

Appendix to Current Policy Perspectives 25-15:

‘The Beige Book’s Value for Forecasting Recessions’

Mary A. Burke and Nathaniel R. Nelson

November 6, 2025

1. Constructing Sentiment Indexes Using Beige Book Text

Filippou et al. (2024) were the first to use a large language model (LLM) to construct quantitative indexes of economic sentiment based on Beige Book reports.¹ The specific LLM used, FinBERT-Tone, was designed expressly to discern sentiment from financial texts (Huang, Wang, and Yang 2023), building on the original BERT model introduced by researchers at Google in 2018.² FinBERT-Tone classifies entire sentences as either positive, negative, or neutral using a model that discerns context and the relationships between words.³ For a given body of text, such as the Beige Book report for a given Federal Reserve District and a given month, a numerical measure of its overall sentiment, denoted by S , is constructed as a function of the respective numbers of positive and negative sentences it contains, as follows:

$$S = (N^{pos} - N^{neg}) / (N^{pos} + N^{neg})$$

In the above, N^{pos} stands for the total number of positive sentences in the text, and N^{neg} stands for the total number of negative sentences. Sentences deemed neutral are discarded from the calculation. Given a text with all positive sentences (after discarding the neutral ones), the index takes the value 1; given all negative sentences, the index takes the value -1 ; given equal numbers of positive and negative sentences, the index would equal 0.

Values of the sentiment index are calculated for each Beige Book cycle for each of the 12 individual Fed Districts. Separately, *national sentiment* is calculated for each Beige Book report using just the text from the national summary portion of the report.

¹ We independently validate the Beige Book sentiment measures used in Filippou et al. (2024) by feeding archived Beige Book text into FinBERT-Tone (available at <https://huggingface.co/yiyanghkust/finbert-tone>) and calculating the index values ourselves. We replicate relevant portions of Table 1 of Filippou et al. (2024) in Table A1. We accessed Beige Book text from the Minneapolis Fed’s Beige Book Archive: <https://www.minneapolisfed.org/region-and-community/regional-economic-indicators/beige-book-archive>

² BERT stands for bidirectional encoder representations from transformers and was originally published by Devlin et al. (2019).

³ FinBERT-Tone (Huang, Wang, and Yang 2023) is a variant of the original FinBERT (Araci 2019) that received additional training on manually annotated financial texts.

2. Assigning Months to Beige Book Reports

The Beige Book’s irregular release schedule creates a challenge when matching monthly economic data to the report. Following Gascon and Martorana (2024) and Balke and Peterson (2002), we assign each Beige Book to the month that best coincides with its information-collection period and match other economic data from that assigned month. Starting with the Beige Book released in June 1987, the assignment is based on the “closing date” given in each report, which indicates the end of its information-collection period. (The month of the report’s release date either matches or directly follows the month of its closing date.) For closing dates less than two weeks into the month, we assign the report to the preceding month; otherwise, we assign the month of the closing date. For Beige Book texts from May 1970 (the first one) through June 1987, closing dates are not available, and we assign each Beige Book to the month directly preceding the month of its release (following Gascon and Martorana 2024). Based on these procedures, our assigned month either matches the month printed on the report’s cover or directly precedes it. Some months can’t be associated with a report, and we do not impute sentiment values to such months.

As examples of such month assignments, Table A1 of Balke and Peterson (2002) gives the assigned months (column 2) for Beige Book reports with cover dates from July 1983 through January 1997 (column 3). The Beige Book, then called the *Red Book*, was circulated internally within the Federal Reserve System from May 1970 through May 1983, and starting in July 1983, it was released to the public as the Beige Book.

3. Principal Components Analysis of District-level Sentiment

To obtain a low-dimension summary of the District-level reports, we run a principal components analysis of the 12 District-level sentiment scores. The first principal component explains 53 percent of the sum of the variances of the individual District scores. The second explains an additional 7 percent, and each of the remaining components explains 6 percent or less. See Table A4 for details. As shown in Table A5, the first principal component loads very similarly onto each of the individual District sentiment indexes, whereas the second component (and higher components) load differently in different Districts.⁴ Scores of the second and higher principal components were found to add little to no extra value for forecasting recessions and are omitted from the analysis.

4. Out-of-sample Recession Forecasting

⁴ Balke and Peterson (2002) conduct a principal components analysis based on their scores of individual District Beige Book reports. Their results are remarkably similar to ours even though they use a scoring system that is different from ours and a more limited date range of Beige Book reports.

