JMP Academic Case Study 061

Quantifying Sentiment in Economic Reports

Sentiment Analysis

Produced by

Ross Metusalem, JMP Statistical Discovery
ross.metusalem@jmp.com

Muralidhara A, JMP Statistical Discovery
muralidhara.a@jmp.com

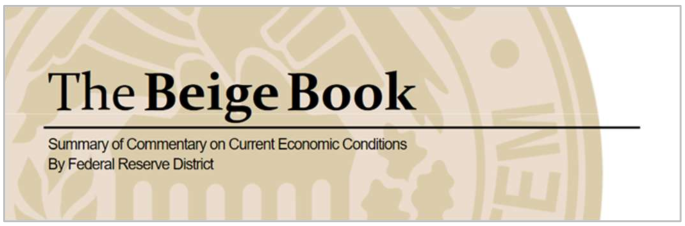
Quantifying Sentiment in U.S. Federal Reserve Economic Reports
Sentiment analysis

Key Ideas

Sentiment analysis, sentiment dictionary, sentiment terms, negators, intensifiers, document scores.
(This case study requires JMP® Pro 16 or later version.)

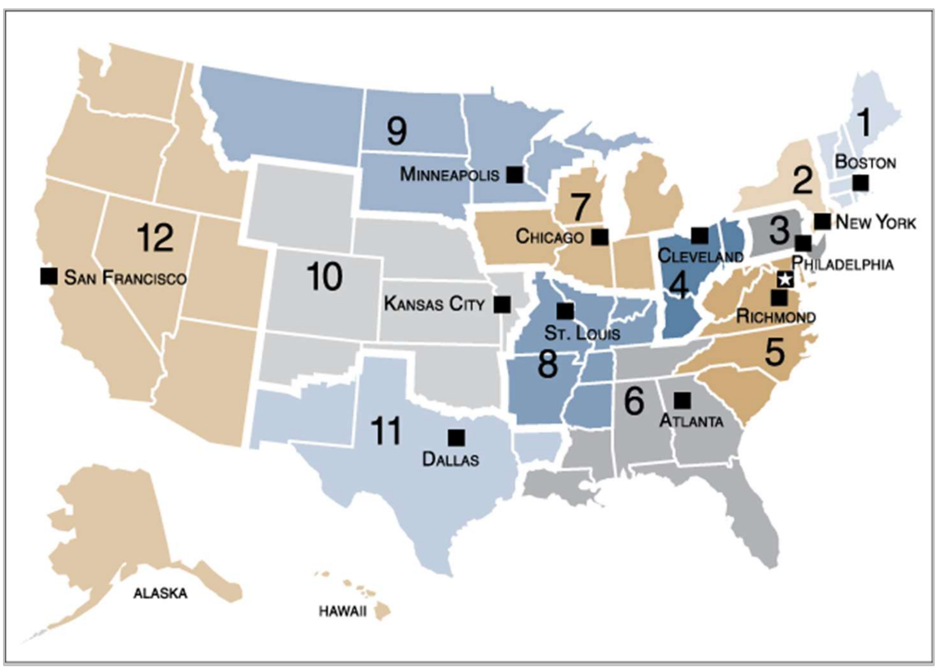
Background

The Beige Book reports



(A partial cover from a Beige Book report)

The Federal Reserve System is the central bank of the United States. It is comprised of 12 districts, each overseen by a board of directors. Eight times per year, each board of directors issues a narrative report of the state of the economy in their district. These reports, collectively known as the Beige Book for their traditional beige-colored covers, summarize the conditions in many economic sectors, such as housing, retail, agriculture, tourism, and finance. Each report presents anecdotal information gleaned through interviews with business leaders, industry experts, and other sources.



(The 12 districts of the U.S. Federal Reserve.)

The reports frequently use language that conveys sentiment – feelings, attitudes, or emotions – about the economy; for example, a report might state that “retailers are generally upbeat” (positive sentiment) or “commercial construction continued to report dismal conditions” (negative sentiment). The Beige Book thus contains valuable information about how experts *feel* about the state of the economy, a valuable source of information not readily available in other economic indicators. Yet, this sentiment is contained within unstructured text data, which is not amenable to quantitative analysis in its raw form.

Lexical sentiment analysis

In text mining, sentiment analysis refers to a collection of techniques for quantifying sentiment in unstructured text data. The goal is to assign a sentiment label or score to each *document*, where a document can be a product review, social media post, email, newspaper article, or in this case, a economic narrative report. Documents typically are labeled or scored for positive versus negative sentiment. The sentiment classifications or scores then form the basis for quantitative analysis.

