Michigan SAS Conference
June 7, 2017

Lag Models with Social Response Outcomes

David J. Corliss, PhD
Director, Peace-Work
OUTLINE

Lag Models

Social Media Data

Linking Solar Storms & Tweets

Hate Tweets and Violence

Google Trends

Summary
# Lag Models – An Example

<table>
<thead>
<tr>
<th>Lead / Lag</th>
<th>Correlation</th>
<th>(None)</th>
<th>-3 Weeks</th>
<th>-2 Weeks</th>
<th>-1 Week</th>
<th>Even</th>
<th>+1 Week</th>
<th>+2 Weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.399</td>
<td>0.718</td>
<td>0.747</td>
<td>0.526</td>
<td>0.399</td>
<td>0.193</td>
<td>-0.355</td>
</tr>
<tr>
<td>Week 1</td>
<td>Y₁, Y₂</td>
<td>0.22%</td>
<td>0.22%</td>
<td>0.22%</td>
<td>0.22%</td>
<td>0.22%</td>
<td>0.22%</td>
<td>0.22%</td>
</tr>
<tr>
<td>Week 2</td>
<td>Y₁, Y₂</td>
<td>0.526</td>
<td>1.12</td>
<td>0.89</td>
<td>1.12</td>
<td>1.00</td>
<td>0.89</td>
<td>1.00</td>
</tr>
<tr>
<td>Week 3</td>
<td>Y₁, Y₂</td>
<td>0.56%</td>
<td>0.56%</td>
<td>0.56%</td>
<td>0.56%</td>
<td>0.56%</td>
<td>0.56%</td>
<td>0.56%</td>
</tr>
<tr>
<td>Week 4</td>
<td>Y₁, Y₂</td>
<td>-0.50%</td>
<td>-0.50%</td>
<td>-0.50%</td>
<td>-0.50%</td>
<td>-0.50%</td>
<td>-0.50%</td>
<td>-0.50%</td>
</tr>
<tr>
<td>Week 5</td>
<td>Y₁, Y₂</td>
<td>-0.18%</td>
<td>-0.18%</td>
<td>-0.18%</td>
<td>-0.18%</td>
<td>-0.18%</td>
<td>-0.18%</td>
<td>-0.18%</td>
</tr>
<tr>
<td>Week 6</td>
<td>Y₁, Y₂</td>
<td>0.15%</td>
<td>0.15%</td>
<td>0.15%</td>
<td>0.15%</td>
<td>0.15%</td>
<td>0.15%</td>
<td>0.15%</td>
</tr>
<tr>
<td>Week 7</td>
<td>Y₁, Y₂</td>
<td>-1.01%</td>
<td>-1.01%</td>
<td>-1.01%</td>
<td>-1.01%</td>
<td>-1.01%</td>
<td>-1.01%</td>
<td>-1.01%</td>
</tr>
<tr>
<td>Week 8</td>
<td>Y₁, Y₂</td>
<td>-1.74%</td>
<td>-1.74%</td>
<td>-1.74%</td>
<td>-1.74%</td>
<td>-1.74%</td>
<td>-1.74%</td>
<td>-1.74%</td>
</tr>
<tr>
<td>Week 9</td>
<td>Y₁, Y₂</td>
<td>-1.24%</td>
<td>-1.24%</td>
<td>-1.24%</td>
<td>-1.24%</td>
<td>-1.24%</td>
<td>-1.24%</td>
<td>-1.24%</td>
</tr>
<tr>
<td>Week 10</td>
<td>Y₁, Y₂</td>
<td>-0.18%</td>
<td>-0.18%</td>
<td>-0.18%</td>
<td>-0.18%</td>
<td>-0.18%</td>
<td>-0.18%</td>
<td>-0.18%</td>
</tr>
<tr>
<td>Week 11</td>
<td>Y₁, Y₂</td>
<td>0.40%</td>
<td>0.40%</td>
<td>0.40%</td>
<td>0.40%</td>
<td>0.40%</td>
<td>0.40%</td>
<td>0.40%</td>
</tr>
<tr>
<td>Week 12</td>
<td>Y₁, Y₂</td>
<td>0.10%</td>
<td>0.10%</td>
<td>0.10%</td>
<td>0.10%</td>
<td>0.10%</td>
<td>0.10%</td>
<td>0.10%</td>
</tr>
<tr>
<td>Week 13</td>
<td>Y₁, Y₂</td>
<td>0.05%</td>
<td>0.05%</td>
<td>0.05%</td>
<td>0.05%</td>
<td>0.05%</td>
<td>0.05%</td>
<td>0.05%</td>
</tr>
<tr>
<td>Week 14</td>
<td>Y₁, Y₂</td>
<td>-0.20%</td>
<td>-0.20%</td>
<td>-0.20%</td>
<td>-0.20%</td>
<td>-0.20%</td>
<td>-0.20%</td>
<td>-0.20%</td>
</tr>
</tbody>
</table>
Creating Lagged Variables

%macro lag(var1, var2, lag_n);
  data lag_data;
    set source_data;

    x1_lag0 = &var1;
    x2_lag0 = &var2;

    %do i = 1 %to &lag_n;
      %let j = %eval(&i - 1);
      x1_lag&i = lag(x1_lag&j);
      x2_lag&i = lag(x2_lag&j);
    %end;
  run;
%mend;

%lag(gdp, unemployment, 3);
A Model With Lagged Variables

```plaintext
proc reg data=ht_data;
  model count = count_tminus1 forecl_change_rate PCI rtw_state Poverty_pct_tminus2 ;
run;
```
Social Media Data

Different Social Media channels have unique characteristics

Twitter: Short duration and highly reactive

Facebook: medium term, extended conversation threads

Google Trends: May be more predictive. The Social Media equivalent of Durable Goods.