The probit models (described in the main text) can be used to generate predicted values for the probability of a recession, such as a 40 percent probability, rather than binary indicators of whether the economy will or will not be in a recession within three (or 12) months. Results of in-sample probit regressions are given in Tables A2 and A3. To assess out-of-sample forecasting success, we estimate models using roughly the first third of the observation history and use the results to predict recession status over the rest of the history. (See notes to Figures 2 and 3 for details on training and prediction periods.) To do that, we need to translate a given model’s predicted recession probabilities into binary indicators of predicted recession status for the given horizon. The choice of mapping from recession probabilities into binary recession indicators is somewhat arbitrary, and a variety of reasonable decision rules could be used. We adopt the criteria used by Gascon and Martorana (2024), which require that the model give a recession probability of at least 50 percent for two months in a row before a “predicted recession” is indicated, where the start of the recession is backdated to the first month in which the probability reached 50 percent or more. (Of necessity, we ignore months in which there is no associated Beige Book and apply the criteria to two adjacent, non-missing months.) If the predicted recession probability is 25 percent or less, the recession indicator is set to zero. For predicted recession probabilities strictly between 25 and 50 percent, we assign the recession indicator from the preceding (non-missing) month. If the recession probability equals 50 percent or greater in an isolated month, the recession indicator is set to its value as of the preceding month. The analysis is somewhat sensitive to this choice of criteria, but, in qualitative terms, the relative forecasting performance of the different models does not change if we adopt different (reasonable) thresholds for declaring a recession, or if we do a more comprehensive assessment based on the area under the receiver operating curve (AUROC).

5. Sensitivity and Specificity

Sensitivity is defined as the true positive rate of a prediction model. In our case, that means the share of actual “recession-preceding” months (for a given horizon) in which the model is predicting a recession over that horizon. Whether the model is predicting a recession is based on the criteria described in Section 4. Taking the example of the model predicting a recession within three months, we count the number of months in which the model predicts the economy will be in a recession within three months and divide that by the number of months in which the economy is actually in a recession either one, two, and/or three months ahead of that date, regardless of the true recession status in the “current” month. Specificity is defined as the true negative rate, or the share of non-recession-preceding months in which the model is predicting there will not be a recession within three months.

Table A.1: Partial Replication of Table 1 of Filippou et al. (2024)

	(1) National	(2) Current Expansion	(3) Expansion in 3 Months	(4) Expansion in 6 Months
National		-0.03 (-0.02)	-1.52 (-1.23)	-0.60 (-0.56)
Boston	0.11*** (2.77)	1.30 (1.36)	3.44*** (3.49)	4.34*** (5.24)
Chicago	0.15*** (3.24)	2.53** (2.00)	2.23** (1.97)	1.75* (1.76)
Atlanta	0.07** (2.06)	1.20 (1.29)	1.67* (1.76)	1.64* (1.89)
Cleveland	0.06 (1.03)	-0.30 (-0.29)	1.31 (1.36)	1.16 (1.43)
Dallas	0.08** (2.31)	1.11 (1.46)	0.27 (0.36)	-0.17 (-0.25)
Kansas City	0.08* (1.95)	-0.35 (-0.36)	0.24 (0.30)	0.25 (0.38)
Minneapolis	0.04 (0.93)	2.04** (2.20)	-0.31 (-0.45)	-0.77 (-1.08)
New York	0.12*** (3.79)	1.28 (1.15)	0.32 (0.35)	-0.16 (-0.15)
Philadelphia	0.05 (1.49)	1.79*** (2.67)	-0.01 (-0.01)	-1.08* (-1.82)
Richmond	0.12*** (3.14)	2.84** (2.31)	1.56* (1.73)	0.86 (1.13)
San Francisco	0.13*** (3.71)	2.92** (2.41)	1.16 (1.23)	-0.96 (-1.08)
St. Louis	0.03 (1.18)	-0.41 (-0.48)	0.79 (0.98)	1.67** (2.55)
Constant	-0.00 (-0.34)	5.51*** (6.11)	4.63*** (8.49)	4.03*** (10.38)
R-squared	0.69	0.67	0.50	0.38
Observations	455	454	453	450
National = 0		0.98	0.22	0.57
Districts = 0		0.00	0.00	0.00

Source: Authors' calculations based on data from Mitchell et al. (2025) and National Bureau of Economic Research/Haver Analytics.

Notes: Data pertain to the period from May 1970 through March 2024, excluding months with no Beige Book report. Column (1) gives results of a linear regression of the national sentiment index on the sentiment indexes in each of the 12 Federal Reserve Districts (e.g. Boston, Chicago, etc.), where the sentiment index values are calculated using the methods described in Section 1 of this Appendix. Columns (2) through (4) show results of logit models relating indicators of economic expansion (per NBER) at the indicated horizon to national sentiment and sentiment in each District. t-statistics in parentheses; in column (1) those are based on HAC standard errors with five lags. *** indicates statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level. The row marked "National=0" reports p-values of the respective tests of the statistical significance of the national sentiment index in each regression; the row marked "Districts=0" reports p-values of the respective tests of the joint statistical significance of the 12 District-level sentiment indexes.

Table A.2: Probability of Recession within 3 or 6 Months; Marginal Effects from Probit Models Including Beige Book Sentiment and/or Term Spreads and Stock Returns.