Numerous methods exist for assigning sentiment scores to raw text. This case study utilizes a common technique known as *lexical sentiment analysis*. In this method, the researcher or analyst specifies a *dictionary* (aka *lexicon*) of sentiment terms and corresponding numeric scores, and the values in this dictionary are used as the basis for calculating an overall sentiment score for each document. We discuss the dictionary and document scoring in detail in the following sections.

The Task

We use lexical sentiment analysis to explore changes in sentiment in the Beige Book reports before and after the Great Recession of December 2007-June 2009. Precipitated by a complex set of factors, most notably the collapse of the U.S. housing market, the Great Recession saw a loss of $2 trillion USD from the global economy. Bankruptcies, business closures, unemployment, and personal hardship spread across the world, and governments took drastic measures to avert a global economic collapse. By many metrics, the recession was the most severe economic downturn since the Great Depression of the 1930s.

We use sentiment analysis to determine the timing and extent of changes in Beige Book sentiment in the time period encompassing the Great Recession. Exploring this relationship represents a first step in understanding how the sentiment conveyed in economic reports might be useful as an indicator of economic recession and recovery.

The Data **BeigeBook_Oct2003-Sep2013.jmp**

The file contains the full text of 960 regional Beige Book reports from October 2003 to September 2013, a 10-year period encompassing the Great Recession. Each row contains the full text of a single report. These data were obtained from the Federal Reserve's online Beige Book archives at https://www.federalreserve.gov/monetarypolicy/beige-book-archive.htm.

Date	The issue date of the report
District	The Federal Reserve district that issued the report
Report Text	The full text of the report

Performing lexical sentiment analysis

Overview of sentiment dictionaries and document scoring

Lexical sentiment analysis requires a dictionary of sentiment terms and corresponding scores. Consider the example dictionary in Exhibit 1, which utilizes a -100 to 100 scale, in which 0 indicates neutral sentiment. The magnitude of a term's score corresponds to the strength of the sentiment conveyed; for example, "benefit" has a score of 40 because it is moderately positive, while "dismal" has a score of -80 because it is strongly negative. Note that neutral or non-sentiment terms, such as "indifferent" or "economy," are omitted from the dictionary, which is equivalent to assigning them a score of 0.

Exhibit 1 Example sentiment terms and their scores

Sentiment term	**Score**
good	30
bad	-30
benefit	40
dismal	-80

The dictionary also specifies *negators*, such as "not," and *intensifiers*, such as "very." These combine with sentiment terms to form sentiment *phrases*, such as "no benefit" or "very good." Negators reverse the sign of the sentiment score (e.g., "no benefit" receives a score of -40), while intensifiers multiply the sentiment score (e.g., "very good" receives a score of 30 x 1.6 = 48). Intensifiers with a score greater than 1 therefore amplify a term's sentiment, while those with a value less than 1 diminish it.

Exhibit 2 Example intensifiers and their multipliers and negators

Intensifier	**Multiplier**
very	1.6
exceptionally	1.8
partial	0.4
almost	0.8

Negator
no
not
don't
without

The dictionary is used to assign an overall sentiment score to each document. This case study uses *scaled* scoring, in which a document's overall score is simply the mean score of all its sentiment terms and phrases. For example, if a document contained "dismal" (-80), "not bad" (30), and "partial benefit" (16), its overall score would be -11.3.

Specifying the dictionary

We first launch Text Explorer, which uses white space and punctuation to delineate each individual word (or *token*) in each document. The initial report, shown in Exhibit 3, presents terms and phrases ordered by their frequency in the 960 documents. "Sales" is the most frequent term, occurring 11,022 times. Exhibit 3 also shows a word cloud, which represents term frequencies graphically: each term's size is determined by its frequency, with more frequent terms being larger. Notably, the most frequent terms in these data would not be considered sentiment terms, in that they do not imply positive or negative emotion. Many text mining techniques require declaring irrelevant terms as *stop words* to exclude them from further analysis. In lexical sentiment analysis, however, declaring non-sentiment terms as stop words is not necessary, because any term not included in the sentiment dictionary will not affect the sentiment scores. However, many other text mining procedures rely on declaring stop words, among other data preparation steps. Text Explorer has many data preparation tools under the red triangle menu.

