Quick and cheap relative to surveys
- Better at capturing turning points?
- Permit retrospective analysis because beliefs and actions are “archived”
Michigan Indicator in Brief

- Hand-selected Tweet n-grams related to job loss: expert knowledge versus machine learning
- Extract job loss index: First principal component of job loss Tweets
- Rolling fit of index to UI claim ground truth
## Signals (Tweet n-grams)

<table>
<thead>
<tr>
<th>Category</th>
<th>Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lost job signals</td>
<td>axed, canned, downsized, outsourced, pink slip, lost job, fired job, been fired, laid off</td>
</tr>
<tr>
<td>Unemployment signal</td>
<td>unemployment, unemployed</td>
</tr>
</tbody>
</table>
## Factor Loadings on Job Loss Signals

<table>
<thead>
<tr>
<th>Variable</th>
<th>Factor 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Axed</td>
<td>0.65</td>
</tr>
<tr>
<td>Canned</td>
<td>0.68</td>
</tr>
<tr>
<td>Downsized</td>
<td>0.60</td>
</tr>
<tr>
<td>Outsourced</td>
<td>0.52</td>
</tr>
<tr>
<td>Pink slip</td>
<td>-0.11</td>
</tr>
<tr>
<td>Lost job</td>
<td>0.78</td>
</tr>
<tr>
<td>Fired job</td>
<td>0.77</td>
</tr>
<tr>
<td>Been fired</td>
<td>0.51</td>
</tr>
<tr>
<td>Laid off</td>
<td>0.83</td>
</tr>
<tr>
<td>Standard deviation of factor</td>
<td>0.76</td>
</tr>
<tr>
<td>Cumulative fraction of variance</td>
<td>4.27</td>
</tr>
</tbody>
</table>
Michigan Indicator: Initial Success

- Tracked initial claim ground truth
- Tracked shock, e.g., hurricane sandy, govt shutdown
- Incremental predictive power
Factor 1 and Initial UI Claims

Initial Claims (left scale)
Social Media (right scale)
Factor 1 and Initial UI Claims

Initial Claims (left scale)
Social Media (right scale)

Hurricane Sandy
Factor 1 and Initial UI Claims

Initial Claims (left scale)
Social Media (right scale)

California reporting problem
Factor 1 and Initial UI Claims

Government shutdown
## Predicting Initial Claims

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant</strong></td>
<td>90.24</td>
<td>30.16</td>
<td>98.46</td>
<td>31.57</td>
<td>43.00</td>
<td>19.87</td>
<td>23.60</td>
</tr>
<tr>
<td></td>
<td>(21.98)</td>
<td>(23.09)</td>
<td>(22.93)</td>
<td>(23.56)</td>
<td>(21.51)</td>
<td>(22.48)</td>
<td>(22.72)</td>
</tr>
<tr>
<td><strong>Lagged initial claims</strong></td>
<td>0.75</td>
<td>0.05</td>
<td>0.48</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.16)</td>
<td>(0.07)</td>
<td>(0.15)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Consensus Forecast</strong></td>
<td></td>
<td>0.92</td>
<td>0.86</td>
<td>0.67</td>
<td>0.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.06)</td>
<td>(0.18)</td>
<td>(0.10)</td>
<td>(0.21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Social Media: Factor 1 (Scaled)</strong></td>
<td></td>
<td>0.73</td>
<td>0.40</td>
<td>0.27</td>
<td>0.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.06)</td>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(0.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Adjusted R²</strong></td>
<td>0.57</td>
<td>0.64</td>
<td>0.53</td>
<td>0.63</td>
<td>0.65</td>
<td>0.66</td>
<td>0.66</td>
</tr>
</tbody>
</table>

Sample period: July 16, 2011 – November 2, 2013
Dependent variable: preliminary ICSA
Predicting Initial Claims

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>90.24</td>
<td>30.16</td>
<td>98.46</td>
<td>31.57</td>
<td>43.00</td>
<td>19.87</td>
<td>23.60</td>
</tr>
<tr>
<td></td>
<td>(21.98)</td>
<td>(23.09)</td>
<td>(22.93)</td>
<td>(23.56)</td>
<td>(21.51)</td>
<td>(22.48)</td>
<td>(22.72)</td>
</tr>
<tr>
<td>Lagged initial claims</td>
<td>0.75</td>
<td>0.05</td>
<td>0.48</td>
<td></td>
<td>0.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.16)</td>
<td>(0.07)</td>
<td>(0.15)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus Forecast</td>
<td>0.92</td>
<td>0.86</td>
<td>0.67</td>
<td>0.48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.18)</td>
<td>(0.10)</td>
<td>(0.21)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Media: Factor 1 (Scaled)</td>
<td></td>
<td>0.73</td>
<td>0.40</td>
<td>0.27</td>
<td>0.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.06)</td>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(0.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.57</td>
<td>0.64</td>
<td>0.53</td>
<td>0.63</td>
<td>0.65</td>
<td>0.66</td>
<td>0.66</td>
</tr>
</tbody>
</table>

Sample period: July 16, 2011 – November 2, 2013
Dependent variable: preliminary ICSA
## Predicting Initial Claims