	1–3 Months			1–6 Months		
	(1)	(2)	(3)	(4)	(5)	(6)
National Sentiment	0.021 (0.064)		0.018 (0.053)	-0.024 (0.086)		-0.081 (0.069)
District Sentiment	-0.070*** (0.009)		-0.032*** (0.009)	-0.072*** (0.011)		-0.020* (0.010)
Term Spread		0.012 (0.014)	0.011 (0.011)		-0.007 (0.020)	-0.010 (0.015)
Lagged Term Spread		-0.085*** (0.014)	-0.051*** (0.015)		-0.096*** (0.019)	-0.059*** (0.019)
S&P 500 Return		-0.007*** (0.001)	-0.003*** (0.001)		-0.008*** (0.001)	-0.004*** (0.001)
Observations	460	460	460	458	458	458
Pseudo R-squared	0.719	0.768	0.773	0.647	0.768	0.758
Beige Books = 0	0.000		0.014	0.000		0.045

Source: Authors' calculations based on data from Mitchell et al. (2025), National Bureau of Economic Research/Haver Analytics, Standard & Poor's/Haver Analytics, Federal Reserve Board/Haver Analytics, and Haver Analytics.

Notes: Data pertain to the period from April 1970 (the month assigned to the earliest Beige Book) through November 2024, looking ahead to February 2025, excluding months with no assigned Beige Book report. (See Section 2 of the Appendix for details on assigning months to Beige Book reports.) The dependent variable is an indicator of whether the US economy was in an NBER-dated recession within one to three months of the given month (first three columns), or within one to six months of the given month (second three columns). "National Sentiment" refers to the sentiment index based on the Beige Book national summary, "District Sentiment" refers to the first principal component score of the 12 District-level sentiment indexes. See Section 1 of the Appendix for details on the sentiment indexes. "Term Spread" refers to the yield spread (in percentage points per year) between 10-year and 3-month Treasuries. "Lagged Term Spread" is the six-month lagged value of the term spread. "S&P 500 Return" refers to the percentage change in the S&P 500 stock index in the preceding 12 months. Coefficients represent marginal effects from a probit regression; Newey-West standard errors are in parentheses. "Beige Books=0" reports p-values of the respective tests of the joint statistical significance of the National Sentiment and District Sentiment in each regression. *** indicates statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level. Pseudo-R-squared values are calculated according to McKelvey and Zavoina (1975).

Table A.3: Probability of Recession within 12 Months: Marginal Effects from Probit Models Including Beige Book Sentiment and/or the Excess Bond Premium

	1-12 Months		
	(1)	(2)	(3)
National Sentiment	-0.026 (0.147)		-0.104 (0.109)
District Sentiment	-0.083*** (0.017)		-0.035*** (0.013)
Term Spread		-0.131*** (0.023)	-0.090*** (0.019)
Excess Bond Premium		0.291*** (0.055)	0.130*** (0.047)
Observations	424	424	424
Pseudo R-squared	0.537	0.644	0.709
Beige Books = 0	0.000		0.006

Source: Authors' calculations based on data from Mitchell et al. (2025), National Bureau of Economic Research/Haver Analytics, Favara et al. (2016).

Notes: Data pertain to the period from January 1973 through February 2024, looking ahead to February 2025, excluding months with no assigned Beige Book. (See Section 2 of the Appendix for details on assigning months to Beige Book reports.) The dependent variable is an indicator of whether the US economy was in an NBER-dated recession within one to 12 months of the given month. "National Sentiment" refers to the sentiment index based on the Beige Book national summary, "District Sentiment" refers to the first principal component score of the 12 District-level sentiment indexes. See Section 1 of the Appendix for details. "Excess Bond Premium" refers to the portion of a corporate bond credit spread not attributable to expected default risk (Favara et al. 2016). Coefficients represent marginal effects from a probit regression; Newey-West standard errors are in parentheses. "Beige Books=0" reports p-values of the respective tests of the joint statistical significance of National Sentiment and District Sentiment in each regression. *** indicates statistical significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level. Pseudo-R-squared values are calculated according to McKelvey and Zavoina (1975).

Table A.4: Share of Combined Variation in the Twelve District Sentiment Indexes Explained by Each Principal Component of District Sentiment

Share of Combined Variation Explained	
Component 1	0.529
Component 2	0.069
Component 3	0.057
Component 4	0.050
Component 5	0.049
Component 6	0.043
Component 7	0.042
Component 8	0.038
Component 9	0.034
Component 10	0.032
Component 11	0.029
Component 12	0.028

Source: Authors' calculations based on data from Mitchell et al. (2025).

Notes: These results were generated using the Stata command 'pca' with default options.

Table A.5: Loading of Principal Components onto Individual District Sentiment Indexes

	1st Principal Component	2nd Principal Component
Atlanta	0.308	-0.217
Boston	0.273	-0.200
Chicago	0.304	0.200
Cleveland	0.309	-0.044
Dallas	0.233	0.760
Kansas City	0.291	0.220
Minneapolis	0.286	-0.054
New York	0.285	-0.175
Philadelphia	0.282	-0.395
Richmond	0.302	-0.037
San Francisco	0.308	0.187
St Louis	0.273	-0.136

Source: Authors' calculations based on data from Mitchell et al. (2025).

Notes: These results were generated using the Stata command 'pca' with default options.