Exhibit 3 The initial Text Explorer report, including a word cloud

Text Explorer for Report Text

Number of Terms	Number of Cases	Total Tokens	Tokens per Case	Number of Non-Empty Cases	Portion of Non-Empty Cases
10220	960	1258171	1310.59	960	1.0000

Term and Phrase Lists

Term	Count
sales	11022
reported	9261
year	8826
contacts	8716
prices	8075
activity	7980
demand	7765
district	7080

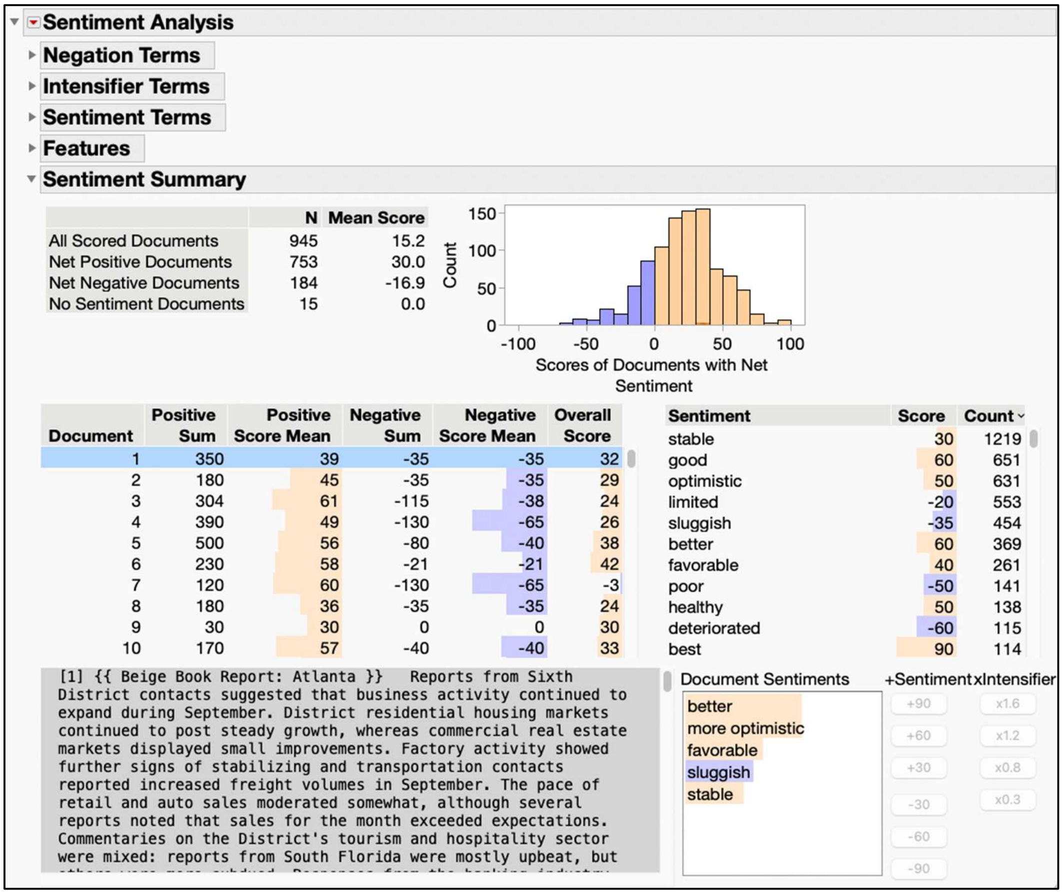
Phrase	Count	N
real estate	4468	2
year ago	2205	2
commercial real estate	1747	3
commercial real	1747	2
last report	1627	2
since the last	1428	3
beige book	1336	2
contacts reported	1286	2

Word Cloud

(Go to Analyze > Text Explorer, place the Report Text column into the Text Columns role, and click OK. In the Text Explorer report, go to the red triangle and select Display Options > Show Word Cloud. In the red triangle next to the word cloud, select Layout > Centered.)

To begin specifying the sentiment dictionary, we first launch the sentiment analysis tool. This produces full analysis results, as well as reveals tools for specifying the sentiment dictionary, shown in Exhibit 4. JMP Pro's Text Explorer has automatically applied a simple default dictionary, and the analysis results under Sentiment Summary are based on this dictionary. However, **the default dictionary must be inspected and modified before interpreting results!** The default dictionary contains a relatively small list of common sentiment terms, intensifiers, and negators, but specific text data are likely to use sentiment terms beyond (or in contradiction with) the default dictionary. For example, the term "strong" may be positive when used in an economic report ("sales were strong") but neutral or negative when used in a customer review of a perfume ("the scent was strong"). Linguistic idiosyncrasies like this are common, so it is critical to ensure that the dictionary is appropriate for the data before interpreting results.