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant</strong></td>
<td>90.24</td>
<td>30.16</td>
<td>98.46</td>
<td>31.57</td>
<td>43.00</td>
<td>19.87</td>
<td>23.60</td>
</tr>
<tr>
<td></td>
<td>(21.98)</td>
<td>(23.09)</td>
<td>(22.93)</td>
<td>(23.56)</td>
<td>(21.51)</td>
<td>(22.48)</td>
<td>(22.72)</td>
</tr>
<tr>
<td><strong>Lagged initial</strong></td>
<td>0.75</td>
<td></td>
<td>0.05</td>
<td>0.48</td>
<td></td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td><strong>claims</strong></td>
<td>(0.06)</td>
<td></td>
<td>(0.16)</td>
<td>(0.07)</td>
<td></td>
<td>(0.15)</td>
<td></td>
</tr>
<tr>
<td><strong>Consensus</strong></td>
<td></td>
<td>0.92</td>
<td>0.86</td>
<td>0.67</td>
<td>0.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Forecast</strong></td>
<td></td>
<td>(0.06)</td>
<td>(0.18)</td>
<td>(0.10)</td>
<td>(0.21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Social Media:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Factor 1 (Scaled)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.73</td>
<td></td>
<td>0.40</td>
<td>0.27</td>
<td>0.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td></td>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(0.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Adjusted R²</strong></td>
<td>0.57</td>
<td>0.64</td>
<td>0.53</td>
<td>0.63</td>
<td>0.65</td>
<td>0.66</td>
<td>0.66</td>
</tr>
</tbody>
</table>

Sample period: July 16, 2011 – November 2, 2013
Dependent variable: preliminary ICSA
## Predicting Initial Claims

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>90.24</td>
<td>30.16</td>
<td>98.46</td>
<td>31.57</td>
<td>43.00</td>
<td>19.87</td>
<td>23.60</td>
</tr>
<tr>
<td></td>
<td>(21.98)</td>
<td>(23.09)</td>
<td>(22.93)</td>
<td>(23.56)</td>
<td>(21.51)</td>
<td>(22.48)</td>
<td>(22.72)</td>
</tr>
<tr>
<td>Lagged initial</td>
<td>0.75</td>
<td></td>
<td>0.05</td>
<td>0.48</td>
<td></td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>claims</td>
<td>(0.06)</td>
<td></td>
<td>(0.16)</td>
<td>(0.07)</td>
<td></td>
<td>(0.15)</td>
<td></td>
</tr>
<tr>
<td>Consensus Forecast</td>
<td>0.92</td>
<td>0.86</td>
<td>0.67</td>
<td>0.48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.18)</td>
<td>(0.10)</td>
<td>(0.21)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Media: Factor</td>
<td>0.73</td>
<td>0.40</td>
<td>0.27</td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (Scaled)</td>
<td>(0.06)</td>
<td>(0.07)</td>
<td>(0.08)</td>
<td>(0.09)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.57</td>
<td>0.64</td>
<td>0.53</td>
<td>0.63</td>
<td>0.65</td>
<td>0.66</td>
<td>0.66</td>
</tr>
</tbody>
</table>

Sample period: July 16, 2011 – November 2, 2013
Dependent variable: preliminary ICSA
Things fall apart

After about a year of successful tracking....

.... Initial claims ground truth starts tracking down

.... Job loss tweets do not track declining new claims
Things Fall Apart:
Groundtruth trends down beginning in early 2014
Rule Out Signal Processing Issues

Hypothesis 1: Flawed normalization because of changing Twitter use
Rule Out Signal Processing Issues

Hypothesis 1: Flawed normalization because of changing Twitter use
Hypothesis 2: “Grey swans”

- Do “infrequent but not rare” non-economic events change signal?
  - Movie “friends with benefits” is released
  - Real-life warfare influences “fired”
  - “pink slip” advertising campaign

- These one-off events are orders of magnitude larger than economic-related Tweets

- Solution: Windowed Median Filtering
Resolution: Shifting Economic Relationship

Developed Twitter model in early part of economic recovery:

• “Lost job” was bad news
• Job loss leads to UI claim

Things fall apart

• “Lost job” less bad news in healing economy
• Job losses do not lead to UI claim
Resolution: Shifting Economic Relationship

In retrospect, job loss and UI claims coming apart not not surprising

• UI claiming rate known to decline in recoveries
• NBER WP discussed shifting Beveridge curve, but had limited data span
Resolution: Shifting Economic Relationship

Use Tweets to document shifting relationship between job loss and UI claiming

New 5-factors model:

job loss [original factor]
quit
job posting
hiring
search
Loadings of Twitter Factors on UI Claims

New Model: Job Loss

A. 1-factor Index
Job loss (involuntary separations)
Coefficients
2011 2012 2013 2014 2015
-10.0
-7.5
-5.0
-2.5
0.0
2.5
5.0
7.5
10.0

B. 5-factor Index
Job loss (involuntary separations)
Coefficients
2011 2012 2013 2014 2015
-10.0
-7.5
-5.0
-2.5
0.0
2.5
5.0
7.5
10.0

C. 5-factor Index
Quit (voluntary separations)
Coefficients
2011 2012 2013 2014 2015
-10.0
-7.5
-5.0
-2.5
0.0
2.5
5.0
7.5
10.0

D. 5-factor Index
Posting
Coefficients
2011 2012 2013 2014 2015
-10.0
-7.5
-5.0
-2.5
0.0
2.5
5.0
7.5
10.0

E. 5-factor Index
Hiring
Coefficients
2011 2012 2013 2014 2015
-10.0
-7.5
-5.0
-2.5
0.0
2.5
5.0
7.5
10.0

F. 5-factor Index
Search
Coefficients
2011 2012 2013 2014 2015
-10.0
-7.5
-5.0
-2.5
0.0
2.5
5.0
7.5
10.0
New 5-factor model