YouTube – the #2 Social Media channel - and #3 Instagram: –no effective means to mine images at this time
Solar Eruptions and Twitter Data

Data Source: Geomagnetic World Data Center / Kyoto

Solar Storm
DST < (μ - 2σ) = -38.2

Calm

Data Source: Geomagnetic World Data Center / Kyoto
#techfail: Tweets & Astrophysics

**Solar Storms**
(DST < 2σ below long-term baseline)

#techfail Count
(Tweets)
#techfail: Tweets & Astrophysics

Time Window: 2 to 4 days following a storm

16.0% of all days follow a solar storm

Top #techfail days: 57.9% follow a solar storm

Burst of #techfail tweets 3.6 times more likely

Statistically significant at > 99% confidence
Mining Twitter Feeds: A Process

1. Register with Twitter as a developer to get access

2. API to deploy search terms and receive tweets

3. Parse plain text tweet stream into a SAS dataset

4. Exploratory Data Analysis to find best search terms

5. High-volume search using results from #4
1. Register With Twitter

https://dev.twitter.com/oauth/overview
2. Twitter API: Obtain Tokens

1. consumerKey: Twitter User ID

2. consumerSecret: Twitter login password

3. Bearer Token: Used in SAS code to access tweets
2. Twitter API: SAS Code

/* Code from Isabel Litton and Rebecca Ottesen, California Polytechnic State University. Paper: "%GrabTweet: A SAS® Macro to Read JSON Formatted Tweets", WUSS 2013. Modified by D. Corliss 2017 to parameterize TYPE */

%MACRO grab_tweet_100(search_term, type, target_amount);
   %LET type = %NRSTR(&result_type=&type);
   %LET num_tweet = %NRSTR(&count=100);
   %IF &target_amount < 100 %THEN %LET
      num_tweet = %NRSTR(&count=)&target_amount;
   %ELSE %LET num_tweet = %NRSTR(&count=100);

   /* Location of Token File */
   filename auth "C:\dcorliss\SAS\Twitter\token.txt";

   /* Location of Output File */
   filename twtOut "C:\dcorliss\SAS\Twitter\Tweets.txt";

   /* Set search parameters: COUNT = and TYPE = Popular, Recent, or Mixed */

   /* Location of Token File */
   filename auth "C:\dcorliss\SAS\Twitter\token.txt";

   /* Location of Output File */
   filename twtOut "C:\dcorliss\SAS\Twitter\Tweets.txt";

   /* Set search parameters: COUNT = and TYPE = Popular, Recent, or Mixed */

   %LET num_tweet = %NRSTR(&count=100);
   %IF &target_amount < 100 %THEN %LET
      num_tweet = %NRSTR(&count=)&target_amount;
   %ELSE %LET num_tweet = %NRSTR(&count=100);

   /* Location of Token File */
   filename auth "C:\dcorliss\SAS\Twitter\token.txt";

   /* Location of Output File */
   filename twtOut "C:\dcorliss\SAS\Twitter\Tweets.txt";

   /* Set search parameters: COUNT = and TYPE = Popular, Recent, or Mixed */

   %LET type = %NRSTR(&result_type=&type);
3. Parse Plain Text Tweet Stream

Raw Twitter Text Output:
3. Parse Plain Text Tweet Stream

/* Code from Isabel Litton and Rebecca Ottesen, California Polytechnic State University. Paper: "%GrabTweet: A SAS® Macro to Read JSON Formatted Tweets", WUSS 2013 */

INPUT

"created_at":' created_at1
"id":' tweet_id
"text":' text
"user":{"id":' user_id
"name":' name
"screen_name":' screen_name
"location":' location
"created_at":' created_at2
"lang":' lang
"contributors":null,' retweeted @@
;
4. Exploring Search Terms

**** Mining Twitter for Hate Speech ****

%grab_tweet_100(Sharia America hate, recent, 100);