Exhibit 4 The full Sentiment Analysis report, before customizing the dictionary

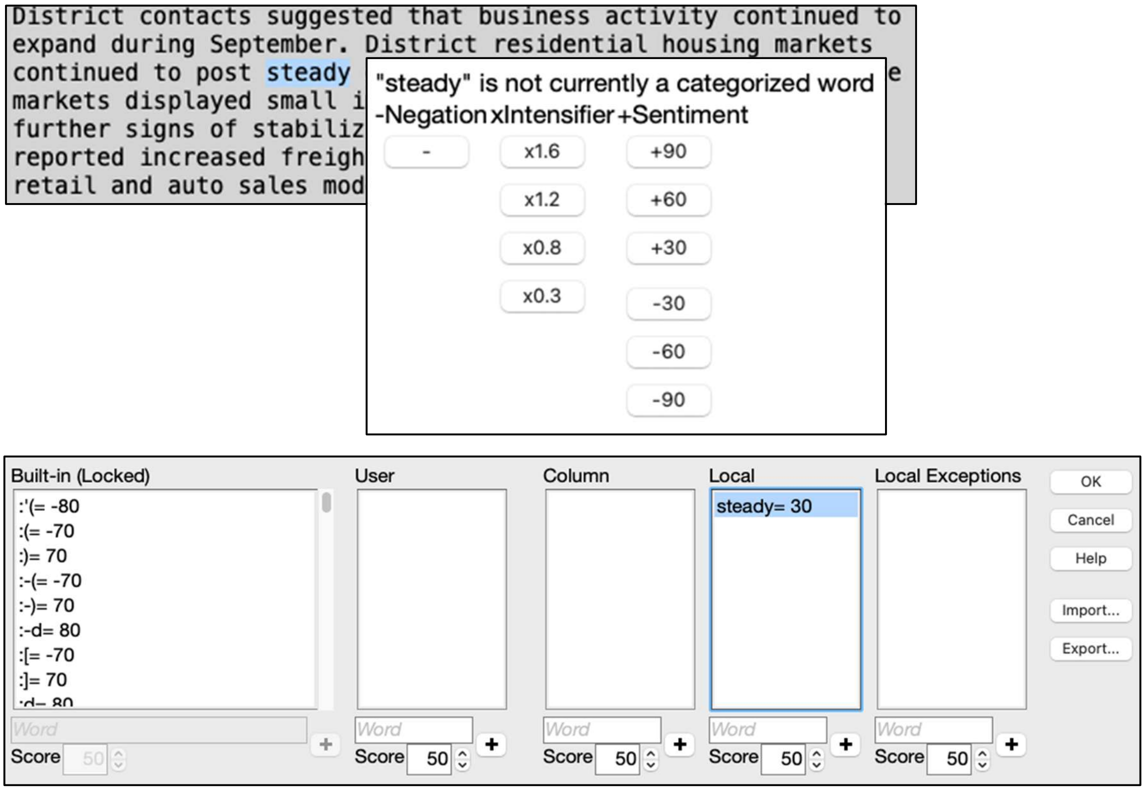


(Go to the Text Explorer red triangle and select Sentiment Analysis. A progress bar may appear, indicating that Text Explorer is parsing grammatical structures of the sentences in each document to assign negation and intensification more accurately. You can select "Continue without parsing" to skip this step, but the sentiment scores will be less accurate.)

We begin by reading several Beige Book reports to identify sentiment terms to add to the dictionary. Document 1 is selected in the interactive document table in the center of the Sentiment Summary, and the text of that document is displayed in the gray box below. The Document Sentiments list on the bottom right shows all the default sentiment terms found in Document 1, with negative terms in purple and positive terms in orange. Hover the pointer over any term or phrase to see its score.

We identify several sentiment terms in Document 1: for example, we find the term "steady" in the second sentence, which in our data typically implies desirable economic stability. In Exhibit 5, we add "steady" to the dictionary with a score of 30. Note that this and all other modifications to the sentiment dictionary are ultimately subjective and should be made using domain expertise. "Steady" now appears in the Document Sentiments list, and the scores of all documents have been updated to reflect the addition of this term to the dictionary. In addition, adding other grammatical forms of this term to the dictionary should be considered: for example, "steadied" or "steadying."

Exhibit 5 Two methods for assigning “steady” a score of 30 in the dictionary.



(Top: Hover the pointer over “steady” and click the 30 button in the pop-up window that appears.

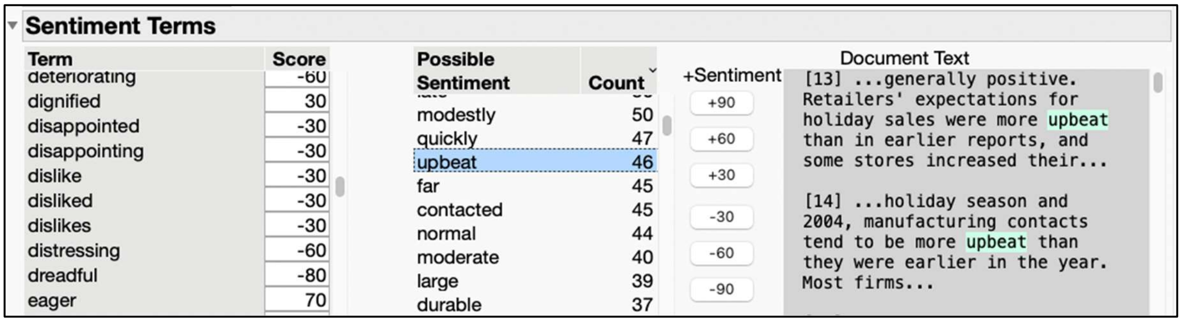
Bottom: Go to the Sentiment Analysis red triangle and select Manage Sentiment Terms. In the Local section, type “steady” into the text field labeled Word, and enter a score of 30 below. Click the plus button, and then click OK. Adding a term to the Local list will apply that term only for this instance of Sentiment Analysis. Click the Help button to learn more about the other Manage Sentiment Terms options, including the ability to import and export dictionaries. Comparable tools exist for managing intensifiers and negators.)

Reading a random selection of documents reveals many possible sentiment terms. Each time, after judging whether the term conveys positive or negative sentiment on average in the data, it is added to the dictionary. After adding many terms, we review the dictionary and scores in the Sentiment Terms section of the report, shown in Exhibit 6. The scores can be viewed and adjusted here, and the Possible Sentiment table to the right algorithmically identifies potential sentiment terms that we may want to add to the dictionary. After carefully considering these terms, we add several more, for example, assigning a score of 40 to “upbeat,” which occurs 46 times in our data.

In practice, we would spend significant time customizing the sentiment dictionary, ideally working with a team of other domain experts to arrive at consensus on terms and scores. We also would clearly document any potentially controversial judgments that we make so that we can justify them when reporting our results. We would do this for the sentiment terms as well as the intensifiers and negators, though they are omitted here for the sake of brevity.

Once the dictionary is finalized, we move on to exploring the results.

Exhibit 6 The Sentiment Terms section, including the current dictionary and a list of possible sentiment terms to add

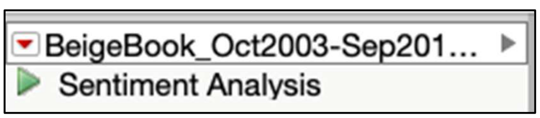


(Click the gray triangle next to Sentiment Terms to open this report section. Manually adjust scores of existing sentiment terms by typing into the text fields to the left. Use the center list to identify possible sentiment terms for consideration. Click on any term to view example uses of that term in the gray box to the right. You also can search for a term by right-clicking on the table header and selecting Show Filter. To add the term to the dictionary, click on the desired score buttons. If you require a different score than those available via the buttons, consider adding the term via the Manage Sentiment Terms tool shown in Exhibit 5.)

Examining the results

At this point, we run Sentiment Analysis from a script saved to the data table; see Exhibit 7. This automatically loads a prespecified sentiment dictionary, saving time and ensuring consistency with the numerical results presented in the remainder of the case study.

Exhibit 7 Launching the table script



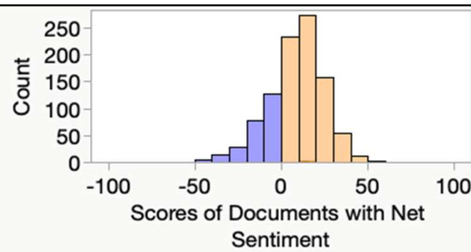
(Click the green arrow to launch Sentiment Analysis with the dictionary used for the remainder of the case study.)

The top of the Sentiment Summary, shown in Exhibit 8, provides a histogram and statistics summarizing the distribution of sentiment scores across the 960 reports. The histogram reveals substantial variability, with individual sentiment scores ranging from approximately -50 to 50. This is unsurprising, given that the 10-year period of analysis encompasses an economic meltdown and subsequent recovery. The summary table shows that the mean across all scored documents is 8.3, a mildly positive average sentiment. We also find that a strong majority of the reports (720 of 960, or 75%) had a positive overall score. Every report contained at least one sentiment term, so there are zero No Sentiment Documents.

The documents table shows the scores of each individual document, and as we saw previously, we can click on any row in the table to view the corresponding document. Document 1 has received an overall score of 12, with a mix of positive and negative sentiments expressed. The score of 12 is substantially different from the score of 32 assigned to this document by the default dictionary (as shown in Exhibit 4); it's a good thing we customized our dictionary! Hovering our mouse over the document's overall score reveals a pop-up window that shows how the value of 12 is calculated: it is the mean of 36 individual sentiment terms and phrases appearing in the document.