<table>
<thead>
<tr>
<th>created_at</th>
<th>tweet_id</th>
<th>text</th>
<th>user_id</th>
<th>name</th>
<th>screen_name</th>
<th>location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thu Jun 08 08:52:55 +0000 2017</td>
<td>872738119677455944</td>
<td>@pleadour_You hate America &amp; support Sharia here. Your time will come,</td>
<td>1223623118</td>
<td>Michael Erwin</td>
<td>michaelerwin</td>
<td>California, USA</td>
</tr>
<tr>
<td>Thu Jun 08 08:52:00 +0000 2017</td>
<td>872662468714624001</td>
<td>RT @LMARocks @pleadour Not a hate crime, and guess what... America is on to you, and your stand is now promotion. Your strategy backfired</td>
<td>4223111532</td>
<td>PorkChop Express</td>
<td>China_is_here</td>
<td>Wing Kong Exchange</td>
</tr>
<tr>
<td>Mon Feb 23 03:30:48 +0000 2009</td>
<td>8727630305043443856</td>
<td>RT @WayneRoot: Agreed our schools are run by creeme &amp; traitors who hate America &amp; its Christian tradition. What exactly do you call that?</td>
<td>1216557579</td>
<td>Karen D. Scopics</td>
<td>KDScopics</td>
<td></td>
</tr>
<tr>
<td>Wed May 06 21:43:22 +0000 2009</td>
<td>872645955080608780</td>
<td>RT @lockman#mehouse: Many New Manners hate America</td>
<td>3290719405</td>
<td>Maine4Trump</td>
<td>Maine4Trump</td>
<td>#ME02 Moscow, Maine</td>
</tr>
<tr>
<td>Wed Feb 19 03:09:53 +0000 2014</td>
<td>872659755656407929</td>
<td>RT @lockman#mehouse: Many New Manners hate America</td>
<td>81640981851660</td>
<td>Maine First Media</td>
<td>MaineFirstMedia</td>
<td>Waterville, ME</td>
</tr>
<tr>
<td>Thu Jun 08 03:42:52 +0000 2017</td>
<td>872645070016428884</td>
<td>RT @WayneRoot: Agree our schools are run by creeme &amp; traitors who hate America &amp; its Christian tradition. What exactly do you call that?</td>
<td>77952044287848</td>
<td>AndersGop</td>
<td>AndersGop</td>
<td>Maine, USA</td>
</tr>
<tr>
<td>Wed Feb 19 03:09:53 +0000 2014</td>
<td>872659755656407929</td>
<td>RT @WayneRoot: Agree our schools are run by creeme &amp; traitors who hate America &amp; its Christian tradition. What exactly do you call that?</td>
<td>81640981851660</td>
<td>Maine First Media</td>
<td>MaineFirstMedia</td>
<td>Waterville, ME</td>
</tr>
<tr>
<td>Thu Jun 08 02:19:06 -0500 2017</td>
<td>872639291027514763</td>
<td>RT @WayneRoot: Agree our schools are run by creeme &amp; traitors who hate America &amp; its Christian tradition. What exactly do you call that?</td>
<td>2897854619</td>
<td>Sammie Snickers</td>
<td>Sammie_Snickers</td>
<td></td>
</tr>
<tr>
<td>Wed May 06 21:43:22 +0000 2009</td>
<td>8726103223983507056</td>
<td>RT @lockman#mehouse: Many New Manners hate America</td>
<td>81640981851660</td>
<td>Maine First Media</td>
<td>MaineFirstMedia</td>
<td>Waterville, ME</td>
</tr>
<tr>
<td>Wed Feb 19 03:09:53 +0000 2014</td>
<td>872659755656407929</td>
<td>RT @WayneRoot: Agree our schools are run by creeme &amp; traitors who hate America &amp; its Christian tradition. What exactly do you call that?</td>
<td>81640981851660</td>
<td>Maine First Media</td>
<td>MaineFirstMedia</td>
<td>Waterville, ME</td>
</tr>
<tr>
<td>Thu Jun 08 02:24:21 -0500 2017</td>
<td>872610213991501828</td>
<td>Many New Manners hate America</td>
<td>2351033034</td>
<td>Larry Lockman</td>
<td>lockman4mehouse</td>
<td></td>
</tr>
<tr>
<td>Wed Jun 07 23:38:02 +0000 2017</td>
<td>8726598556407929</td>
<td>RT @lockman#mehouse: Many New Manners hate America</td>
<td>81640981851660</td>
<td>Maine First Media</td>
<td>MaineFirstMedia</td>
<td>Waterville, ME</td>
</tr>
</tbody>
</table>
5. **High-Volume Search**

1. Litton-Ottesen macro to access Twitter-imposed limit of 100 tweets at a time

2. Macro wrapper to repeat until specified count, going backwards from end of previous iteration

3. Data transformation required for time series dataset
Analytic Results
Analytic Results

- **Search Terms: Kill Muslims**
- **Search Terms: Kill Jews**
- **Search Terms: Kill Gays**
- **Search Terms: Kill Immigrants**
Analytic Results
Analytic Results

Search Terms: Kill Muslims

Frequency Count vs. date (19 May 2017 to 6 June 2017)
Olathe, Kansas
Attack
2/22/2017
Google Trends: A Leading Indicator?

Google Trends
Search Term: Sharia
Summary

Social media data can be matched to event data for Time Series Analysis, including lead / lag modeling.

Different social channels have different characteristics, including lead or lag times.

Analytics methods optimized for big data methods may be necessary.

Twitter data can be mined and analyzed at high volume.

Analysis of Hate Crimes with social media data shows promise but better crime data is needed.
Questions

David J. Corliss
davidjcorliss@peace-work.org
www.peace-work.org