Exhibit 8 Sentiment score summary table and histogram, as well as the first few entries in the documents table and the sentiment terms table

	N	Mean Score
All Scored Documents	960	8.3
Net Positive Documents	720	15.5
Net Negative Documents	239	-11.0
No Sentiment Documents	0	0.0



Document	Positive Sum	Positive Score Mean	Negative Sum	Negative Score Mean	Overall Score
1	845	35	-404	-34	12
2	510	43	-310	-28	9
3	1173	49	-694	-30	10
4	630	45	-542	-30	3
5	1450	38	-722	-33	12

Sentiment	Score	Count
strong	60	3025
steady	30	2231
mixed	-10	1523
improved	30	1313
modest	-20	1267
weak	-60	1260

We can use the documents table to identify positive or negative reports for further inspection. By clicking on the Overall Score column header to sort the table in ascending order, we see that Document 517 is the most negative, with a score of -50. Clicking on this row reveals the report text, as well as selects the corresponding row in the data table. We find that row 517 contains an Atlanta district report from early 2009, in the heart of the Great Recession. Reading the report, we note many negative sentiments, especially related to real estate, housing, and construction, as might be expected given the history of the Great Recession.

The sentiment terms table, shown in Exhibit 9, summarizes the frequencies of all sentiment terms that appear in the data. We see that "strong" is the most frequent, occurring 3,025 times across all documents, an average of more than three times per document. We can click on any sentiment term, and the gray text pane will display examples of that term in context. For example, we can click on "steady" and verify that its usage does appear mildly positive on average, as we had judged when we assigned this term a score of 30 in the dictionary.

Exhibit 9 List of sentiment terms appearing in the data, ordered by frequency

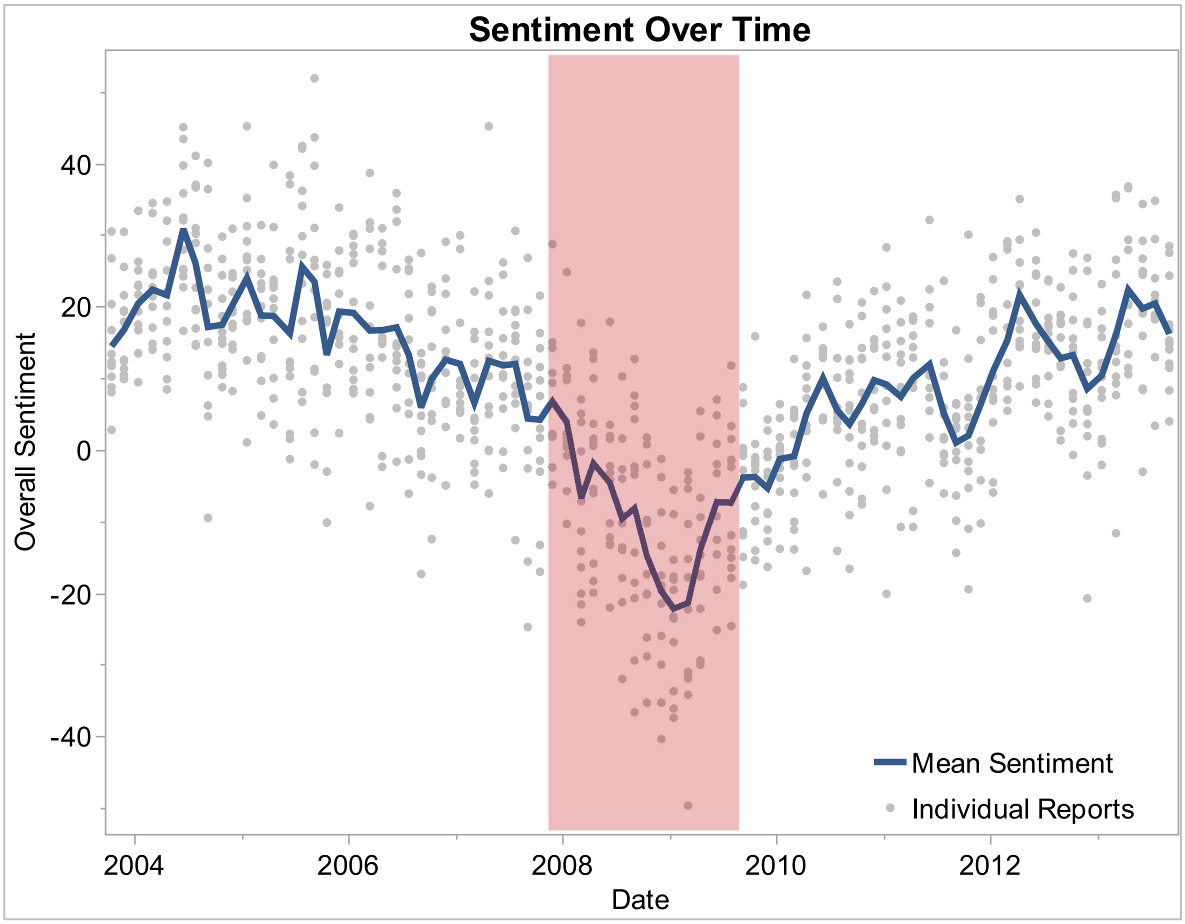
Sentiment	Score	Count
strong	60	3025
steady	30	2231
mixed	-10	1523
improved	30	1313
modest	-20	1267
weak	-50	1260
stable	30	1219
slow	-20	839
positive	30	833
improvement	30	791
solid	60	782

Sentiment decline and recovery during the Great Recession

With sentiment now quantified, we address our primary question: What was the timing and extent of changes in Beige Book sentiment for the time period encompassing the Great Recession? We answer that question with a simple graph of mean sentiment over time, shown in Exhibit 10. The graph shows a gradual decline in sentiment beginning in 2004, nearly three years before the start of the recession.

Sentiment does not reapproach its early 2004 peak until early 2012. Notably, a steep "V" shape appears from mid-2007 to mid-2010. Sentiment began declining rapidly just before the onset of the Great Recession, and it also began to rise rapidly just before the end. However, the time lags between these rapid changes in sentiment and the start and end of the recession are rather short, and in real time, the changes probably would not have been detected until the recession had already started or ended. It appears that Beige Book sentiment does track with the recession and recovery, but it may not be useful as a leading indicator.

Exhibit 10 Mean Beige Book sentiment over time, with the Great Recession shaded in red



(Go to the Sentiment Analysis red triangle and select Save Document Scores to save the sentiment scores to the data table. Launch Graph Builder (Graph > Graph Builder). Place Date on the X-axis and Overall Sentiment on the Y-axis. Deselect the smoother element, then click and drag the line element into the graph. Finally, double-click the X-axis to open the axis settings, and use the Reference Lines controls to add a shaded region from December 2007 to June 2009. See the Graph Builder documentation at www.jmp.com/help for details on changing the graph aesthetics.)

Narrowing the domain

The Beige Book reports, like many types of documents, express sentiment related to numerous topics or domains, such as housing, finance, and retail. If we are interested in a particular domain, we can rescore documents to reflect sentiment expressed specifically with respect to that domain. Construction is a good example here, in that the Great Recession was spawned in large part by a collapse of the U.S. housing market, and the fallout of the housing collapse was particularly bad for the construction industry.

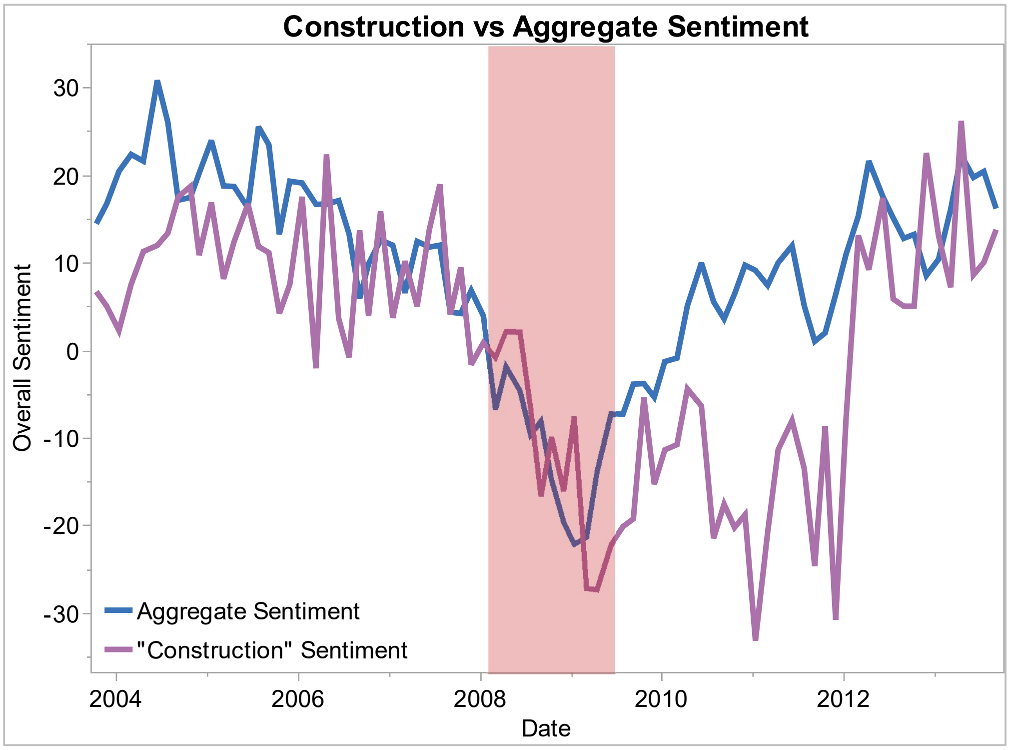
We use the Features section of the Sentiment Analysis report to rescore documents with respect only to the term "construction." This section utilizes an algorithm that identifies sentiment terms that appear in close proximity to "construction" and includes only those in the calculation of document scores. Exhibit 11 shows the Features section and updated summary table and histogram. The summary table shows a nearly neutral mean overall score, with approximately half of the reports expressing net negative sentiment (239 of 464). This histogram shows several spikes – for example, at +/-60 – but if we consult the sentiment terms table, we see that terms with these scores are commonly used to describe construction. The spikes are ultimately an artifact of our dictionary and scoring system, and their locations would change if we were to assign different scores to those common sentiment terms. Still, we see relative balance between positive and negative “construction” sentiment.

Exhibit 11   Using the Features section to rescore documents for “construction” sentiment

(Open the Features section of the report and click Search. Find “construction” in the Possible Features list and click Score Selected Features.)

However, a graph tells an interesting story about how “construction” sentiment changed over time. The graph in Exhibit 12 overlays “construction” sentiment onto the graph from Exhibit 10. The two lines exhibit a similar decline from 2007 through the trough in late 2008, but while aggregate sentiment rapidly recovers, “construction” sentiment remains depressed on average until 2012, before suddenly rebounding. According to our analysis, sentiment related to the construction industry took approximately three additional years to recover relative to sentiment about the economy in aggregate.

Exhibit 12 “Construction” sentiment versus aggregate sentiment over time



(After scoring “construction” sentiment, use the Sentiment analysis red triangle to save the new sentiment scores to the data table. Revisit the graph from Exhibit 10, and drag the new column capturing overall “construction” sentiment (renamed “Construction Sentiment” here) into the Y axis drop zone that lies just inside the graph frame. Deselect the points element in the ribbon of graph elements at the top of the Graph Builder window.)

Summary

Statistical insights

Lexical sentiment analysis uses a dictionary of sentiment terms and scores to quantify positive and negative emotion expressed in unstructured text. Sentiment terms can combine with negators and intensifiers to form sentiment phrases. A document's overall score can be calculated as the mean score of all sentiment terms and phrases contained in the document, though other scoring methods exist. Sentiment in unstructured text is thereby quantified and can be used in subsequent statistical analysis.

Managerial implications

Sentiment analysis can quantify positive or negative emotion in any unstructured text, from popular media to technical reports to product reviews. Sentiment scores can be used to quantify average sentiment, to examine changes in sentiment over time, to identify strongly positive or negative documents for review, or to use sentiment scores as variables in other analyses. However, care must be taken to specify a complete and accurate sentiment dictionary before interpreting or using sentiment scores.

JMP features and hints

This case study demonstrated the use of Sentiment Analysis, first introduced in JMP Pro 16. It showed several methods for specifying the sentiment dictionary and for interactively exploring the results. It also demonstrated how to use the Features section to score documents according to specific subdomains, as well as how to save sentiment scores for further graphing and analysis.

Exercises

1. Creating the sentiment dictionary is a subjective enterprise. Find at least three things you would change about the dictionary (terms to add or remove, scores to change, etc.). Justify your choices in writing, and then implement them using the sentiment term management tools.
2. We did not modify the default intensifiers or negators in this case study. Make at least two modifications to each. Again, justify your choices in writing.
3. Has the distribution of sentiment scores changed notably? If yes, how has it changed? How have the changes you made to the dictionary affected the sentiment scores? If you didn't observe notable changes in sentiment scores, why not?
4. Recall the analysis of "construction" sentiment. Identify another domain of interest, rescore the documents for sentiment related to that domain, and recreate the plot from Exhibit 12. What is the relationship between sentiment in your chosen domain and aggregate sentiment?
5. In this case study, we did not explore differences across the districts. Using the results from your newly modified dictionary, recreate the graph from Exhibit 10 but with District in Graph Builder's Overlay or Wrap drop zones. Which district appears to have been least affected by the Great Recession (i.e., experienced the smallest drop in sentiment)? Which appears to have been the most affected? Hint: If the line graph is hard to interpret, try using the smoother instead.